The role of the native language in non-native perception and spoken word recognition:

English vs. Spanish learners of Portuguese

Jaydene Deanne Elvin

BArts (Hons) Spanish, Languages and Linguistics

Masters of Arts (TESOL)

A thesis submitted for the Degree of

Doctor of Philosophy

The MARCS Institute for Brain, Behaviour and Development

ARC Centre of Excellence for the Dynamics of Language

Western Sydney University

March, 2016

Principal Supervisor: A/Prof. Paola Escudero

Associate Supervisors: Dr. Daniel Williams

Dr. Jason Shaw

Prof. Catherine T. Best

Abstract

The ultimate goal for adult learners of a second language (L2) is successful communication. If learners cannot perceive, recognise and produce sounds and words in the L2 they may struggle to understand speakers of that language, who may in turn struggle to understand L2 learners. Not all learners will attain the same level of proficiency and even when immersed in the L2 environment, difficulties in L2 speech perception, spoken word recognition and L2 production persist. These difficulties in L2 speech are often attributed to the influence of the native language on the acquisition of L2 speech perception, spoken word recognition and production. This thesis investigates the role of the native language in Australian English (AusE) and Iberian Spanish (IS) listeners' non-native vowel perception and spoken word recognition of Brazilian Portuguese (BP) and the interrelation between these two abilities. The thesis also investigates whether or not individual listeners follow different developmental patterns when perceiving and recognising BP. Rather than focusing on beginner, intermediate or advanced learners, this thesis investigates naïve listeners (i.e., AusE and IS monolinguals) with no prior knowledge of BP in order to establish the onset (or initial state) of learning.

This thesis comprises four content chapters, as well as an Introduction and Discussion. Chapter 2 reports on Elvin et al. (2014) which is a preliminary investigation of AusE and IS listeners' discrimination of BP vowels and investigates vowel inventory size and acoustic similarity as predictors of discrimination difficulties. Chapter 3 reviews previous acoustic analyses of IS and BP and introduces a new comprehensive acoustic analysis of AusE vowels (Elvin, Williams, & Escudero, 2016) that can be used to predict L2 difficulty in future cross-linguistic studies. The study in Chapter 4 (Elvin, Escudero, Williams, Shaw & Best, under review a) investigates of the role of acoustic similarity in predicting individual listeners' non-native categorisation and discrimination patterns. Chapter 5 (Elvin, Escudero,

Williams, Shaw & Best, under review b) investigates whether individual listeners' difficulties in spoken word recognition are the same as their non-native discrimination difficulties.

The findings indicate that the native language plays an important role in predicting non-native speech perception and spoken word recognition. Specifically, acoustic similarity largely predicted listeners' non-native categorisation and discrimination patterns, and that difficulty in discrimination positively correlated with word recognition difficulty. Interestingly, AusE and IS listeners, whose native languages differ in terms of the size of their vowel inventories, found the same BP vowel contrasts equally easy/difficult to perceive and recognise. Finally, the analyses of individual differences in both abilities were essential for understanding the original group findings, particularly in spoken word recognition where individual variation accounts for part of the identified group effect.

In loving memory of Deanne and Michael Roberts

Acknowledgements

Firstly, I would like to thank my Principal Supervisor, Associate Professor Paola Escudero for her encouragement, guidance, advice and friendship throughout my PhD candidature. I am very grateful for all the training she has given me and I have learned a lot from her. I would also like to thank my supervisory panel, Dr. Daniel Williams, Dr. Jason Shaw and Professor Catherine Best for their guidance, support and advice throughout my candidature. I am particularly grateful to Daniel Williams for providing excellent guidance with some of my data analyses.

Secondly, I would like to thank The MARCS Institute for providing me with technical support, research equipment and a wonderful work environment. I would like to acknowledge The MARCS Institute Higher Degree Research Training Scheme, the ARC Centre of Excellence for the Dynamics of Language (project number CE140100041) and an Australian Postgraduate Award for supporting my research financially. To my colleagues at MARCS, thank you for your friendship and support and the many wonderful memories. You have all helped make this experience so enjoyable and I could not have done it without you.

I would also like to acknowledge all those who have participated in my experiments in Sydney and Madrid. I would like to thank Professor Carmen Victoria Marrero Aguiar for hosting me during my fieldwork at the Universidad de Educación a Distancia. A special thanks to Miguel Jimenez Bravo for his invaluable help in recruiting native Spanish participants and for his assistance in language interpreting and running the experiments in Madrid. I'm also grateful to "mi familia española" for supporting me during my stay in Madrid.

To my amazing friends and housemates, who have been with me during various stages of my PhD; thank you for being a part of this journey and for helping me stay sane throughout this process. Finally, I am eternally grateful to my parents, Dayna and Rob Elvin for always encouraging me to achieve the best. I am truly grateful to them and my brothers, Clint and Jake, for their continued patience, love and support.

Statement of Authentication

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

[redacted]

……………………………………….

Author Contributions

Introduction (Chapter 1)

Some parts of this chapter will appear in the following book chapter:

Production and Perception in the Acquisition of Spanish and Portuguese
Jaydene Elvin, Polina Vasiliev and Paola Escudero

M. Gibson & J. Gil (eds) Romance Phonetics and Phonology, Oxford University Press

The sections from this chapter which were included in this book chapter were parts of the general introduction and the descriptions of the models of L2 speech perception as well some of the studies that investigate L2 Portuguese. JE wrote these sections with advice from PE.PV wrote up sections on L2 Spanish acquisition; however these sections do not appear in this thesis chapter.

Spanish is better than English for discriminating Portuguese vowels: acoustic similarity versus vowel inventory (Chapter 2)

Jaydene Elvin, Paola Escudero and Polina Vasiliev

Frontiers in Psychology, 5(OCT), 1–8. http://doi.org/10.3389/fpsyg.2014.01188

JE posed the research question with valuable input from PE. The stimuli were the same as Vasiliev (2013) and were selected from Escudero et al. (2009). The scripts to run the experiments were provided by PV. JE recruited participants and ran the experiment in Sydney and PE and PV recruited participants and ran the experiment in Madrid. Data analyses were conducted by JE with valuable input from PE. JE wrote the first version of the text and JE, PE and PV rewrote the text into its final version.

Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English (Chapter 3)

Jaydene Elvin, Daniel Williams and Paola Escudero

Journalof the Acoustical Society of America, 140, 576–581. http://doi.org/10.1121/1.4952387

JE posed the research question and designed the experiment with valuable input from PE and DW. Jason Shaw and Catherine Best also provided advice regarding the specific words used in the experiment. JE recruited participants, ran the experiment and segmented the data. DW and a research assistant also assisted with final vowel segmentation and boundary adjustments. JE extracted formant measurements and JE and DW analysed the data with advice from PE. JE wrote the first version of the text and JE, DW and PE rewrote the text into its final version.

The role of acoustic similarity and non-native categorisation in predicting non-native discrimination: Brazilian Portuguese vowels by English vs. Spanish listeners (Chapter 4)

Jaydene Elvin, Paola Escudero, Daniel Williams, Jason Shaw and Catherine T. Best

JE posed the research question with valuable input from PE. JE designed the experiment with PE and advice from JS and CTB. The stimuli were selected from Escudero et al. (2009) provided by PE and edited by JE. JE programmed the experiments in E-Prime, recruited participants and ran experiments in Sydney. JE ran experiments in Madrid with assistance from Miguel Jimenez Bravo. JE segmented all of the production data with some assistance from a Masters of Research student. JE analysed the data with valuable input from DW. JE wrote the first version of the text and JE, PE, DW, JS and CTB rewrote the text into its final version.

The role of speech perception and individual differences when learning novel words in a non-native language: English vs. Spanish learners of Portuguese (Chapter 5)

Jaydene Elvin, Paola Escudero, Daniel Williams, Jason Shaw and Catherine T. Best

JE posed the research question with valuable input from PE. JE and PE designed the experiment with from JS and CTB. The stimuli were selected from Escudero et al. (2009) provided by PE and edited by JE. PE provided the E-Prime script and JE modified it into its final version. JE recruited participants and ran experiments in Sydney. JE ran experiments in Madrid with assistance from Miguel Jimenez Bravo. JE analysed the data with valuable input from PE and DW. JE wrote the first version of the text and JE, PE, DW, JS and CTB rewrote the text into its final version.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction[1]

1.1. General Introduction

Listening, comprehending and speaking in a second language (L2) are often difficult and daunting tasks for adult learners. Even when surrounded by the L2, native-like proficiency in the attainment of perception, recognition and production of L2 sounds is rather difficult. These communicative abilities are essential for the mastery of any given language. If listeners cannot perceive sounds or recognise words in speech, they are unable to understand others and the same goes for those with whom they speak. If they cannot accurately produce sounds and words in the second language, they may not be able to make themselves understood (Díaz, Mitterer, Broersma, Escera, & Sebastián-Gallés, 2015). These difficulties in L2 speech are often attributed to the influence of the native language on the acquisition of the L2. That is, the first language (L1) influences how a learner will perceive, understand and ultimately produce sounds and words in the L2 (Best, 1995; Escudero, 2005, 2009; Flege, 1995).

The present thesis investigates the role of the native language or first language (L1) in Australian English (AusE) and Iberian Spanish (IS) listeners' non-native vowel perception and spoken word recognition of Brazilian Portuguese (BP), as well as the interrelation between these communicative abilities. Most studies that investigate L2 learning do so based on learners who are either beginner, intermediate and advanced learners of that particular language, and focus on group performance rather than on each learner as an individual.

[1] Some sections of this chapter will appear in the following book chapter: Elvin, J., Vasiliev, P. & Escudero, P. (in press). 'Production and Perception in the Acquisition of Spanish and Portuguese' in M. Gibson & J. Gil (eds) Romance Phonetics and Phonology, Oxford University Press.

However, it is also important to consider perception and word recognition at the onset of learning for each learner so that language teaching can be tailored to the specific difficulties an individual of a particular language may face. Therefore in this thesis, I investigate non-native listeners with no prior knowledge of BP in order to establish the onset of learning, or in other words, their initial state as well as the individual variation among listeners at this very beginning of learning.

This chapter is designed to provide an overview of the literature that is relevant to this thesis. In § 1.2 a general overview of the relevant literature pertaining to L2 speech learning is presented, followed by a review of literature pertaining to the role of the native language, which is reported in § 1.3, and in § 1.4, three theoretical models of speech perception are reviewed. § 1.5 includes a literature review of individual variation in L2 learning and why it is important to make predictions for L2 performance at the individual level. The research aims and questions of the present thesis are presented in § 1.6, and the chapter concludes with an outline for the remainder of this thesis.

1.2. L2 speech learning

Foreign-accented speech is a distinctive feature of L2 learners because it reflects the influence of the L1 on the L2, and many learners struggle to "lose" their foreign accent (O'Brien & Smith, 2010). The speech of an L2 learner tends to be slower and more hesitant, consisting of pauses, errors and repairs (Temple, 2000). While one of the main goals for L2 learners is to be able to speak in the second language without a foreign accent, this goal is unfortunately rarely achieved (Rallo Fabra & Romero, 2012). This is because there ere are numerous factors that affect the degree of foreign-accented speech by L2 learners. One contributing factor to foreign-accented speech is the age at which one begins to learn a L2. In fact, it is said that a "critical period" exists for one to acquire the sounds of a new language and when this period has passed the ability to attain native-like pronunciation is lost (Piske, MacKay, & Flege, 2001). For example, Flege (1995) stated that according to the "critical period" hypothesis, in order for language acquisition to be effective, it must occur before the establishment of the hemispheric specialisation of language functions. This critical period is generally thought to last through childhood, with the cut-off being puberty and once this period has passed; language development by adults will be much slower. This claim is supported by many studies that show that adults seem to have much stronger foreign accents than L2 speakers who have learned their second language in childhood (e.g., Flege, Munro, & Mackay, 1995; Flege, 1991, 1995; Johnson & Newport, 1991; Piske et al., 2001). However, the "critical period" hypothesis is not supported by all, as some studies have shown that some L2 learners are able to speak the L2 without any detectable accent (e.g., Bongaerts, Van Summeren, Planken, & Schils, 1997; Moyer, 1999).

Another important factor in foreign-accented speech is language experience, that is, how long learners have been speaking the L2 and how much L2 input learners have received. Language experience is often reflected in the individual's length of residence in an L2-speaking region. It is thought that the more experience that the learner has in the target language or the longer time spent in the L2 environment the more native-like the L2 pronunciation is likely to be. However, empirically, conflicting results have been found; for example, some studies have shown an association between amount of language experience and the degree of foreign accent (e.g., Baker & Trofimovich, 2006; Bohn & Flege, 1990; Flege, 1995). Baker & Trofimovich (2006) found that the relationship between perception and production differed depending on the participants' age of acquisition and length of residence in the L2 environment. In particular, the authors suggest that perception and production may be aligned at initial and advanced stages of learning, but in the intermediate stages there is a misalignment between the two. They further suggest that this misalignment could be related to the variations in the learner's amount and type of L2 experience. Their findings also supported the claim that the degree of foreign accent depends on the age of acquisition because they found that learners who were exposed to the L2 in childhood were the only learners to perceive and produce the L2 sounds with native-like accuracy. However, not all studies have found this same association between language experience and the degree of foreign accented speech (e.g., Flege, 1988; Moyer, 1999). Piske et al. (2001) suggest that these discrepancies could be related to the fact that length of residence only provides a rough index of overall L2 experience and a more longitudinal design may be necessary.

Language use is another contributing factor in the degree of foreign accented speech. That is, learners who speak their L1 frequently are likely to have a much stronger foreign accent than those who use it infrequently. For example, Flege, Frieda and Nozawa (1997) found that Italian speakers who continued to frequently use their L1 spoke English with a stronger foreign accent than those who rarely spoke their L1. But as with findings relating to length of residence, not all have found a similar effect of language use with some studies (e.g., Elliott, 1995) finding little to no effect of language use.

Other factors such as motivation and language learning aptitude are also thought to contribute to the degree of foreign accented speech. According to Piske et al. (2001), most studies that examine the effect of motivation have found some influence of motivation (e.g., professional motivation, social motivation, desire to produce the L2 correctly, etc.), however these studies have not been able to provide evidence that these factors automatically lead to accent-free L2 speech. In regards to language learning aptitude, Piske et al. (2001) also state that an aptitude for language learning as a result of musical ability or the ability to mimic has not yet been identified as a significant and independent predictor of the degree of foreign accented speech and that studies are needed that address whether an ability to learn languages is something one is born with or develops as a result of factors not yet identified.

Most important to the present study is the fact that, the strength or degree of a foreign accuracy is indeed strongly influenced by the learner's L1 and in particular their ability to perceive sounds in the L2. If they are unable to perceive a particular sound in the L2 it is very likely that they will not be able to produce that sound. To date, much of the research investigating an interrelation between L2 abilities has predominantly focused on the relationship between perception and production. That is, pronunciation problems are thought to be the result of difficulty in discriminating the sounds of the new language as they are ultimately perceived and categorised according to the sounds that exist in the native language. In fact many of the models of speech perception account for the link between the two abilities. For example, the Speech Learning Model (SLM, Flege, 1995) was developed to account for the limitations of a learner's ability to perceive and produce native-like sounds due to experience and age related limitations. According to the SLM, a learner's ability to produce native-like sounds is largely dependent on how these sounds are perceived according to their L1. The Perceptual Assimilation Model (PAM, Best, 1995) and its extension to PAM-L2 (Best & Tyler, 2007) posits that learners are able to detect articulatory information in the speech that they perceive and therefore it assumes that a common articulatory metric is shared between perception and production. Furthermore, the Second Language Linguistic Perception model (L2LP, Escudero, 2005) posits that at the initial state of learning, an individual's perception of L2 sounds should closely match the acoustic properties of the sounds as they are produced in the learner's native language. Thus, if a learner perceives these L2 sounds as examples of their own L1 sounds, they should also produce those L2 sounds using acoustic properties similar to their L1 sounds.

There are a number of studies supporting the aforementioned theoretical claims as they have indeed identified a link between perception and production (e.g., Flege, Bohn, & Jang, 1997; Levy & Law, 2010; Morrison, 2003, 2006; Rallo Fabra & Romero, 2012; Rauber, Escudero, Bion, & Baptista, 2005) and suggest that native language perceptual patterns do seem to influence L2 production patterns (Levy & Law, 2010; Llisterri, 1995; Morrison, 2003). However the relation between the two is rather complex and a specific link between the two is still a matter of debate as it is not always a well-established link in terms of empirical evidence (Levy & Law, 2010). In fact some studies have even found an opposite pattern where production precedes perception. An example of this is documented in Sheldon and Strange (1982) where they found that some Japanese learners of English were more accurate in producing the English /r/-/l/ contrast and in fact had lower accuracy scores when perceiving the contrast.

A growing body of research has also shown perceptual difficulties to be linked to difficulties in spoken word recognition (e.g., Broersma, 2002; Escudero & Wanrooij, 2010; Escudero, Broersma, & Simon, 2013; Escudero, Hayes-Harb, & Mitterer, 2008; Escudero, Simon, & Mulak, 2014; Escudero, 2006; Pallier, Colomé, & Sebastián'Gallés, 2001; Weber & Cutler, 2004). In order to understand spoken language, the sounds which are heard must be connected to their meaning. Spoken word recognition refers to the process of mapping an incoming speech sound (or spoken word) with its lexical representation in the mental lexicon, thus providing a bridge between sound and meaning (Marslen-Wilson, 1987). In this way, spoken word recognition differs from speech perception, a process that generally involves categorisation i.e., mapping incoming speech signal to its phonological representation(s) and/or discrimination, i.e., distinguishing between two speech sounds. When a listener hears a spoken word, a number of different possible word forms may be activated in the mental lexicon. The listener must then select one of these activated words as the lexical representation of the spoken word, based on the likelihood of it being correct (Auer, 2009; McQueen, 2007). L2 spoken word recognition difficulty increases when there are more active candidates resulting in more competition between words and slower word recognition

(Vroomen & de Gelder, 1995; Weber & Cutler, 2004). When words sound similar across the two different languages, this may cause multiple language activation and there will be more competitor words activated (Weber & Cutler, 2004).

When two words differ by only one sound (e.g., /sɪt/ and /sæt/) they are considered a minimal pair. When minimal word pairings contain L2 phonological contrasts (e.g., /i/ and /ɪ/) that are perceived as the same sound(s) in the L1, learners may struggle to perceive them as two separate sounds. When these two sounds occur in a minimal pair contrasts such as /bit/ and /bɪt/, learners may as a result find it difficult to recognise these words as two separate words. A number of studies have in fact found an increase in word recognition difficulty for L2 learners when minimal word pairings contain contrasts that are difficult to discriminate (Broersma, 2002; Escudero, et al., 2013; Escudero, et al., 2008; Pallier et al., 2001; Weber & Cutler, 2004). For example, Broersma (2002) found that Dutch learners of English judged non-words more often as words than native listeners when the word contained a sound from a contrast that was difficult for learners to perceive. Specifically, Dutch learners struggle to perceive a difference between the English /æ-ɛ/ contrast, which influences their spoken word recognition as words which contain this contrast (e.g., "flash" and "flesh") are perceived as homophones rather than two different words (Cutler & Broersma, 2005). Additionally, the results from a word recognition task indicated that L1 dominant Spanish-Catalan bilinguals processed words containing minimal pairs as homophones, which in turn provided direct evidence that word recognition uses language-specific phonological representations (Pallier et al. 2001). Escudero et al. (2008) also found that auditory confusable novel words (sound contrasts in words that are difficult to discriminate) can lead to confusion in word recognition.

In sum, the native language strongly influences how well a sound is perceived and ultimately produced. Additionally, the sounds which are difficult to perceive are also difficult to recognise in a word context. Below, I will provide further detail relating to the role of the native language and how it shapes phonological development when learning a new language.

1.3. The role of the native language in L2 phonological development

Cross-language or non-native speech perception refers to how we perceive the sounds of an unfamiliar language. Infants have a remarkable ability to discern the differences between phonetic units across many different languages (Kuhl et al., 2006; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). However, by six months of age infants begin to lose the ability to discriminate those speech sounds which do not occur in the ambient language (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Kuhl, 2000; Tsao, Liu, & Kuhl, 2004). Languages differ on both a phonetic and phonological level (Best, Halle, Bohn, & Faber, 2003). Phonologically, differences in the type and number of phonemes (both consonants and vowels) may affect how the learner perceives, comprehends and ultimately produces non-native or L2 sounds. The strength or degree of a foreign accent across different groups of learners is generally a result of differing phonemic inventories across languages. Consequently, cross-linguistic difficulty is not uniform. Similarities and differences between the native and target phonemic inventories influence development in L2 speech acquisition. Studies have shown that if a listener's native language lacks some of the sounds with similar phonetic properties in the target language they will have greater difficulties perceiving and producing those sounds. For example, Japanese learners of English struggle with the English /r-l/ contrast (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Goto, 1971; Miyawaki et al., 1975), whereas German learners of English do not (Iverson et al., 2003).

Vowels are particularly difficult to acquire because unlike consonants, there are no sharp boundaries between one type of vowel and another and they are distinguished in a more continuous manner (Ladefoged & Johnson, 2011). A well-studied example of this is the

English /i-ɪ/ contrast, which poses numerous difficulties for listeners of a number of different L1 backgrounds including Spanish (Escudero & Boersma, 2004; Escudero, 2005; Flege, et al., 1997; Fox, Flege, & Munro, 1995; Morrison, 2009), Mandarin (Flege et al., 1997), Portuguese (Rauber et al., 2005) and Russian (Kondaurova & Francis, 2008). The difficulty in the discrimination of this contrast for these listeners is attributed to the fact that the vowel inventories of the aforementioned languages does not contain /ɪ/, making it difficult for these learners to distinguish between the two sounds. On the other hand, studies have shown that German learners, who do have the same contrast in their L1 vowel inventory, have less difficulty in perceiving this particular English contrast (e.g., Bohn & Flege, 1990; Bohn & Flege, 1992; Flege, Bohn, et al., 1997; Iverson & Evans, 2007).

These above findings lead to the suggestion that the size of the native vowel inventory in comparison to the target language may be predictive of L2 development. For example, non-native listeners' categorisation of stimuli to English labels was largely predicted by these same listeners' non-native categorisation to vowel categories in their L1. According to Fox et al. (1995), this finding suggests that vowel identification is dependent on the size and nature of the listener's native vowel inventory. This suggestion is supported by research conducted by Bradlow (1995) who found that English listeners' categorisation of Spanish /i/-/e/ and /o/-/u/ synthetic continua was indeed affected by the presence of extra American English categories. Fox et al. (1995) compared monolingual English and Spanish bilinguals' vowel perception and found that English listeners use more phonetic features than Spanish listeners when distinguishing vowels. The authors showed that the structure of a listener's vowel space is affected by their L1 native vowel inventory, as English listeners used three underlying dimensions (vowel height, vowel backness and vowel centrality), whereas Spanish listeners used only two dimensions.

Researchers have suggested that learners with a larger native vowel inventory than the target language should be more successful at learning new sounds than learners with smaller L1 vowel inventories than the target language. Specifically, Iverson and Evans (2007, 2009) show that listeners with a larger vowel inventory (e.g., German and Norwegian) than the L2 were more accurate and had higher levels of improvement post training when perceiving L2 vowels (e.g., English) than those with a smaller vowel inventory (e.g., Spanish). The findings from these studies seem to suggest that having a larger and more complex vowel system may facilitate new learning. This is supported by Diaz Granado (2011) who found that L2 and L3 American English learners of Brazilian Portuguese who displayed difficulties in producing the Brazilian Portuguese /e/-/ɛ/ contrast, were in fact significantly better at discriminating Brazilian Portuguese /e/-/ɛ/ and at assimilating this contrast to the closest English contrast (as represented by the words 'bait'-'bet') than American English listeners who had no experience with Brazilian Portuguese. This finding indicates that, unlike Spanish listeners, American English listeners' initial difficulty with this contrast diminishes with experience.

In contrast to the findings of Diaz Granado (2011), Vasiliev (2013) found that American English listeners from California with no experience with Brazilian Portuguese had hardly any difficulty with BP /e/-/ɛ/, as shown by discrimination accuracy at above 90%. Yet they had considerable difficulty with the BP /i/-/e/ and /o/-/u/ contrasts (accuracy between 60 and 70%). As a follow up to Vasiliev (2013), Elvin and Escudero (2014) presented the same BP discrimination task to Australian English and Spanish listeners who had English as their second language. The authors found that despite their differing vowel inventory sizes, both groups found the same Brazilian Portuguese contrasts difficult or easy to discriminate. Furthermore, the authors found that neither group outperformed the other, which may indicate that a larger vowel inventory may not always prove advantageous for discriminating L2 sounds. It might therefore be useful to take into consideration the acoustic similarity between native and non-native vowels when predicting L2 difficulty. In fact, the Second Language Linguistic Perception model (described below; Escudero, 2005, 2009; van Leussen & Escudero, 2015) explicitly proposes that non-native vowel categorisation and discrimination patterns can be reliably predicted through detailed acoustic comparisons of target and native sound categories.

Numerous studies have found evidence that acoustic similarity can be a reliable predictor of L2 success, particularly if one uses acoustic measurements that are comparable across languages (e.g., Escudero & Boersma, 2004; Escudero & Chládková, 2010; Escudero, Simon, & Mitterer, 2012; Escudero & Vasiliev, 2011; Escudero & Williams, 2011, 2012; Gilichinskaya & Strange, 2010). For example, Escudero and Vasiliev (2011) tested monolingual Peruvian Spanish listeners' perception of Canadian French and Canadian English /ɛ/ and /æ/. The authors found that the acoustic similarity between native and target vowels (as measured by a linear discriminant analysis) was a very good predictor of context-specific perceptual mapping.

In addition, the acoustic analyses presented in Escudero, Sisinni and Grimaldi (2014) predicted different perceptual difficulties for the categorisation of Southern British English vowels despite the fact that the two listener groups (Salento Italian and Peruvian Spanish) shared the same phonemic inventory of five vowels. This finding suggests that even when languages have similar vowel inventories, cross-language acoustic similarity has an important role in predicting L2 perceptual difficulty, as only acoustics predicted the observed differences in non-native vowel perception between the two listener groups.

1.4. Review of Theoretical Models

A number of theoretical models have been developed to describe and account for the L2 phonological development and in particular, L2 speech perception, spoken word recognition and production. The influence of the native language on L2 performance is a core feature of most of the models that have been proposed to explain learners' difficulty when acquiring L2 sounds. For example, the Speech Learning Model (SLM; Flege, 1995), the Perceptual Assimilation Model (PAM; Best, 1995) and PAM-L2 (Best & Tyler, 2007) and the Second Language Linguistic Perception model (L2LP; Escudero, 2005, 2006, 2009) are all theoretical models that investigate non-native and L2 speech perception and focus on the influence of the L1 on the acquisition of L2 speech perception.

In spoken word recognition, models such as the Cohort model (Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980), TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997) and its extension to Shortlist-B (Norris & McQueen, 2008), the neighbourhood activation model (Luce & Pisoni, 1998) and its extension to PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000) all share the common assumption that spoken word recognition involves multiple word activation and the subsequent competition between activated words (Weber & Scharenborg, 2012).

However, unlike the previous models of speech perception and spoken word recognition, there is no specific "model" of L2 speech production. Instead, researchers have drawn on theories of phonetics, phonology and L2 speech acquisition (Colantoni et al., 2015, p. 57). For example, the Articulatory settings theory (Honikman, 1964) investigates cross-linguistic influences and the effect of previously acquired languages on L2 speech production and proposes that all languages are characterised by a series of general articulatory positions that involve the lips, tongue, cheeks, jaw and pharynx (Honikman, 1964; Colantoni et al., 2015, p. 58). The Markedness Differential Hypothesis (Eckman, 1977) is based on theories of markedness, which is a measure of relative phonological complexity (Colantoni et al., 2015, p. 63). The hypothesis proposes that target language structures which differ from the L1 and are more implicationally marked will be difficult to acquire in comparison to those which are less marked or considered a universal, namely those structures widely used cross-linguistically (Eckmann, 1977; Colantoni et al., 2015, p.64). Finally, The Ontogeny-Phylogeny model (Major, 2001) posits an interaction between previous two theories and proposes that language learning is shaped by a learner's L1, the target language and universals (Major, 2001; Colantoni et al., 2015, p. 66).

Below I provide a detailed review of three models of speech perception (SLM, PAM and L2LP. I review these theoretical models in depth as the present study is primarily concerned with the influence of the native language on non-native speech perception as weel as the influence of non-native speech perception on spoken word recognition. SLM, PAM, PAM-L2 and L2LP all claim that the similarity and dissimilarity between native and target language sounds is predictive of how accurate target sounds will be perceived, however they all differ in how they account for these difficulties as described below in the review of each of these models.

1.4.1. The Speech Learning Model (SLM; Flege, 1995)

The SLM (Flege, 1995) was developed as a means of accounting for the limitations of a learner's ability to perceive and produce native-like sounds due to experience and age-related limitations. The model addresses how adults acquire the phonological segments of the L2 with a focus on production by experienced L2 speakers (Best, McRoberts, & Goodell, 2001; Flege, 1995). Furthermore, it was designed with a focus on L2 learning and considers changes in L2 learning that occur during the life span (Guion, Flege, Akahane-Yamada, & Pruitt, 2000). Originally designed to address L2 production, SLM also explores L2 perception as it suggests that the learners' production is largely dependent on the influence of their L1. According to the SLM, a learner's ability to create phonetic categories (long term memory representations of speech sounds) remains intact across a life span (Guion et al., 2000). In the SLM, L1 and L2 sounds are perceptually related at the position-sensitive allophonic level and as a result, L1 and L2 sounds are assumed to occur on the same phonological space (Flege, 1995).Thus, the model predicts the learner's ease or difficulty in learning new L2 sounds by comparing where they are produced in the acoustic space. That is, acoustic comparisons between the two languages can be made to predict how learners will perceive and produce new L2 sounds. In relation to L2 learning, the model claims that the acquisition of L2 sounds is dependent on the perceived similarity and dissimilarity between L1 and L2 sounds (Flege, 1995; Kluge, Rauber, Reis, Bion, & Program, 2007). For example, if there is a reasonable difference between the target L2 phone and the closest phone in the learner's L1 inventory then acquisition is dependent on the learner's ability to establish a new L2 phonetic category. In other words, the greater the perceived phonetic distance between the L2 phone and the closest L1 phone (or the more dissimilar the L2 phone is from the closest L1 phone), the more likely the learner is to detect a phonetic difference and establish a new phonetic category for it, resulting in more native-like perception (Best et al., 2001; Flege, 1995; Guion et al., 2000; Strange, 2007). However, when a new L2 is very similar to an existing L1 sound, the learner will assimilate the L1 category to the L2 category resulting in lower perceptual accuracy (Flege, 1995). When listeners are unable to notice a difference between the L2 sound and the closest L1 sound, the formation of a new category is blocked by "equivalence classification", that is when the perceived L2 speech sound is continually assimilated as the closet L1 sound even after years of L2 use (Flege, 1995; Fowler, Sramko, Ostry, Rowland, & Halle, 2008; Mackay, Piske, & Schirru, 2001). Thus, if a learner has not successfully formed a new category for the new L2 sound they will then produce it according to the acoustic properties of the closest L1 sound (Flege, 1995).

In the SLM, when two L2 sounds are matched to different L1 sounds, they will be discriminated (and therefore produced) in a native-like fashion; however, Flege (1995) also argues that there is no guarantee that they will be fully perceived in the same way as a native speaker. The model also states that the older an individual is at the onset of learning the L2, the greater difficulty the learner will have discerning phonetic contrasts (Flege, 1995). According to the model, the authors predict that adult L2 learners are likely to perceive phonetic differences between some L1 and L2 vowels particularly when the L1 has fewer vowels in its inventory than the target language (Flege, 1995). Furthermore, the more phonetically distant an L2 vowel is from the closest L1 category the more likely new phonetic categories for the L2 vowels will be created which is especially true for L1's which have fewer vowels in their inventory than those of the target language. These L2 vowels will eventually be produced according to their phonetic category representation. This means that if the learner has not successfully formed a new category for the new L2 sound they will then produce it according to the acoustic properties of the closest L1 category, which then results in foreign accented speech (Flege, 1995). In sum, SLM posits that sounds which are similar in the L2 to an existing L1 category should be more difficult to acquire than those in which new categories are formed that do not closely resemble any L1 category (Colantoni, Steele, & Escudero, 2015; Flege, 1995).

Evidence for the SLM's hypotheses has been identified in various studies. For example MacKay, Flege, Piske & Schirru (2001) examined Italian-English bilinguals' production of the English /b/ and their perception of /b,d,g/ tokens to test for phonetic learning in the absence of category formation. Recall that in the SLM, when the perception and production of an L2 sound is considered native-like, a new category is thought to be formed as there is less influence of the L1 on the L2 sound (Flege, Mackay, & Meador, 1999). Whereas category formation is blocked when equivalence classification occurs, that is, when the L2 sound is assimilated as an L1 category (Flege, 1995; Fowler et al., 2008). However, according to the SLM, the absence of category formation does not mean that phonetic learning cannot take place because it is hypothesised that a merged category of the

L1 and L2 speech sound will develop over time (Flege, 1995; MacKay et al., 2001). The findings from MacKay et al. (2001) are in line with the SLM claims as they suggest that phonetic learning for these sounds even in the absence of category formation did take place, thus supporting the SLM claims. Furthermore, early bilinguals were able to perceive English /b, d, g/ and produce English /b/ better than late bilinguals. However, the authors acknowledge that their study does not provide direct evidence of the establishment of new phonetic categories by early bilinguals and instead suggest that the early bilinguals may have performed better than late bilinguals because they had received more phonetic input from native English speakers. In addition, Aoyama et al. (2004) tested the SLM hypothesis that if two L2 sounds differed in perceived dissimilarity from the closest L1 sound, the more dissimilar of the two sounds would manifest a greater amount of learning. The authors tested this hypotheses with Japanese learners of the English [ɹ] and [l] by native Japanese speakers and found that native Japanese children did indeed show a substantial amount of improvement over the course of a year and as expected showed more improvements for the English [ɹ] than [l].

1.4.2. The Perceptual Assimilation Model (PAM; Best, 1994, 1995)

The Perceptual Assimilation Model (PAM; Best, 1994, 1995) was designed to specifically address non-native perception, i.e., listeners who are naïve to the target language and not active L2 learners. Similar to the SLM, PAM claims that a non-native learner's perception of the target language is strongly influenced by their L1. However, unlike SLM, which focuses on individual phonemes in the target language (Colantoni et al., 2015), PAM investigates naïve listeners' ability to discriminate non-native phonological contrasts as influenced by their L1 (Best & Tyler, 2007). Furthermore, while SLM's predictions are based on comparisons of acoustic properties, PAM predicts that listeners detect articulatory properties in the speech signal, such as tongue height and backness, in determining cross-language phonetic similarities (e.g., Best, 1995; Best & Tyler, 2007) The model has also been extended to PAM-L2 (Best & Tyler 2007) in order to incorporate L2 development. PAM-L2 predicts that perceptual learning is influenced by the individual's entire language learning experience (Bundgaard-Nielsen, Best, & Tyler, 2011). PAM and PAM-L2 make predictions for discrimination difficultly based on articulatory-phonetic similarity and posit three main patterns of assimilation of L2 phonemes by listeners. The first pattern of assimilation consists of the L2 phoneme being categorised as a native phoneme, with its goodness of fit ranging from excellent to poor. The second pattern of assimilation involves the L2 phoneme being an uncategorised sound that falls somewhere between native phonemes. The third pattern occurs when the sound is categorised as non-speech, namely a phoneme which cannot be assimilated because it bears no similarity to any native phoneme (Best, 1994; 1995).

In the first pattern of assimilation, in which the L2 phoneme is considered categorised, three assimilation types can be identified, namely Single-Category, Category Goodness and Two-Category. Single-Category assimilation describes a situation in which two non-native sounds are classified as a good example of one single L1 category and is predicted to lead to poor discrimination of the non-native contrast (Best, 1994). Category Goodness (CG) occurs when two non-native phones are assimilated to the same native category, where one phoneme is considered to be more acceptable than the other, more deviant phoneme. Discrimination is predicted to be moderate to very good, depending on the level of difference in category goodness for each of the non-native sounds (Best, 1995). In contrast, Two-Category assimilation describes a pattern in which two non-native sounds are perceived as two different native phonemes, leading to good to excellent discrimination of non-native sounds (Best 1994). PAM's uncategorised assimilation can also be divided into three types, namely focalised, clustered or dispersed categorisation. Faris et al. (2016) explain that phones which are considered "uncategorised" are those that are not consistently assigned to a single L1 category above a predetermined threshold, for example, 50%. Thus in a *focalised* response type, non-native sounds are predominately perceived as one single L1 category, yet below the categorisation threshold. Uncategorised assimilation that is *clustered* occurs when an uncategorised non-native sound is perceived as being similar to a small set of L1 categories. The third type of uncategorised assimilation is *dispersed* where listeners select a range of different L1 categories which seems to reflect random responses as none of the available L1 categories seem similar to the non-native sound (Faris et al, 2016).

In order to test PAM's predicted patterns of assimilation, perceptual assimilation data and category goodness ratings are collected. The perceptual assimilation task is considered an effective method for predicting discrimination difficulties (Strange, 2007) as listeners are asked to classify the sounds of the target language which they hear as one of their own native phonemes. PAM and PAM-L2 predictions have been tested and supported in a number of studies (e.g., Best et al., 2001; Best & Strange, 1992; Guion et al., 2000; Strange, Trent, & Nishi, 2001). For example, Best and Strange (1992) tested the identification and discrimination of three synthetic series of American English approximant contrasts by both native American English and Japanese learners of English. The findings were in line with PAM predictions; inexperienced Japanese learners had difficulty identifying the English /r-l/ contrast resulting in poor performance in discrimination. Guion and colleagues (2000) found that perceptual assimilation patterns of English vowels to Japanese L1 categories successfully predicted discrimination difficulties for native Japanese learners of English. The authors explained that the degree of perceptual difficulty seemed to depend on the extent to which the two members of a consonant contrast were identified as a pattern of single category assimilation (Guion et al., 2000). Further support for PAM predictions is seen in Best et al. (2001) who showed that the assimilation patterns found in the identification of Zulu sounds by English listeners successfully predicted performance in the discrimination of these contrasts.

1.4.3. The Second Language Linguistic Perception model (L2LP; Escudero 2005, 2006, 2009; van Leussen & Escudero, 2015)

The Second Language Linguistic Perception model (L2LP; Escudero, 2005, 2006, 2009) is the most recent model to explain cross-linguistic speech perception and considers individual learners at all stages of L2 learning. That is, the model explains L2 development from initial contact with the L2 (i.e. the onset of language learning) right through to their ultimate attainment (Colantoni et al., 2015). The model explains and predicts L2 difficulty through comprehensive acoustic comparisons of listeners' native L1 sound categories and the sound categories of the target language (Colantoni et al., 2015). It further posits that, at the initial stage, learners will initially perceive (and produce) the sounds of the L2 in the same way they perceive (and produce) the sounds in their L1 (Escudero, 2005; 2006; 2009). This can therefore lead to difficulties in L2 learning, as learners whose native language has fewer sounds than the target language must learn new sounds (Escudero & Chládková, 2010) and learners with more sounds than the target must learn that some sounds in their native language do not exist in the L2 (Escudero & Boersma, 2002).

According to L2LP, L2 learners will learn to categorise new target language sounds through distributional learning, a type of human learning mechanism whereby perception is affected by the phonetic distribution of speech sounds along an acoustic continuum that encompass the two sound categories (Escudero, Benders, & Wanrooij, 2011; Maye, Werker, & Gerken, 2002). Escudero et al. (2011) revealed that distributional learning results in improved L2 sound perception and that it is beneficial for learners classifying difficult L2 sounds. Learners will then eventually adjust their perceptual mapping to match those of the target L2 with the help of their lexical (word) representations (Escudero, 2005). The model thus proposes that the comparison of acoustic properties will determine how new sounds are perceived, and that non-native perception is, in turn, a reliable indicator of later L2 development (Escudero & Chládková, 2010).

Acoustic comparisons should ideally be numerical and the L2LP model makes these comparisons through the *optimal perception hypothesis*. This hypothesis suggests that the perceptual mapping of a speech signal depends on the particular characteristics of the listener's production environment (Escudero, 2005). The hypothesis states that experience with the way in which sounds are produced results in optimal perception whereby listeners learn to identify auditory inputs as the sounds that the speaker most likely intended by paying attention to the acoustic cues that are most reliable in the perceptual environment (Escudero, 2005, 2006, 2009). In order to make predictions, the L2LP model suggests that the optimal perception of the listener must first be established, followed by a measurement of how the optimal perception of the listener will handle the acoustic values of the tokens of the new language. Computing the optimal perception of the L1 allows predictions for the listener's initial state of L2 learning, which is the perceptual system that learners will use at the starting point of L2 learning (Escudero, 2005). The model further claims that the differences in the productions of both the native and target language will lead to the differences in their respective optimal perception (Escudero, 2009). Thus by computationally analysing the acoustic values of the L1 and the L2, perceptual assimilation patterns and discrimination difficulties can be predicted.

The L2LP model predicts similar patterns of assimilation as SLM and PAM, however, in the L2LP theoretical framework these are referred to as "learning scenarios". For example, when two non-native sounds are perceived and categorised as a single native sound, known as the NEW scenario, L2LP predicts difficulty in discrimination. When two non-native sounds are perceived and categorised as two separate native categories, L2LP refers to this as the SIMILAR scenario, and discrimination is predicted to be easier in this scenario. It is important to note that although SLM and L2LP use similar terminology, they differ in their use. For instance, in SLM a new sound is one that will not be perceptually assimilated to any L1 category, thus resulting in the creation of a new category which unlike L2LP's definition of the scenario poses little difficulty for learners. Whereas a similar scenario in SLM occurs when a sound is similar to an L1 category and unlike L2LP is predicted to result in difficulties for the learner.

A third scenario in the L2LP theoretical framework is known as the SUBSET scenario (also known as multiple category assimilation) and occurs when the two sounds in the L2 contrast are perceived and categorised as two or more native L1 categories. While some studies suggest that this learning scenario is not problematic for L2 learners (e.g., Gordon, 2008; Morrison, 2009; Morrison, 2003), Escudero and Boersma (2002) suggest that this pattern may be challenging. The subset problem occurs when a learner must realise that certain features or sounds in the target language do not exist and that they cannot process them in the same manner as in their L1, a task that may pose considerable difficulty. Some studies have shown the SUBSET scenario to lead to difficulties in discrimination (e.g., Escudero & Boersma, 2002), particularly when a perceptual or acoustic overlap occurs (e.g., Vasiliev, 2013).

The L2LP model also considers a link between L2 perception and spoken word recognition (Escudero & Boersma, 2004; Escudero, 2005, van Leussen & Escudero, 2015). That is, phonetic contrasts which are confused in speech perception should also be confused in spoken word recognition. However the nature of the link may not always be straightforward. This is because there are a number of factors (e.g., the learning strategies, different task demands, training effects, etc.) that could possibly affect the nature of the relationship between the two abilities.

Specifically, the model claims that while initial L2 perception is based on the *optimal perception hypothesis*, perceptual learning takes place because it is *message-driven* (or *meaning-driven* in the revised L2LP; van Leussen & Escudero, 2015). That is, learners have no direct access to the phonological categories that naïve speakers of the L2 possess, therefore they infer these categories based on how well they are able to understand the meaning intended by the speaker and then attempt to improve recognition by updating their lexical representations. The revised L2LP model (van Leussen & Escudero, 2015) sets out to improve previous computational implementations of the L2LP theoretical proposal for meaning-driven perceptual learning (e.g., Escudero & Boersma, 2004; Weiand, 2007) as well as the specific connections between possible activated candidates, specifically in the SUBSET scenario. To this end, van Leussen & Escudero (2015) modelled the acquisition of the Spanish /i/-/e/ contrast by simulated Dutch learners. Additionally, following long-standing debates in psycholinguistics with respect to the information flow from the speech signal to the lexicon, the authors explored two possibilities in recognising L2 categories: it may be either strictly bottom-up (sequential, as previously described in the original version of L2LP) or it may allow lexical feedback to perception (interactive).

The results from van Leussen and Escudero's (2015) simulations indicate that meaning-driven learning successfully predicts the developmental path of L2 phoneme perception, as outlined by L2LP's SUBSET scenario, and that performance increases in line with the more input a learner receives. With respect to the two types of processing (sequential or interactive), both ultimately lead to correct L2 recognition, but different predictions are made regarding phonetic representations and, consequently, the amount of exposure to L2 words required to acquire the L2 contrast. The results showed that simulated sequential learners needed a larger amount of input data (exposure to L2 words) to achieve the same level of success with the L2 contrast as interactive learners. Van Leussen and Escudero (2015) acknowledge that either strategy may occur in real L2 learners and would be identifiable in the individual differences between learners.

1.4.4. Summary of theoretical models

Table 1 summarises the L2 learning scenarios suggested by the three theoretical models, the predictions that are made and the way in which these models predict L2 ease or difficulty. As shown in the table, although each model addresses some aspects of perception and production, the L2LP model is the only model that explicitly links perception to both word recognition and production. The model predicts that individual learners at their initial state of learning will perceive, recognise, and produce L2 vowels in the same way as their L1. The model predicts that the learner’s L2 productions will match the acoustic properties of their L1. Furthermore, when two non-native categories are perceived as a single native category, the learner’s L2 lexicon will contain the same lexical representation for the two words, that is, they will be perceived as homophones (Escudero, 2005). The fact that L2LP explicitly connects all three abilities makes it a good candidate to test the model predictions in establishing the initial state for AusE and IS listeners’ non-native perception and recognition of Brazilian Portuguese.

Table 1.1 SLM, PAM and L2LP models' focus and predictions for L2 success

Model	Population covered	Aspect of L2 speech	How it describes L2 Speech	Manner of Predicting difficulty	Predictions
SLM	Experienced learners	Production and perception	Acoustic properties	Acoustic comparisons Comparing vowel inventories (using acoustics and not IPA symbols)	<u>*Similar sounds:*</u> difficult to discriminate <u>*Dissimilar* sounds:</u> easy to discriminate
PAM	Non-native L2 learners	Perception	Articulation	Perceptual assimilation patterns	<u>*SC:*</u> poor discrimination <u>*CG:*</u> moderate - very good discrimination <u>*TC:*</u> good -excellent discrimination *UC:* good or poor discrimination depending on uncategorised type
L2LP	Learners from the initial to final state	Perception, word recognition and production	Acoustic properties	Detailed acoustic comparisons of listeners' own vowel productions and those of the L2.	<u>*NEW:*</u> poor discrimination <u>*SIMILAR:*</u> good – excellent discrimination <u>*SUBSET:*</u> good discrimination or poor discrimination depending on whether or not a *subset* problem is identified.

Of the theoretical models described above, SLM investigates differences among L2 learners at the level of extra-linguistic factors such as age of acquisition, language exposure and language use (e.g., Flege, Mackay, & Meador, 1999). However its approach is to group learners according to these factors prior to comparing their performance (Colantoni et al., 2015). Studies investigating perception under the framework of PAM also acknowledge the existence of individual difference among listeners; however few studies are yet to explain such differences. In fact, Tyler et al. (2014) suggests that individual variation should be considered when predicting L2 difficulties. A key component of the L2LP model is that it specifically considers that individuals, even those from the same language group, have different acoustic realisations of their native vowels and this may influence their L2 development at all stages of learning and across different learning abilities (i.e., perception, word recognition and production). The model further posits that listeners follow their own developmental path and as a result predictions for L2 development are made for individual learners based on detailed acoustic comparisons of their L1 categories and the categories of the specific target language variety (Colantoni et al., 2015). Below I will discuss some of the empirical research that emphasise the importance of investigating individual variation in L2 speech.

1.5. Individual variation in non-native and L2 acquisition

Learners vary in terms of how successful they are in learning a language. Not only do they differ in relation to the strength of their foreign accent as well as the speed of their L2 acquisition but the ultimate level of achievement also varies across individuals (Ellis, 2004). The growing body of literature has found language aptitude, motivation and differing learner strategies and styles (Dörnyei & Shekhan, 2003; Ellis, 2004) to be key factors that explain individual variation in second language learning. Despite the growing body of literature for individual differences, only a few studies investigate individual differences in non-native and L2 phonetics and phonology. In fact most empirical studies on L2 vowel perception and spoken word recognition have investigated learners at group level rather than on an individual basis as researchers are generally interested in how populations behave. While group data is normally sufficient for such investigations, Smith and Hayes-Harb (2011) warn that researchers should be careful when drawing general conclusions about typical performance patterns for L2 listeners based on group averages as individual data may prove useful for interpreting group results, especially given the large variety of situations that influence L2 learning.

Of the studies that do consider individual differences, Golestani and Zatorre (2009) investigated individual variation in perceptual training for English listeners' identification and discrimination of the Hindi dental-retroflex contrast. Results showed evidence of a relationship between the amount of improvement on identification and discrimination across participants, where those who improve on one task were also likely to improve on the other and vice versa. There was also considerable variability on individual listeners' performance pre and post training with only half of the participants showing significant learning. They further found that linguistic experience did not predict any of the pre-test, post-test or training measures suggesting that individual differences may not be explained by differences in language background. The authors suggest that the observed individual differences across the two tasks could be attributed to differences in working memory which differentially influences identification and discrimination. The differences could also be related to the differing abilities of listeners to create long-term memory representations of sounds and/or their ability to later retrieve these representations when classifying or discriminating sounds. The authors further speculate that individual differences in performance on these different aspects of processing might predict different aspects of brain structure and that they might also arise from individual differences in the critical periods underlying these different processing abilities.

Perrachione, Lee, Ha & Wong (2011) also demonstrated that learning a phonological contrast for pitch depends on the interaction between individual differences in perceptual abilities and the design of the training program. In particular, the study indicated that different training paradigms produced markedly different levels of learning achievement depending on learners' pre-training aptitudes. Importantly, the authors found that individuals with weaker pitch perception abilities were disproportionately impaired in a high-variability training environment in which their perceptually stronger peers excelled. Thus, the study demonstrates how individual differences in pre-training measurements of speech and language learning aptitudes interact with the design of training paradigms to the benefit or detriment of learning outcome.

Furthermore, Díaz, Mitterer, Broersma, & Sebastian-Gallés (2012) found that even when factors such as age of acquisition, language exposure and motivation are controlled, individual variation among listeners persists. The study evaluated patterns of individual differences in late bilinguals across different phonological processing tasks (i.e., categorisation, word identification and lexical decision). The authors found that when analysing group results, L2 listeners were less accurate than native listeners, yet on an individual level, almost half of the L2 listeners' performed as well as native listeners in categorisation, with a small percentage also performing within a native range in the identification and lexical decision tasks. They further found distinct proficiency levels to be a function of the phonological processes involved in the task. The authors suggested that their results implied that language teaching programs should adapt to the individual differences in students' learning pace and that successful phonetic training must cover a range of phonological tasks.

A relationship between speech perception abilities and individual differences has also been identified in a series of studies that assess participants' sensitivity to acoustic and speech changes by measuring an event-related brain potential, the mismatch negativity (MMN; e.g., Díaz, Baus, Escera, Costa, & Sebastián-Gallés, 2008; Díaz et al., 2015; Y. Jin, Díaz, Colomer, & Sebastián-Gallés, 2014; Sebastián-Gallés, Soriano-mas, & Baus, 2012). In these studies, the MMN is elicited when the auditory perceptual system detects a mismatch between stimuli that has been frequently repeated and stimuli differing in at least one feature (Díaz, et al, 2015). For example, Sebastián-Gallés et al (2012) found anatomical brain differences to be related to individual differences in vowel perception and Díaz et al (2015) found that variability in the acquisition of non-native speech sounds derives from variability in a speech-specific capability, regardless of the age of onset of L2 learning. Additionally, Jin et al. (2014) also investigated the brain oscillatory dynamics in the theta band, the spectral correlate of the MMN, and found it to underpin success in phoneme learning.

Findings from the above studies suggest that individual differences go beyond the commonly studied factors of age, language experience and motivation. Individual differences persist even in apparently homogenous groups of listeners. The fact that there are individual differences among listeners from similar populations provides additional evidence that not all leaners will master a language equally well. Although L2 difficulty in perception and spoken word recognition may be explicable at a group level, the group examination may fail to account for important aspects of L2 speech perception as well as how native speakers of a target language might perceive speech produced by an L2 speaker (Smith & Hayes-Harb, 2011). For instance, Smith and Hayes-Harb (2011) found that the patterns identified at a group level did not apply for all L1 Mandarin, German and English listeners at an individual level in relation to their judgements of the voicing of final consonants produced by other L1 and L2 speakers of English. Importantly the authors stress the importance of taking care when forming descriptive accounts of theories pertaining to the process of L2 speech perception when they are based on group averages. The authors also caution researchers to not overlook non-native listeners who do not follow the typical pattern as one cannot assume that another subset of participants would not perform in a similar manner.

As previously mentioned, the L2LP model specifically accounts for individual variation and there are a number of studies that use the L2LP theoretical framework to investigate individual differences (e.g., Escudero & Boersma, 2004; Mayr & Escudero, 2010; Wanrooij, Escudero, & Raijmakers, 2013). For example, Escudero and Boersma (2004) found that native Spanish learners of English behave differently when acquiring the English *sheep-ship* contrast depending on whether their target dialect was Scottish English or British English. In particular, those whose target dialect was Scottish behaved in a similar manner as native Scottish speakers, however, those whose target dialect was Southern British behaved in a manner that was distinct from both native Scottish and native Spanish patterns. Escudero, Benders & Lipski (2009) investigated the use of vowel spectrum and duration as cues for the categorisation of Dutch vowels by the groups: L1 Dutch, L2 Spanish and Spanish learners of Dutch. The authors found a large amount of individual variation in cue weighting across the three groups. Interestingly, 14 of the 38 Spanish learners of Dutch weighted spectral cues heavier than duration cues. The authors claim that this finding seems to suggest that some non-native speakers rely on spectrum rather than duration when categorising non-native vowels and that future investigations with larger participant numbers may highlight different cue weighting strategies among individual learners.

Perhaps one of the most comprehensive accounts of individual variation according to the L2LP theoretical framework is that of Mayr and Escudero (2010). The authors investigated whether native English learners of German follow different developmental learning paths in their perceptual development. The study involved two experiments, the first collected perceptual assimilation patterns of L2 vowels which were subsequently used to predict leaners' performance in an L2 identification task. The results indicated that despite some inconsistencies, individual performance in the L2 identification task was largely consistent with the predictions made based on their individual perceptual assimilation patterns. Mayr and Escudero (2010) further found that the variability in the perceptual assimilation patterns was systematically constrained. That is, despite variation in terms of exact classification percentages, native English listeners' were consistent in selecting only a subset of the 13 native response options available to them when assimilating the six German vowels to native vowel categories.

Additionally, Wanrooij et al. (2013) demonstrated that distributional vowel training can help individual learners to improve their classification of difficult non-native contrasts. Interestingly, the findings revealed that the changes in the use of perceptual cues after training were related to the participants' own listening strategies (i.e., a specific way of using acoustic cues or acoustic-cue weighting) prior to training. In order to identify the different listening strategies, the authors introduced a new method of analysis known as the *latent class regression analysis*. This method of analysis is used to identify groups of participants who have similar latent or non-overt individual characteristics in a statistically reliable way. Importantly, listener strategies were extracted by determining the degree to which acoustic cues predict an individual's vowel classifications, regardless of whether those classifications are correct or incorrect. The authors explain that an acoustic vowel dimension that is a statistically significant predictor of a participant's classification is considered a cue that this particular individual used and therefore a significant part of that particular listener's strategy.

The above findings all support the L2LP claim that individuals differ in terms of their L2 developmental sequences even when common factors that influence individual variation are controlled. However, to adequately test some of the L2LP model's core assumptions, the initial state of learning should be established for individual listeners based on acoustic predictions using their own native vowel productions. The fact that individuals may vary in their acoustic realisations of native vowels should also be considered when investigating non-native categorisation, discrimination and word recognition patterns as this may predict individual differences in perception performance, or in other words, the within-category variation in production may influence perception.

1.6. Research Aims and Questions

The above subsections have introduced the relevant literature pertaining to L2 speech learning, followed by a review of literature pertaining to the role of the native language. They also introduced the three theoretical models of speech perception and a literature review of individual variation in L2 learning highlighting the importance of making predictions for L2 performance at the individual level.

The overall aim of this thesis is to investigate how the native language shapes individual differences in Australian English and Iberian Spanish listeners' non-native speech perception and spoken word recognition of Brazilian Portuguese. This thesis also investigates the interrelation between these two communicative abilities. To address the goals of this thesis, the following specific research questions have been identified and will be investigated in one or more of the experimental chapters.

1. Do Australian English and Iberian Spanish listeners, whose languages differ in terms of vowel inventory size, have similar problems when perceiving and recognising Brazilian Portuguese nonce words? (Chapters 2,4,5)
2. Is acoustic similarity between the native and target language a reliable predictor of non-native perception for listeners at the onset of learning (i.e., the initial state)? (Chapters 2 and 4; Chapter 3 is also applicable in the sense that it reports on data which is necessary to measure acoustic similarity)
3. Do listeners follow different (individual) developmental paths in non-native speech perception and spoken word recognition? (Chapters 4 and 5)
4. Is there an interrelation between the difficulties identified in individual listeners' non-native vowel categorisation / discrimination patterns and spoken word recognition? (Chapter 5)

Australian English, Iberian Spanish and Brazilian Portuguese were the languages chosen because they have different vowel inventory sizes. Australian English has thirteen monophthongal vowels in its inventory, namely, /iː, ɪ, ɪə[2] , e, eː, ɜː, ɐ, ɐː, æ, oː, ɔ, ʊ and ʉː/

[2] The choice of the phonetic labels for the AusE vowels follows the conventions used for describing Australian English in Harrington, Cox & Evans (1997). Traditionally, the AusE /ɪə/ vowel is considered a diphthong and this is the symbol used by researchers of AusE to depict this vowel. However, as mentioned in Chapter 3, studies (e.g., Harrington et al., 1997; Cox, 2006) demonstrate that this vowel has in fact become more monophthongal, particularly in a closed CVC context. The acoustic data presented in Chapter 3 were produced in a CVC context and the results indicate that this vowel does indeed appear more monophthongal, and that therefore it is important to consider this vowel a monophthong and a possible response option in L2 speech perception (Chapter 4), especially because the stimuli were presented in a CVCV context.

and is representative of a language that has a larger vowel inventory than the target language. On the other hand, Iberian Spanish was selected because of its small vowel inventory which consists of five vowels /i, e, a, o, u/. Finally Brazilian Portuguese was selected as the target language because its inventory of seven oral vowels, /i, e, ɛ, a, o, ɔ, u/, falls in between Australian English and Iberian Spanish in terms of its size. The above literature review has shown that the size of the native vowel inventory plays an important role in predicting L2 difficulty. Interestingly, there are very few studies that investigate the acquisition of Portuguese by speakers with a larger vowel inventory (Diaz Granado, 2011; Kendall, 2004; Oliviera, 2006; Vasiliev, 2013). This is mainly because learning to perceive and produce novel vowel contrasts has been demonstrated to be very difficult for L2 learners (e.g., Flege, et al., 1997). However some studies (Escudero & Boersma, 2002; Vasiliev, 2013) have shown that difficulties can indeed arise for learners acquiring a smaller vowel inventory than their native language. In fact, as discussed above, a number of studies have shown that vowel inventory size alone may not be the best predictor of L2 difficulty and therefore suggest that acoustic similarity between the two languages may instead be a more reliable predictor.

When addressing all four research questions, the L2LP theoretical framework is the most applicable to this thesis. This is because it was specifically designed to account for individual variation among non-native speakers at all stages of learning from the onset or initial state of learning right until ultimate attainment. In particular, the model claims that different listeners have different developmental patterns and as a result, predictions based on the L2LP model are made for individual learners based on detailed acoustic comparisons of their L1 categories and the categories of the specific target language variety. Importantly the present thesis has been designed in such a way that it would be an adequate test of the above L2LP model's claims as it investigates whether acoustic similarity reliably predicts individual listeners' own non-native categorisation and discrimination patterns. While most studies that investigate individual differences focus on factors such as age of acquisition, length of residence, language use, language learning aptitude, motivation etc., in this thesis, I focus on the individual differences in native vowel production, which may account for individual differences L2 performance. This approach more adequately tests some of the L2LP model's proposal more adequately than it has been done in previous studies because rather than using previously reported acoustic data to predict L2 difficulty as has been done in many previous studies, the present study collected native vowel production data from the same participants who completed the perception and spoken word recognition tasks. By using the same participants across the different tasks, the present study also adequately tests the L2LP model's claims that there is an interrelation between individual listeners' discrimination and spoken word recognition difficulties. Investigations into the interrelation between differing communicative abilities have previously been problematic due to methodological reasons (Levy & Law, 2010). In particular, most studies use different techniques and different participants when assessing the different communicative abilities which have different task demands. Furthermore, data analyses have been conducted on groups, rather than individual learners. Crucially, in order to bridge the gap between the results found for these communicative abilities, an analysis of individual performance in these different tasks with comparable and controlled methodology is needed. In this thesis, and in particular the Elvin, Escudero, Williams, Best and Shaw (under review a) and Elvin, Escudero, Williams, Best and Shaw (under review b) studies reported in Chapters 4 and 5, the methodology was controlled as much as possible. In particular, native vowel production data, as well as non-native categorisation, discrimination and word recognition data were also collected from the same individuals, using the same stimuli across all non-native perception and spoken word recognition tasks. Participants were also selected based on strict criteria to ensure that all listeners were of similar ages, with similar educational backgrounds and were selected from one geographical location. Thus, all participants were university students aged between 18 and 30 and were selected from one geographical location, namely Western Sydney for Australian English, Madrid for Iberian Spanish and São Paulo for Brazilian Portuguese. The Australian English participants had little to no knowledge of any other language including Portuguese; however, the Iberian Spanish participants did have little-intermediate knowledge of English as it was unavoidable due to the nature of their education system in Spain. The Brazilian Portuguese speakers whose data was used as stimuli across the different tasks also reported no knowledge of any foreign language. By controlling for variation relating to language experience, age and native background the L2LP model claims and individual differences in native production and non-native perceptual relationships can be adequately tested.

This thesis is therefore, to my knowledge, one of the first to predict non-native perception (both non-native categorisation and discrimination) based on the listeners' own native vowel productions. It is also one of the few that predicts word recognition difficulty based on each individual's own discrimination results. For these reasons, it provides a novel and more adequate test of some of L2LP's core assumptions.

1.7. Chapter Summary and thesis outline

This chapter has introduced the overall aim of the thesis which is to investigate how the native language influences Australian English and Iberian Spanish listeners' non-native speech perception and spoken word recognition of Brazilian Portuguese at both a group and individual level and their subsequent interrelation. The chapter has introduced some general findings about non-native speech perception and spoken word recognition as well as a review of the studies that investigate the role of the native language in L2 speech, the theoretical models of speech perception and studies that specifically investigate individual variation in L2 perceptual learning.

In relation to the overall goals of the thesis, the literature review has highlighted the importance of investigating the role of the native language in predicting L2 difficulty. The mixed findings relating to the effect of vowel inventory size on vowel perception have suggested that a comparison of the acoustic properties of vowels in both the native and target languages might provide reliable predictions of L2 perceptual difficulties. In addition, the review of individual variation in second language learning has highlighted the need for more studies to consider individual variation when reporting group data in the field of non-native and L2 speech perception, word recognition and speech production. Therefore, as previously mentioned the present study will explore individual variation in relation to the differences in the listeners' acoustic realisations of their native vowels and how within-category variation in native vowel production may influence non-native perception and spoken word recognition.

Finally, this chapter has highlighted some of the unique attributes of this thesis. In particular, this thesis includes data collected from the same participants across a number of experiments using a very strict and controlled method. These factors makes this thesis appropriate for using the L2LP model as a theoretical framework and for testing some of its core assumptions, specifically those relating to the measurement of the acoustic similarity between the listeners' own native vowel production and the target language in order to establish the initial state of L2 learning. The fact that the same participants and stimuli are used across the tasks in Chapters 4 and 5 also allows for an appropriate examination of the interrelation between perception and spoken word recognition.

Chapters 2 to 5 constitute the experimental chapters of this thesis. Chapter 2 introduces Elvin et al. (2014) which was designed to address the first two research questions and is a preliminary investigation on Australian English and Iberian Spanish listeners' discrimination of Brazilian Portuguese vowels. In Elvin et al. (2014), it is predicted that acoustic similarity is a good predictor of L2 difficulty at a group level; however, more thorough predictions can be made if they are based on detailed acoustic comparisons that specifically relate to the populations intended for testing. For this reason, Chapter 3 will include a review of two detailed acoustic analyses, one for Iberian Spanish and another for Brazilian Portuguese. A new study (Elvin, Williams & Escudero, 2016) will then be introduced that presents a detailed acoustic analysis of Australian English vowels. In Chapters 4 and 5 the core research aims will be addressed. Specifically, Chapter 4 introduces Elvin et al. (under review a), which was designed to address the first three research questions and is a comprehensive investigation of the role of acoustic similarity in predicting individual listeners' non-native categorisation and discrimination patterns. Chapter 5 (Elvin et al. under review b) also addresses the first and third research questions, in addition it addresses the final research question as it investigates the relationship between non-native perception and spoken word recognition as well as how the relationship between these two abilities may differ across individuals. Finally, Chapter 6 will conclude with a general discussion of the key findings from this thesis as well as implications and suggestions for future research.

Chapter 2

Spanish is better than English for discriminating Portuguese vowels: acoustic similarity versus vowel inventory (Elvin, Escudero & Vasiliev, 2014)[3]

2.1. Introduction

It is widely recognised that second language (L2) learners are often unable to distinguish sound contrasts that are not present in their native language (L1). A well-known example is the English /i-ɪ/ vowel contrast which is discriminated poorly by listeners of many L1 backgrounds including Spanish (Escudero & Boersma, 2004; Escudero, 2005; Flege, Bohn, & Jang, 1997; Fox, Flege, & Munro, 1995; Morrison, 2009), Mandarin (James Emil Flege, Bohn, et al., 1997), Portuguese (Rauber et al., 2005) and Russian (Kondaurova & Francis, 2008). However, not all contrasts that are absent in the L1 are equally difficult to discriminate. Models of non-native and L2 sound perception, such as the Second Language Linguistic Perception model (L2LP, Escudero, 2005; 2006; 2009) and the Perceptual Assimilation Model (PAM, Best, 1994; 1995) and its extension to L2 acquisition (PAM-L2, Best & Tyler, 2007) claim that perceptual similarity between native sounds and target language contrasts predicts how accurately naïve listeners and L2 learners will identify the members of those contrasts.

Both L2LP and PAM predict high difficulty in discrimination of target language contrasts that do not exist in the listener’s native language, which is commonly the case when

[3] A version of this chapter appeared in Frontiers in Psychology, Elvin et al. (2014)

the L1 has a smaller sound inventory than the L2. This results in many target language contrasts being assimilated to a single native category, which is known as Single-Category assimilation in PAM (Best, 1994; 1995) and as the NEW scenario in L2LP (Escudero, 2005). On the other hand, target language sounds that are mapped to two different native categories (Best's Two-Category assimilation and L2LP's SIMILAR scenario) are less problematic for learners (Best, Faber, & Levitt, 1996; Escudero & Boersma, 2004). A third scenario, referred to as uncategorised assimilation in PAM and the SUBSET scenario or multiple category assimilation for L2LP, occurs when two L2 vowels in a binary contrast are perceived as belonging to more than two vowel categories in the L1 (Escudero & Boersma, 2002). This scenario usually occurs when the vowel inventory of the target language is smaller than that of the L1. Discrimination in this scenario is expected to be less problematic for learners than the case of Single-Category assimilation (Bohn, Best, Avesani, & Vayra, 2011; Escudero, 2005). However, Escudero and Boersma (2002) suggest that multiple category assimilation may be problematic when it leads to a subset problem where the learner needs to realise on the basis of positive evidence alone that some features or vowels of their own language do not exist in the target language and may find it difficult not to perceive the extra L1 category.

The present study aims at testing the explanatory power of two possible predictors of non-native vowel discrimination accuracy, namely vowel inventory size versus vowel acoustic properties. To this end, we compare how naïve Australian English (AusE) and Iberian Spanish (IS) listeners discriminate vowels in Brazilian Portuguese (BP). These three languages were chosen because they have different vowel inventory sizes: IS has the smallest number of vowels with only five stressed monophthongs /i, e, a, o, u/, BP has a slightly larger inventory of seven stressed oral monophthongs, /i, e, ɛ, a, o, ɔ, u/, and AusE has the largest vowel inventory with 12 monophthongs, /iː, ɪ, e, eː, ɜː, ɐ, ɐː, æ, o, ɔ, ʊ, ʉː/. The two listener groups were chosen because vowel inventory size is likely to determine the specific learning scenarios, from those mentioned above, that a listener will experience when confronted with a new language. Specifically, AusE listeners who have a large vowel inventory are likely to accurately discriminate most BP vowel contrasts, as they all exist in their native language. They may perceive some BP vowels as multiple AusE vowels but as mentioned above, substantial difficulty for this learning scenario is not expected. Conversely, Spanish learners who have a smaller vowel inventory will face Single-Category assimilation scenarios for BP /e/-/ɛ/ and /o/-/ɔ/ as these contrasts are not present in Spanish. Below we will review the evidence supporting vowel inventory size as a successful predictor of non-native and L2 discrimination accuracy, together with findings suggesting that a comparison of the acoustic properties of the listeners' native vowels and those of the target language may be a better predictor.

To investigate the effect of L1 vowel inventory size on L2 perception, Scholes (1968) had six non-native speakers of English classify synthetic vowels sounds firstly in terms of their own native vowels and then in terms of English vowels. Scholes found that listeners' categorisation of the stimuli using English labels in the second condition was largely predictable by their native language responses from the first condition. Fox, Flege and Munro (1995) have interpreted Scholes' (1968) findings as an indication that vowel identification depends in part on the number and nature of the listener's native vowel categories. Bradlow (1995) also found that listeners' categorization of Spanish /i/-/e/ and /o/-/u/ synthetic continua was strongly affected by the presence of extra American English categories. Fox et al. (1995) compared vowel perception of monolingual English speakers and Spanish bilinguals and found that English listeners use more phonetic features to distinguish vowels. Specifically, the authors showed that the structure of a listener's vowel space is affected by their L1 native vowel inventory, as English listeners used three underlying dimensions (vowel height, vowel backness and vowel centrality), whereas Spanish listeners used only two dimensions.

Other studies suggest that learners with a larger L1 vowel inventory than the target language should be better at learning new vowel categories than learners with smaller L1 vowel inventories than the target language. For example, Iverson and Evans (2007) found that when identifying English vowels, German and Norwegian listeners, who have a larger L1 vowel inventory than English, were more accurate than Spanish and French listeners, whose L1 vowel inventory is smaller than English, despite the fact that both groups used the same acoustic cues to identify the English vowels. In a more recent study, Iverson and Evans (2009) found more improvement after an auditory training with English vowels for German than Spanish listeners, which led the authors to conclude that having a larger and more complex vowel system (German) may facilitate vowel learning.

Based on the above findings supporting the predictive role of vowel inventory size in non-native perception (Bradlow, 1995; Fox et al., 1995; Iverson & Evans, 2007, 2009; Scholes, 1968), Spanish listeners should be less accurate at discriminating BP vowels than AusE listeners. As mentioned above, BP /e/-/ɛ/ and /o/-/ɔ/ should be the most difficult as they are likely to be perceived as a single Spanish category, given that /ɛ/ and /ɔ/ are not present in Spanish. Previous studies have indeed shown that Spanish natives, including those who began learning the target language at an early age, have substantial difficulty perceiving the Catalan mid-vowel contrasts /e/-/ɛ/ and /o/-/ɔ/ (e.g., Mora, Keidel, & Flege, 2010; Pallier, Bosch, & Sebastián-Gallés, 1997; Pallier, Colomé, & Sebastián'Gallés, 2001; Sebastián-Gallés, Echeverrı, & Bosch, 2005; Sebastián-Gallés & Soto-Faraco, 1999). If vowel inventory is a good predictor for non-native vowel perception, Spanish listeners should experience a similar level of difficulty with the same mid-vowel contrasts in BP.

On the other hand, AusE listeners should perform better overall than Spanish listeners due to their larger vowel inventory and they should experience fewer problems with BP mid-vowel contrasts as their larger vowel inventory contains similar contrasts, namely /e/-/ɜː/ and

/o/-/ɔ/. Although little is known regarding AusE listeners' perception of Portuguese vowels, a number of studies have examined American English (AE) learners' perception of Portuguese vowels. For example, Diaz Granado (2011) observed that while L2 and L3 AE learners of BP had difficulties producing the BP /e/-/ɛ/ contrast, they were significantly better at discriminating BP /e/-/ɛ/ and at assimilating this BP contrast to the closest English contrast (as represented by the words 'bait'-'bet') than AE listeners who had no experience with BP. This finding indicates that unlike Spanish listeners, AE listeners' initial difficulty with this contrast diminishes with experience, supporting the claim that a larger and more complex vowel inventory may facilitate vowel learning (Iverson & Evans, 2009). However, in contrast to the findings of Diaz Granado (2011), Vasiliev (2013) found that AE listeners from California with no experience with BP had hardly any difficulty with BP /e/-/ɛ/, as shown by a discrimination accuracy at above 90%, while they had considerable difficulty with BP /i/-/e/ and /o/-/u/ (accuracy between 60 and 70%). This finding suggests that examining differences in the number and type of vowel categories between the L1 and the target language may not be sufficient to fully account for differences in non-native perception.

Thus, it seems important to consider the role of acoustic properties in explaining findings such as those reported in Vasiliev (2013). Unlike PAM and PAM-L2, which rely on perceptual assimilation results to predict discrimination accuracy, the L2LP model (Escudero, 2005, 2009) explicitly proposes that non-native vowel discrimination can be reliably predicted with detailed acoustic comparisons of the target language and native sound categories. The model puts forward that the perception of native sounds is optimal because native listeners' perception matches the specific acoustic properties of native sounds (Escudero, Sisinni, & Grimaldi, 2014; Escudero, 2005). Therefore, according to the L2LP model, a listener's initial non-native sound perception should closely match the acoustic properties of sounds as they are produced in the listener's native language (Escudero &

Boersma, 2004; Escudero, 2005; Escudero, Simon, & Mitterer, 2012; Escudero et al., 2014; Escudero & Williams, 2012). The model also advances that as a result of this direct link between production and perception, if languages or dialects differ in their productions of the same phonemes, those differences should be evident in cross-dialectal and cross-linguistic perception (Escudero & Boersma, 2004; Escudero, 2005).

The validity of this cross-linguistic and cross-dialectal proposal was first demonstrated empirically by the differential perception of the same tokens of /i/ and /ɪ/ in native Standard Scottish English (SSE) and Standard Southern British English (SSBE) listeners (Escudero & Boersma, 2004), and in monolingual Peruvian Spanish (PS) listeners (Escudero, 2005). A growing body of recent studies (Escudero & Chládková, 2010; Escudero et al., 2014; Escudero & Vasiliev, 2011) further supports the L2LP proposal, demonstrating that the specific acoustic properties of a language or a particular dialect substantially affect non-native vowel perception. However, some studies have found that acoustic similarity between the native and target language does not always predict native and non-native perception. For example, Strange, Bohn, Trent, & Nishi (2004) compared the acoustic properties of AE and Northern German (NG) vowels using linear discriminant analysis models and found that the models' classifications did not accurately predict NG listeners' perceptual assimilations of AE vowels and that the consonantal context in which NG vowels were produced did not affect AE listeners' classifications, despite the fact that there were significant differences in acoustic properties of the NG vowel when produced in the different contexts.

Escudero and Vasiliev (2011) directly tested Strange et al.'s context-independent hypothesis on PS perception of Canadian English (CE) and Canadian French (CF) /ɛ/ and /æ/ and found that context-specific acoustic differences in the production of the two sounds between CE and CF resulted in differences in PS listeners' assimilation of these phones to native categories. Furthermore, linear discriminant analysis revealed that acoustic similarity between native and target language vowels was a very good predictor of context-specific perceptual mappings. Discriminant analyses including native and target language vowel acoustics have also been shown to successfully predict assimilation patterns for Russian listeners of AE (Gilichinskaya & Strange, 2010) and differences in L2 (English) vowel perception due to dialectal differences in native (Dutch) vowel productions by North Holland versus Flanders speakers (Escudero et al., 2012).

Although not all studies have shown that acoustic comparisons predict non-native or L2 perception, there are indeed many recent studies that have shown that cross-linguistic acoustic similarity can successfully predict difficulty in non-native vowel perception for one group of English listeners (Vasiliev, 2013) and for two groups of listeners with Spanish and Italian as native languages (Escudero et al., 2014). Given that the present study compares two listener groups, the findings of Escudero et al. (2014) are particularly relevant. Their acoustic analyses predicted different perceptual difficulties for the categorisation of Southern British English vowels despite the fact that the two listener groups (Salento Italian and Peruvian Spanish) shared the same phonemic inventory of five vowels. This finding suggests that even when languages have the same vowel inventories, cross-language acoustic similarity has an important role in predicting L2 perceptual difficulty, as only acoustics predicted the observed differences in non-native vowel perception between the two listener groups. We therefore also examined the explanatory power of a comparison of vowel acoustic properties for predicting IS and AusE listeners' discrimination accuracy of BP contrasts.

Figure 1 shows the F1 and F2 values of the seven vowels of BP (Escudero, Boersma, Rauber, & Bion, 2009), together with the five vowels of IS (Chládková, Escudero, & Boersma, 2011) and the twelve AusE monopthongs (Cox, 2006). Although AusE has a larger vowel inventory than BP as well as contrasts that may be comparable to the BP contrasts /e/-

/ɛ/ and /o/-/ɔ/ that are not present in IS, visual inspection of the figure shows that AusE and IS vowels compare similarly to BP vowels, which would predict similar non-native vowel discrimination across these two listener groups. For example, the BP contrasts /i/-/e/ and /o/-/u/ should be more problematic for both IS and AusE listeners than the other four contrasts as a result of single category assimilation for IS listeners and single and multiple category assimilation for AusE listeners, while the other four contrasts have vowels that visually appear closest to two different native vowels. That is, both BP /o/ and /u/ are acoustically close to one native category /u/ for IS, and /ʊ/ for AusE. In the case of the BP /e/ and /i/ both vowels seem to be acoustically close to one native IS category /i/, yet multiple categories (/iː/ and /ɪ/) for AusE.

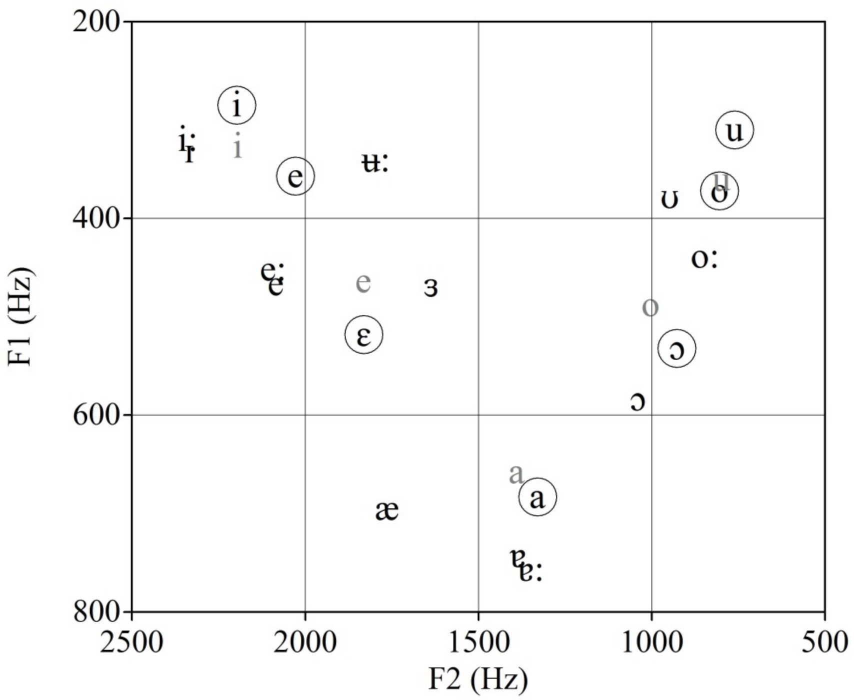


Figure 2.1 Male speakers' average F1 and F2 values for BP (black with circles), AusE (black) and IS (grey).

While plotting the vowels of each language acoustically provides insight for cross-linguistic differences in the location of vowels within the F1-F2 acoustic space, the calculation of the Euclidean Distances (ED)[4] between target vowel contrasts (BP) and native (IS or AusE) vowels can be used as a quantitative measure of cross-linguistic similarity. Table 2.1 shows the EDs between the six BP vowel contrasts considered in this study and the first and second acoustically closest IS or AusE vowel as well as the difference in ED between the first and second acoustically closest vowels. For all BP contrasts, the two vowels involved are acoustically closer to an IS than to an AusE vowel, as shown by the smaller EDs. Additionally, the differences in ED between the first and second acoustically closest vowels are much smaller for AusE than for IS, which suggests that this second native category is a likely choice for AusE but not for IS listeners. Thus, an acoustic comparison predicts overall higher accuracy for IS than AusE listeners. This is because a single IS vowel is acoustically similar to a corresponding BP vowel, while for AusE at least two competing native vowels are in close proximity (neither of which is as close to the BP vowel as the closest IS vowel), which may at least slow discrimination and even lead to confusion.

[4] F1 and F2 values for both BP and L1 vowels were used for computing Euclidean distances. The following equation was used to measure the distance in Bark between the two vowels: d (p,q)= √((p1-q1)^2+(〖p2-q2)〗^2) or d (TV,L1v)= √((TVF1-L1v1)^2+(TVF2-L1v2) ^2), where d stands for Euclidean distance, TV for target vowel, L1v for native vowel, and F1 and F2 for this vowel's average F1 and F2 values.

Table 2.1 Euclidean distances (ED) between the acoustic closest (1st) and second closest (2nd) native vowel (IS and AusE) and each of the two vowels in the six BP contrasts as well as the difference in ED between the 1st and 2nd closest native vowels

BP vowel	**IS: 1st / 2nd closest vowel**						**AusE: 1st / 2nd closest vowel**					
/A-B/	**to A**	**ED**	**ED_diff**	**to B**	**ED**	**ED_diff**	**to A**	**ED**	**ED_diff**	**to B**	**ED**	**ED_diff**
/a/-/ɔ/	a / o	0.34 / 2.37	2.03	o / a	0.60 / 2.71	2.11	ɐ / ɐː	0.52 / 0.55	0.03	ɔ / oː	0.81 / 0.97	0.16
/a/-/ɛ/	a / o	0.34 / 2.37	2.03	e / a	0.47 / 2.15	1.68	ɐ / ɐː	0.52 / 0.55	0.03	ɜ / e	0.87 / 0.98	0.11
/i/-/e/	i / e	0.61 / 1.21	0.6	i / e	0.43 / 2.12	1.69	eː / ɪ	0.92 / 0.98	0.06	iː /ɪ	0.55 / 0.63	0.08
/o/-/u/	u / o	0.11 / 1.66	0.55	u / o	0.57/2.3	1.73	oː / ʊ	0.69 / 0.94	0.25	oː / ʊ	1.38 / 1.41	0.03
/e/-/ɛ/	i / e	0.61 / 1.21	0.6	e / a	0.47 / 2.15	1.68	eː / ɪ	0.92 / 0.98	0.06	ɜ / e	0.87 / 0.98	0.11
/o/-/ɔ/	u / o	0.11/ 1.66	0.55	o / u	0.60 / 1.77	1.17	oː / ʊ	0.69 / 0.94	0.25	ɔ / oː	0.81 / 0.97	0.16

The EDs reported in Table 1 also support the predictions based on Figure 1 regarding the relative discrimination difficulty of BP contrasts and are in line with the findings for American English listeners in Vasiliev (2013). For IS, the EDs confirm that IS /i/ is acoustically the closest vowel to BP /e/ and /i/ and that IS /u/ is acoustically close to both BP /o/ and /u/, which will lead to discrimination difficulty as a result of Single-Category assimilation. For AusE listeners, difficulty in discrimination is also predicted when there is a neutralisation of a L2 contrast caused by multiple category assimilation (L2LP's SUBSET scenario). That is, the two target language vowels are each acoustically close to the same two or more native language vowels, resulting in a partial or total acoustic overlap. For instance, although Figure 1 suggests that only AusE /ʊ/ is acoustically close to both BP /o/ and /u/, the values presented in Table 1 show that in addition to /ʊ/, AusE /oː/ is also acoustically close to the two BP vowels, resulting in a total acoustic overlap for BP /o/-/u/, which will lead to difficulty in discrimination. For BP /i/ and /e/, at first inspection of the EDs, it may seem that the closest AusE vowels are /iː/ and /eː/ respectively, suggesting possible Two-Category assimilation and no difficulty in discrimination. However, the second closest AusE vowel to both BP /i/ and /e/ is AusE /ɪ/, which, due to its acoustic proximity, may well be a competing attractor for BP /i/ and /e/, suggesting a partial acoustic overlap which could lead to confusion and discrimination difficulty for this contrast. Conversely, multiple category assimilation is unlikely to be problematic for AusE listeners in cases like the BP /a/-/ɛ/ contrast where no acoustic overlap occurs.

In sum, if vowel inventory size is a good predictor of non-native vowel discrimination, AusE listeners should be more accurate at discriminating BP vowels than IS listeners because the probability of having vowels which are phonetically similar to the BP vowel system is higher for speakers of larger vowel inventories than speakers of smaller vowel inventories. In particular, the BP contrasts /e/-/ɛ/ and /o/-/ɔ/ should be more difficult to discriminate for IS than AusE listeners, as the lack of /ɛ/ and /ɔ/ in Spanish may result in single-category assimilation. Alternatively, if acoustic similarity measures (as shown in Figure 1 and Table 1) determine success in non-native vowel discrimination, following the L2LP model's acoustic hypothesis, both groups should find the same vowel contrasts equally difficult or easy to discriminate and that in particular, the BP /i/-/e/ and /o/-/u/ contrasts should be most difficult for both groups. Furthermore, if acoustic values are a good predictor of non-native discrimination accuracy IS listeners should be overall more accurate in discriminating BP vowels than AusE listeners.

2.2. Method

2.2.1. Participants

Listeners were 16 AusE and 15 IS functional monolinguals aged between 19 and 55 (mean age, 25.8 for AusE and 25.9 for IS). The AusE participants were tested at the Western Sydney

University and reported little to very basic knowledge of any foreign language and no knowledge of Portuguese. The IS participants were all tested at the Universidad Complutense and at the Universidad Nacional de Educación a Distancia, both in Madrid. They reported a basic to intermediate knowledge of English but did not use English in their daily lives. They also reported very little knowledge of any other foreign language and no knowledge of Portuguese, which suggests that they are functional monolinguals. All participants provided informed consent in accordance with the Western Sydney University Human Research Ethics Committee.

2.2.2. Stimuli

Listeners were presented with 70 Brazilian Portuguese (BP) isolated vowel tokens produced by five male and five female monolingual speakers of BP from Sao Paulo, which were selected from a larger corpus reported in Escudero et al. (2009). The 7 BP isolated vowel (V) tokens (i, e, ɛ, a, o, ɔ, u) were extracted from nonce words in the /fVfe/ context produced in a carrier phrase. We also used seven synthetic tokens representing each of the seven BP vowels for the A and B stimuli in the XAB categorical discrimination task that will be described below. These tokens were synthesised using the computer program Praat (Boersma and Weenink, 2015) and were based on the average F1 and F2 values for BP vowels shown in Figure 1. Figure 2 shows the male and female F1 and F2 values for the natural vowel tokens in relation to the synthesised BP prototypes.

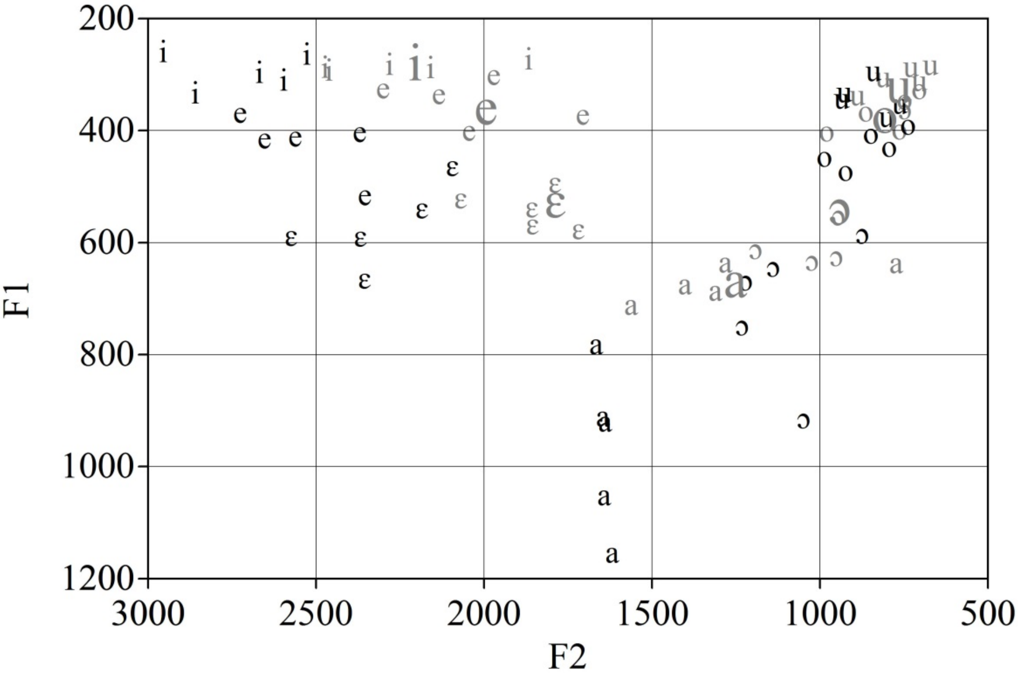


Figure 2.2 F1 and F2 values for the male (grey, small font) and female (black, small font) natural BP vowel tokens, and for the synthetic vowel tokens (grey, large font)

2.2.3. Procedure

Participants were tested in a sound-attenuated room in Sydney and in a sound-proof booth in Madrid. Following the same procedure as Escudero et al. (2009), Escudero and Wanrooij (2010), and Escudero and Williams (2012), participants were presented with an auditory discrimination task in the XAB format, which was run on a laptop computer using Praat. Testing consisted of six categorial discrimination tasks, with each task containing one of six BP contrasts, /a/-/ɔ/, /a/-/ɛ/, /i/-/e/, /o/-/u/, /e/-/ɛ/, and /o/-/ɔ/. In each trial, listeners were presented with three vowel tokens, one after the other, and were asked to decide whether the first vowel (X) sounded more like the second (A) or the third (B) by clicking with a mouse on the corresponding options (either “2” or “3”) on the screen. There were 44 trials for each contrast, and in each trial, the order for the A and B response was counterbalanced, namely XAB and XBA. The X sounds were the natural tokens and the A and B responses were always the two synthetic tokens described above, which mimic the acoustic properties of the specific BP vowels, involved in each of the six XAB tasks. One synthetic token of each of the two vowels was presented twice as the X stimulus to ensure that listeners understood the task and were able to match acoustically equal tokens.

We used synthetic stimuli with mean values of naturally produced Brazilian Portuguese vowels (from Escudero et al., 2009) in order for listeners to make their discrimination decision based on a comparison between individual tokens and average or prototypical values. This results in a categorical discrimination task, where listeners are expected to base their decision of whether A and B are more similar to X on phonemic rather than acoustic differences, as they have to compare different types of stimuli (synthetic versus natural) with different acoustic properties (individual tokens versus average values). The phonemic nature of this XAB task is further strengthened with the use of an inter-stimulus interval (ISI) of 1.2 seconds to ensure phonological processing (Escudero, Benders, & Lipski, 2009). Our task, stimuli types and ISI are identical to those of previous studies which have successfully shown differences between native and non-native listeners for specific vowel contrasts (Escudero & Wanrooij, 2010; Escudero, Benders, & Wanrooij, 2011; Escudero & Williams, 2012; 2014). These studies have also shown that this task avoids listeners' reliance on native orthography, which has been shown to affect their non-native perception. Oral instructions were given in the listeners' native language (English or Spanish). As in Escudero and Wanrooij (2010), a practice session was conducted using a fairly easy contrast, namely /i/-/u/. The experiment took approximately 30 minutes to complete as listeners took around 5 minutes to complete each individual XAB task.

2.3. Results

Table 2 shows the percentage correct with which the AusE and IS monolingual listeners discriminated the BP vowel contrasts.

Table 2.2 AusE and IS monolingual listeners' accuracy scores for the 6 BP contrasts. Standard Error (SE) and Lower (L) and Upper (U) bound confidence intervals of the means are also given.

	/a/-/ɔ/	**/a/-/ɛ/**	**/i/-/e/**	**/o/-/u/**	**/e/-/ɛ/**	**/o/-/ɔ/**
IS	83.18 SE: 2.89 L: 77.25 U: 89.08	98.5 2.03 94.36 102.64	73.67 2.71 68.13 79.2	64.87 3.24 58.15 71.58	82.5 2.99 76.38 88.62	89.33 2.3 84.63 94.04
AusE	75.63 SE: 2.80 L: 69.90 U: 81.35	92.19 1.96 88.18 96.2	66.25 2.62 60.89 71.61	65.94 3.18 59.43 72.44	82.81 2.9 76.88 88.74	80.31 2.23 75.76 84.87

The table shows that, with the exception of BP /o/-/u/ and /e/-/ɛ/, IS listeners had higher discrimination accuracy than AusE listeners. A repeated measures ANOVA with group as a between-subjects factor and contrast as a within-subjects factor revealed main effects of group (F [1, 29] = 5.457, p = .027, η_p^2=.158) and contrast (F [5, 80] = 37.764, p =<0.001, η_p^2 = .566), but no interaction between contrast and * listener group (F [5, 80] = 1.550, p =.178, η_p^2 = .051). This indicates that both groups found the same contrasts equally easy or difficult, but that IS listeners had higher accuracy overall.

To compare accuracy across BP contrasts, paired samples t-tests including all listeners pooled together were conducted for each possible comparison of the six contrasts, with α = .0033 (15 comparisons). The results indicated that /i/-/e/ and /o/-/u/ had significantly lower accuracy than the remaining four contrasts (ts(30) = 7.705 – 12.805, all ps(two-tailed) <0.001), indicating that they were the most difficult to discriminate. The paired t-test that compared accuracy for /i/-/e/ and /o/-/u/ did not yield significance (t(30) = 1.583, p(two-tailed) = 0.124), indicating that these two contrasts were equally difficult. Finally, /a/-/ɔ/, /e/-

/ɛ/ and /o/-/ɔ/ had comparable accuracy (*ts*(30) = 0.803 – 2.204, *ps*(two-tailed) = 0.035 – 0.428), but were more difficult than /a/-/ɛ/ (*ts*(30) = 5.890 – 8.163, all *ps*(two-tailed) <0.001). We present our listeners' ranking of the order of difficulty in the same way that Escudero and Wanrooij (2010) present their order of difficulty. Thus, the ranking of difficulty for both listener groups, ranging from the most to the least difficult BP vowel contrast, is as follows: /o/-/u/ ~ /i/-/e/ > /a/-/ɔ/ ~ /e/-/ɛ/ ~ /o/-/ɔ/ > /a/-/ɛ/, where "~" means equal or comparable difficulty and ">" means higher difficulty.

2.4. Discussion

The present study tested the explanatory power of two possible predictors for non-native discrimination difficulty, namely vowel inventory size versus a detailed comparison of acoustic properties across native and non-native languages. To this end, the discrimination of Brazilian Portuguese (BP) vowels by Australian English (AusE) and Iberian Spanish (IS) listeners was compared. Following predictions based on vowel inventory sizes, AusE listeners, whose native vowel system includes all of the BP contrasts, should outperform IS listeners, who only have five native vowels and lack two of the mid vowels (/ɛ/ and /ɔ/) present in BP, which should result in Single-Category Assimilation and poor discrimination. Alternatively, following the L2LP model's acoustic hypothesis, if a comparison of vowel acoustic properties (see Figure 1 and Table 1) successfully predicts non-native vowel discrimination, IS listeners should have higher accuracy overall in the discrimination of BP vowels. Acoustic properties also predict that both groups will have the same level of difficulty for all contrasts.

The findings are in line with the L2LP model's acoustic hypothesis and the corresponding predictions based on the detailed comparison of BP, IS and AusE vowels that was presented in the Introduction. That is, IS listeners did have higher overall accuracy than

AusE listeners, and relative ease or difficulty for each BP contrast for both groups was largely predictable based on the acoustic comparisons presented in the Introduction (see Figure 1 and Table 1). In particular, the BP contrasts /i/-/e/ and /o/-/u/, where the SUBSET scenario was identified for AusE listeners and the NEW scenario for IS listeners, were indeed the most difficult, and /a/-/ɛ/ was the easiest to discriminate for both groups. It is interesting to note that unlike previous studies of Catalan, which shares a similar vowel inventory to BP, the mid-vowel contrasts /e/-/ɛ/ and /o/-/ɔ/ were not as difficult as the high-vowel contrasts /i/-/e/ and /o/-/u/ for IS listeners. Likewise, the findings of the present study were not in line with those of Diaz Granado (2011), yet they were comparable to those for Californian English (CE) listeners in Vasiliev (2013), as CE listeners also found /i/-/e/ and /o/-/u/ to be substantially difficult. However, unlike AusE listeners, CE listeners found /a/-/ɔ/ as difficult as /o/-/u/, and had significantly lower accuracy scores for /a/-/ɔ/ than for /e/-/ɛ/ and /o/-/ɔ/. As shown in Williams and Escudero (2014), differences in non-native vowel perception between native listeners with different English dialects are also explained by dialectal differences in English vowel production. In that respect, ongoing research comparing AusE, CE and native BP listeners suggest that acoustic properties may be at the heart of the differential non-native patterns. This new study also demonstrates the validity of the BP stimuli used in the present study, as native BP listeners tested in Sao Paulo, Brazil, had accuracy scores of above 83% for the six BP vowel contrasts. Interestingly, a preliminary analysis also shows different levels of accuracy across vowel contrasts and that acoustic proximity is likely to explain this variability in native vowel perception.

The fact that IS listeners had overall higher accuracy despite their vowel inventory lacking the same contrasts that are present in both BP and AusE seems to suggest that vowels which are acoustically closer to the target vowels with no activation of competing categories are easier to discriminate. According to the acoustic predictions described in the Introduction, AusE listeners may use all the vowel categories that are acoustically close to the target BP vowel. This is likely to cause confusion because of the multiplicity of possible response options, resulting in the poorer performance shown in the present study. In other words, our acoustic predictions and discrimination findings seem to suggest that the number of mental representations (i.e., vowel categories) available to the listener influences native and non-native vowel perception. Further evidence for this claim was provided by Benders, Escudero and Sjerps (2012), who investigated the influence of stimulus range (i.e., different subsets of the Spanish /i/-/e/ continuum) and the number of available response categories on vowel categorisation. The authors investigated the influence of the number of response categories by giving half of the participants /i/ and /e/ as responses and the other half, /i/, /e/, /a/, /o/ and /u/. The results showed that listeners who only chose from two response categories were more sensitive to broad and local acoustic contexts than listeners presented with five response categories. Listeners with only two response options were able to shift their boundary between /i/ and /e/ early, while listeners with five responses required more time. The authors argued that the delay in the boundary shift was caused by the availability of extra response options, causing them to be less precise in their responses (Benders et. al., 2012). Although the participants were listening to their own native language, not all tokens presented were native-like, as they were part of a continuum, and so it seems that having more response options or a larger vowel inventory with more mental representations to choose from may result in difficulty in both native and non-native vowel perception.

If the number of mental representations is affecting the AusE listeners' overall performance in discriminating BP vowels, this may indeed suggest an effect of multiple category assimilation ie., L2LP's SUBSET scenario, which can be problematic in vowel discrimination, as demonstrated with Dutch learners of Spanish whose multiple category assimilation patterns were reflected in their poorer classification of Spanish front vowels

(Escudero & Boersma, 2002). Following from our acoustic comparisons, it may be that the AusE listeners' lower overall discrimination scores are a result of multiple category assimilation affecting how well they discriminate BP contrasts. Recall from the values in Table 1 that the difference in ED between the first and second acoustically closest vowels are much smaller for AusE than for IS, which suggests that this second native category is a likely choice for AusE but not for IS listeners. Furthermore, we predicted that discrimination would be difficult when an acoustic overlap (partial or total) was involved. Therefore the AusE listeners' overall lower accuracy scores could be explained by these smaller differences in ED between the first and second acoustically closest vowels.

In order to test whether multiple category assimilation is a factor contributing to the overall lower discrimination accuracy by AusE listeners, we used general linear mixed modelling (run in R version 3.1.1) with the difference in ED between the vowel category of the X stimulus (the BP vowel category in that contrast) and the first and second closest native vowel (referred to as ΔED) as a predictor. We thus fit a binomial mixed model to our accuracy data using the *glmer* function (binomial family). We predicted that the smaller the ΔED (which results in multiple category assimilation), the lower the discrimination accuracy for that particular trial. ΔED was included as a fixed effect and participant and speaker as random effects (both slopes and intercepts). The model confirmed that ΔED predicted discrimination accuracy ($\beta = 0.4395$, $SE = 0.1989$, $z = 2.210$, $p = 0.0271$), with the positive β coefficient indicating that the larger the ΔED, the higher the accuracy. We therefore conclude that ΔED can account for the overall lower performance by the AusE participants, as a smaller ΔED is representative of multiple category assimilation and the resulting lower discrimination accuracy[5]. However, this is only for contrasts that result in complete or partial

[5] A second model including both ΔED and Contrasts showed that these two factors are even better predictors of discrimination accuracy, suggesting that ΔED can also explain the difference in difficulty across the different BP vowel contrasts.

neutralisation in non-native perception (e.g. for /i/-/e/ and /o/-/u/), whereas for contrasts involving multiple category assimilation, but no neutralisation (e.g. /a/-/ɛ/), no difficulty in discrimination is found for either one of the two group of listeners.

In sum, the present study demonstrates that vowel inventory size (even when acoustic similarity is assumed) may not be sufficient for accurately predicting L2 discrimination difficulty unless detailed acoustic comparisons (e.g., EDs) are made as these comparisons yield more successful predictions (as previously shown in Escudero & Chládková, 2010; Escudero et al., 2012, 2014; Escudero & Vasiliev, 2011; Escudero & Williams, 2011, 2012). Despite differences in vowel inventory size, both groups found the same BP contrasts equally difficult or easy to discriminate, this is because cases of single category assimilation (i.e., L2LP's NEW scenario) were identified for the same contrasts where multiple category assimilation (L2LP's SUBSET scenario) was identified for AusE listeners. Both of these scenarios were predicted to be difficult because they exhibit a case of acoustic overlap (i.e., the two BP vowels in the contrast were mapped acoustically close to the same native vowel(s). This highlights the importance of identifying the specific learning scenario a learner may face in L2 vowel acquisition and this can done by considering both the size of the native vowel inventory in relation to the target language, as well as the acoustic similarity between the native and target language. Thus, ongoing research aims at demonstrating whether acoustic properties, vowel inventory or a combination of both explains different levels of discrimination for BP vowel contrasts across listeners from different English dialects. Furthermore, as the present study is only applicable to vowels, future research is necessary for testing whether the L2LP acoustic hypothesis could also be applied to predicting difficulty for L2 consonants.

2.5. Chapter Summary

In this chapter, a preliminary investigation (Elvin et al., 2014) of Australian English and Iberian Spanish listeners' discrimination of Brazilian Portuguese vowels was presented. The results of the study indicate that the findings were more in line with predictions based on acoustic similarity between Australian English, Iberian Spanish and Brazilian Portuguese. Despite their differing vowel inventory sizes, Australian English and Iberian Spanish listeners found the same Brazilian Portuguese contrasts equally easy or difficult to discriminate. Furthermore, in line with acoustic predictions, the Iberian Spanish listeners had higher accuracy overall. The overall lower accuracy for Australian English listeners was attributed to multiple category assimilation (or the SUBSET scenario) which contributed to an acoustic overlap in some Brazilian Portuguese contrasts. That is, the two vowels in a given contrast were acoustically similar to the same native categories.

Although the findings for discrimination were in line with predictions based on acoustic similarity between the native vowel inventories, it is important to note that the predictions for Australian English could be improved. Recall that the acoustic comparison for Australian English was based on data reported in Cox (2006). This corpus includes vowel productions by adolescent speakers recorded in the 1990's using a different methodology to those reported for Spanish and Portuguese. For more accurate predictions, a thorough acoustic analysis of Australian English vowels that analyses production data from similar populations intended for testing and using a methodology that could identify the dynamic properties of Australian English vowels. For this reason, Chapter 3 reviews the two comprehensive acoustic analyses for Iberian Spanish and Brazilian Portuguese. It then introduces a new study (Elvin et al., 2016), which presents a comprehensive acoustic analysis of Australian English that could be beneficial for future cross-dialectal, cross-linguistic and L2 studies.

Chapter 3

Comprehensive acoustic descriptions of Brazilian Portuguese, Iberian Spanish and Australian English vowels.

3.1. Introduction

In the previous chapter, a preliminary investigation (Elvin et al. 2014) into the discrimination of Brazilian Portuguese vowels by Australian English and Iberian Spanish listeners was introduced. In this study, the predictive power of vowel inventory size versus vowel acoustics was examined with the findings appearing to be in line with those predictions made by an investigation of the acoustic similarity between the native and target language(s). According to the L2LP model, in order to predict a learner's initial state, the optimal perception of the listener must first be established. That is, detailed acoustic analyses of the native and target languages are required in order to determine cross-language acoustic similarity. The acoustic predictions in Elvin et al. (2014) were based on the measurements of Euclidean Distances between F1 and F2 native and target vowels based on reported averages in Escudero et al. (2009) for Brazilian Portuguese, Chládková et al. (2011) for Iberian Spanish and Cox (2006) for Australian English. Although Escudero et al. (2009) and Chládková et al. (2011) are cross-linguistically comparable because they used the same method of data collection and analysis, Cox (2006) examined vowel productions by adolescent speakers, collected in the 1990's, with different procedures of data collection and analyses. Although the findings reported in Elvin et al. (2014) predicted AusE listeners' discrimination difficulty successfully, acoustic predictions could be improved with an up-to-date study that is comparable cross-dialectally and importantly, cross-linguistically comparable to more recent studies. Importantly, the L2LP model claims that for the most accurate predictions of L2 difficulty, comparisons of acoustic similarity should ideally be based on the same listeners intended for testing, or at the very least, the same population.

In this chapter, a review of the studies used to acoustically compare Iberian Spanish and Brazilian Portuguese is presented. This will be followed by a new, comprehensive acoustic analysis of Australian English vowels. Given the aforementioned limitations of Cox (2006), we propose that this new acoustic description of Australian English vowels is more appropriate for future studies designed to predict L2 difficulty for adult speakers, as it ensures that method are the same across languages and dialects.

3.2. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese (Escudero et al., 2009)

Escudero et al. (2009) presents a cross-dialectal acoustic analysis of European Portuguese (EP) and Brazilian Portuguese (BP). Both dialects in this study share a vowel inventory that contains seven oral vowels, namely, /i, e, ɛ, a, ɔ, o, u /, which are produced in a stressed position. The authors explain that apart from the central low vowel /a/, the BP vowel inventory has an internal symmetry with three unrounded front vowels (i, e, ɛ) and three rounded back vowels (u, o, ɔ). These can be further divided into three pairings, specifically, high (i-u), higher-mid (e-o) and lower-mid (ɛ-ɔ).

3.2.1. Overview of Escudero et al. (2009)

This study examined four acoustic correlates of vowel identity, namely, the first and second formants, duration, as well as fundamental frequency, in BP and EP. One of the study's research aims was to determine whether or not Portuguese follows the cross-linguistic tendency of vowel intrinsic duration and the universal tendency of vowel intrinsic pitch. As previous studies on other Romance languages with a comparable vowel inventory e.g., French (Landick, 1995) and Catalan (Recasens & Espinosa, 2009) have found signs of lower-mid vowels merging with higher-mid vowels, the study aimed at investigating whether signs of a future merger were also observed for European and Brazilian Portuguese. Finally, the study aimed at looking at gender differences across the two dialects.

3.2.2. Methodology and main findings

Participants: Twenty male and twenty female speakers from both dialects were selected to participate in the study. They were university undergraduate students under the age of 30, selected from the largest metropolitan area in each country, specifically, Lisbon for EP and São Paulo for BP. All participants were reported to have lived in either Lisbon or São Paulo throughout their lives and did not speak any foreign language with a proficiency of 3 or more on a scale from 0 (" I don't understand a single word") to 7 ("I understand like a native speaker").

Data collection: Participants produced the target vowels, /i, e, ɛ, a, ɔ, o, u/, which were orthographically presented as *i, ê, é, a, ó, ô and u*. These vowels were embedded in disyllabic nonce words which were produced in isolation and also in a carrier phrase. Each target vowel was produced in a CVCV sequence where each consonant was an identical voiceless stop (produced in one of five consonantal contexts, p, t, k, f and s), the first vowel was the stressed target vowel and the final vowel was always the unstressed /e/ or /o/. An example of a trial was as follows: *Em pêpe e pêpo temos ê*, which is translated as *In pepe and pepo we have e*.

Data analysis: The start and end points of each vowel token were labelled manually in order to get duration measurements. Both points were selected at the zero crossing of the waveform. PRAAT (Boersma,& Weenink, 2015) was used to measure the F0 curves of all recordings using a cross-correlation method. The pitch range of the analysis was set to 60-400 Hz for males and 120-400 Hz for females. The authors further explain that to get a robust measure of the F0 of each vowel token, the median F0 value was taken of values measured in steps of 1 ms in the central 40% of the vowel. A new method of analysis, *the optimal ceiling method*, was used to investigate the first and second vowel formants. In this method, for every vowel, per speaker, the *optimal ceiling* was chosen as the one that yields the least amount of variation for both the first and second formant values within the set number of tokens for each target vowel. Thus, the formant ceilings ranged between 4500 and 6500 Hz for females and 4000 and 6000 Hz for males.

Main findings: In line with studies of other languages, the authors found that Portuguese vowels do indeed exhibit vowel intrinsic pitch and vowel intrinsic duration. They further found that the size of the F1 and F2 vowel space was larger for women than for men and that F0 and formant values were higher for females. The study confirmed the symmetric nature of the Portuguese vowel inventory, however, back vowels were found to have slightly higher F1 vowels than front vowels. Interestingly, the results indicated that Portuguese speakers seem to have turned vowel duration into a cue for vowel identity to an extent that it goes beyond the automatic lengthening of open vowels. They further found that BP vowels were on average longer than EP vowels and that the vowel intrinsic pitch effect was greater in BP than it was in EP. They also found that the lower-mid vowel /ɛ/ was higher in EP than it was BP and that it was in fact closer to EP /e/ than what was observed in BP. The authors suggest that this might mean that a future merger is occurring, however a larger investigation was required to determine if that was indeed the case.

Importantly, the differences identified between the two dialects would need to be considered when investigating L2 perception to ensure that predictions are applicable to the populations tested. That is, if a group of listeners consist of European and Brazilian speakers, predictions should ideally be based on their own native dialect and not Portuguese in general. Furthermore, the Escudero et al. (2009) study provides published average values of vowel duration, F0, F1, F2, F3 and formant ceilings for the seven EP and BP vowels which can and have since been used in cross-linguistic studies. In particular, the published formant and duration values have since been used in studies (Elvin, Escudero, & Vasiliev, 2014; Elvin & Escudero, 2014; Vasiliev, 2013) to test whether acoustic similarity predicts L2 perceptual difficulties. The optimal ceiling method also presented in this study has since been used in other cross-dialectal studies (e.g., Chládková, Escudero, & Boersma, 2011) thus ensuring the cross-linguistic comparability of these studies.

3.3. Context-specific acoustic differences between Peruvian and Iberian Spanish vowels (Chládková, Escudero & Boersma, 2011)

Chládková et al. (2011) present a cross-dialectal acoustic analysis of Peruvian Spanish (PS) and Iberian Spanish (IS). Both dialects in this study share a vowel inventory that contains five vowels, namely, /i, e, a, o, u /. The inventory can be divided into two high vowels, /i/ and /u/, two mid vowels /e/ and /o/ and a single low /a/ vowel.

3.3.1. Overview of Chládková et al. (2011)

This study presents a thorough acoustic analysis of Iberian (IS) and Peruvian Spanish (PS), similar to the acoustic analysis presented in Escudero et al. (2009). Similar to Escudero et al. (2009) the study investigates whether or not the universal tendency for vowel intrinsic duration and vowel intrinsic pitch are also applicable to IS and PS. It further investigates whether there are between-dialect differences in the strength of this effect. In addition, the study considers how consonantal and phrasal context affects vowels in these two Spanish dialects. The study is cross-linguistically comparable to Escudero et al. (2009) in the sense that its vowels are produced in the same consonantal contexts with the same data collection and analysis procedures.

3.3.2. Methodology and main findings

Participants: In line with Escudero et al. (2009) participants were 20 IS and 20 PS university students aged between 19 and 28, from Madrid and Lima. To control for an effect of gender within and between dialects there was an equal number of male and female speakers in each dialect, however the authors acknowledge that the data of one PS female participant could not be included due to issues during recording. All speakers were monolingual speakers of Spanish with no knowledge of any language other than Spanish and English and were raised by monolingual parents. Although the participants did report some knowledge of English, on a scale of 0-7 with 7 being highly proficient all rated themselves as having a self-estimated proficiency below 2.

Data collection: Following the methodology of Escudero et al (2009) participants read aloud nonce words containing one of the five target monophthongs (/i, e, a, o, u/) which were orthographically present on the screen. The nonce words were of a CVCV format where C was one of five consonantal contexts, namely /pVpV/, /tVtV/, /kVkV/, /fVfV/ and /sVsV/. The first vowel was always one of the target vowels and because Spanish does not have a schwa vowel, the second vowel was always /e/ or /o/. An example of a trial would be: *Pipe. En pipe y pipo tenemos i*; which is translated as, *Pipe. In pipe and pipo we have i*.

Data analysis: As with Escudero et al. (2009), duration measurements for each vowel were gathered by manually labelling the start and end point of each vowel token and were always labelled at the zero crossing for each point. The analysis of F0 and the first and second formants were also the same as Escudero et al. (2009). In particular, the *optimal ceiling method* was used as this method of formant analysis has been shown to minimise between-speaker variation which results in a more reliable measurement of vowel formants.

Main findings: The results from the acoustic analysis indicated that in line with other languages, Spanish also has several near-universal phenomena such as vowel intrinsic duration and F0. High vowels had shorter durations than lower vowels which corresponded to previous findings for Portuguese (Escudero et al., 2009), as both languages have no phonological length contrast. Also in line with the findings for Portuguese (Escudero et al., 2009) was the fact that higher vowels also exhibit a higher F0 than lower vowels. When comparing vowel productions across genders, the results indicated that women have higher F0, F1 and F2 than men and that their measured vowels were longer than males. The authors further found that vowels produced in isolation were longer than those produced in a sentence context which corresponded to stressed-timed languages even though Spanish is typically described as syllable-timed[6]. In regards to the Spanish-specific findings the study found that the locations of the five vowels in the F1-F2 vowel space are different to other languages with five monophthongs. The authors also found that voiceless fricatives /f/ and /s/ caused longer vowels in women than in men even after normalisation. In terms of dialect specific findings, the authors found that sentences were longer in PS than IS, the Spanish vowel /a/ had a higher F1 in IS than in PS and that the two mid-vowels /e/ and /o/ have more peripheral F2 values in PS than in IS. The authors concluded that their findings may have a number of implications for future cross-dialectal cross-linguistic and second-language research with participants from different Spanish speaking countries. For example, given the differences in vowel production across the dialects (e.g., the lower F1 value for /a/ in PS than in IS), future studies should control for dialectal variation rather than pooling all subjects together. The published duration and formant values in this study are beneficial for studies investigating acoustic similarity as a predictor of L2 difficulty (e.g., Escudero & Williams, 2011; 2012).

[6]Although Spanish has often been referred to as a syllable-timed language (e.g., Hockett, 1955), it is important to note that research (e.g., Grabe & Low, 2002; White & Mattys, 2007) has shown that rhythmical classification is in fact more complex than the stress-timed vs. syllable-timed distinction referred to here in the summary of Chládková et al., (2011).

3.4. Toward a comprehensive acoustic analysis of Australian English vowels

The two studies reviewed above are comprehensive acoustic analyses of Spanish and Portuguese which have since been used by studies in L2 perception to measure acoustic similarity between native and non-native languages. The formant values from these two studies were particularly useful in Elvin et al. (2014) to determine the acoustic similarity between Iberian Spanish and Brazilian Portuguese. The investigation of acoustic similarity successfully predicted IS listeners' discrimination difficulty. The corpus that was used in Elvin et al. (2014) to investigate acoustic similarity between Australian English and Brazilian Portuguese was based on data collected from adolescent speakers produced in the hVd context in the 1990's. While this analysis did predict Australian English listeners' discrimination difficulties, for the most accurate predictions of acoustic similarity, the acoustic comparisons should ideally be based on data collected from a similar population intended for testing, and in the case of Australian English, using a methodology that appropriately captures the dynamic properties of Australian English vowels.

Many studies, including Escudero et al. (2009) and Chládková et al. (2011) characterise monophthongs by their relatively "steady-state" central portions of the vowel targets, with formant measurements obtained at a single time-point, usually the mid-point of the vowel (Williams & Escudero, 2014). However, some acoustic analyses of English monophthongs and diphthongs have found that an investigation of dynamic properties or vowel inherent spectral change, which is typically used to characterise diphthongs, may also be relevant to monophthongs. For example, Williams & Escudero (2014) conducted a cross-dialectal study of Northern and Southern British English vowels. Using a similar systematic approach to Escudero et al. (2009) and Chládková et al. (2011) in the fact that the vowel production data was produced by adult university students or recent graduates aged between 18 and 30. The study can be considered cross-linguistically comparable to Portuguese (Escudero et al., 2009) and Spanish (Chládková et al., 2011) as it uses the same method of data collection with vowels produced in a variety of consonantal contexts. However, the study differs from the above two acoustic studies as it considers the vowel inherent spectral change in monophthongs and diphthongs. Interestingly, the study found that an analysis of vowel inherent spectral change was an important feature in differentiating the two regional dialects of British English.

Importantly, studies of Australian English (e.g., Harrington & Cassidy, 1994; Watson & Harrington, 1999) have also found that this dynamic approach to monophthongs is particularly useful in classifying Australian English vowels. Below a new acoustic analysis of Australian English vowels is presented (Elvin et al., 2016). The study is cross-dialectally comparable to Williams & Escudero (2014) as it follows the exact same methodology, with one additional consonantal context (namely, hVd). It is also cross-linguistically comparable to Brazilian Portuguese and Iberian Spanish as it investigates participants with similar characteristics as well as the same procedure of data collection. Finally, Elvin, Williams and Escudero (under review) provides a relatively concise, yet comprehensive and up-to-date account of Australian English vowels, which can serve as a reference point for future studies, such as those using acoustic comparisons to predict non-native and L2 difficulty (perception, word recognition and production) for AusE learners of different L2s or speakers of other languages acquiring AusE.

3.5. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English (Elvin, Williams & Escudero, 2016)[7]

3.5.1. Introduction

Interest in investigating Australian English (AusE) has steadily increased since its recognition as a variety of English separate from British English. Mitchell (1946) provided one of the earliest accounts of AusE which acknowledges this distinction by utilising phonetic symbols for phonemic vowel labels that reflect more typical AusE realisations than those conventionally used for the British standard. With respect to the AusE vowel system, there are a handful of detailed acoustic descriptions. One of the most notable is that of Harrington, Cox & Evans (1997) whose analysis of AusE vowels from the Australian National Database of Spoken Language resulted in a modification of Mitchell's (1946) use of phonetic symbols in order to more accurately represent individual AusE vowel categories. Cox (2006) provided a more recent acoustic description encompassing data on 18 AusE stressed vowels in the /hVd/ environment produced by 60 male and 60 female adolescents from the Northern Beaches in Sydney. Importantly, this study is one of the first to include vowel productions by both male and female AusE speakers, as past studies mainly investigated only male speakers, and given its comprehensive scope, it has since become a point of reference for acoustic information on AusE vowel production.

Unlike the study of other varieties of English where the mainstay of phonetic variation research has been on regional accents, variation in AusE has conventionally been discussed with reference to a socio-stylistic continuum consisting of Broad, General and Cultivated accents. However, in the face of phonetic change over time, studies such as

[7] A version of this chapter was published in JASA.

Harrington et al. (1997) have since found a considerable amount of phonetic overlap among the three registers. Such findings have given rise to a shift in attention to more general phonetic variation in AusE, especially regional variation, and a small body of literature has since emerged (e.g., Billington, 2011; Bradley, 2008; Cox & Palethorpe, 1998, 2003). These studies have found that regional variation in AusE does exist, however, it may manifest itself more subtly than in other varieties of English, e.g., North American (Clopper & Pisoni, 2006) or British English (Williams & Escudero, 2014). For example, regional differences have been found in speakers from New South Wales and Victoria in the acoustic realizations of /u/ and /æ/ and particularly in the merger of /e/ and /æ/ pre-laterally (Cox & Palethorpe, 2003). Billington (2011) found some evidence of regional variation among speakers from Sydney, Adelaide and Melbourne, with the main differences found in the extremities of the vowel space. The findings from the study suggest that Melbourne vowels seem to be lower than Sydney and Adelaide vowels, with a lower and more retracted /æ/ vowel being one feature that distinguished female Melbourne speakers from Sydney and Adelaide speakers. Cox & Palethorpe (1998) also found differences in the realisation of AusE vowels by female adolescents from Western Sydney, Northern Sydney and the Northern Beaches. Interestingly, the study found that Western Sydney speakers displayed a spatial vowel arrangement that differed to speakers from the other two regions of Sydney. For example, /e/, /æ/, /ɔ/ and /ɜː/ are more phonetically raised for Western Sydney speakers with the /ɔ/ and /ʊ/ vowels being more retracted. These findings have important implications for research relating to regional variation as it suggests that phonetic variation in AusE occurs not only across states and territories but also within a single urban centre.

Among the many applications of acoustic-phonetic descriptions and corpora like those mentioned above is to test theoretical claims about the predictability of second-language (L2) learners' speech perception and production. The Second Language Linguistic Perception model (L2LP; see the model's latest version and reference to previous versions in van Leussen & Escudero, 2015), for instance, claims that individuals' non-native and L2 difficulties can be predicted from detailed acoustic comparisons between their native language (L1) and the non-native target language because patterns of acoustic (dis)similarity have repeatedly been shown to approximate patterns of perceptual (dis)similarity between the two languages (e.g., Escudero & Williams, 2011; Gilichinskaya & Strange, 2010).

Predictions on non-native and L2 speech perception and production based on acoustic comparisons are optimised when the specific native and target varieties (i.e., dialects) of the individuals involved are represented. With respect to AusE individuals' vowel perception, Elvin & Escudero (2015) collected a small corpus of speech production data on AusE by speakers from Western Sydney to test whether acoustic comparisons can predict AusE listeners' perception of non-native Brazilian Portuguese (BP) vowels. In line with L2LP's claim, acoustic (dis)similarity between native AusE and non-native BP vowels was shown to be a good predictor of AusE listeners' BP vowel discrimination (Elvin et al., 2014).

The present study reports a set of newly collected acoustic data (duration and formants) of Western Sydney Australian English (AusE) vowels. Given the fact that regional variation does exist in AusE, we selected a single variety of Australian English (namely, AusE). This will allow us to provide a relatively concise yet comprehensive and up-to-date account which can serve as a reference point for future studies, such as acoustic comparisons for predicting non-native and L2 difficulty (perception, word recognition and production) for AusE learners of different L2s or speakers of other languages acquiring AusE. Vowel tokens representing the 18 AusE categories /iː, ɪ, e, eː, ɜː, ɐ, ɐː, æ oː, ɔ, ʊ, ʉː, ɑe, æɪ, æɔ, oɪ, ʊə, əʉ, ɪə/[8] were collected in a wider range of consonantal frames and phrasal positions than previous studies on AusE (e.g., vowel tokens came from only isolated /hVd/ words in Cox, 2006, and

[8] /ʊə/ is sometimes reported for AusE as a phonological category, though it is seldom used and /oː/ is more common (Cox, 2006); therefore it does not feature in the present study.

Watson & Harrington, 1999) in order to convey both between- and within-speaker phonetic variability.

Conventionally, frequencies of vowel formants may be sampled in different ways: at a single time point for monophthongs and at multiple time points for diphthongs to examine spectral changes over time or the vowel inherent spectral change (VISC). In the present study, we adopt the latter approach for both nominal monophthongs and diphthongs because previous research has reported that this is especially relevant for English as compared to other languages, such as Dutch which has a similarly large vowel inventory (Williams, van Leussen, & Escudero, 2015). For instance, a number of studies have reported that incorporating measures of VISC enhances the acoustic classification of vowels (including nominal monophthongs) in varieties of North American English (Hillenbrand, Getty, Clark, & Wheeler, 1995; Jacewicz & Fox, 2013; Zahorian & Jagharghi, 1993). Additionally, VISC has found to be an important feature in differentiating between regional dialects of North American English (e.g., Jacewicz & Fox, 2013) and British English (Williams & Escudero, 2014). The vowels of AusE have also been investigated with reference to VISC (e.g., Harrington & Cassidy, 1994; Watson & Harrington, 1999). In particular, Watson & Harrington (1999) found this approach to be useful for accurately classifying AusE nominal diphthongs as well as improving the within-class separation of the AusE tense monophthongs (i.e., /iː, ɜː, ɐː, oː, ʉː/) from their lax counterparts (i.e., /ɪ, e, ɐ, æ, ɔ, ʊ/). Furthermore, spectral change patterns may provide phonetic detail in AusE vowels that is relevant in second-language (L2) learning (Jin & Liu, 2013; Morrison, 2009) and, therefore, may prove beneficial for predicting L2 difficulty.

3.5.2. Method

3.5.2.1. Speakers

Participants were 19 (12 female) monolingual speakers of Australian English from Western Sydney who reported little to very basic knowledge of any foreign language. All participants were university students, with the majority recruited through Western Sydney University's psychology pool and aged between 18 and 30. Participants were included in the analysis if their parents were Australian English speakers and they were born and raised in greater Western Sydney (or moved to the region for they were 12 months old). We do acknowledge that three participants also reported living in the Lower Blue Mountains, a region adjacent to and often coupled with greater Western Sydney and two reported living in another state (for no more than 2 years). Data for eight of the participants were previously reported in Elvin & Escudero (2015).

3.5.2.2. Materials and recording procedure

Recordings took place in a sound-attenuated booth at Western Sydney University using a Shure SM10A-CN headset microphone and an Edirol Quad-Capture UA-55 sound card with a 44.1 kHz sampling rate. Participants were presented with the 18 target vowels /iː, ɪ, e, eː, ɜː, ɐ, ɐː, æ, oː, ɔ, ʊ, ʉː, ɑe, æɪ, æɔ, oɪ, əʉ, ɪə/ on a computer screen in isolated words and sentences and were instructed to read the words and sentences aloud at a normal conversational rate. The words containing each target vowel had one of six consonantal contexts, namely, /bVp/, /dVt/, /fVf/, /gVk/, /hVd/ and /sVs/. The inclusion of the /hVd/ context was motivated by the fact that it is the common phonetic context studied in previous studies on AusE. The remaining five consonantal contexts matched those used in Williams & Escudero's (2014) study on British English vowels, which were selected to represent a mix of fricative and stop consonant environments, in order to allow for future cross-dialectal comparisons. As English is not orthographically transparent, participants first read an example real word, followed by the target vowel in isolation. They then read each target vowel in isolated /CVC/ and /CVCə/ words in one of the aforementioned consonantal contexts. These words were then read in a carrier sentence. Thus, a trial would consist of the following format: "*See – ee – deet – deeta - In deet and deeta we have ee*" to elicit tokens of /iː/ in a /dVt/ frame. As each of the 18 target vowels was produced twice (both in an isolated word and a carrier sentence) in six different consonantal contexts, there were 12 tokens per vowel for each of the 19 speakers, yielding a total of 4,104 vowel tokens (vowels produced in the /CVCə/ frame were not included in the present analysis).

3.5.2.3. Acoustic analysis

The start and end boundaries for each vowel token were first determined automatically in *WebMaus* (Kisler, Schiel, & Sloetjes, 2012) and then hand-corrected to correspond to the first and last positive zero crossings of the quasi-periodic waveform associated with the vowel. Vowel duration was measured as the time (ms) between these start and end boundaries. Following Williams & Escudero (2014), first, second and third formant frequency estimates (F1, F2, F3) were obtained at 30 equally-spaced time points from the central (20% to 80%) portion of each vowel token in *Praat* (Boersma & Weenink, 2015).

In order to characterise F1, F2 and F3 trajectories, each set of 30 formant values was transformed using the discrete cosine transform (DCT). This procedure produces a series of values that are amplitudes of increasing frequency cosines which make up the original sequence of values; in this way, each coefficient characterises an aspect of a formant trajectory's shape based on a cosine. Specifically, the zeroth coefficient corresponds to a straight line proportionate to the formant trajectory's mean, the first coefficient to half a cosine, equivalent to the magnitude and direction of change from the mean (i.e., half a "U" shape which indicates trajectory slope) and the second coefficient to a full cosine (i.e., a full

"U" shape which indicates trajectory curvature); higher-order coefficients represent the amplitudes of higher frequency cosines and therefore correspond to increasingly fine-grained information of a trajectory's shape. Typically, the first and second DCT coefficients are sufficient for characterising formant trajectory shape in vowels (Morrison, 2013).

3.5.3. Results

Table 3.1 presents average durations and formant frequencies (Hz) of the AusE monophthongs and diphthongs produced by Western Sydney speakers.

Table 3.1 Duration (ms) and formant values (Hz) averaged across sentence and isolated positions... The formant values T1 and T2 refer to means of measurements from the third (24.1% duration) and 28th time points (75.9% duration), respectively. We also include means of median values across the 30 time points to provide a single time point measurement as is commonly reported for monophthongs.

							Monophthongs											**diphthongs**			
		iː	**ɪ**	**e**	**eː**	**ɜː**	**ɪə**	**æ**	**ɐ**	**ɐː**	**ɔ**	**oː**	**ʊ**	**ʉː**			**oɪ**	**æɪ**	**əʉ,**	**æɔ**	**ɑe**
M	**dur**	168	101	117	204	195	205	150	125	217	129	200	108	148	**M**	**dur**	176	178	165	203	1967
	F1^{Ave}	358	378	575	553	537	369	776	688	711	601	484	413	342		**F1^{T1}**	487	661	588	644	663
	F1^{T1}	407	365	547	527	513	348	731	672	705	590	480	423	363		**F1^{T2}**	412	432	454	692	676
	F1^{T2}	323	386	574	568	540	405	773	678	710	601	492	401	333							
	F2^{Ave}	2202	2079	1820	1915	1475	2147	1609	1226	1173	994	798	991	1736		**F2^{T1}**	960	1701	1359	1734	1095
	F2^{T1}	2051	2094	1842	1939	1501	2165	1619	1259	1205	1009	814	1069	1711		**F2^{T2}**	1870	2154	1577	1212	1578
	F2^{T2}	2280	2049	1814	1871	1473	2069	1629	1258	1184	1057	859	981	1749							
	F3^{Ave}	2869	2732	2642	2684	2467	2792	2519	2552	2627	2564	2635	2478	2214		**F3^{T1}**	2517	2506	2326	2519	2552
	F3^{T1}	2774	2759	2644	2674	2471	2845	2511	2538	2571	2553	2591	2479	2256		**F3^{T2}**	2457	2803	2220	2532	2468
	F3^{T2}	2907	2702	2633	2680	2476	2714	2565	2585	2641	2567	2638	2461	2205							
F	**dur**	174	112	129	212	205	210	148	132	216	130	202	124	166	**F**	**dur**	187	182	175	215	207
	F1^{Ave}	454	469	744	711	715	484	972	879	916	742	608	520	445		**F1^{T1}**	600	769	713	875	939
	F1^{T1}	484	457	699	672	680	460	916	852	903	715	582	511	453		**F1^{T2}**	499	514	554	876	836
	F1^{T2}	404	472	739	726	717	535	940	846	898	728	637	507	420							
	F2^{Ave}	2580	2529	2048	2141	1844	2489	1768	1522	1422	1272	1033	1192	2130		**F2^{T1}**	1258	2114	1692	1923	1510
	F2^{T1}	2503	2523	2077	2158	1862	2523	1836	1550	1462	1280	1057	1210	2101		**F2^{T2}**	2383	2421	2031	1432	1952
	F2^{T2}	2548	2431	1976	2063	1834	2366	1742	1546	1450	1303	1134	1249	2158							
	F3^{Ave}	3022	3008	2813	2813	2879	2982	2581	2877	2829	2980	3073	2930	2636		**F3^{T1}**	2902	2784	2661	2723	2777
	F3^{T1}	2959	3022	2824	2842	2866	3036	2589	2882	2843	2964	3020	2894	2661		**F3^{T2}**	2937	2975	2639	2883	2780
	F3^{T2}	3075	2975	2813	2818	2884	2925	2643	2931	2873	3009	3009	2955	2654							

3.5.3.1. Vowel duration and formant values

Figure 3.1 shows a clear length distinction in the AusE vowel system, often corresponding to tense and lax vowel pairs (e.g., /iː-ɪ/, /ʉː-ʊ/, /ɐː- ɐ/ and /oː-ɔ/). AusE follows the cross-linguistic trend for some open vowels (e.g., /ɐː/) to exhibit longer durations (Chládková, et al., 2011), as confirmed by relatively weak but significant Pearson's correlations between duration and F1 0^{th} DCT coefficients (i.e., mean F1 trajectory frequencies) on all vowel tokens (ρ = 0.164, p < 0.001 for males and ρ = 0.158, p < 0.001 for females).

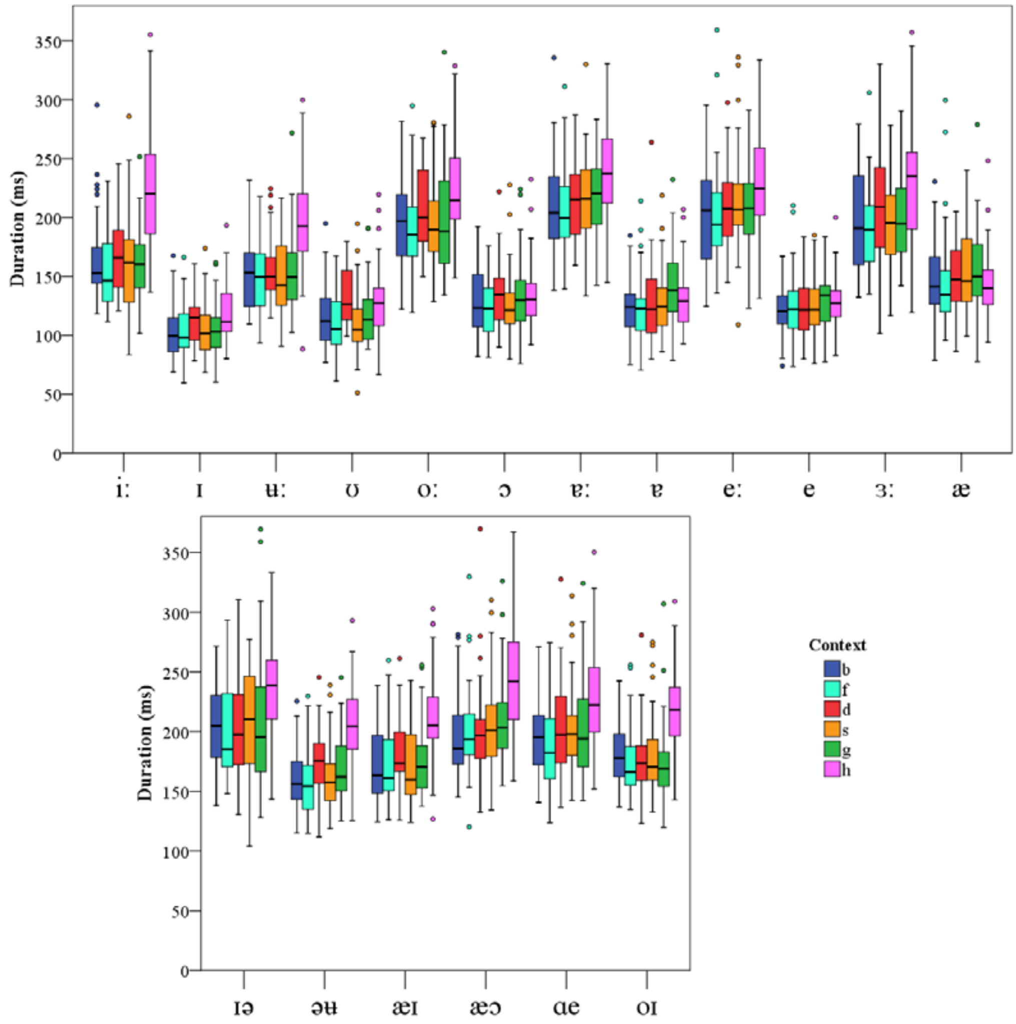


Figure 3.1 Average vowel duration (ms) per AusE vowel (isolated and sentences positions combined) in six consonantal contexts. Monophthongs are presented in the upper panel and diphthongs in the lower panel.

Figure 3.2 plots average formant (F1 and F2) trajectories for the 18 vowels in the six consonantal contexts in F1 × F2 vowel spaces produced by male (left) and female (right) speakers. Visual inspection suggests that our data is generally in line with previous studies (e.g., Watson & Harrington, 1999) whereby AusE nominal monophthongs (uppermost panels) display less spectral change than nominal diphthongs (lower panels). With respect to AusE specifically, Cox & Palethorpe (1998) found adolescent and young adult females produced some notable variants of AusE vowels. While the present results do not address differences within the Sydney city region, it appears present-day young adult AusE speakers are participating in sound change occurring in AusE, such as /e/-lowering (Cox & Palethorpe, 2008). For example, Cox & Palethorpe (1998) report midpoint F1 values in the range of approximately 400-600 Hz for AusE /e/, whereas our female AusE speakers produced this vowel with at least approximately 700 Hz.

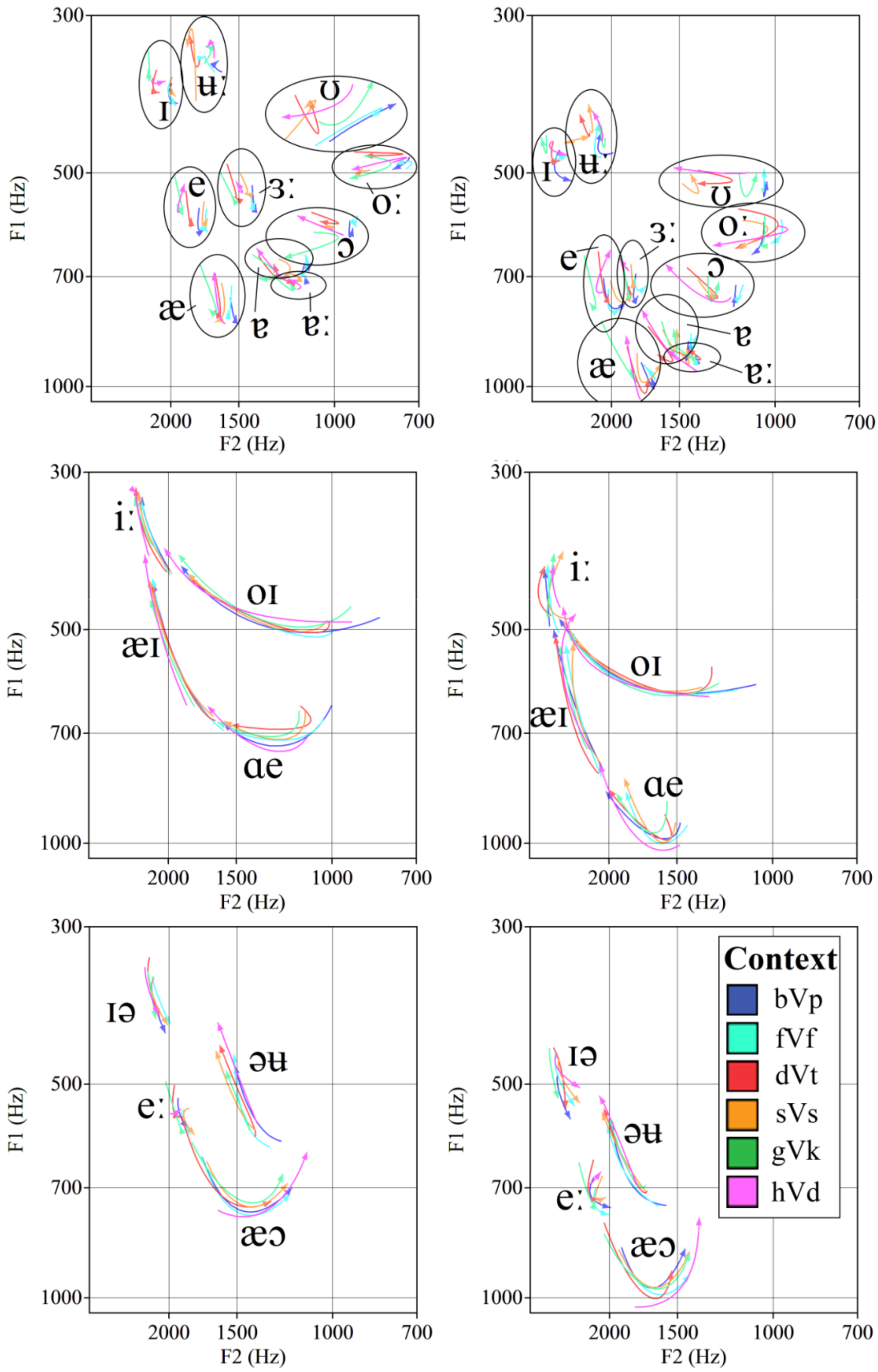


Figure 3.2 F1 and F2 values (Hz) from the 20 time points per AusE vowel in six consonantal contexts averaged over productions in the isolated and sentence contexts by 7 male (left) and 12 female speakers (right). The 18 AusE vowels are spread over F1 × F2 vowel plots on three rows and the vowels in the uppermost panel are circled for clarity.

In order to investigate which acoustic parameters best represent the 18 vowels, we conducted discriminant analyses (DAs) in which different combinations of acoustic measures were used to obtain discriminant functions and centres of gravity to classify the tokens into the 18 discrete vowel categories. As a control measure, tokens were separated according to speaker gender, position (isolation or sentence) and consonantal context. DAs with 0^{th} DCT coefficients (for F1, F2, F3) correctly classified only 60.1% of vowel tokens and adding duration values improved this by 14.8% (74.9% correct). Also including 1^{st} DCT coefficients led to a 16.3% improvement (91.2% correct), whereas adding 2^{nd} DCT coefficients increased accuracy by just 1.6% (92.7% correct). Thus, in addition to formant trajectory means, duration and magnitude and direction of formant trajectory slope are essential acoustic parameters for representing the 18 AusE vowels, whereas formant curvature may not be necessary.

As has been reported previously for AusE (e.g., Harrington et al., 1997; Cox, 2006), Figure 3.2 (lowermost panels) shows AusE /ɪə/ has much less spectral change than other nominal diphthongs. This raises the question as to how acoustically distinguishable it is from the spectrally close /iː/ and /ɪ/ vowels (Table 3.1 and Figure 3.2) that are contrasted by duration and spectral change. To assess the acoustic separability of /iː, ɪ, ɪə/, DAs were run as before but on these tokens only. Table 2 shows that duration can distinguish the vowel, /ɪ/, relatively well and 1st DCT coefficients can distinguish the vowels /iː/ and /ɪə/ relatively well. Both of these parameters together result in considerable improvement and are better than either 0th DCT coefficients or 0th DCT coefficients and duration. While using all parameters (bottom row in Table 3.2) shows the best performance, the pattern of other results indicates formant trajectory slope and duration are crucial. Thus, even if /ɪə/ is becoming more monophthongal, its duration and spectral change are still distinct from the spectrally close /iː-ɪ/ contrast.

Table 3.2 Percentage correct classification of /iː, ɪ, ɪə/ from DAs run on duration, 0th and 1st DCT coefficient values (percentages for vowels produced in a sentence and isolated position are combined in this table).

Acoustic measure	**iː**	**ɪ**	**ɪə**	**Mean**
Duration	54.1	93.3	63.4	70.2
0^{th} DCT	63.7	51.3	43.6	52.9
1^{st} DCT	89.6	72.8	70.0	77.5
Duration, 0^{th} DCT	68.2	93.0	68.8	76.7
Duration, 1^{st} DCT	93.9	94.9	90.6	93.1
Duration, 0^{th} DCT, 1^{st} DCT	95.9	97.6	92.6	95.4

3.5.3.2. Consonantal context

Variation in duration and formant trajectories according to consonantal context can be observed in Figures 1 and 2, respectively. Most striking are the longer durations for vowels in the /hVd/ context, which is unsurprising given that there is a qualitative difference of the voiced /d/ coda and the voiceless codas of thee other consonantal contexts (House, 1961), as well as the differences in F2 trajectories across the consonantal contexts, because F2 often determines degree of coarticulation in CV sequences (Herrmann, Cunningham, & Whiteside, 2014). These differences in F2 trajectories in the /hVd/ context are predominantly found in the central and back vowels, i.e., / ɐ, ɐː, æ, oː, ɔ, ʊ/, where the paths of their formant trajectories proceed in the opposite direction to the same vowels produced in the other five consonantal contexts. However, for the front vowels, namely, /iː, ɪ, e, eː, ɜː, ɪə/, formant trajectories in the /hVd/ context more closely resemble those in the other consonantal contexts.

To test the generalisability of the acoustic values of the 18 vowels in a single consonantal context to different consonantal contexts, we conducted DAs (in the same manner as reported above) to compare how well vowel tokens from one consonantal context (one of /bVp/, /fVf/, /dVt/, /sVs/, /gVk/, /hVd/) can be classified in each of the other five contexts separately as well as the other five contexts combined. The results in Table 3.3 show that, unsurprisingly, tokens tested on the DA trained on tokens from the same context were most accurately classified. Notably, the DA trained on tokens from the /hVd/ context, commonly used in previous studies (e.g., Watson & Harrington, 1999; Cox, 2006), consistently produced the least accurate classifications when vowel tokens from the other five contexts were tested on it (bottom row in Table 3.3), suggesting /hVd/ vowels are least comparable to other contexts. The /gVk/-trained DA also classified vowel tokens relatively poorly.

Table 3.3 Percentage correct classification of vowel tokens trained on one consonantal context and tested with tokens from the other five the same context (bold) and tokens from the other five contexts. "All other" refers to vowel tokens from contexts combined test

Test / Training	/bVp/	/fVf/	/dVt/	/sVs/	/gVk/	/hVd/	All other
/bVp/	**92.4**	86.4	82.7	83.9	79.8	78.8	81.3
/fVf/	89.7	**93.0**	81.1	82.6	75.9	73.6	80.6
/dVt/	83.8	82.0	**91.1**	81.2	82.1	73.3	80.5
/sVs/	85.5	86.1	83.8	**90.0**	79.6	73.9	81.8
/gVk/	79.6	77.0	84.1	76.2	**89.1**	68.3	77.0
/hVd/	75.0	75.0	73.2	74.5	74.3	**92.1**	74.4

3.5.4. Discussion

The aim of the present study was to present a new acoustic analysis of Australian English monophthongs and diphthongs produced by speakers from Western Sydney that can also be used to predict L2 difficulty. Recall that the L2LP model suggests that for the most accurate predictions, acoustic analyses should match the intended target population; therefore the results from the present study would be appropriate for predicting L2 difficulty for Australian English learners from Western Sydney. Our findings are in line with previous studies with English vowels (e.g., Hillenbrand et al., 1995; Jacewicz & Fox, 2013; Watson & Harrington, 1999; Williams & Escudero, 2014; Zahorian & Jagharghi, 1993) that find evidence for the relevance of spectral change in analysing both nominal monophthongs and diphthongs. In accordance with Watson & Harrington (1999), nominal monophthongs displayed less spectral change than diphthongs. In particular, we find that, in addition to formant trajectory means, duration and magnitude and direction of formant trajectory slope are essential acoustic parameters for representing the 18 AusE vowels, whereas formant curvature may not be necessary. As reported previously for AusE (e.g., Harrington et al., 1997; Cox, 2006), our findings indicate that the nominal diphthong /ɪə/ exhibits much less spectral change than other nominal diphthongs. Interestingly, even if /ɪə/ is becoming more monophthongal, we find that both duration as well as formant trajectory slope help to distinguish it from the vowels in the spectrally close /iː-ɪ/ contrast.

As expected, vowel durations and formant trajectories are affected by the consonantal context in which they were produced. Most striking are the longer durations for vowels in the /hVd/ context, which is partly expected due to the voiced /d/ coda and differing coarticulatory effects. Overall, we found that the discriminant analysis trained on tokens produced in the /hVd/ context consistently yielded the least accurate vowel classifications when vowel tokens from the other contexts were tested on it. This may indicate that the phonological voicing status of the following coda may significantly alter both the duration and spectral properties of AusE vowels. Of course, further investigation with a greater number of voiced coda environments is required to determine the nature of such acoustic differences.

As previously mentioned, among the applications of acoustic-phonetic descriptions and vowel corpora like the one reported here is to test theoretical claims about L2 learners' speech perception and speech production. Recall that the L2LP model suggests that for the most accurate predictions, acoustic analyses should match the intended target population; therefore the results from the present study would be appropriate for predicting L2 difficulty for learners from Western Sydney (e.g., refining Elvin et al. 2015's predictions for AusE learners of Brazilian Portuguese).

Our findings show that, AusE vowel duration and formant trajectories differed across the six consonantal contexts. Notably, the /hVd/ context – which is used in the Cox (2006) and is also commonly used in acoustic analyses of English vowels due to its predicable coarticulatory effects (Cox & Palethorpe, 2004) – could be unsuitable in future cross-linguistic acoustic comparisons. Firstly, AusE /hVd/ vowels are acoustically least similar from those in the other consonantal contexts used in this study. This may be due to vowel lengthening triggered by the voiced coda /d/, which is a well-known phenomenon in English but may not occur in other languages. Secondly, /h/ is not a word-initial consonant in some languages, including Romance languages such as Spanish and Portuguese. Furthermore, the reported averages in Table 3.1 differ from those reported in Cox (2006), primarily because they are based on data averaged across all consonantal contexts, whereas those in Cox (2006) were based on vowels produced in the hVd context only. These factors considered together lend support to the L2LP model (van Leussen & Escudero, 2015) and previous studies (e.g., Levy & Law, 2010) that advocate acoustic predictions of L2 difficulty should be based on data from the same phonetic context.

Last, capturing aspects of VISC may influence L2 vowel acquisition as demonstrated in previous studies demonstrated in previous studies (e.g., Morrison, 2009; Jin & Liu, 2013). For example, Morrison (2009) investigated the effects of VISC on Mexican-Spanish listeners' acquisition of the English /i/-/ɪ/ contrast. The results indicated that at the initial stage L1 Mexican-Spanish listeners distinguish the English /i/-/ɪ/ contrast in a multidimensional manner, using all the available acoustic cues (or VISC), and not in a unidimensional way using duration cues only. Additionally, Jin & Lu (2013) investigated VISC in English vowels spoken by non-native speakers and found that the learners' native language influenced the degree of VISC in their non-native English productions. Therefore, capturing aspects of VISC may be a relevant factor for future studies on L2 acquisition, given its role in separating AusE vowels.

3.6. Chapter Summary

This chapter presented a review of the comprehensive acoustic analyses of Brazilian Portuguese and Iberian Spanish that were used to measure acoustic similarity in Chapter 2. For Australian English, the vowel acoustic description presented in Chapter 2 used the Cox (2006) corpus to measure acoustic similarity and predict non-native discrimination difficulty. However, this corpus was limited in the fact that it was based on vowel production data produced by adolescent speakers in the hVd context in the 1990's. Importantly, in the new study (Elvin et al., 2016) presented in Chapter 3, AusE vowel duration and formant trajectories differed across the six consonantal contexts with the hVd context being the most distinct. Thus, for cross-linguistic studies, such as those presented in this thesis, the Cox (2006), which is based on vowels produced in the hVd context is not comparable to the Spanish and Portuguese data reviewed in 3.2 and 3.3. Given the fact that the Spanish vowels in Chladkova et al. (2011) and Portuguese vowels in Escudero et al. (2009) were produced in contexts where the closing consonant is voiceless, the study introduced in this chapter would be more appropriate for cross-linguistic analyses as it investigates vowels produced in consonantal contexts which are comparable to Spanish and Portuguese.

In Chapter 4, Australian English and Iberian Spanish listeners' native vowel production and non-native categorisation of Brazilian Portuguese is investigated. Recall that the L2LP model was specifically designed to account for individual variation among non-native speakers at all stages of learning and across different learning abilities (i.e., perception, word recognition and production). As a result, L2LP predictions are made for individual learners based on detailed acoustic comparisons of their own L1 categories and the categories of the specific target language variety. Chapter 4 reports on a study (Elvin et al. under review a), which takes an approach that specifically investigates Australian English and Iberian Spanish listeners' perception of Brazilian Portuguese, with the native production data collected from the same listeners who participated in the study's non-native categorisation and discrimination tasks.

Chapter 4

The role of acoustic similarity and non-native categorisation in predicting non-native discrimination: Brazilian Portuguese vowels by English vs. Spanish listeners (Elvin, Escudero, Williams, Shaw & Best, under review a)[9]

4.1. Introduction

It is well known that learning to perceive and produce the sounds of a new language can be a difficult task for many second language (L2) learners. Models of speech perception such as Flege's Speech Learning Model (SLM; Flege, 1995), Best's Perceptual Assimilation Model (PAM, Best, 1994; 1995), its extension to L2 acquisition PAM-L2 (Best & Tyler, 2007) and the Second Language Linguistic Perception model (L2LP; Escudero, 2005; 2009; van Leussen & Escudero, 2015) claim that both the phonological and articulatory-phonetic (PAM, PAM-L2), or acoustic-phonetic similarity (SLM, L2LP) between the native and target language are predictive of L2 discrimination patterns. This suggests that discrimination difficulties are not uniform across groups of L2 learners, at least at the initial stage of learning, as a result of their differing native (L1) phonemic inventories.

Specifically, contrasts that are present in the native language inventory may be easier to discriminate in the L2 than contrasts that are not present. This has been demonstrated in Spanish listeners' difficulty to perceive and produce the English /i/-/ɪ/ contrast (Escudero & Boersma, 2004; Escudero, 2001; 2005; Flege, Bohn, & Jang, 1997; Morrison, 2009). This may be attributed to the fact that the Spanish vowel inventory does not contain /ɪ/ and

[9] A version of this chapter is currently under review.

Spanish listeners often perceive both sounds in the English contrast as one native sound category. When non-native sounds are perceived and/or categorised according to native categories, this is known as a "learning scenario" in the L2LP theoretical framework and as "perceptual assimilation patterns" in PAM and as "equivalence classification" in SLM. However, it is important to note that L2LP and PAM explore these learning scenarios or assimilation patterns by investigating L2 or non-native sound contrasts, SLM focuses on the similarity or dissimilarity between individual sound categories, rather than contrasts. Thus, the above example of Spanish listeners' difficulty with the English /i/-/ɪ/ contrast would be an example of the NEW[10] scenario in L2LP and single-category assimilation in PAM. Both models predict this type of learning scenario to be difficult as learners must either create a new L2 category or split an existing L1 category (van Leussen & Escudero, 2015).

In contrast, German learners, who have the same /i/-/ɪ/ contrast in their L1 vowel inventory, have fewer difficulties when perceiving the same English contrast than Spanish learners (Bohn & Flege, 1990; Flege et al., 1997; Iverson & Evans, 2007). It is likely that this is an example of the SIMILAR learning scenario in L2LP, and PAM's two-category assimilation, whereby the two non-native sounds in the contrast are mapped to two separate native vowel categories. Both PAM and L2LP would predict that a scenario of this type would be less problematic for listeners to discriminate as their existing L1 categories are replicated and adjusted so that their boundaries match those of the L2 contrast (van Leussen & Escudero, 2015). A third scenario, known as the SUBSET scenario (i.e., multiple category assimilation) in L2LP, occurs when the two sounds in the L2 contrast are perceived as split between two or more native L1 categories. This scenario may be comparable to focalised,

[10] The L2LP terms "NEW" and "SIMILAR" should not be confused with SLM's use of these terms. SLM posits that when listeners are presented with an L2 phoneme that does not closely resemble any L1 category they form a *new* category which should be easier to acquire than an L2 phoneme that is similar to an existing L1 category, which despite the phonetic differences, should be more difficult to acquire (Colantoni et al., 2015). In contrast, as mentioned above, a SIMILAR scenario in L2LP is predicted to be much easier to acquire than a NEW scenario.

clustered or dispersed uncategorised assimilation in PAM (Faris et al., 2016). While some studies suggest that this learning scenario is not problematic for L2 learners (e.g., Gordon, 2008; Morrison, 2009; 2003); other studies have shown the SUBSET scenario to lead to difficulties in discrimination (Escudero & Boersma, 2002), particularly when a perceptual or acoustic overlap occurs (Bohn et al., 2011; Elvin et al., 2014; Tyler, Best, Faber, & Levitt, 2014; Vasiliev, 2013).

As the above theoretical models claim, it is the similarity of L2 sounds to native categories that determines L2 discrimination accuracy. It could be the case that individuals whose L1 vowel inventory is larger and more complex than that of the L2 may be faced with relatively less difficulty discriminating L2 vowel contrasts simply because there are many native categories available onto which the L2 vowels can be mapped. Indeed, Iverson & Evans (2007; 2009) found that listeners with a larger vowel inventory (e.g., German & Norwegian) than the L2 were more accurate and had higher levels of improvement post training at perceiving L2 vowels (e.g., English) than those with a smaller vowel inventory (e.g., Spanish). However, other studies have found that having a larger native vowel inventory than the L2 does not always provide an advantage in L2 discrimination. For instance, recent studies have shown that Australian English listeners do not discriminate Brazilian Portuguese (Elvin et al., 2014) or Dutch (Alispahic, Escudero, & Mulak, 2014) vowels any more or less accurately than Spanish listeners, despite the fact that the Australian English vowel inventory is larger than Brazilian Portuguese and approximately similar in size to that of Dutch, while the Spanish vowel inventory is smaller than those of both Brazilian Portuguese and Dutch. In fact, the findings in Elvin et al. (2014) indicate that Australian English and Spanish listeners found the same Brazilian Portuguese contrasts perceptually easy or difficult to discriminate despite their differing vowel inventory sizes, and that overall, Spanish listeners had higher discrimination accuracy scores than English listeners.

Thus it seems that vowel inventory was a good predictor of L2 discrimination performance for some of the aforementioned studies, but not all, which may suggest that vowel inventory size alone is not sufficient for predicting L2 discrimination performance. After all, theoretical models such as L2LP and PAM claim the acoustic-phonetic or articulatory-phonetic similarity between the vowels in the native and target languages predict L2 discrimination performance rather than phonemic inventory. In fact, the L2LP model claims that any acoustic variation in native and target vowel production can influence speech perception (Williams & Escudero, 2014). Specifically, the model proposes that the listener's initial perception of the L2 vowels should closely match the acoustic properties of vowels as they are produced in the listener's first language (Escudero & Boersma, 2004; Escudero, 2005; Escudero, et al., 2014; Escudero & Williams, 2012). In this way, the L2LP model proposes that both L2 and non-native categorisation patterns and discrimination difficulties can be predicted through a detailed comparison of the acoustic similarity between the sounds of the native and target languages.

This L2LP hypothesis is supported by a number of studies which show that acoustic similarity successfully predicts non-native and L2 categorisation and/or discrimination e.g. (Elvin et al., 2014; Escudero & Chládková, 2010; Escudero, et al., 2014; Escudero & Vasiliev, 2011; Gilichinskaya & Strange, 2010; Williams & Escudero, 2014). For example, acoustic comparisons successfully predicted that Salento Italian and Peruvian Spanish listeners would categorise Standard Southern British English vowels differently, despite the fact that their vowel inventories contain the same five vowel categories, because the realisations of the five vowels are not identical across the two languages (Escudero, et al., 2014). Furthermore, as previously mentioned, Elvin et al. (2014) investigated Australian English and Iberian Spanish listeners' discrimination accuracy for Brazilian Portuguese vowels and found that a comparison of the type and number of vowels in native and non-native phonemic inventories was not sufficient for predicting L2 discrimination difficulties, and that accurate predictions can be achieved if acoustic similarity is considered. Specifically, the L2LP model posits that for the most accurate predictions, the acoustic data should be collected from the same group of listeners intended for testing. The present study investigates the non-native categorisation and discrimination of Brazilian Portuguese (BP) vowels by Australian English (AusE) and Iberian Spanish (IS) listeners. Similar to Elvin et al. (2014) these language groups were chosen on the basis of their differing inventory sizes. The AusE vowel inventory contains thirteen monophthongal vowels, namely, /iː, ɪ, ɪə , e, eː, ɜː, ɐ, ɐː, æ, oː, ɔ, ʊ and ʉː/, and is larger than BP, which has seven oral vowels, /i, e, ɛ, a, o, ɔ, u/, while IS has the smallest vowel inventory of the three languages containing five vowels, /i, e, a, o, u/.

In the aforementioned Elvin et al. (2014) study the authors investigated two predictors of non-native discrimination difficulty namely, vowel inventory size and acoustic similarity. On the basis of vowel inventory size, the authors predicted that AusE listeners whose vowel inventory is larger than IS should have had higher accuracy overall as a larger vowel inventory is thought to facilitate learning. On the other hand, a comparison of the acoustic similarity between the languages indicated that despite their differing inventory sizes, both Spanish and English are similar in regards to how their vowels acoustically compare to BP vowels. Consequently and following the L2LP model, the authors predicted that this would lead to both groups finding the same contrasts easy and/or difficult to discriminate. In addition, the authors found that IS vowels were acoustically closer to BP vowels with fewer competing categories and therefore predicted that IS listeners should have higher accuracy overall. The findings were in line with acoustic predictions revealing that both groups had the same pattern of discrimination difficulty and that IS listeners' performed better overall.

While most empirical research in L2 vowel perception including Elvin et al. (2014), investigate L2 development for groups of learners, the present study investigates non-native perception at both a group and individual level. Studies typically focus on learner groups rather than individuals because they are particularly interested in how populations behave. Despite the fact that many researchers are aware that some variability does exist among individuals (Smith & Hayes-Harb, 2011) the group data obtained is generally sufficient for their purposes of demonstrating differences between groups. Importantly, however, other studies (e.g., Díaz, Mitterer, Broersma, & Sebastian-Gallés, 2012; Smith & Hayes-Harb, 2011) have shown that an investigation of individual differences can be important for understanding L2 development. For example, Smith and Hayes-Harb (2011) warn that researchers need to be careful in drawing general conclusions about typical performance patterns for L2 listeners based on group averages, as individual data may be crucial to interpreting group results, especially given the large variety of situations that influence L2 learning by individual learners.

Most of the studies that investigate individual differences focus predominately on factors such as age of acquisition; length of residence; language use; motivation; etc. (Escudero & Boersma, 2004; Flege et al., 1995). In particular, much of the research conducted under the SLM theoretical framework (e.g., Flege et al., 1997; 1995) investigated the above extra-linguistic factors as a means of explaining the degree of foreign accent in an L2 learner. However, even when these factors are controlled, individual differences still seem to persist (Jin et al., 2014; Sebastián-Gallés & Díaz, 2012). For example, Díaz et al. (2012) found that as a group, L2 listeners were less accurate than native listeners across three phonological tasks (categorisation, identification and lexical decision), yet on an individual level, almost half of the L2 listeners' performed as well as native listeners in categorisation, with a small percentage also performing within a native range in the identification and lexical decision tasks. Their findings highlighted the importance of investigations of individual differences, as they found evidence of strong individual variability even when factors such as age of acquisition and language environment were controlled. They further found distinct proficiency levels to be a function of the phonological processes involved in the task. The authors suggested that their results imply that language teaching programs should adapt to the individual differences in students' learning pace and that successful phonetic training must cover a range of phonological processes.

The fact that individual differences persist even when possible factors that influence such variations are controlled suggests that language learners, even those at the initial stage (i.e., the onset of learning), may follow different developmental paths to successful acquisition of L2 speech. While SLM investigates differences among L2 learners at the level of extra-linguistic factors such as age of acquisition and language exposure, the approach is to group these learners according to these factors prior to comparing their performance (Colantoni et al., 2015). Studies investigating perception under the framework of PAM also acknowledge the existence of individual difference among listeners; however few studies are yet to explain such differences. In fact, Tyler et al. (2014) found individual differences in assimilation of non-native vowel contrasts, and proposed that individual variation should be considered when predicting L2 difficulties, but did not examine the sources of the individual differences they had observed. This is where the L2LP model may be particularly helpful: it was specifically designed to account for individual variation among non-native speakers at all stages of learning and across different learning abilities (i.e., perception, word recognition and production). As a result, L2LP predictions are made for individual learners based on detailed acoustic comparisons of their L1 categories and the categories of the specific target language variety (Colantoni et al., 2015, p.44). To date, however, most studies that test L2LP's (e.g., Elvin et al., 2014; Escudero et al., 2014; Escudero & Vasiliev, 2011; Escudero & Williams, 2012; Vasiliev, 2013) theoretical claims have done so based on results from group data. An exception is Mayr and Escudero (2010) who investigated whether native

English learners of German follow different developmental learning paths in their perceptual development. The study involved two experiments, the first of which collected perceptual assimilation patterns of L2 vowels; these patterns were subsequently used to predict leaners' performance in an L2 identification task. The results indicated that individual performance in the L2 identification task was largely consistent with the predictions made based on their individual perceptual assimilation patterns. Mayr and Escudero (2010) further found that learners followed similar trends in their perceptual assimilation patterns. That is, despite variation in terms of exact classification percentages, native English listeners were consistent in selecting only a subset of the 13 native response options available to them when assimilating the six German vowels to native vowel categories. This finding indicates that the individual variations in the listeners' perceptual assimilation patterns were not random but highly constrained and systematic (Mayr & Escudero, 2010).

The present study aims to directly test the L2LP model by determining (1) whether detailed acoustic comparisons using the IS and AusE participants' own native production data successfully predicts their non-native categorisation of BP vowels and (2) whether acoustic similarity and the L2LP learning scenarios identified in non-native categorisation subsequently predict their BP discrimination patterns at both a group and individual level. In our investigation of individual variation, we focus specifically on the fact that individuals from the same native language background may have different acoustic realisations of vowels and this factor may predict individual differences in perceptual performance. That is, the within-category variation in native production may influence non-native categorisation and discrimination. Very few studies including Elvin et al. (2014) collected vowel productions from the same listeners that they tested, which according to the L2LP model, is an essential ingredient for accurate predictions of L2 difficulty and for the identification of any individual variation that may be caused by individuals' different acoustic realisations of their own native vowels. Thus, although representative acoustic measurements from the listener populations have successfully explained L2 perceptual difficulty, such comparisons may not account for the individual variation among listeners.

The present study improves on Elvin et al. (2014) and many previous non-native perception studies in that it includes native acoustic production as well as non-native categorisation and discrimination data from the same participants across all tasks. Furthermore, the BP acoustic data that we use to measure acoustic similarity are the same data that we use as stimuli in the non-native categorisation and discrimination tasks. We also control for variation within languages and speakers by ensuring participants, as well as the speakers in our target BP dialect, were all naïve listeners of similar ages selected from a single urban area within each of their respective countries. By controlling for variation relating to language experience, age and native background we are able to conduct a carefully controlled investigation of individual differences in L1 native production and non-native perceptual relationships.

The present study builds on the findings from Elvin et al. (2014) as the acoustic data are based on the listeners' own productions of native vowels in the fVf context, the same context used as stimuli in non-native categorisation and discrimination. This differs from Elvin et al's (2014) AusE acoustic data that were based on the Cox (2006) corpus, which contained acoustic measurements of adolescent speakers from the Northern Beaches, collected in the 1990's and extracted from an hVd context. Interestingly, Elvin et al., (2016) found that vowel duration and formant trajectories varied depending on the consonantal context in which they were produced. In particular, vowels produced in the hVd context were acoustically the least similar to the vowels produced in all of the remaining consonantal consonants. Thus /hVd/ may not be the most representative phonetic context for predicting L2 vowel perception difficulty; instead, ideal predictions should be based on native vowels produced in the same phonetic context used in testing. To measure acoustic similarity between vowels, Elvin et al (2014) used Euclidean Distances between the reported F1 and F2 averages for each vowel. However, because native production data were available for the present study we instead used cross-language discriminant analyses as a method of measuring acoustic similarity, to use in predicting performance in the perceptual tasks (Experiments 2-3). This should improve predictions of acoustic similarity over those from simple Euclidean Distance, as we are able to include more detailed acoustic information relevant for vowel perception, namely F1, F2 and F3 and duration values as well as formant trajectory, as input parameters for each individual participant.

Considering that patterns of non-native categorisation underlie discrimination difficulties, which according to the L2LP model is predictable based on acoustic properties, the inclusion of non-native categorisation data in the present study further allows for an investigation of whether or not listeners' individual categorisation patterns do in fact predict difficulty in discrimination. The incorporation of a categorisation task also allows us to investigate whether the L2LP learning scenarios at the onset of learning (unfamiliar BP stimuli) are similar across the two listener groups (IS and AusE). Although cases of the NEW learning scenario are common in languages such as Spanish who have fewer vowels in their L1 inventory than the target language, and are expected in the present study, it is possible that the same scenario exists in languages that have more opportunities for the SIMILAR learning scenario and vice versa. In fact, there are many opportunities for the vowels of languages with a larger L1 inventory to be categorised to multiple categories which, when contributing to a perceptual or acoustic overlap (i.e., when the two non-native sounds in the contrast are acoustically similar to or categorised to two or more of the same native categories), creates a learning scenario that is comparable to that of the NEW learning scenario (Elvin et al., 2014; Levy, 2009; Vasiliev, 2013).

Although Elvin et al. (2014) already report discrimination accuracy for the same language groups investigated in the present study, it was essential that we repeated the discrimination task with this new set of participants who also completed the native production and non-native categorisation tasks, in order to adequately test the individual difference assumptions of the L2LP model. The L2LP model explicitly states that different listeners have different developmental patterns and it is important to conduct all tasks on the same set of listeners. To this end, we selected naïve listeners in both non-native groups who represent the initial stage of language learning in the L2LP framework. Their inclusion provides a good opportunity for assessing differences in language learning ability that are not confounded by other factors that vary widely among actual L2 learners. Given the aforementioned improvements on Elvin et al. (2014), the present study provides the first direct test of the L2LP prediction that the acoustic comparisons between native production and target language indeed account for the differences in discrimination accuracy between two groups of listeners.

The discrimination task in the present study further differs from that reported in Elvin et al. (2014) in that the vowels are presented in a nonce word context rather than as vowels in isolation. We made this change because, outside of the laboratory, learners are faced with words rather than vowels in isolation. The L2LP model assumes continuity between lexical and perceptual development, specifically positing that perceptual learning is triggered when learners attempt to improve recognition by updating their lexical representations (van Leussen & Escudero, 2015). Thus the presentation of the vowels in the context of a nonce word not only reflects learning that is closer to a real world situation, but also the discrimination accuracy patterns could be used to predict listeners' performance in word learning and word recognition.

The present study is therefore, to our knowledge, one of the first to evaluate predictions about L2 perception (both non-native categorisation and discrimination) based on the listeners' own native vowel productions, thereby providing a novel test of one of L2LP's core assumptions. In Experiment 1, native AusE and IS listeners' native vowel productions are compared to the BP production data that are used as stimuli in Experiments 2 and 3. Results from Experiment 1's cross-language acoustic comparisons are used to predict the non-native categorisation patterns in Experiment 2 and the discrimination results in Experiment 3.

4.2. Experiment 1: Cross-language acoustic comparisons

4.2.1. Participants

Twenty Australian English (AusE) monolingual listeners from Western Sydney and twenty Iberian Spanish (IS) monolingual participants from Madrid participated in this study. All participants were second generation Australian English or Iberian Spanish listeners currently residing in Greater Western Sydney or Madrid, respectively, and aged between 18 and 30 years old. The AusE participants reported little to no knowledge of any foreign language. The IS participants reported little to intermediate knowledge of English and little to no knowledge of any other foreign language. AusE participants were recruited through the Western Sydney University psychology pool, or from the Greater Western Sydney region and received $40 for their participation. IS participants were recruited from universities and institutes around the Universidad Nacional de Educación a Distancia and received €30 for their time. All participants were part of a larger scale study that looked at the interrelations among non-native speech perception, spoken word recognition and non-native speech production. All participants provided informed consent in accordance with the ethical protocols in place at the Universidad Nacional de Educación a Distancia and the Western Sydney University Human Research Ethics Committee.

4.2.2. Stimuli and procedure

AusE and IS completed a native production task in which participants read pseudo-words containing one of the 13 Australian English vowels, namely, iː, ɪ, ɪə , e, eː, ɜː, ɐ, ɐː, æ, oː, ɔ, ʊ and ʉː, or one of 5 Iberian Spanish vowels, i, e, a, o, u, in the fVf (AusE) and fVfo (IS) context. Each vowel was repeated 10 times per vowel for a total of 130 tokens for AusE and 50 tokens for IS. For our acoustic analysis of BP vowels, we used the pseudo-words with the fVfe context produced by five male and five female speakers from São Paulo, selected from the Escudero et al. (2009) corpus. These words were the same BP words used as stimuli in Experiments 2 and 3. These BP pseudo-words were produced in isolation and within a carrier sentence e.g., "Fêfe. Em fêfe e fêfo temos ê" which translates to: "Fêfe. In fêfe and fêfo we have ê"(Escudero et al., 2009). The vowel in the first syllable was always stressed and corresponded to one of the seven Portuguese vowels /i, e, ɛ, a, o, ɔ, u/. WebMaus (Kisler et al., 2012) was used to segment vowels within each target word in each language (AusE, IS and BP). The automatically generated start and end boundaries were checked and manually adjusted to ensure accuracy. Vowel duration was measured as the time (ms) between these start and end boundaries. Formant measurements for each vowel token were extracted at three time points (25%, 50%, 75%) following the optimal ceiling method reported in Escudero et al. (2009) to ensure that our acoustic predictions are comparable across both the target and native languages. In the optimal ceiling method, for every vowel, per speaker, the "optimal ceiling" is chosen as the one that yields the least amount of variation for both the first and second formant values within the set number of annotated tokens for the vowel. Formant ceilings ranged between 4500 and 6500 Hz for females and 4000 and 6000 for males.

4.3. Results and discussion

4.3.1. Cross-language acoustic comparisons

Figure 4.1 shows the average male (right) and female (left) F1 and F2 values of the thirteen AusE (black) and five IS (grey) vowels, together with the average F1 and F2 values for both BP male (blue, circled) and female (green, circled) vowels selected from Escudero et al. (2009) and used as stimuli in the present study. Although we considered normalising the data to plot both male and female vowels together, we instead plot native male or female vowels together with both non-native male and female vowels in order to reflect the discriminant analyses we conducted for each individual.

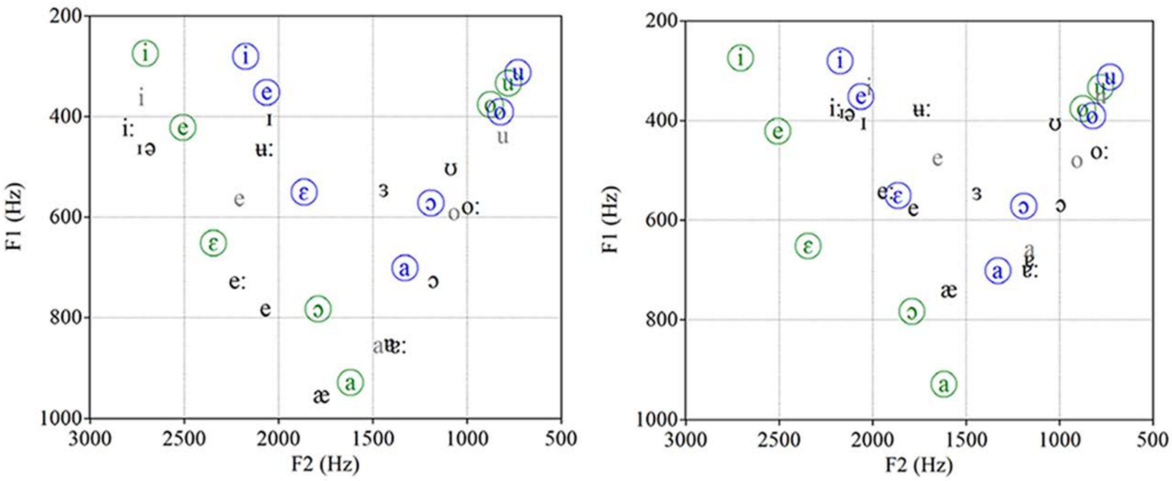


Figure 4.1 Average male (right) and female (left) F1 and F2 values (Hz) for the thirteen AusE (black) and five IS (grey) vowels together with both BP male (blue, circled) and female (green, circled) vowels.

Visual inspection of the plot reveals that although AusE has many more vowels in its native vowel inventory than IS, the vowels of both languages fall in and around the same locations within the acoustic space. Following Strange et al. (2004) and Escudero & Vasiliev (2011), we conducted a series of discriminant analyses to determine acoustic similarity and use it to predict listeners' non-native categorisation patterns. We first performed separate linear discriminant analyses (cross-validation method) on male and female native IS and AusE vowel sets to uncover how a trained AusE or IS discriminant analysis model would classify tokens from the same native language. These analyses were conducted to determine the underlying acoustic parameters that predict the vowel categories for test tokens from the BP corpus. The input parameters were F1, F2 and F3 Bark values measured at the vowel midpoint (i.e., 50%) as well as formant trajectory length for a measure of spectral change (i.e., the Euclidean distance (Bark) in the F1/F2space between F1 and F2 measurements at the 25 and 75% time points). A second set of analyses included the same aforementioned input parameters as well as duration values. The best fitting model, i.e., the one that categorised the native AusE or IS vowels most accurately, was the one that included duration values, in which the IS model yielded 95 % (females) and 96.7% (males) correct classifications and the AusE model yielded 89.2% (females) and 86.9 % (males) correct classifications.

We then conducted a cross-language discriminant analysis, using F1, F2 and F3 values (measured at 50%), formant trajectory length and duration as input parameters to determine how likely the BP vowel tokens would be categorised in terms of AusE and IS vowel categories. In a typical discriminant analysis, the vowels in the test corpus (in our case BP) are categorised with respect to the vowel centres of gravity established in the input corpus (Strange et al., 2004). In other words, the discriminant analysis tests how well the BP tokens fit with centres of gravity of the input corpus tokens (AusE or IS), providing a predicted probability that each vowel will be categorised as one of the native vowel categories. The native vowel category that receives the highest probability for a given BP vowel indicates the native vowel that is acoustically closest to the non-native vowel.

Given the fact that we only have one token per BP speaker (5 male and 5 female), rather than reporting the percentage of times a BP vowel was categorised as a native vowel as is commonly reported, we instead report the probabilities of group membership averaged across the BP vowel tokens: For each individual BP vowel token we report the predicted probability of it being categorised as any of the 13 native AusE or 5 native IS vowels and average these probabilities over all tokens of that BP vowel. The benefit of reporting probabilities in the present study is that it takes into account that some BP tokens may be acoustically close to more than one vowel, which can be masked by categorisation percentages. The predicted probabilities averaged across individuals for AusE and IS are shown in Tables 4.1 and 4.2 respectively.

Table 4.1 Average probability scores of predicted group membership for male and female BP tokens tested on each individual AusE model. Probabilities are averaged across the individual discriminant analysis for each speaker. The native vowel category with the highest probability appears in bold and those probabilities below chance (i.e. 0.08) appear in grey.

AusE vowels	**BP vowels**						
	i	**e**	**ɛ**	**a**	**o**	**ɔ**	**U**
iː	**0.55**	**0.30**	0.04			0.01	
ɪ	0.10	0.25	0.14			0.02	
ɪə	0.13	0.15	0.02			0.01	
e			0.24	0.01		0.14	
eː		0.02	**0.25**	0.01		0.06	
ɜː			0.03	0.12		0.01	
ɐ			0.04	0.23		0.12	
ɐː				0.01		0.04	
æ				**0.48**		0.07	
ɔ			0.01	0.14	0.10	**0.29**	0.05
oː					0.16	0.07	0.13
ʊ			0.01		**0.73**	0.15	**0.80**
ʉː	0.20	0.28	0.23				

We take the probability of group membership to correspond to the degree of acoustic similarity, i.e., a high probability indicates a high level of acoustic similarity. Visual inspection of Table 1 shows the averaged probabilities of predicted group membership that is representative of the “average listener”. The table above reveals that most BP vowels were acoustically similar to two or more native AusE vowels. An acoustic categorisation overlap can be observed for BP contrasts /i/-/e/ and /o/-/u/, where each vowel in the BP contrast is acoustically similar to the same native AusE vowel(s). In the case of BP /i/-/e/, there is a 0.55 probability that BP /i/ will be categorised as AusE /iː/ and a 0.30 probability that BP /e/ will also be categorised as AusE /iː/. There is also a 0.10 probability that BP /i/ will be categorised as AusE /ɪ/, and a probability of 0.25 that BP /e/ will be categorised as AusE /ɪ/. There is also a 0.13 probability that BP /i/ will be categorised as AusE /ɪə/ as well as a 0.20 probability that it could be categorised as AusE /ʉː/. Likewise, there is a 0.15 probability that BP /e/ will also be categorised as AusE /ɜː/ and a 0.28 probability that it will be categorised as AusE /ʉː/ For BP /o/-/u/ there is a 0.73 probability that BP /o/ will be categorised as AusE /ʊ/ and a 0.80 probability for BP /u/ to be categorised as AusE /ʊ/. We also observe a 0.31 probability that BP /u/ will be categorised as AusE /oː/ and a 0.16 probability that BP /o/ will also be categorised as AusE /oː/. An acoustic overlap is also observed in BP /e/-/ɛ/, with a 0.14 probability that BP /ɛ/ will be categorised as AusE /ɪ/ and a 0.25 probability for BP /e/. Furthermore, there is a 0.23 probability that BP /ɛ/ will be categorised as AusE /ʉː/ and a 0.28 probability for BP /e/. Partial acoustic overlapping is also observed in the BP /o/-/ɔ/ contrast with a 0.73 probability that BP /o/ will be categorised as /ʊ/ and a 0.15 probability for BP /ɔ/.

Table 4.2 Average probability scores of predicted group membership for BP male and female tokens tested on a IS model. Probabilities are averaged across the individual discriminant analysis for each speaker. The native vowel category with the highest predicted probability appears in bold and those probabilities below chance (i.e. 0.20) appear in grey.

IS vowels	BP vowels						
	i	**e**	**ɛ**	**a**	**O**	**ɔ**	**U**
i	**0.99**	**0.82**	0.07	0.02		0.02	0.01
e	0.01	0.18	**0.92**	0.05	0.03	0.25	
a			0.01	**0.81**		**0.36**	
o				0.12	0.25	0.26	0.06
u					**0.71**	0.11	**0.93**

Table 4.2 shows the BP tokens tested on the IS model. The results indicate that BP /i/, /ɛ/, /a/ and /u/ are each acoustically similar to different single native categories, while the remaining three BP vowel categories (/e/, /o/ and /ɔ/**)** are acoustically close to two or more native vowels. In the cases where a BP vowel is acoustically similar to two native IS vowels, the similarity is not equal across the two vowel categories with the probability scores indicating a greater likelihood of classification of one vowel over the other. For example, there is a 0.82 probability that BP /e/ would be categorised as IS /i/. BP /e/ also has a 0.18 probability of being categorised as IS /e/, however this probability is below chance (i.e. 0.20). There is also a 0.71 probability that BP /o/ will be categorised as IS /u/ and only a 0.25 probability that it will be categorised as IS /o/. In the case of BP /ɔ/, the probabilities are spread across three vowel categories, with a 0.36 probability that it will be categorised as IS /a/, a 0.26 probability that it would be categorised as IS /o/, and finally a 0.25 likelihood that it will be categorised as IS /e/.

An acoustic overlap can also be identified in the IS categorisations for BP contrasts /i/-/e/, /o/-/u/ and /o/-/ɔ/. BP /i/ and /e/ are both acoustically similar to IS /i/, BP /o/ and /u/ are both acoustically similar to IS /u/ and BP /o/ and BP /ɔ/ are acoustically similar to both IS /o/.

A partial acoustic overlap is also observed in the BP contrast /a/-/ɔ/, as BP /ɔ/ had a probability of 0.20 of being categorised as IS /a/.

4.3.2. L2LP Predictions for non-native categorisation

Figures 4.2 and 4.3 present the predicted non-native categorisation patterns for each BP contrast based on the reported averaged group results of the discriminant analyses.

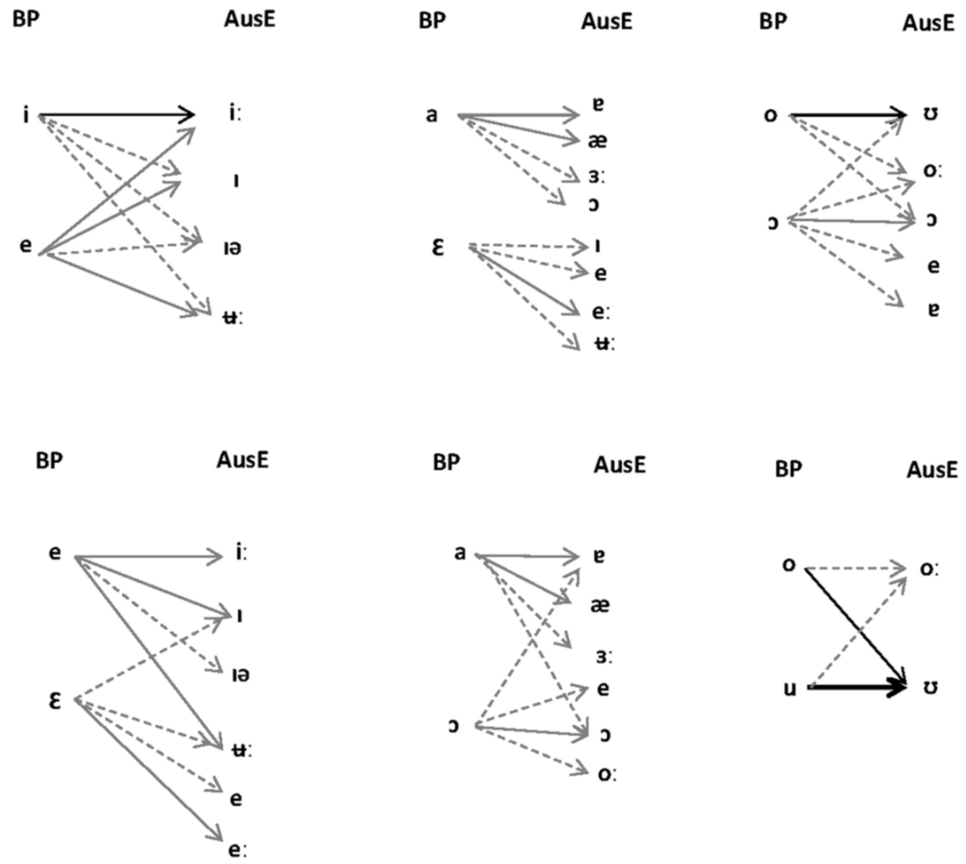


Figure 4.2 AusE predicted non-native categorisation patterns based on the averaged predicted probabilities for the group. Probabilities above 75% are represented with a thick black line, those between 50% and 75% are represented with a thinner black line, those between 25% and 50% with a thin grey line and those between 0.08% and 25% are shown with a thin dotted grey line. Probabilities below chance (0.08) are not shown.

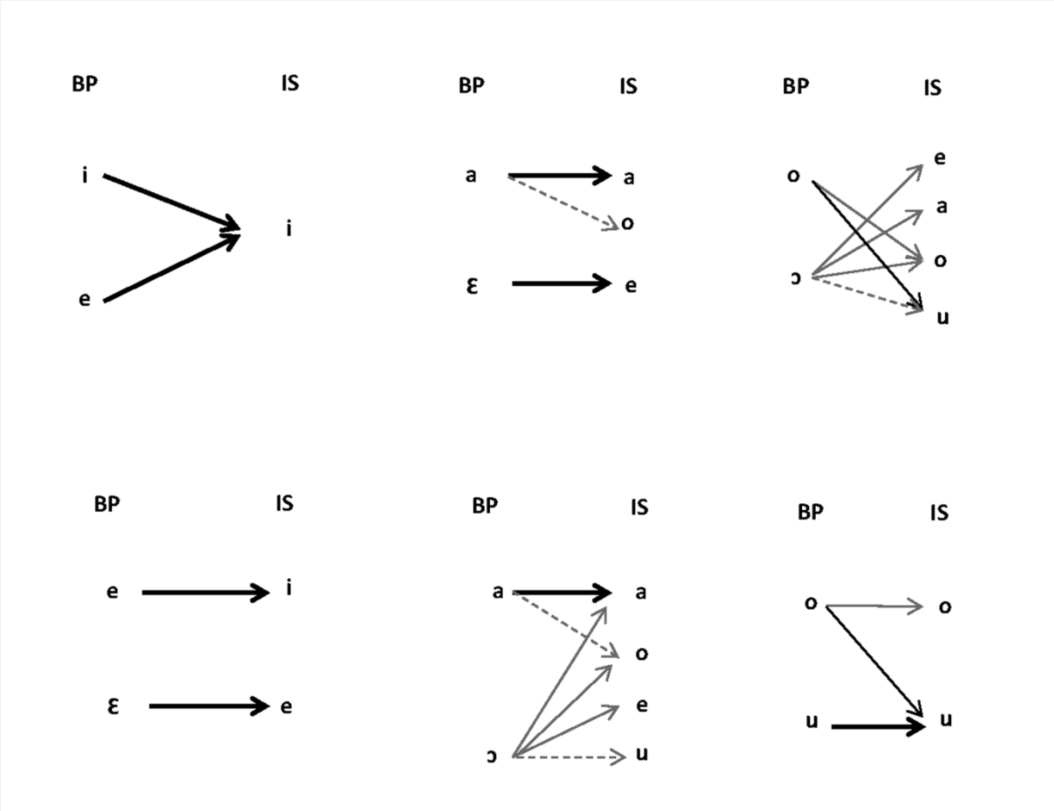


Figure 4.3 IS predicted non-native categorisation patterns based on the averaged predicted probabilities for the group. Probabilities above 75% are represented with a thick black line, those between 50% and 75% are represented with a thinner black line, those between 25% and 50% with a thin grey line and those between 0.08% and 25% are shown with a thin dotted grey line. Probabilities below chance (0.20) are not shown.

The acoustic similarity as determined by the probability scores from our discriminant analyses are used to predict perceived phonetic similarity in a categorisation task. For AusE listeners, we predict cases of L2LP's SUBSET scenario, in which all BP vowels will be categorised as more than one native category. We also predict that there will be instances of perceptual overlap in the non-native categorisation patterns for the BP contrasts /i/-/e/ and /o/-/u/, and partial perceptual overlapping in BP /e/-/ɛ/ and /o/-/ɔ/ as these were the contrasts in which we identified acoustic overlapping.

For the IS listeners, we predict most BP vowels should be categorised as one single native category. In particular, we expect that BP /ɛ/, /a/ and /u/ will be categorised as /e/, /a/,

/u/ respectively. Furthermore, a case of L2LP's NEW scenario is predicted, as it is likely that BP /i/ and BP /e/ will both be categorised as IS /i/. It is likely that BP /o/ will be categorised to two native categories as there is a 0.71 probability that it will be categorised as IS /u/ and a 0.25 chance it will be categorised as IS /o/. Finally BP /ɔ/ may be categorised to a number of IS categories, as there was a 0.26 probability that it would assimilate to IS /o/, and a 0.25 and 0.26 probability that it could be assimilated to IS /e/ and /a/.

4.3.3. L2LP Predictions for non-native discrimination

The L2LP model claims that discrimination difficulty can be predicted by acoustic similarity between the native and target language, unlike PAM which makes predictions based on articulatory-phonetic similarity and collects perceptual assimilation data and category-goodness ratings to test its predictions. As a result, we can use the predicted categorisation patterns to predict discrimination accuracy. In particular, we would predict that the perceptually easy contrasts for both groups of listeners to discriminate would be those with little to no acoustic overlap. However, the BP contrasts with a large amount of acoustic overlap should be perceptually difficult to discriminate.

For a quantitative method of predicting non-native discrimination difficulty we calculated acoustic overlap scores based on Levy's (2009) "cross language assimilation overlap" method. This method provides a quantitative score of overlap between the members of a non-native contrast and native categories. Although originally designed to compute perceptual overlap scores based on listeners' perceptual assimilation patterns for testing predictions in PAM, we apply the same method to our acoustic similarity data to yield acoustic overlap scores in order to test the L2LP model's acoustic predictions. Each overlap score was calculated by adding the smaller probabilities where the two vowels in the BP contrast were acoustically similar to the same native categories. For example, on the BP /i/-/e/ contrast, as observed in Table 4.1 there was a probability that both BP /i/ and BP /e/ would be categorised as AusE /iː/, /ɪ/ and /ɪə/ and /ʉː/. To calculate the acoustic overlap score for this contrast, we calculate the smaller proportion of when both BP vowels had a probability of being categorised as the same AusE vowel category. Thus in the case of AusE /iː/ there was a 0.55 probability that BP /i/ would be categorised as this vowel and a 0.30 probability that BP /e/ would also be categorised as AusE /iː/. The smaller proportion of the overlapping of BP /i/ and /e/ to AusE /iː/ in this case would be 0.30 for BP /e/ which was included in the calculation of the acoustic overlap. The calculation of acoustic overlap for BP /i/-/e/ was as follows: AusE /iː/ 0.30 + AusE /ɪ/ 0.10 + AusE /ɪə/ 0.13 + AusE /ʉː/ 0.20 = 0.73 acoustic overlap.

Table 4.3 shows the acoustic overlap scores for each language.

Table 4.3 Acoustic overlap scores for AusE and IS listeners. Only non-zero overlaps are shown.

Group	**BP vowel contrast**					
	/a/-/ɔ/	**/a/-/ɛ/**	**/i/-/e/**	**/o/-/u/**	**/e/-/ɛ/**	**/o/-/ɔ/**
AusE	0.37	0.14	0.73	0.91	0.45	0.32
IS	0.55	0.08	0.83	0.77	0.25	0.39

Based on the acoustic overlap scores in Table 4.3, we would predict that the BP contrasts with little to no acoustic overlap would be perceptually easier to discriminate than those with higher overlap scores. In particular, both groups should find BP /i/-/e/ and /o/-/u/ difficult to discriminate and BP /a/-/ɛ/ easy to discriminate. In addition, however, these acoustic comparisons do predict group differences: AusE listeners should find BP /a/-/ɔ/ easier to discriminate than IS listeners, whereas BP /e/-/ɛ/ should be easier for IS than AusE listeners to discriminate.

4.4. Experiment 2: Non-native categorisation

4.4.1. Participants

Participants in Experiment 2 were the same as those previously reported in Experiment 1, § 4.2.1. However, the non-native categorisation results of five IS participants were excluded due to an error that occurred during testing.

4.4.2. Stimuli and procedure

Participants were presented with the same naturally produced BP pseudo-words in the fVfe context selected from Escudero et al.'s (2009) corpus that was reported and analysed in Experiment 1. In this experiment there was a total of 70 fVfe target words (7 vowels x 10 speakers), as well as three additional nonsense words by each speaker (pipe, kuke and sase), included as distractors. Thus, in the non-native categorisation task we had a total of 100 BP word tokens (70 target and 30 distractors).

A non-native categorisation task (also known as a perceptual assimilation task with the PAM model) programmed in E-Prime, and similar to those reported in Vasiliev and Escudero (2011) and Escudero and Williams (2011), followed the discrimination task (Exp. 3). In the categorisation task, participants identified the stressed vowel sound of each target BP word (i.e., the target vowel) to one of their own 13 AusE or 5 IS vowel categories. Unlike Spanish, English is not orthographically transparent. Thus, while the IS listeners saw the 5 vowel categories (i, e, a , o, u) on the screen, the AusE vowels were presented in one of the 13 keywords, *heed*, *hid*, *head*, *heared*, *haired*, *heard*, *hud*, *hard*, *had*, *hoard* , *hod*, *hood* and *who'd*, which correspond to the AusE phonemes /iː, ɪ, ɪə, e, eː, ɜ, ɐ, ɐː, æ oː, ɔ, ʊ, ʉː/. Participants were required to choose one of their own native response options on each trial, even when unsure. Participants received a short practice session before beginning each task and took about 10 minutes to complete the task.

4.4.3. Results and discussion

Tables 4.4 and 4.5 present the percentage of times each BP vowel was categorised as a native and IS vowel.

Table 4.4 Australian English listeners' classification percentages (percentages are rounded to the nearest whole number). The native vowel category with the highest classification percentage appears in bold and those percentages below chance (i.e. 8%) appear in grey.

AusE vowels	BP vowels i	e	ɛ	a	o	ɔ	U
iː	**43**	**23**	5				1
ɪ	**43**	20	2				
ɪə	6	15	6	2		1	1
e	6	14	14				
eː	1	**23**	**58**	6	1	4	
ɜː			1	1	8	5	3
æ	1	3	7	38		1	
ɐː	1	2	8	**50**		13	
ɐ				3	4		6
ɔ					9	15	7
oː		2	1	2	**53**	**58**	6
ʊ					23	3	**65**
ʉː					4	2	12

Table 4.5 Iberian Spanish listeners' classification percentages (percentages are rounded to the nearest whole number). The native vowel category with the highest classification percentage appears in bold and those percentages below chance (i.e. 20%) appear in grey.

IS vowels	BP vowels i	e	ɛ	a	o	ɔ	U
i	**100**	**56**	1				
e		44	**94**				
a			5	**100**		2	
o					**69**	**97**	1
u					31	1	**99**

The categorisation percentages reported in Tables 4.4 and 4.5 are in line with acoustic predictions that most BP vowels would be categorised as more than two native categories by AusE listeners, and that most BP vowels would instead be categorised as one single native category by IS listeners.

Indeed, in keeping with acoustic predictions, all BP vowels were categorised as two or more native AusE categories, and there was some degree of perceptual overlap between relevant pairs of BP vowels. In particular, we found evidence of the SUBSET scenario in the BP contrasts /i/-/e/, /e/-/ɛ/, /o/-/ɔ/ and /o/-u/, where both vowels in the contrast were categorised as two or more of the same native AusE vowels. For example, BP /i/ was categorised as AusE /iː/ and as well as AusE /ɪ/, 43% of the time for both vowels, which was in fact predicted by acoustic similarity. Our discriminant analysis indicated the probability that BP /e/ would be categorised across four native vowel categories, namely, /iː/, /ɪ/, /ɪə/ and /ʉː/. This prediction was largely consistent with AusE listeners' non-native categorisation patterns, with /e/ being categorised as AusE /iː/ (23%), /ɪ/ (20%) and /ɪə/ (15%). But although our acoustic analysis indicated that BP /e/ would be categorised as AusE /ʉː/, this was not the case, as listeners instead rather substantially categorised the vowel to two unpredicted AusE vowels, /e/ (14%, i.e. as often as to /ɪə/) and /eː/ (23%, i.e., equal to the actual choices of the top acoustically predicted choice /i:/). The BP /ɛ/ tokens were expected to be predominately categorised as AusE /e/ and /eː/, with a probability that they should also be categorised as AusE /ɪ/ and /ʉː/. Indeed, BP /ɛ/ was categorised as AusE to /eː/ 58% of the time and to /e/ 14% of the time. However, again, actual perception failed to meet acoustic predictions for some AusE categorisations: listeners failed to choose the acoustically predicted /ɪ/ or /ʉː/ for BP /ɛ/. Partially consistent with the patterns of acoustic similarity, BP /a/ was categorised as /æ/ 38% of the time. However, instead of the predicted moderate level of choice of AusE /ɐ/ (as in *hut*) for BP /a/, the long AusE /ɐː/ (as in *heart*) was instead selected most frequently at

50% of the time. A large majority of BP /o/ tokens were expected to be categorised as AusE /ʊ/ with a probability that some would also be categorised as AusE /oː/. However, the listeners actually reversed the balance between the two AusE vowel categories: the large majority BP /o/ tokens were instead categorised as AusE /oː/ (53% of the time), and as AusE /ʊ/ only 23% of the time. Furthermore, our acoustic predictions suggested that the majority of BP /ɔ/ tokens would be categorised as AusE /ɔ/ with some also being categorised as /ʊ/, /e/ and / ɐ /. However, the non-native categorisation patterns indicate that the great majority of BP /ɔ/ tokens were categorised as AusE /oː/ (58% of the time), with only 15% being categorised as / ɔ / and 13% to /ɐː/, and none selected as the predicted top choice of AusE /ɔ/. Finally, acoustics successfully predicted the majority of BP /u/ tokens to be categorised as AusE /ʊ/ (65%), however none were categorised to AusE /oː/ as was predicted, instead 12% were categorised /ʉː/, which was not predicted to be a listener choice.

Acoustic similarity was largely consistent in predicting IS listeners' non-native categorisation patterns. As expected, BP /i/, /ɛ/, /a/ and /u/ were each categorised as the native IS vowel category identified in acoustics. In accordance with our acoustic analyses, BP /o/ was categorised as both IS /o/ and /u/, with similar percentage scores. BP /e/ was also categorised as IS /i/ and /e/ as expected. Our acoustic analysis predicted that listeners would favour IS /i/ over IS /e/; however this was not the case as the categorisation percentages were largely similar across the two vowels. Finally, despite the fact that our acoustic analysis predicted that BP /ɔ/ could potentially be categorised as three different native categories, with the highest likelihood that it would be categorised as IS /o/, the results indicate that BP /ɔ/ was consistently categorised as IS /o/.

Based on the above findings it appears that at a group level, the non-native categorisation patterns for IS listeners were largely in line with predictions based on acoustic similarity reported in Experiment 1. For AusE listeners, acoustic similarity predicted some non-native categorisation patterns but not all. For example, in line with the predictions based on acoustic similarity, all BP vowels were categorised as two or more native AusE categories and cases of the SUBSET scenario were identified in the BP contrasts /i/-/e/, /e/-/ɛ/, /o/-/ɔ/ and /o/-u/. However, there were a couple of instances where some AusE listeners did not categorise some BP vowels to categories which were predicted to be acoustically similar. For example, our acoustic analysis indicated that AusE /ʉː/ may be a possible response category for listeners' categorisation of BP /e/ and /ɛ/, however this was never the case. Additionally, our linear discriminant analysis indicated that listeners should predominately classify some of the BP /a/ tokens as AusE /ɐ/ (as in *hut*), however they instead predominately selected the long AusE /ɐː/ (as in *heart*) vowel.

A possible explanation that may account for the discrepancies between the acoustically predicted categorisations and those actually observed may be a result of the limitations of the statistical tool that we used to predict acoustic similarity, namely the cross-language discriminant analyses. An important aspect of the L2LP model is its strong emphasis on acoustic and auditory cue-weighting (the relative importance of acoustic cues in the learner's native and target languages). This important tenet of the L2LP model involves modelling production, perception and word recognition with more sophisticated means such as computational Stochastic Optimality Theory, Latent Logistic Regression models or neural networks, all using the Gradual Learning Algorithm. Thus, it may be that listeners weigh certain cues more than others, as has been shown in previous studies (e.g., Curtin, Fennell, & Escudero, 2009), a factor which cannot be predicted in a linear discriminant analysis. However, modelling native production using a more sophisticated model might show that the relative importance of individual cues can vary, even across different vowels.

Furthermore, the above acoustic predictions and subsequent findings of non-native categorisation patterns are based on group means and as previously mentioned, the L2LP model posits that there are individual differences among learners that could affect their non-native categorisation patterns on an unfamiliar language. AusE and IS listeners' non-native categorisation patterns were largely predicted by the probability scores based on acoustic similarity reported in Experiment 1. We now turn to an analysis of individual differences based on the relationship L2LP predicts between the acoustic properties of *individuals'* own native vowel productions and their individual categorisation patterns for the BP vowels.

Figure 4.4 shows the relationship between native production and perception for each participant. The horizontal axis shows the predicted probability of a vowel category given its acoustic properties. This was determined by the discriminant analysis, as described in Experiment 1, for each speaker. The vertical axis shows the percentage of times that the corresponding vowel was selected in the categorisation task. The left panel shows the results for AusE with regression lines fitted through each listener's set of data points and the same is shown for IS on the right panel. The positive slope of the regression lines indicates that, for all subjects, there was a relationship between predictions based on acoustic similarity and the listeners' actual non-native categorisation data. The plots show each AusE (left) and IS (right) listener's predicted probability scores against their own native-categorisation percentages with regression lines fitted through each listener's set of data points. They indicate clearly that all listeners display a positive relationship between acoustic similarity and non-native categorisation. For all but three AusE listeners there was a significantly positive correlation between acoustic similarity and non-native categorisation (β = .425, SD = 0.10, all *p*s = <0.013). Marginal positive correlations were identified for two AusE listeners (β = .202, SD = 0.002, *p*s = 0.053 – 0.057) and only one AusE listener showed no correlation (β = .099, *p* = .352). The results for IS listeners indicate a positive correlation between acoustic similarity and non-native categorisation for all participants (β = .773, SD = 0.08, all *p*s = <0.001). Based on these results and as observed in Fig 3, the regression slopes are generally shallower and show a wider range of variation for AusE listeners than IS. This finding is consistent with our findings at a group level where acoustic similarity was not as good of a predictor for AusE as it was for IS. These findings at an individual level support and provide further insight into those reported at a group level, suggesting that acoustic similarity has a greater influence on non-native vowel categorisation for IS listeners than for AusE listeners. This is perhaps not too surprising because IS listeners had fewer response options than AusE listeners, thus it is likely that the greater the number of response options in the task, the more opportunities for individual variation among listeners. Interestingly, despite the variation among listeners both listener groups seem to display a similar trend whereby the BP vowels that had high probabilities of being classified as a particular native category were the same vowels with larger classification percentages. This is compatible with L2LP's predictions that acoustic similarity between the native and target language is predictive of listeners' non-native categorisation patterns at an individual level.

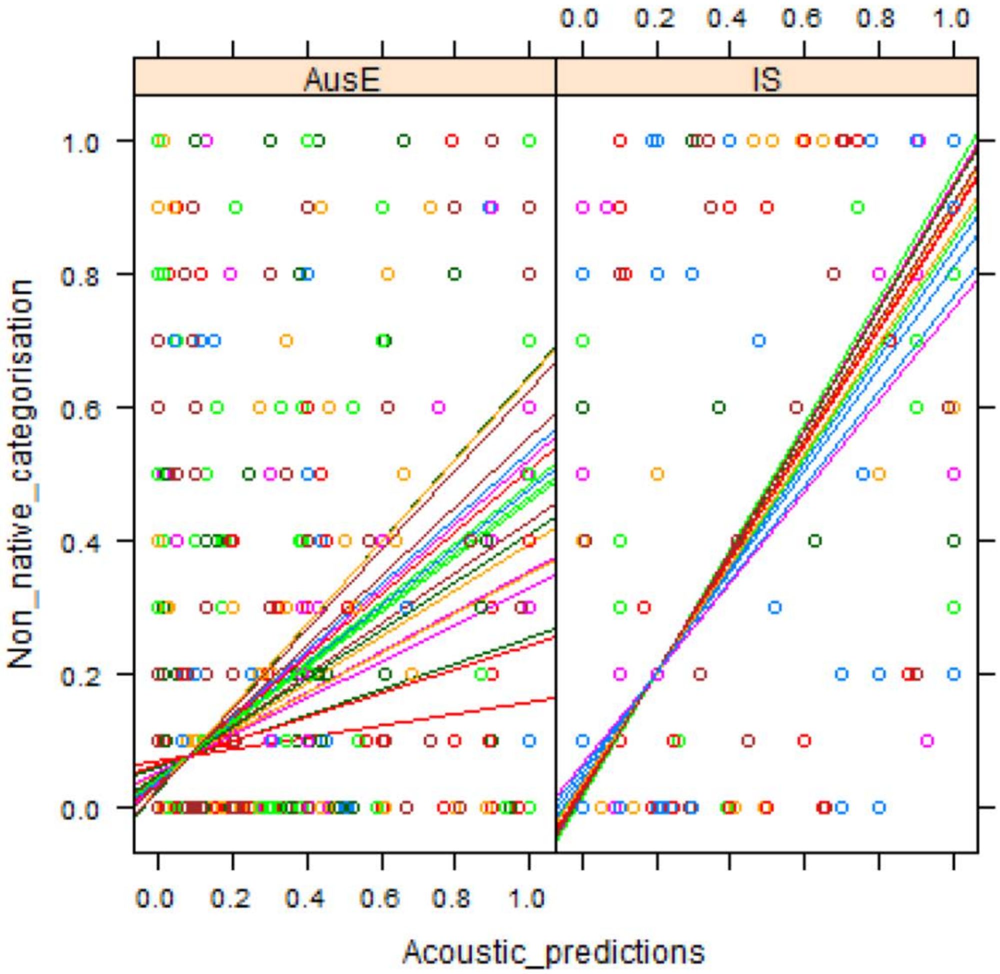


Figure 4.4 AusE (left) and IS (right) individual acoustic predicted probabilities from the discriminant analyses (x-axis) plotted against individual non-native categorisation percentages (y-axis) for each BP vowel. Regression lines have been fitted for each individual. Colours represent each individual's data points and regression line.

4.4.3.1. Predictions for discrimination accuracy

The results from the non-native categorisation task are largely consistent with the predictions based on acoustic similarity at a group level and in particular at the individual level, with both groups showing positive correlations between acoustic similarity and non-native categorisation. Interestingly, the strength of the relationship is much stronger for IS listeners than it is for AusE listeners at a group and an individual level. As shown in Table 4.6, we also computed perceptual overlap scores following Levy (2009) for our categorisation data to determine how predictions of discrimination difficulty based on non-native categorisation would compare with our predictions based on our acoustic comparisons (see Table 3). As mentioned in Experiment 2, to determine an overlap score, based on the categorisation percentages reported in Tables 4.4 and 4.5, we calculate the smaller percentage of when both BP vowels in a given contrast are categorised as the same native vowel category.

Table 4.6 Perceptual overlap scores for AusE and IS listeners expressed as proportions.

Group	**BP Contrasts**					
	/a/-/ɔ/	**/a/-/ɛ/**	**/i/-/e/**	**/o/-/u/**	**/e/-/ɛ/**	**/o/-/ɔ/**
AusE	0.22	0.11	0.58	0.47	0.57	0.73
IS	0.02	0.05	0.56	0.32	0.45	0.70

When comparing the predictions for discrimination difficulty based on perceptual overlap scores with those based on acoustic overlap (see Table 4.3), the predictions are rather similar for BP /a/-/ɛ/. That is both acoustic similarity and non-native categorisation patterns predict that this contrast should be perceptually easy for both groups of listeners to discriminate. Acoustic similarity and non-native categorisation patterns are also similar in that they both predict BP /i/-/e/ to be perceptually difficult to discriminate. For AusE this contrast had the second highest acoustic and perceptual overlap score and in the case of IS it had the highest acoustic overlap score and second highest perceptual overlap score. For the remaining four contrasts, predictions based on acoustics and non-native categorisation differ. For example, acoustic similarity predicts that both groups should find BP /o/-/u/ to be one of the most difficult contrasts to discriminate, whereas perceptual overlap scores suggest that /o/-/ɔ/ should be the most difficult to discriminate. Interestingly, perceptual overlap scores predict that /a/-/ɔ/ should be perceptually easy to discriminate, yet according to acoustic overlap scores, this contrast should actually be more difficult than /o/-/ɔ/. From these findings, two possible predicted patterns of difficulty can be identified. In line with the L2LP model, predictions based on acoustic similarity would suggest the following order of difficulty (ranging from the lowest acoustic overlap score to the highest) for AusE: /a/-/ɛ/ > /o/-/ɔ/ > /a/-/ɔ/ > /e/-/ɛ/ > /i/-/e/ > /o/-/u/ and for IS /a/-/ɛ/ > /e/-/ɛ/ > /o/-/ɔ/ > /a/-/ɔ/ > /o/-/u/ > /i/-/e/. On the other hand, non-native categorisation patterns which would be considered to be in line with PAM's predictions for discrimination difficulty would predict the following pattern for difficulty for both AusE and IS: /a/-/ɛ/ > /a/-/ɔ/ >/o/-/u/ > /e/-/ɛ/ > /i/-/e/ > /o/-/ɔ/. An examination of the pattern of difficulty in the results for discrimination accuracy will shed some light as to whether discrimination difficulty is in line with predictions based on acoustic similarity consistent with the L2LP model or those based on non-native categorisation patterns, consistent with PAM or a combination of both models.

4.5. Experiment 3: Non-native discrimination

4.5.1. Participants

Participants in Experiment 3 were the same participants previously reported for Experiments 1 and 2.

4.5.2. Stimuli and procedure

Listeners were presented with the same 70 naturally produced BP fVfe target words (7 vowels x 10 speakers), selected from Escudero et al.'s (2009) corpus, previously reported and analysed in Experiments 1 and 2.

To test for discrimination accuracy, participants completed an auditory 2 alternative forced choice (2AFC) task in the XAB format, similar to that of Escudero & Wanrooij (2010), Escudero & Williams (2012) and Elvin et al., (2014), however, it was run on a laptop using the E-Prime 2.0 software program.

Participants were instructed to listen to the three words using headphones and were required to make a decision as to whether the first word they heard sounded more like the second or the third. That is, three stimulus items were presented per trial. The second (A) and third (B) items were always from different BP vowel categories and the first item (X) was the target item about which a matching decision was required. In each trial, X was always one of the 70 target BP words and the A and B stimuli were always the 7th male and 7th female speaker from the Escudero et al. (2009) corpus to avoid any confusion of overlapping target and response categories, this differs from the Elvin et al. (2014) study where the A and B stimuli were synthetic. Furthermore, the order of the A and B responses was counterbalanced (namely, XAB and XBA). For the first ten participants for each language group, testing consisted of six blocks of categorical discrimination tasks, with a short break permitted between blocks. Each block consisted of 40 trials with one of the six BP contrasts, namely /a/-/ɔ/, /a/-/ɛ/, /i/-/e/, /o/-/u/, /e/-/ɛ/ and /o/-/ɔ/, with the blocks presented in a randomised order. To determine whether discrimination accuracy differs when stimuli are blocked by contrast or randomised, the remaining 10 participants completed the same discrimination task, with the same breaks, but with the stimulus contrasts presented in random order, unblocked.

4.5.3. Results and discussion

We conducted a repeated measures Analysis of Variance (ANOVA) with contrast as a within-subjects factor and condition (blocked, randomised) as a between-subjects factor, to evaluate whether blocking by BP contrast has an effect on overall performance. The results yielded no significant effect of condition on performance, [$F(1, 38) = 1.905$, $p = .176$, $\eta p^2 = .048$], suggesting that listeners had similar accuracy scores regardless of the condition (blocked vs. randomised). Thus, Figure 4.5 shows discrimination accuracy for the AusE and

IS groups, including their individual variation, across the six BP vowel contrasts for both conditions pooled together.

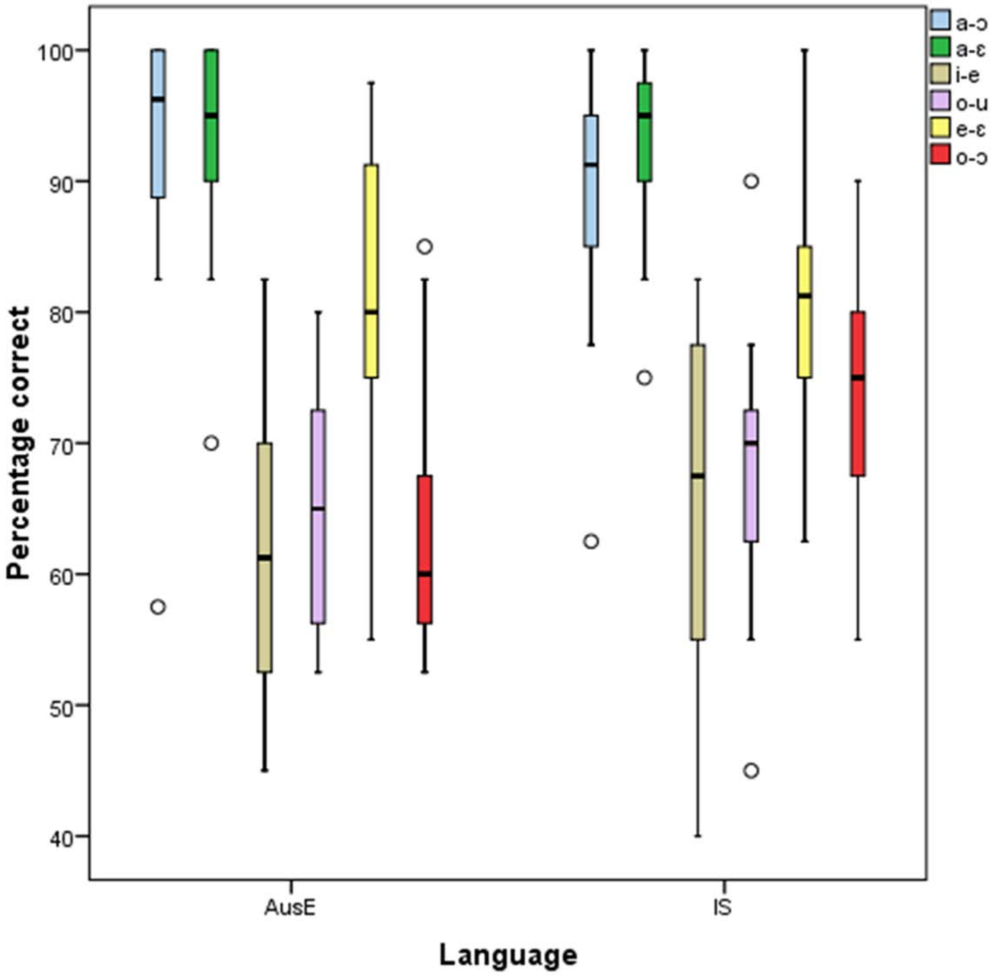


Figure 4.5 Overall discrimination accuracy including individual variation for AusE (left) and IS (right). Thick black lines within the boxes represent median values, the boxes and whiskers represent the interquartile range and 1.5 times the interquartile range, respectively, and circles are outliers outside this range.

The figure shows that the average accuracy scores are comparable across the two language groups. Both groups appear to have highest accuracy for /a/-/ɔ/ and /a/-/ɛ/ and lowest accuracy for /i/-/e/, /o/-/ɔ/ and /o/-/u/, with intermediate accuracy on /e/-/ɛ/.

In order to test for differences across the contrasts and between the two groups, a repeated measures ANOVA with language group as a between-subjects factor and BP contrast as a within-subjects factor revealed a main effect of contrast [$F(5, 190) = 66.559$, $p = <.001$, $\eta_p^2 = .637$]. There was no significant effect for language group [$F\ (1, 38) = 2.726$, $p = .107$, $\eta_p^2 = .067$] nor was there an interaction of BP contrast*group [$F\ (5, 190) = 2.024$, $\mathrm{p} = .077$, $\eta \mathrm{p}^2 = .051$]. This confirms that discrimination accuracy varies depending on the BP contrast and that there are no reliable differences between AusE and IS scores on any of the contrasts.

Given the significant effect of contrast and the non-significant group effect and interaction, as in Elvin et al. (2014), we conducted paired samples t-tests including listeners from both language groups pooled together to compare accuracy across the six BP contrasts, with $\alpha = .0033$ (Bonferroni correction for 15 comparisons, i.e., 0.05/15 paired t-tests, i.e. each BP contrast paired with the remaining five contrasts). Results indicated that /i/-/e/, /o/-/u/and /o/-/ɔ/ had significantly lower accuracy scores than the remaining three contrasts (*t*s(39) = -5.259 – 14.900, all *p*s (two-tailed) <0.001), but had comparable levels of difficulty (*t*s(39) = -.874 – -1.820, *p*s (two-tailed) = 0.076 – 0.388). While BP /e/-/ɛ/ was not easier than /a/-/ɔ/ or /a/-/ɛ/ (*t*s(39) = 4.645 – 7.412, all *p*s (two-tailed) = <0.001), it was significantly easier than the remaining three contrasts. Finally, the overall higher accuracy scores for /a/-/ɔ/ and /a/-/ɛ/ indicate that these contrasts were the easiest to discriminate, however the paired t-test that compared accuracy between these two vowel contrasts was not significant (*t* (39) = -1.421, *p* = 0.163). The order of difficulty from least difficult to most difficult is as follows: /a/-/ɔ/ ~ /a/-/ɛ/ > /e/-/ɛ/ > /i/-/e/ ~ /o/-/ɔ/ ~ /o/-/u/, where “~” means equal or comparable difficulty and “>” signifies higher accuracy.

Recall that we predicted two possible patterns of discrimination difficulty, depending on whether or not findings would be more consistent with predictions based on acoustic similarity or those based on non-native categorisation patterns as determined by the degree of perceptual overlap. As predicted by acoustic similarity and non-native categorisation patterns, BP /a/-/ɛ/ was indeed perceptually easy to discriminate. The results for BP /i/-/e/ and /o/-/u/ were in line with predictions based on acoustic similarity. In comparison, the results for BP /a- ɔ/, /e/-/ɛ/ and /o/-/ɔ/ were in line with predictions based on the perceptual overlap scores of non-native categorisation patterns. These findings were therefore consistent with predictions from a combination of both PAM and L2LP.

Turning now to an analysis of individual differences, recall that L2LP predicts that in addition to categorisation patterns, discrimination difficulty can be predicted through an analysis of acoustic similarity between the languages in question. In Experiments 1 and 2 we predicted that discrimination difficulty would depend on the degree of acoustic similarity (Experiment 1) and non-native categorisation patterns (Experiment 2). We therefore ran a Spearman's rank correlation using the acoustic overlap scores reported in Table 4.3 and perceptual overlap scores reported in Table 4.6 to test whether the degree of acoustic and perceptual similarity (determined by non-native categorisation patterns) among the listeners' native vowel system in a given BP contrast correlated with overall discrimination accuracy. We use a non-parametric way to assess correlations because the acoustic and perceptual overlap scores cannot be considered interval measures. In the analysis that investigated the correlation between acoustic overlap and discrimination accuracy we included all 20 participants from both groups, however, for the correlation between non-native categorisation and discrimination error, we include the individual data for the 20 AusE participants and 15 IS participants as we were unable to use the categorisation data of five IS listeners due to an error that occurred during testing. The results at a group level indicate a positive correlation between acoustic overlap scores and discrimination accuracy for the six BP contrasts for AusE listeners (r_s = .771, p = .036), although less reliable for IS (r_s = .714, p = .055). When both groups are analysed together we do get a reliable positive correlation (r_s = .713, p = .005). There was also a strong positive correlation between perceptual overlap scores and discrimination scores for IS listeners (r_s = .870, p = .012) and, albeit less reliable, for AusE listeners (r_s = .714, p = .055). There was also a significant correlation when both groups were analysed together (r_s = .760, p = .002). These findings indicate that in most cases, the larger the percentages of acoustic or perceptual overlapping of native vowel categories in a given BP contrast, the lower its discrimination accuracy. However, it seems that acoustics is a better predictor of discrimination difficulty for AusE listeners, whereas perceptual overlap scores based on non-native categorisation is a more reliable predictor of discrimination difficulty for IS listeners.

We further tested listeners' discrimination accuracy at an individual level. Figure 4.4 shows the individual variation among AusE (left) and IS (right) accuracy scores for the six BP contrasts. It can be observed that while individuals in both language groups differ in terms of their accuracy scores, the variation appears to be systematically constrained, in that they generally found the same contrasts perceptually easy/difficult to discriminate. That is, individuals in both groups have higher accuracy scores for /a/-/ɔ/ and /a/-/ɛ/ than for /i/-/e/ and /o/-/u/.

We tested whether the degree of acoustic or perceptual overlapping of BP vowels to native categories also correlated with discrimination accuracy at the individual level. We did so by calculating individual acoustic and perceptual overlap scores for each BP contrast (i.e., calculating the smaller percentage of times that both BP vowels in a given contrast were categorised as the same native vowel category) and plotted them against individual discriminations scores. Thus, we would expect that a large degree of overlap (acoustic or perceptual) should correspond with increased discrimination errors. Figure 4.6 shows the individual variations of acoustic (left) and perceptual (right) overlap scores, plotted against the percentage of discrimination errors.

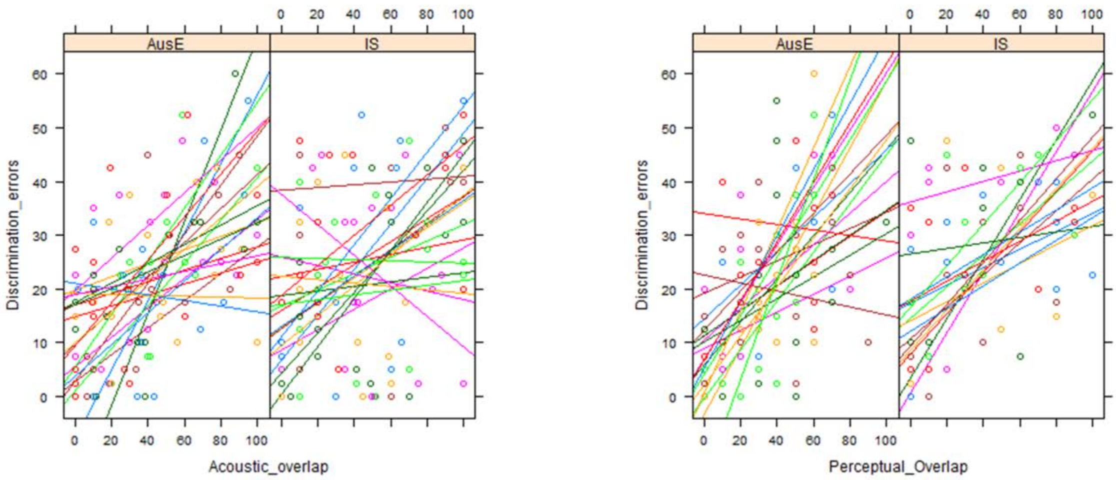


Figure 4.6 Individual variation for acoustic (x-axis, left) and perceptual (x-axis, right) overlap scores, plotted against the percentage of discrimination errors (y-axis) for each BP vowel contrast. Regression lines have been fitted for each individual

Figure 4.6 indicates a large amount of individual variation in the relationship between the degree of acoustic overlap and discrimination difficulty as well as perceptual overlap and discrimination difficulty. Visual inspection of the plots indicates that the strength of the relationship between the degree of acoustic and/or perceptual overlap and discrimination difficulty varies between individuals. In fact, we even have a few instances where there is a negative relationship in which a higher degree of acoustic and/or perceptual overlap corresponded with lower error scores. Despite some inconsistencies for a couple of participants, an overall trend can be observed in which a large percentage of perceptual and acoustic overlap seems to correspond with increased errors in non-native discrimination. We again ran a Spearman's rank correlation for each individual participant to determine the relationship between acoustic similarity and perceptual overlap and discrimination based on their overlap and discrimination scores. Results indicate a positive correlation between acoustic similarity and discrimination difficulty for all but two AusE listeners (one with no correlation and the other with a negative correlation), and for all but four IS listeners (three with negative correlations and one with no correlation). In addition the results indicate a positive correlation between perceptual overlap and discrimination difficulty for all but two AusE listeners and all but one IS listener. When averaging the individual correlations for each participant, the results indicate a positive correlation between individual acoustic overlap scores and discrimination accuracy for the six BP contrasts (AusE: r_s = .569, SD: .315; IS: r_s = .366, SD: .350 and AusE: r_s = .601, SD: .289; IS: r_s = .441, SD: .315 when excluding participants with a negative correlation). Positive correlations are also identified indicating a link between perceptual overlap scores and discrimination scores (AusE: r_s = .526, SD: = .359; IS: r_s = .609, SD: = .301 and AusE: r_s = .608, SD: = .272; IS: r_s = .688, SD: = .207 when excluding participants with negative correlations), with a stronger correlation for IS listeners. When, comparing the correlations at a group and individual level, it seems that the results for group reveal slightly stronger correlations between the degree of acoustic and perceptual overlap and discrimination difficulty than for correlations at an individual level. This is not unsurprising given the fact that we do observe individual variation for both groups as well as some instances where individual participants do not follow the trend. However, our results indicate that the implicational relationships that characterise the different groups also characterise individuals, even as overall accuracy varies from one person to the next. Importantly, for the majority of our listeners' their discrimination accuracy correlates with predictions based on acoustic similarity and perceptual similarity at both a group and individual level. Interestingly, based on observations of Figure 4.6 and the Spearman's rank correlations for individual differences, it seems that some participants show a stronger positive correlation for either acoustic similarity or perceptual overlap and in some cases both. This particularly evident among IS listeners who show a stronger positive correlation for the relationship between perceptual overlap and discrimination difficulty and AusE listeners who show a stronger positive relationship between acoustic similarity and discrimination difficulty. This finding indicates that while acoustic similarity and non-native categorisation are linked to discrimination difficulty, the strength of this relationship varies among individuals and across language groups. Although the strength of the relationship between acoustic similarity and perceptual similarity differs across the two-language groups in the present study, we do not exclude the possibility that this may be irrelevant in the long run (i.e. in future studies of non-native vowel perception) as it may be possible that the acoustic metrics used to investigate the similarity between the native and target languages in the present study may be limited.

4.6. General Discussion

The present study directly tested the L2LP model by determining whether detailed acoustic comparisons using the participants' own native production data successfully predicts AusE and IS listeners' non-native discrimination and categorisation patterns at both a group and individual level. Our findings indicate that acoustic similarity based on a comparison of the listeners' own native productions and the same L2 production data used as stimuli largely consistent with the listeners' non-native categorisation and discrimination patterns. In particular, for AusE listeners, acoustic similarity successfully predicted cases of the SUBSET scenario in that each BP vowel was categorised to multiple categories of L1 vowels, as expected. In line with our acoustic analyses, we also find cases where this SUBSET scenario contributed to perceptual overlapping in the L2 contrast. In line with acoustic predictions, instances of L2LP's NEW scenario, in which two L2 categories were mapped on to the same native category, were identified for IS listeners. Interestingly, the acoustic comparison also successfully predicted that BP /e/ and BP /o/ would be mapped to two native IS categories, which contributed to perceptual overlapping.

We do find some discrepancies between our acoustic predictions and the group results of the non-native categorisation experiment (Experiment 1), particularly for the AusE listeners, in that some responses that were given in categorisation were not predicted by acoustic similarity. Furthermore, we observe a stronger relationship between the predicted probabilities and non-native categorisation patterns at the individual level for IS listeners than we do for AusE listeners. These discrepancies could be a result of the limitations of the cross-language discriminant analysis that we used to predict acoustic similarity, which does not factor in the acoustic and auditory cue-weighting, which is necessary because in non-native categorisation, listeners may weigh some acoustic and auditory cues more than others. These differences might also be explained in terms of the influence of orthographic labels in non-native categorisation or the number of response options. For instance, our acoustic analyses are not influenced by orthography, whereas listeners were presented with orthographic labels to represent each native vowel category in the non-native categorisation task. The influence of orthography on vowel perception has been demonstrated in Escudero & Wanrooij (2010) where Spanish learners of Dutch exhibited different patterns of vowel categorisation across an auditory only and auditory with orthography task (for a full review see: Escudero & Wanrooij, 2010). Additionally, given their larger vowel inventory, the AusE listeners had a greater number of options to choose from when categorising BP vowels and as a result an increased chance of electing a greater number of native categories per vowel. For example, Benders et al. (2012) investigated the influence of the number of response categories in vowel categorisation and found that listeners presented with two response categories were more sensitive to broad and local acoustic contexts and were able to shift their boundaries between the /i/ and /e/ vowel categories quicker than listeners presented with five response categories. Therefore it may be the case that although acoustic similarity partially predicted cases of the SUBSET scenario for AusE learners, the instances where responses given in non-native categorised differed to those predicted by acoustics could be a result of the listeners making use of the multiple response options provided, or because they are weighing some cues more than others. Nevertheless, we find that, overall; both acoustic overlap and perceptual overlap scores positively correlate with discrimination accuracy. Specifically, we found that the more acoustic or perceptual overlap of native categories present in a BP contrast, the more difficult that contrast will be to discriminate. The finding of a positive correlation between acoustic overlap and discrimination accuracy, which was comparable to the perceptual overlap correlation, indicates that measuring acoustic similarity between the native and target language is an effective method of predicting discrimination difficulty as posited by the L2LP model.

Interestingly, our study found that the Spanish listeners also displayed a case of the SUBSET scenario, in that there was a perceptual overlap for the BP /o/-/u/ contrast: both BP /o/ and /u/ were acoustically similar to and categorised as the same two native vowel categories, namely IS /o/ and /u/. This finding suggests that the SUBSET scenario, which is commonly found in cases where the L1 is larger than that of the L2, can also be found in cases where the L1 is smaller than that of the L2. This finding has important implications for research relating to L2 perceptual difficulty with vowels, as it provides further evidence that L1 vowel inventory size alone cannot account for relative discrimination accuracy. Instead, discrimination accuracy is largely dependent on non-native categorisation patterns, which we have shown are correlated with the acoustic and/or perceptual similarity between the native and target language vowels.

Both the present study and Elvin et al., (2014) predict and subsequently find that both groups found the same BP contrasts easy or difficult to discriminate. However, the exact patterns of discrimination differed slightly across the two studies as observed in Table 4.7. Furthermore, unlike Elvin et al. (2014), IS listeners did not have higher accuracy overall. Importantly, the minor discrepancies between the two studies are in fact accounted for in the predictions based on both acoustic similarity and non-native categorisation. Recall that the L2LP model indicates that listeners have different developmental paths. The fact that we do see some difference between these listeners and those from Elvin et al. (2014) suggests the importance of considering individual variation when predicting L2 difficulty and that that such predictions should be tailored to the individuals or groups actually tested.

Table 4.7 Reported patterns of discrimination accuracy in Elvin et al. (2014) and the present study beginning from perceptually easy to perceptually difficult.

	Order of discrimination difficulty
Elvin et al., 2014	/a/-/ɛ/ > /a/-/ɔ/ ~ /e/-/ɛ/ ~ /o/-/ɔ/ > /i/-/e/ ~ /o/-/u/
The present study	/a/-/ɛ/ ~ /a/-/ɔ/ > /e/-/ɛ/ > /o/-/ɔ/ ~ /i/-/e/ ~ /o/-/u/

In our study we did find that at a group level, predictions of discrimination difficulty for the BP contrasts /i/-/e/ and /o/-/u/ were in line with predictions based on acoustic similarity and predictions for /a/-/ɔ/, /e/-/ɛ/ and /o/-/ɔ/ were more in line with those based on perceptual overlap scores which were, in turn, based on non-native categorisation data. Interestingly, the BP contrasts that were among the most difficult to discriminate and successfully predicted by acoustic similarity were those containing corner vowels. Although acoustic similarity predicted discrimination difficulty for BP /i/-/e/ and /o/-/u/, another framework that could explain this difficulty is the Natural Referent Vowel (NRV) framework (Polka & Bohn, 2011). The framework assumes that directional asymmetries may influence discrimination difficulty and that discrimination of a vowel contrast which involves a change from a peripheral vowel to a more central vowel should be more difficult to detect than discriminating a change from a central to peripheral vowel. We tested this framework on our data with a repeated-measures ANOVA which indeed showed an effect of directional asymmetries; however our analysis of each contrast only partially supported the NRV hypothesis. That is, we did find that directional asymmetries led to AusE listeners' lower discrimination accuracy for BP /i/-/e/, but not for IS and not for BP /o/-/u/ for either group. Surprisingly, we also found an unexpected effect for /a/-/ɔ/ (p = 0.007) where AusE listeners had *lower* discrimination accuracy when the more central vowel was presented first. Further investigations would be required to fully test whether the NRV framework successfully explains AusE and IS listeners' discrimination of BP vowels.

The inclusion of individual native production data in the present study proved useful for the examination of the relationship between acoustic similarity, non-native categorisation and discrimination at an individual level. In particular, we found that the individual listeners' non-native categorisation patterns were largely consistent with the predictions based on acoustic similarity between their own native vowel productions and the target BP vowels. In addition, we found that the amount of acoustic and perceptual overlapping in a given BP contrast correlated with individual listeners' performance in the discrimination task. Crucially, we find that acoustic comparisons based on the individuals' own native production data confirms the L2LP model claims and also replicates previous studies (e.g., Levy & Law, 2010) that for the most reliable predictions of L2 perception (both non-native categorisation and discrimination), acoustic similarity should be based on data collected from the same listeners intended for testing, using the same consonantal context used as stimuli in the perceptual tasks.

Our analyses of individual differences are also in line with those of Mayr and Escudero (2010) as we found that although perception differs among individuals, trends in relation to the contrasts that are easy and/or difficult to perceive can be found across both populations. Interestingly, in our examination of the relationship between the degree of acoustic similarity and non-native categorisation (or perceptual overlap) at an individual level, it seems that some speakers show a stronger effect of acoustic overlap, while others show a stronger effect of perceptual overlap on discrimination difficulty. In particular, the data suggests that perceptual overlap is a better predictor for IS listeners than acoustic similarity and that the opposite is true for AusE listeners. It therefore seems that findings for IS listeners are more in line with predictions based on PAM's theoretical framework and that findings for AusE listeners are more in line with predictions based on L2LP. This finding may have important theoretical implications for both of these models; however, we would first need to confirm that these findings are not simply related to limitations in the analyses of acoustic similarity and task design for non-native categorisation.

For example, it could be that perceptual overlap scores do not predict as well for AusE listeners because, as previously mentioned, they have many more response options and they may be selecting some of these options at random. Additionally, because there were only a small number of response options for IS in the discriminant analyses it could be that the probabilities of categorisation were higher because the model forces categorisation even if none of the categories are a good match. If AusE listeners are choosing some options at random and the discriminant analysis model is indeed forcing categorisation for IS this might result in reduced perceptual overlap scores for AusE and inflated acoustic scores for IS. Importantly, the discrimination analyses and non-native categorisation task were based on one repetition of each vowel per speaker (7 vowels x 10 speakers). To rule out the possibility of this finding being task or design dependent, it would be worth investigating whether listeners (and in particular AusE listeners) would be more consistent in their responses if they heard multiple repetitions by the same speaker. Furthermore, perhaps the inclusion of multiple repetitions of each vowel by the same speakers in the discriminant analysis might not force the probabilities as much for IS listeners. By doing so, this might result in less inflated acoustic overlap scores for IS listeners and lead to a stronger correlation between acoustic predictions and discrimination difficulty.

If these design limitations are controlled and we still find that for one language group or for some BP contrasts either acoustic or perceptual overlap is a better predictor, then this would indeed have important theoretical implications for models of speech perception such as PAM and L2LP. In particular, one would need to determine why some individuals or language groups indicate a stronger preference for one predictor over the other. It also raises the question as to whether or not discrimination difficulty is dependent on a combination of acoustic similarity between the native and target languages as well as the individual's perception of these non-native sounds.

In sum, our findings indicate that listeners' non-native categorisation and discrimination patterns were largely predicted by a detailed acoustic comparison of the native and target languages, with data collected from the same populations for both vowel productions and perceptual testing. Importantly, we find that AusE listeners do not have an advantage when perceiving non-native vowels despite their native language (English) having a larger and more complex vowel inventory than that of the IS listeners (Spanish). In fact, we find that listeners' discrimination patterns are largely dependent on the L2LP learning scenario identified for each vowel contrast, which were similar across the two language groups. That is, contrasts which contained evidence of L2LP's NEW or SUBSET scenarios (containing an acoustic or perceptual overlap where the two vowels in a given BP contrast are acoustically similar or categorised to two or more of the same native categories) resulted in similar discrimination difficulties for IS and AusE listeners as both scenarios resulted in a neutralisation of the non-native contrast. In addition, both of these learning scenarios are likely to be more difficult than contrasts where only the SIMILAR scenario is present. These findings are also consistent with previous studies (e.g., Bohn et al., 2011; Levy, 2009; Tyler et al., 2014) that test PAM's theoretical predictions that a higher degree of perceived phonetic similarity (i.e., perceptual overlap) between members of a non-native contrast, as observed by perceptual assimilation patterns, is associated with a greater level of discrimination difficulty for that contrast. While research following L2LP's theoretical framework has predominately focused on predicting vowel perception through comparisons of acoustic similarity, the same should also apply for consonants, as has been previously shown in studies that investigate phonetic similarity from PAM's theoretical perspective. Therefore, future studies might also consider testing how well acoustic similarity predicts perceived phonetic similarity for other types of speech sounds such as stops and fricatives. In addition, the fact that our findings at a group level are consistent with both PAM and L2LP predictions indicates that the findings at an individual level may also be consistent w2ith the predictions of both theoretical models. We have shown in the present study that our findings at an individual level are indeed consistent with the L2LP model's claim regarding individual differences; however, future studies examining the hypotheses from PAM may also benefit from an analysis of individual variation as previously proposed by Tyler et al. (2014).

Our results indicate that the same learning scenarios can exist for the same non-native vowel contrasts across participants' languages even if one native language offers more potential opportunities for one learning scenario over another. This is because it is acoustic and/or perceptual similarity which largely determines these learning scenarios and not solely the size of the native vowel inventory. Thus, as seen in the present study, the languages of two learner groups may be comparable in relation to the target language, despite their differing vowel inventory sizes. Furthermore, we acknowledge that acoustic similarity as we have defined it and measured it in this paper would be expected to correlate highly with articulatory similarity, i.e., F1 values reflect tongue height, F2 values reflect the backness of the tongue, and dynamic changes in formant values reflect changes in tongue height/backness over the duration of the vowel. The PAM model in particular predicts that perceivers detect and use such articulatory properties in determining cross-language phonetic similarities (e.g., Best & Tyler, 2007; Best, 1995). Thus, future work might further aim to examine correlations between articulatory and acoustic properties of the target vowels and evaluate their contributions to perceptual assimilation and discrimination patterns.

Finally, in order to adequately test that the L2LP model and other theoretical models such as SLM and PAM that are also applicable to non-native word recognition and production, future studies should compare the data from the present study with these same listeners' spoken word recognition and non-native vowel production data. Indeed, some prior studies (e.g., Broersma, 2002; Escudero, Broersma, & Simon, 2013; Escudero, Hayes-Harb, & Mitterer, 2008; Pallier et al., 2001; Weber & Cutler, 2004) have shown that vowel contrasts that are difficult to perceive are also difficult in spoken word learning and word recognition tasks. Therefore, we would expect that these same listeners' patterns of discrimination difficulty could be used to predict difficulty in a spoken word learning and word recognition task containing the same pseudo-words used in the present study. Furthermore, the findings from the present study could also apply to non-native speech production. In particular, the SLM and L2LP theoretical models claim that listeners' production of non-native or L2 sounds is influenced by their perception of these sounds in the L1. However, to date, most studies (e.g., Diaz Granado, 2011; Flege, Mackay, & Meador, 1999) have used the theoretical framework of SLM to test L2 production. Considering that the L2LP model posits a direct link between non-native and L2 production, it is perhaps surprising that very few studies have tested this claim (e.g., Rauber et al., 2005). Thus the acoustic analyses from the present study and in particular the acoustic and perceptual overlap scores could be used to predict the patterns in these same listeners' non-native vowel productions. In conclusion, future research is required that adequately tests the L2LP model predictions for the role of perception in both word recognition and non-native production, and compares them to other models of non-native and L2 speech perception and production.

4.7. Chapter Summary

This chapter investigated whether detailed acoustic comparisons using the participants' own native production data successfully predicts AusE and IS listeners' non-native discrimination and categorisation patterns at a group and individual level. The results indicate that AusE listeners do not have an advantage when perceiving non-native vowels despite their larger and more complex native vowel inventory. In fact, listeners' discrimination patterns are largely dependent on the L2LP learning scenario identified for each vowel contrast, which were similar across the two language groups. The findings also indicate that acoustic similarity based on a comparison of the listeners' own native productions and the same L2 production data used as stimuli was largely consistent with the listeners' non-native categorisation and discrimination patterns. Specifically, listeners' non-native categorisation patterns were largely consistent with the predictions based on acoustic similarity between their own native vowel productions and the target BP vowels. However, there was a stronger relationship between the predicted probabilities and non-native categorisation patterns at the individual level for IS listeners than for AusE listeners. In addition, the amount of acoustic and perceptual overlapping in a given BP contrast positively correlated with individual listeners' performance in the discrimination task.

In Chapter 5, the interrelation between non-native speech perception and spoken word recognition is investigated. Specifically, individual listeners' performance on non-native discrimination patterns is compared with their performance in word recognition. The interrelation between these two abilities can be reliably tested as the stimuli and participants were consistent across both studies.

Chapter 5

The role of speech perception and individual differences when learning novel words in a non-native language: English vs Spanish learners of Portuguese

(Elvin, Escudero, Williams, Shaw & Best, under review b) [11]

5.1. Introduction

Successful communication in a second language (L2) is largely dependent on the learner's ability to recognise words in spoken language (McQueen, 2007). Developing the ability to recognise the sounds that make up a word and then connecting those sounds together to become meaningful often poses a great degree of difficulty for many L2 learners. Spoken word recognition refers to the process of mapping or connecting an incoming speech signal (or spoken word) with its *lexical* representation in the mental lexicon, thus providing a bridge between sounds and meaning (Marslen-Wilson, 1987). In this way, spoken word recognition differs from speech perception, a process that generally involves categorisation i.e., mapping incoming speech signal to its *phonological* representation(s) and/or discrimination, i.e., distinguishing between two contrasting phonemes. Spoken word recognition is therefore an important stage in L2 development as word forms must first be learned and stored in the mental lexicon before they can be connected to their lexical meaning or recognised in speech.

Various theoretical models have been developed to explain spoken word recognition and by doing so, make assumptions about how lexical representations are stored in the mind. They include: the Cohort model (Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980)

[11] A version of this chapter is under review in the Journal of Memory and Language

TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994; Norris, Mcqueen, Cutler, & Butterfield, 1997) and its extension to Shortlist-B (Norris & McQueen, 2008), the neighbourhood activation model (Luce & Pisoni, 1998) and its extension to PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000) among others. All share the common assumption that spoken word recognition involves multiple word activation and the subsequent competition between activated words (Weber & Scharenborg, 2012). When a listener hears a spoken word, a number of different possible word forms are assumed to be activated in the mental lexicon. The listener must then select one of these activated words as the lexical representation of that spoken word, based on the likelihood of it being correct (Auer, 2009; McQueen, 2007). Spoken L2 word recognition difficulty increases when there are more active candidates resulting in more competition between words and slower word recognition (Vroomen & de Gelder, 1995; Weber & Cutler, 2004).

Increased activation and competition in spoken word recognition can be caused by a number of factors including (but not limited to): lexical stress (Cutler & Pasveer, 2006; Soto-Faraco, Sebastián, & Cutler, 2001), language experience (Escudero, Broersma, & Simon, 2013), orthographic influences (Escudero, Simon, & Mulak, 2014; Hayes-Harb, Nicol, & Barker, 2010; Taft, Castles, Davis, Lazendic, & Nguyen-Hoan, 2008), neighbourhood density (Luce & Pisoni, 1998) and dual activation or similarity between L1 and L2 vocabularies (Dijkstra, Timmermans, & Schriefers, 2000; Dijkstra, Van Jaarsveld, & Brinke, 1998; Weber & Cutler, 2004). For example, Luce and Pisoni (1998) found that adults were slower at recognising words in high density lexical neighbourhoods than those with sparse lexical density. Or in other words recognition was slower when there were a number of items that were highly similar to the target word. Additionally, Weber and Cutler (2004) found that in one of four eye-tracking experiments, Dutch listeners fixated distractor pictures of which the Dutch name resembled the English name of the target picture. Weber and Cutler (2004) also found that non-native listeners' phonetic discrimination difficulties caused inappropriate competitor activation. Specifically, Dutch listeners fixated distractor pictures more often and for a longer amount of time when the English names of the target and distractor pictures contained vowels that for Dutch listeners are confusable.

Difficulty in recognising words in a second language may increase when they form minimal pairs, i.e., two words differing by only one phoneme (e.g., the vowels in /sɪt/ and /sæt/). Minimal word pairings are considered particularly difficult when they contain a phonological contrast that learners struggle to detect. This is supported by a number of studies that have found that when word pairings contain difficult minimal pairs, there is an increase in word recognition difficulty for L2 learners (Broersma, 2002; Escudero, et al., 2012; Escudero, et al., 2008; Pallier et al., 2001; Weber & Cutler, 2004). For example, Japanese learners are known to have difficulties recognising the difference between minimal word pairings that contain the English /r/ and /l/ contrast (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). In addition, Broersma (2002) found that Dutch learners of English, who find the English /æ/-/ɛ/ vowel contrast difficult to perceive, judged non-words containing members of the aforementioned contrast more often as real words than native listeners. This finding has been interpreted to indicate that when the word pairing contains the English /æ/-/ɛ/ contrast, the two words are perceived by L2-English Dutch listeners as homophones, resulting in poorer performance in word recognition (Cutler & Broersma, 2005). Escudero et al. (2008) likewise found that Dutch listeners confused words containing /æ/ and /ɛ/ when they heard the auditory forms only.

From the above examples we see that the native language can influence both L2 speech perception and spoken word recognition. While many of the aforementioned models of spoken word recognition consider the role of speech perception (i.e. L2 phonetic categorisation and discrimination) in spoken word recognition, it is important to consider how the theoretical models that explain L2 speech perception may also be applicable to word learning and spoken word recognition. Three key models of non-native and L2 speech perception are the Speech Learning Model (SLM; Flege, 1995), the Perceptual Assimilation Model's (PAM; Best, 1994, 1995) extension to L2 acquisition (PAM-L2; Best & Tyler, 2007) and the Second Language Linguistic Perception model (L2LP; Escudero, 2005, 2009; van Leussen & Escudero, 2015). These models all claim that the phonetic categories in a learner's first language (L1) largely influence their perception of phonetic categories in the L2. While SLM focuses predominately on *individual* phonetic categories, PAM and L2LP make specific predictions regarding the perceptual development of L2 *contrasts* according to the influence of the L1 (van Leussen & Escudero, 2015). For example, a learning scenario[12] that involves two non-native phonemes that are perceived as one single native category is known as the NEW[13] scenario in L2LP and as single-category assimilation in PAM. Both models predict this scenario to be difficult for naïve and L2 learners as they must either create a new L2 category for one of the contrasting L2 phonemes or split an existing L1 category (van Leussen & Escudero, 2015). An easier learning scenario is known as the SIMILAR scenario in L2LP, or two-category assimilation in PAM. In this scenario, the two L2 phonemes in a contrast are mapped on to two separate native categories. This learning scenario is predicted to be far less problematic than the NEW one as L2LP posits that listeners will simply replicate their existing L2 categories and adjust their boundaries to match those of the target language (van Leussen & Escudero, 2015). Another possible learning scenario is the SUBSET scenario in L2LP which occurs when the two phonemes in the L2 contrast are

[12] The L2LP model refers to patterns of non-native categorisation as "learning scenarios". However the technical term for each scenario does not include the word "learning". Throughout this study we refer to the three scenarios as the NEW scenario, SIMILAR scenario and SUBSET scenario.

[13] The L2LP terms "NEW" and "SIMILAR" should not be confused with SLM's use of these terms. SLM posits that when listeners are presented with an L2 phoneme that does not closely resemble any L1 category they form a *new* category which should be easier to acquire than an L2 phoneme that is similar to an existing L1 category, which despite the phonetic differences, should be more difficult to acquire (Colantoni et al., 2015). In contrast, as mentioned above, a SIMILAR scenario in L2LP is predicted to be much easier to acquire than a NEW scenario.

perceived as two or more native L1 categories. The SUBSET scenario may be comparable to focalised, clustered or dispersed uncategorised assimilation in PAM (Faris et al., 2016). A learning scenario of this type is thought to result in little difficulty if the two L2 vowels are not perceived as the same multiple categories (i.e., no perceptual overlap). If no perceptual overlap occurs, the learning scenario functions in the same way as the SIMILAR scenario. However, if the SUBSET scenario contributes to a perceptual overlap (where the two phonemes in the non-native contrast are perceived as the same multiple native categories) this learning scenario is comparable to the NEW scenario and may be difficult for learners to discriminate.

Given the fact that many difficulties with speech perception are consistent with difficulties identified in spoken word recognition, it is important to investigate how these models of speech perception relate to spoken word recognition. Of the aforementioned theoretical models of L2 speech perception, SLM does not explicitly reconcile how perceptual difficulties relate to word recognition and although PAM-L2, incorporates the role of the lexicon in L2 perception, its consideration of lexical involvement is limited to predicting that vocabulary size determines success in L2 speech perception (Best & Tyler, 2007; Bundgaard-Nielsen, Best, & Tyler, 2011). The L2LP model is the only theoretical model to consider *individual* learners at all stages of development (from the initial state to ultimate attainment) across speech perception, spoken word recognition and speech production. Specifically, the model's *optimal perception hypothesis* states that a learner's initial state of L2 learning is geared towards optimally perceiving and producing the sounds in the particular variety of their native language (Escudero, 2005, 2009), hence the possible learning scenarios for various L2 contrasts outlined above are centred on the individual's native language. While these scenarios specify the learning tasks with which a learner is faced in order to successfully acquire L2 contrasts, they do not in themselves account for what triggers the particular learning trajectory. Crucially, for perceptual development to occur, a learner must be able to recognise L2 phonological categories and how these are used across the L2 lexicon. In this way, the L2LP model states that while initial L2 perception is based on the optimal perception hypothesis, perceptual learning takes place because it is *message-driven* (or *meaning-driven*[14] in the revised L2LP: van Leussen & Escudero, 2015) and is in fact compatible with models of L1 acquisition (e.g., PRIMIR: Werker & Curtin, 2005) as learners attempt to improve recognition by updating their lexical representations.

The revised L2LP model (van Leussen & Escudero, 2015) sets out to improve previous computational implementations of the L2LP theoretical proposal for meaning-driven perceptual learning (Escudero & Boersma, 2004; Weiand, 2007) as well as the specific connections between possible activated candidates, specifically in the SUBSET scenario. To this end, van Leussen and Escudero (2015) modelled the acquisition of the Spanish /i/-/e/ contrast by Dutch learners. The Spanish /i/-/e/ contrast provides an example of the SUBSET scenario, as initially Dutch learners perceive both Spanish /i/ and /e/ as Dutch /ɪ/, but they also perceive Spanish /i/ as Dutch /i/ and Spanish /e/ as Dutch /ɛ/ (Escudero & Boersma, 2002). Van Leussen and Escudero (2015) therefore modelled the learning task required here, namely a reduction in perception of both Spanish phonemes as Dutch /ɪ/ to acquire Spanish /i/ and /e/ as two distinct L2 categories, by simulating Dutch learners on the basis of acoustic data from Dutch and Spanish speakers and Spanish minimal word pairs. Additionally, following long-standing debates in psycholinguistics with respect to the information flow from the speech signal to the lexicon, the authors explored two possibilities in recognising L2 categories: it may be either strictly bottom-up (*sequential*) or it may allow lexical feedback to perception (*interactive*). That is, sequential word recognition involves a two-step process whereby sequential learners will always evaluate the acoustic → phonetic connections or

[14] We will use the revised L2LP terminology of *meaning-driven* perceptual learning from hereon.

mappings of perception (step 1) before the phonetic → phonemic and phonemic → lexical connections of recognition (step 2), whereas interactive processing occurs when all these connections are evaluated together (van Leussen & Escudero, 2015).

The results from van Leussen and Escudero's (2015) simulations indicate that meaning-driven learning successfully predicts the developmental path of L2 phoneme perception, as outlined by L2LP's SUBSET scenario, and that performance improves in line with the more input a learner receives. With respect to the two types of processing (sequential or interactive), both ultimately lead to correct L2 recognition, but different predictions are made regarding phonetic representations and, consequently, the amount of exposure to L2 words required to acquire the L2 contrast. Specifically, sequential learners are predicted to retain an L1 phonetic category for certain "boundary" stimuli (that is, L2 perception will remain filtered by the L1 for intermediate vowels, i.e., those which fall on the border between two L2 categories), while interactive learners will ultimately adapt their vowel system more closely to the L2. Accordingly, the results showed that simulated sequential learners needed a larger amount of input data (exposure to L2 words) to achieve the same level of success with the L2 contrast as interactive learners. Van Leussen and Escudero (2015) acknowledge that either strategy may occur in real L2 learners and would be identifiable in the individual differences between learners, as discussed below. In summary, the updated L2LP model expands on the original model with an improved computational implementation of meaning-driven learning in perceptual development and spoken word recognition.

Although the L2LP model claims that individual learners are likely to follow different developmental paths, most studies that investigate L2 perception and spoken word recognition have predominately focused on groups of learners rather than individuals. Of those studies that do investigate individual differences among L2 learners, the focus is typically on extra-linguistic factors such as age of acquisition and language use and/or experience (e.g., Escudero & Boersma, 2004; Flege, Munro, & Mackay, 1995). However, even when these factors are controlled for, individual variation is still present (Jin et al., 2014; Sebastián-Gallés & Díaz, 2012). In fact, an analysis of individual differences may be beneficial for data that may be difficult to interpret at a group level (Smith & Hayes-Harb, 2011). Some studies have also shown the importance of individual differences for determining and understanding the variable developmental trajectories observed in L2 learning (Elvin, Escudero, Williams, Shaw, & Best, under review a; Mayr & Escudero, 2010; Wanrooij, Escudero, & Raijmakers, 2013). For example, Elvin et al., (under review a) investigated the relationship between acoustic similarity of Brazilian Portuguese (BP) vowels to the listeners' native Australian English (AusE and Iberian Spanish (IS) listeners' vowels and their non-native categorisation and discrimination of BP vowels at both a group and individual level. The authors found that AusE and IS listeners' non-native categorisation patterns were largely predicted by acoustic similarity at both a group and individual level. Furthermore, the results indicate a wider range of individual variation for AusE listeners than for IS listeners, which was unsurprising considering that IS has a smaller vowel inventory and fewer opportunities for variation than AusE. A positive correlation between discrimination and acoustic similarity as well as between discrimination and non-native categorisation (i.e., perceptual similarity) was found at both an individual and group level. Although the strength of the relationship between discrimination difficulty and acoustic similarity as well as non-native categorisation varied between individuals, listeners followed a similar trend in their patterns of discrimination difficulty which were also in line with the findings at a group level.

The present study aims to directly test the L2LP model's predictions that non-native or L2 contrasts which are difficult to perceive are also confused in spoken word learning and recognition at the initial state of L2 learning. In particular, we investigate whether naïve

Australian English (AusE) and Iberian Spanish (IS) listeners' word learning and recognition of spoken Brazilian Portuguese (BP) novel words can be predicted by their own patterns of L2 discrimination difficulty.[15] Both the language groups and target languages were selected on the basis of their differing vowel inventory sizes. AusE has thirteen monophthongal vowels in its inventory, namely, /iː, ɪ, ɪə , e, eː, ɜː, ɐ, ɐː, æ, oː, ɔ, ʊ and ʉː/ and is a representative language that has a larger vowel inventory than the target language, BP, which has seven oral vowels, /i, e, ɛ, a, o, ɔ, u/. Iberian Spanish was selected because it has a smaller vowel inventory than the target language, consisting of five vowels /i, e, a, o, u/.

Previous studies (Elvin et al., 2014; Elvin et al., under review a) have found that despite their differing vowel inventories AusE and IS listeners have similar difficulties in non-native discrimination. Here we investigate whether these similarities also extend to word recognition. This study is, to our knowledge, the first to directly test the L2LP model predictions using word recognition data collected from the same participants for whom speech perception (specifically non-native categorisation and discrimination) data are also available, using the same stimuli across all tasks. This allows for a direct test of the L2LP model's claim that accurate predictions require data collected from the same participants involved in perceptual testing, allowing for the evaluation of L2 development for individual learners. Given the fact that the L2LP model's optimal perception hypothesis explicitly addresses differing developmental paths among individual listeners (based on L1 and L2 relationships), in order to build on the Elvin et al. (under review a) study, we will analyse whether the role of speech perception in spoken word recognition holds at both group and individual levels.

[15] It is important to note that although most traditional spoken word recognition tasks involve the recognition of familiar words; our study involves the learning of and subsequent recognition of novel words. Although our task is in fact a novel word learning and recognition task, we follow previous studies such as Escudero, Simon & Mulak (2014) and Escudero (2015) (which also investigate novel word learning and recognition) and refer to the task that follows the learning phase as a "spoken word recognition task".

Importantly, the L2LP model proposes a link between speech perception and spoken word recognition in that non-native contrasts which are difficult to perceive are the same contrasts which are confused in spoken word recognition. However, this does not necessarily mean that the model would predict a straightforward correlation between these two abilities. This is because factors such as training effects, different task demands and different learning strategies could influence the exact nature of the correlation. For example, numerous studies have shown that speech perception improves after exposure to or training on non-native or L2 sounds and contrasts (e.g., Cenoz & Lecumberri, 1999; Schwab, Nusbaum, & Pisoni, 1985; Strange & Dittmann, 1984). Perceptual training has also been shown to influence L2 production (Bradlow, Pisoni, Akahane-yamada, & Tohkura, 1997). The discrimination task in Elvin et al. (under review a) and the spoken word recognition task in the present study were both conducted in the same experimental session, however, the spoken word recognition task was always presented prior to the non-native discrimination task. Thus, in the present study, we might expect to find possible training effects. That is, learners may have better performance across the six BP contrasts in discrimination than in spoken word recognition because they have already been exposed to the target words.

Task demands have also been shown to play an important role in lexical development (see the PRIMIR model of L1 acquisition: Werker & Curtin, 2005; see also: Escudero, Best, Kitamura, & Mulak, 2014). Differences in the nature of the task have also been shown to affect L2 performance (e.g., Díaz et al., 2012). Even though the present study included the same stimuli used in Elvin et al. (under review), the two tasks (i.e., non-native discrimination and spoken word recognition) differ in regards to the level of processing. While discrimination performance is influenced by the listener's L1, or their optimal perception, as it is a mainly an auditory and phonetic task, spoken word recognition involves meaning-driven learning that comprises higher-order levels of processing because learners must also associate word forms to meaning. Consequently, learners may have difficulties mapping different pre-lexical representations to their lexical counterparts (Escudero, 2011). Thus, a link between non-native speech perception and spoken word recognition may not be completely straightforward due to differing task demands.

Furthermore, there could be discontinuities between speech perception and spoken word recognition (Curtin, Goad, & Pater, 1998; Cutler, Weber, & Otake, 2006; Hayes-Harb & Masuda, 2008; Weber & Cutler, 2004) depending on the particular strategy a learner adopts to follow a perceptual development trajectory. Recall that the revised L2LP model (van Leussen & Escudero, 2015) was formulated to evaluate two proposed routes for lexical processing with computer simulations, namely sequential versus interactive. While simulated learners who applied either processing route were shown to succeed in acquiring the L2 contrast, sequential simulations were slower than interactive simulations as they required a greater quantity of L2 input. Although real listeners may display similar discrimination patterns in a non-native language, they may differ in terms of the strategies they use to improve non-native recognition by updating their lexical representations. We hypothesise that if some of our learners are sequential learners, we would expect them to display lower overall accuracy scores in spoken word recognition than interactive learners, as van Leussen and Esucdero's (2015) modelling demonstrated that the sequential process requires more exposure to the target words than the interactive process to reach the same accuracy level. By examining individual differences in human listeners in the present study, we test this precise possibility, i.e., whether or not listeners adopt different learning strategies in their recognition of BP novel words that are not apparent in speech perception alone.

Although some individual variation among participants is expected, if L2LP's proposal is correct, the findings for the present study should be in line with those reported for speech perception in Elvin et al. (under review a). In particular, in that study, the detailed comparison of acoustic similarity between AusE and IS and BP predicted that perceptually difficult contrasts would be those with a large percentage of acoustic overlap, while those with little to no acoustic overlap was predicted to be easy. An acoustic or perceptual overlap occurs when two L2 sounds that form a phonological contrast are respectively in close acoustic proximity to or are perceived as the same native category or categories. For example, Elvin et al. (under review a) found that both BP /i/ and /e/ were acoustically similar to AusE /iː/, /ɪ/, /ɪə/ and /ʉː/ and categorised as AusE /iː/ and /ɪ/, with an acoustic overlap of 73% and a perceptual overlap of 43%. For IS, BP /i/ and /e/ were acoustically similar to and categorised as IS /i/ with an acoustic overlap of 83% and a perceptual overlap of 56%. Given the larger acoustic and perceptual overlap percentages, the authors predicted that both listener groups would find this contrast difficult to discriminate. On the other hand, the authors predicted BP /a/-/ɛ/ to be perceptually easy to discriminate as both vowels in the contrast were acoustically similar to or categorised as different native categories with a 14% acoustic and an 8% perceptual overlap for AusE listeners and an 8% acoustic and 0% perceptual overlap for IS listeners. The results from the discrimination task indicated that both groups found the same contrasts perceptually easy and/or difficult to discriminate. Based on Elvin et al.'s (under review a) findings for non-native speech perception, the authors presented the following order of discrimination difficulty, ranging from perceptually easy to perceptually difficult: /a/-/ɔ/ ~ /a/-/ɛ/ > /e/-/ɛ/ > /i/-/e/ ~ /o/-/ɔ/ ~ /o/-/u/, where "~" means equal or comparable difficulty and ">" signifies a significant difference in accuracy. If there is a link between non-native speech perception and spoken word recognition, then the order of spoken word recognition difficulty should be comparable to the order of difficulty in discrimination.

5.2. Method

5.2.1. Participants

Twenty Australian English monolingual speakers from Western Sydney (AusE) and twenty Iberian Spanish (IS) monolingual speakers from Madrid participated in the present study. They were the same participants who previously participated in the study reported in Elvin et al (under review a) and were tested in the same experimental session. Participants were university students aged between 18-30 years. The AusE participants were tested at Western Sydney University and reported little to very basic knowledge of any foreign language and no knowledge of Portuguese. The IS participants were tested at the Universidad Nacional de Educación a Distancia, Madrid and although they reported basic to intermediate knowledge of English, they did not use English in their daily lives. Furthermore, they reported little knowledge of any other foreign language, including Portuguese. All participants provided informed consent in accordance with the ethical protocols in place at the Universidad Nacional de Educación a Distancia and Western Sydney University's Human Research Ethics Committee.

5.2.2. Stimuli

The stimuli for this experiment were 14 naturally produced Brazilian Portuguese (BP) novel words which were selected from Escudero et al.'s (2009) corpus focusing on the seven Portuguese oral vowels /i, e, ɛ, a, o, ɔ, u/ produced in a variety of consonantal and phrasal contexts. Of the 14 words, the seven produced in the fVfe context served as target words, namely, *fife, fefe, fɛfe, fafe, fofe, fɔfe* and *fufe*, which were the same target words previously used as stimuli in Elvin et al (under review a), and the remaining seven novel words, namely, *kɔko, kuke, pipe, popo, sase, sɛso* and *teko*, were used as distractor words. In each novel word, the vowel in the first syllable was always stressed and corresponded to one of the seven

Portuguese vowels /i, e, ɛ, a, o, ɔ, u/, while the vowel (either /e/ or /o/) in the second syllable was always unstressed. Ten (5 male, 5 female) of Escudero et al.'s (2009) 20 BP speakers produced each of the 14 novel words in isolation and a carrier sentence (see Escudero et al., 2009 for further details) and tokens from the isolated position were used as stimuli in this experiment.

5.2.3. Procedure

The present study consisted of a learning phase, followed by a testing phase. In both phases, each target and distractor word was paired with a corresponding line drawing selected from the Shatzman and McQueen (2006) corpus. Using a similar procedure to Escudero et al. (2012, 2013), participants were tested in a sound attenuated room with the auditory stimuli being played through a Beyerdynamic DT 297 PV professional headset and their corresponding line drawings appearing on a computer screen. Participants were first presented with a training phase in which they learned to associate the 14 BP novel words with their picture referents and then they were tested on their recognition of these newly learned words in a test phase.

In the training phase, each novel word was presented as a target six times over a total of 84 trials (14 words x 6 trials as target) spoken by one female and one male BP speaker. On each trial, participants were presented with one of the 14 line drawings for 2.0 seconds, coupled with audio presentation of its corresponding novel word, followed by a blank screen (500ms) during which the same corresponding novel word was repeated. Immediately after, participants were presented with the same line drawing paired with another line drawing. Participants were required to select the correct referent image, either the one on the left or right side of the screen, by pressing keys labelled "L" or "R" ("Izda." or "Dcha." i.e., *left* and *right* translated in Spanish) on the keyboard.

In the test phase, each trial consisted of one of the novel words (produced by one of 10 BP speakers) being played as a target word and its associated line drawing and a distractor image being displayed on the screen. Participants were required to indicate whether the word that they heard corresponded to the line drawing on the left or right side of the screen by pressing keys labelled “L” or “R” (“Izda.” or “Dcha.” i.e., *left* and *right* translated in Spanish) on the keyboard. Each novel word was presented as a target 26 times (that is, the novel word was presented twice with each of the other 13 novel words line drawings as the distractor). Following previous studies (Escudero et al., 2012; 2013) the 91 possible unique pairings of the 14 words occurred four times, with each item being presented as the target twice. The total number of trials was 364 (14 words x 26 trials as a target) and comprised 280 non-minimal trials (70 pairs x 4 presentations), 60 perceptually easy trials (15 pairs x 4 presentations) and 24 perceptually difficult trials (6 pairs x 4 presentations). Participants were given six practice trials in the word learning phase and fourteen practice trials (i.e., a practice of each BP vowel by the male and females speakers) before each phase. They then worked through the task at their own pace and took an average of around 20 minutes to complete[16].

5.3. Results

Here we report how often participants selected the correct line drawing associated with each novel word. The average accuracy across all trials for AusE listeners was 91.72% (*SD* = 3.90%) and 95.34% (*SD* = 2.85%) for IS listeners, indicating a high level of accuracy and that listeners were generally able to perform the task successfully. In order to compare the findings from Elvin et al. (under review a) on discrimination accuracy with the present results

[16] It is important to note that this study forms part of a large scale study that examines the interrelation between speech perception, spoken word recognition and speech production. The present experiment took place in the same session as the experiments reported in Elvin et al (under review a). It was always conducted prior to the non-native speech perception and production tasks, which were counterbalanced. The native speech production task, which was used for the acoustic analyses in Elvin et al (under review a) was conducted prior to the experiment in the present study.

on spoken word recognition, we will refer to the trials that contain one of the six BP contrasts investigated in the speech perception study (/a/-/ɔ/, /a/-/ɛ/, /i/-/e/, /o/-/u/, /e/-/ɛ/, and /o/-/ɔ/) as "minimal pair 2 (MP_2)" and all remaining minimal word pairings as "minimal pair 1 (MP_1)". Novel words that differed in more than one phoneme formed non-minimal word pairs. Figure 1 shows word recognition accuracy for non-minimal (NP), MP_1 and MP_2 pair types. In order to investigate whether word pair and native language had effects on accuracy scores, we ran a repeated-measures analysis of variance (ANOVA) with Pair Type (NP, MP_1, MP_2) as a within-subjects factor and Language Group (AusE, IS) as a between-subject factor, with accuracy as the dependant variable. In line with Escudero et al.'s (2012, 2013) results, we found a main effect of Pair Type [F (2, 76) = 176.284, p = <0.001, η_p^2 = .823], indicating accuracy scores varied reliably across the different pairs, and a main effect of Language Group [F (1, 38) =13.511, p = .001, η_p^2 = .262], suggesting the AusE and IS groups' overall accuracy scores differed significantly from one another. We also found a significant Pair Type × Language Group interaction [F (2, 76) = 6.909, p = 0.002, η_p^2 = .154], suggesting patterns of accuracy scores across the three pair types differed for AusE and IS listeners.

Visual inspection of Figure 1 suggests that the above main effect of Pair Type reflects that the NP Type had the highest accuracy scores, followed by MP_1, with MP_2 having overall the lowest accuracy. To confirm which Pair Type differences are indeed statistically significant, we ran paired t-tests (alpha <.016, Bonferroni correction for multiple comparisons, 0.05/3 comparisons, as each Pair Type was compared with the other two Pair Types). The results indicated that for both AusE and IS listeners all Pair Types differed from all others: participants had higher accuracy for NP than MP_1: AusE - t(19), = 5.803, p <0.001, 95% CI [10.42, 22.17] and IS - t(19), = 3.477, p = 0.003, 95% CI [1.38, 1.91]. Participants' accuracy was also higher for NP than MP_2: AusE -, t (19), = 6.546, p <0.001, 95% CI [11.85, 22.99] and IS - *t* (19), = 9.928, *p* <0.001, 95% CI [16.26, 29.95]. The results also indicated higher accuracy for MP_1 than MP_2 for both groups: AusE, *t* (19), = 14.950, *p* <0.001, 95% CI [28.99, 38.43] and IS, *t* (19), = 12.269, *p* <0.001, 95% CI [21.07, 29.74]. We ran an independent-samples t-test (alpha <.016, Bonferroni correction for multiple comparisons, 0.05/3 comparisons) to confirm the pair types whereby AusE listeners' accuracy scores were significantly lower than IS listeners. The results indicate that AusE listeners had significantly lower accuracy scores for both minimal pair types: MP_1 - (*t*(38) = -3.636, p = 0.001) and MP_2 (*t*(38) = -2.844, p = 0.007).

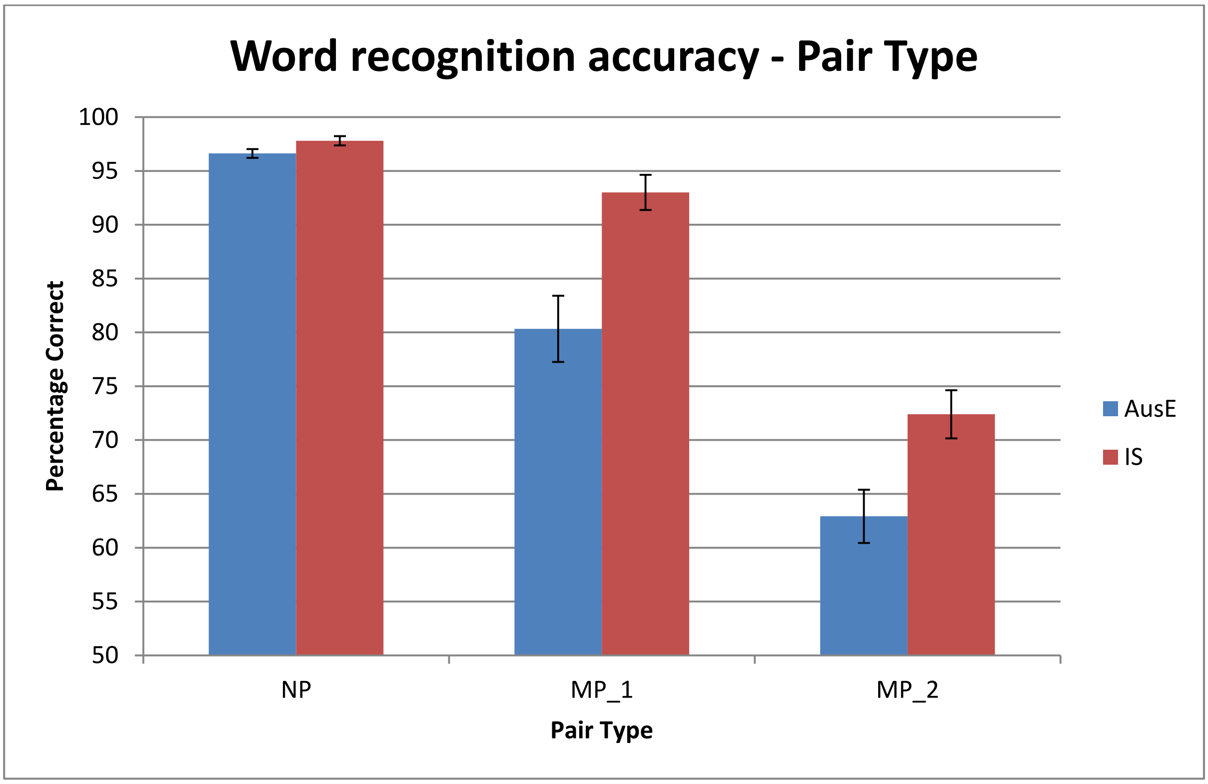


Figure 5.1 Percentage correct (above chance, i.e. 50%) for AusE and IS listeners across non-minimal (NP), minimal-easy (MP_1) and minimal-difficult (MP_2) word pair types.

Regarding word recognition for MP_2, which included the six BP contrasts used in Elvin et al's (under review a) study, visual inspection of Table 5.1 suggests that IS and AusE listeners had comparable accuracy patterns across the six BP contrasts but that IS listeners had higher accuracy overall. A 2x6 repeated measures ANOVA confirmed a main effect of Language Group [F (1, 38) = 8.086, p = 0.007, ηp^2 = 0.175] as well as a main effect of BP Contrast [F (5, 190) = 6.951, p <0.001, ηp^2 = .155], however there was no interaction between BP Contrast and Language Group [F (5, 190) = .467, p = 0.800, ηp^2 = 0.012]. This indicates that both listener groups found the same contrasts easy or difficult to recognise but that the IS listeners did in fact have significantly higher overall accuracy.

Table 5.1 AusE and IS monolingual listeners' mean accuracy scores across the same BP contrasts tested in speech perception (Elvin et al., under review a). Standard Error (SE) and Upper (U) and Lower (L) bound confidence intervals of the means are also given.

	BP Contrast					
	/a/-/ɔ/	**/a/-/ɛ/**	**/i/-/e/**	**/o/-/u/**	**/e/-/ɛ/**	**/o/-/ɔ/**
AusE	*M* 73.75 SE: 6.14	68.75 6.25	55.00 5.62	67.50 5.76	58.75 7.09	53.75 5.22
	U: 85.05 L: 62.45	65.37 58.57	71.20 44.63	71.34 53.66	71.77 45.73	64.61 42.89
IS	*M* 86.25 SE: 4.96	87.50 3.39	61.88 4.58	76.25 5.87	63.13 5.70	59.38 5.51
	U: 97.55 L: 74.95	97.68 77.32	72.25 51.50	66.34 48.87	76.14 50.11	70.23 48.52

To compare accuracy across the six MP contrasts tested in Elvin et al (under review a), we further conducted paired samples t-tests. The accuracy scores for each word pair was compared the other five pairs, with alpha <0.0033 (Bonferroni correction for multiple comparisons, 0.05/15 comparisons) and including all listeners pooled together. The results indicated that the BP word pairings *fafe-fɛfe* and *fafe-fɔfe* had comparable levels of difficulty (t(39) = .433, p (two-tailed) = .667), but were significantly easier to recognise than the all remaining contrasts (ts(38) = 3.388 – 4.662, all ps (two-tailed) < 0.002) except *fofe-fufe* (ts(39) = 1.293 – 1.801, both ps (two-tailed) > 0.079).With the alpha correction of <.0033, *fafe- fɛfe* was not significantly easier than *fefe- fɛfe* $t(38) = 2.922$, p (two-tailed) $= 0.006$. There was no significant difference between the percentage correct among *fofe-fɔfe, fife-fefe* and *fefe-fɛfe* (*t*s(38) = -.377 – .852, all *p*s (two-tailed) > 0.399). Interestingly, with the alpha correction of <.0033, *fofe-fufe* was also not significantly more difficult than *fofe-fɔfe, fife-fefe* and *fefe-fɛfe* (*t*s(39) = -2.335 – 2.547, all *p*s (two-tailed) > 0.015). The order of difficulty of the six minimal pairs from least difficult to most difficult is as follows: /a/-/ɔ/ ~ /a/-/ɛ/ ~ (/o/-/u/)> /e/-/ɛ/ ~ /i/-/e/ ~ /o/-/ɔ/ ~ (/o/-/u/), where "~" means equal or comparable difficulty and ">" signifies a significant difference in accuracy.

While no significant difference was observed between groups in terms of the order of difficulty with the contrasts, we did observe an overall group effect. Therefore, to determine on which contrasts the groups performed differently, we ran an independent-samples t-test (alpha <.003, Bonferroni correction for multiple comparisons, 0.05/15 comparisons). Interestingly, when we apply the alpha correction, a significant difference between the two groups is not observed. However, there is a trend observed for *fafe-fɛfe* ($t(38) = -2.36$, p (two-tailed) $= 0.012$. Thus, it could be that the overall main effect of group could be driven by the fact that IS listeners had much higher spoken word recognition accuracy than AusE participants in the *fafe-fɛfe* contrast.

5.4. Comparison between speech perception and spoken word recognition

The above findings indicate that although IS listeners had overall higher accuracy than AusE listeners, both groups found the same contrasts easy or difficult to recognise. To determine whether or not difficulties were the same for speech perception and word recognition, we compared the results from the present study with these same listeners' discrimination results reported in Elvin et al. (under review a). Table 5.2 compares the order of difficulty of the six BP contrasts tested in discrimination and spoken word recognition.

Table 5.2 Discrimination and word recognition order of difficulty. Both groups reported together as no interaction between contrast and language group was found in either study.

Task	**Order of difficulty**
Discrimination (Elvin et al., under review a)	/a/-/ɔ/ ~ /a/-/ɛ/ > /e/-/ɛ/ > /i/-/e/ ~ /o/-/ɔ/ ~ /o/-/u/
Word recognition	/a/-/ɔ/ ~ /a/-/ɛ/ ~ (/o/-/u/) > /e/-/ɛ/ ~ /i/-/e/ ~ /o/-/ɔ/ ~ (/o/-/u/)

As observed in Table 5.2, the listeners' patterns of discrimination difficulty were indeed almost equal to their patterns of difficulty in spoken word recognition. In particular, of the six BP contrasts, both groups of listeners found /a/-/ɔ/ and /a/-/ɛ/ perceptually easy to discriminate, which was replicated in word recognition. While BP /e/-/ɛ/ was perceptually easier than /i/-/e/, /o/-/ɔ/ and /o/-/u/ in discrimination, it grouped with these three contrasts as equally difficult in spoken word recognition. Additionally, the spoken word recognition results indicate that BP a/-/ɔ/ and /a/-/ɛ/ were not significantly easier to recognise in words than BP /o/-/u/, and although it seems as though there is a trend toward BP /o/-/u/ being significantly easier than the remaining contrasts, with the alpha correction for multiple comparisons, this did not yield any significance.

While it appears that our findings in discrimination were largely consistent with our findings for spoken word recognition, the above comparisons were based simply on order of difficulty. To confirm whether the patterns of difficulty (i.e., on the dependent variable of percentage correct) observed in the spoken word recognition task were comparable to those in the Elvin et al's (under review a) discrimination task, we ran a Spearman's rank correlation. The results based on group means indicate a positive correlation between discrimination accuracy and word recognition accuracy for IS listeners ($r_s(5) = .886$, p = .009), albeit less reliable for AusE: ($r_s(5) = .657$, p = .078). However, when both groups are analysed together a reliable positive correlation was found ($r_s(5) = .538$, $p = .035$). This finding supports our above observation that both AusE and IS listeners found the same BP contrasts equally difficult or easy across both tasks.

The above findings indicate that *at a group level*, in most cases, the BP contrasts that were perceptually easy or difficulty to discriminate pre-lexically, were also easy or difficult to discriminate in spoken word learning and recognition. Additionally, IS listeners were more accurate than AusE listeners in spoken word recognition, whereas IS and AusE listeners did not differ in discrimination. We further tested whether these findings are consistent at an individual level.

Figure 5.2 shows each individual listener's percentage of discrimination error plotted against errors in spoken word recognition. Regression lines were fitted for each individual participant. An upward slope indicates that a high percentage of errors in discrimination predict a high percentage of spoken word recognition errors, suggesting that discrimination performance predicts spoken word recognition accuracy. The steepness of the slope indicates the strength of the relationship between discrimination and spoken word recognition.

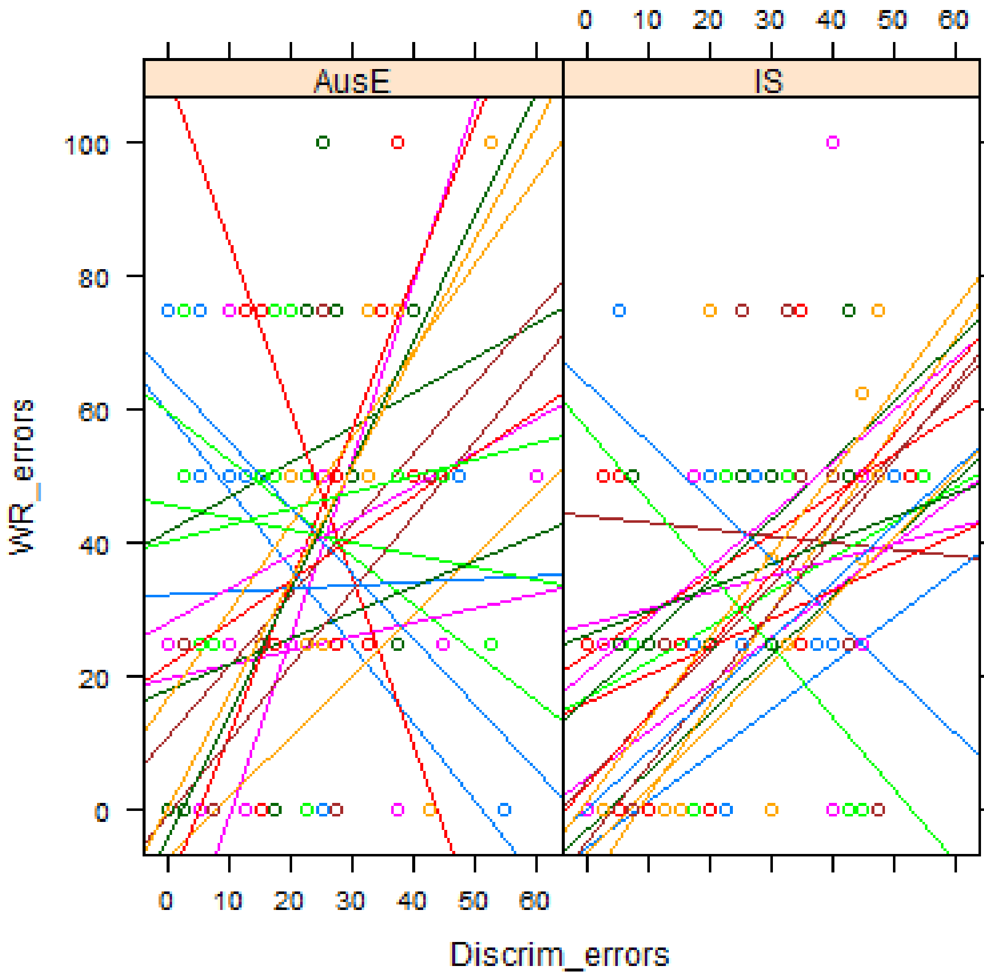


Figure 5.2 AusE (left) and IS (right) individual discrimination accuracy (x-axis) plotted against individual spoken word recognition accuracy (y-axis) for each BP vowel. Regression lines have been fitted for each individual. Colours represent each individual's data.

Visual inspection of Figure 5.2 indicates that the strength of the relationship between discrimination and spoken word recognition errors varies among participants. There is a larger amount of individual variation in slope steepness and direction for AusE listeners than IS listeners. For the most part, the slopes for IS participants are positive and rather steep, indicating a strong relationship between discrimination and spoken word recognition errors. Spearman's rank correlations between individual listeners' discrimination and spoken word recognition accuracy scores for the six BP contrasts for the two listener groups revealed reliable positive correlations (AusE: $r_s(119) = .282$, p = .001; IS: $r_s(119) = .499$, p = <.001), which were stronger for IS than for AusE listeners, a finding that is not unexpected given that the degree of individual variation was smaller for IS listeners.

In addition to slope steepness, the most striking visual patterns in Figure 5.2 are of those participants who display a negative correlation between discrimination and spoken word recognition accuracy, indicating that for these listeners, higher discrimination scores resulted in lower spoken word recognition accuracy. In fact, these listeners appear to perform worse in spoken word recognition than their discrimination accuracy scores would have predicted if they showed the same positive relationship between discrimination and spoken word recognition. For this reason, we examined those eight listeners' performance (six AusE, two IS) across the discrimination and spoken word recognition tasks as compared to the remaining listeners. Table 5.3 shows their accuracy scores for the six BP contrasts in discrimination and spoken word recognition as well as average differences between the two tasks (calculated over the six contrasts per participant).

Inspection of Table 5.3 shows that for these eight participants, spoken word recognition scores of the contrasts in the MP_2 group were on average 10.5 (AusE) and 10% (IS) lower than the remaining participants with positive slopes. However, in discrimination the accuracy scores of these eight participants were comparable to the remaining participants with positive slopes and in the case of AusE they were higher. In addition, it can be seen that lower performance in spoken word recognition was not due to a failure to perform the task because as observed in Table 5.3, these eight participants scored above 90% on NP pairings and above 75% on MP_1 (with the exception of AusE25), which is in line with the remaining participants with positive slopes. Therefore, the individual differences we observe among these participants are specifically related to the same MP contrasts examined in perception.

Table 5.3 Individual accuracy scores across different pair types for the AusE and IS individuals with negative correlations between discrimination accuracy and spoken word recognition. The means for discrimination accuracy and spoken word recognition (WR) accuracy for the MP_2 pairs, as well as NP and MP_1 word pairings, are presented alongside the mean discrimination and WR difference. The averaged difference between discrimination and spoken word recognition accuracy for both languages is also provided.

Participant	**Mean discrimination**	**Mean MP_2**	**Difference between discrimination and WR difficulty**	**NP**	**MP_1**
AusE01	85	58.33	-26.67	99.29	93.33
AusE03	80.83	54.16	-26.67	96.07	81.67
AusE14	76.25	50	-26.25	97.14	81.67
AusE16	81.25	54.16	-27.09	96.43	86.67
AusE25	78.31	58.33	-19.98	96.07	68.33
IS10	79.17	58.33	-20.84	97.5	95
IS21	74.58	58.33	-16.25	92.50	86.67
IS27	70.42	75	4.58	95.71	91.67
Neg. slope AusE	80.33	55	-25.33	97	82.33
Neg. slope IS	74.72	63.89	-10.83	95.24	91.11
Remaining AusE	77.61	65.56	-12.05	96.5	79.67
Remaining IS	76.57	73.90	-2.67	98.26	93.33

Overall, it is very clear that these specific participants had a disproportionately harder time recognising words containing the same MP contrasts which were also tested in non-native perception. Recall that in the introduction we predicted that the link between non-native perception and spoken word recognition might not be as straightforward as one would expect. This is indeed the case for these eight listeners. We propose that the negative correlations for these eight participants could be a result of training effects, different task demands and/or different learning strategies. We consider these factors in more detail in the discussion. Furthermore, as five of the eight participants with negative correlations were AusE listeners, it may be that their disproportionate difficulties could have led to the apparent group overall advantage (i.e., higher spoken word recognition scores) for IS > AusE listeners uncovered previously.

To confirm whether these eight participants' spoken word recognition scores were driving the between-group difference on this task, we reanalysed spoken word recognition accuracy across the six BP contrasts once more but this time excluding the eight participants. A 2x6 repeated measures ANOVA on the remaining participants' scores with contrast as a within-subjects factor and group as a between-subjects factor once more revealed a main effect of BP contrast [F (5, 150) = 10.99, p <.001, ηp^2 = .252] and no interaction between contrast and listener group was found [F (5, 150) = .078, p = 996, ηp^2 = .003]. Interestingly, we do still observe a main effect of group [F (1, 30) = 5.017, p = .033, ηp^2 = .143], however, it is not as strong as the group effect when analysing all participants together. We also ran an additional independent samples t-test, excluding these 8 participants, and again found that with the Bonferonni corrected alpha (<.003, 0.05/15 comparisons), no significant difference between the two groups was observed. Furthermore, the strong trend observed earlier for IS listeners to have higher accuracy than AusE listeners in the *fafe-fɛfe* contrast (p = .012) is not as strong when the 8 participants with negative correlations are excluded (p = .084). This suggests that although there is still an overall advantage for IS listeners in their spoken word recognition accuracy; part of the group effect seems to be driven by the results of the 8 excluded participants.

To further investigate the influence of these eight participants on our correlations, we ran additional Spearman's rank correlation on group means for discrimination and spoken word recognition, excluding the data of these eight participants. As expected, a stronger positive relationship between discrimination accuracy and spoken word recognition is observed for both groups of listeners (AusE: $r_s(5)$ = .829, p = .021; IS: $r_s(5)$ = .943, p = .002). This stronger correlation is also present in a Spearman's rank correlation of the individual means, but excluding those eight participants (AusE: $r_s(119) = .452$, p <.001; IS: $r_s(119) =$ 615, p <.001). Therefore, although the previous apparent advantage for IS listeners in spoken word recognition across the six BP contrasts does seem to be language specific, we do find that the correlation between speech perception and spoken word recognition are much stronger both at a group and individual level, suggesting that these eight participants did have some influence on the correlation between the two abilities.

5.5. General discussion

The present study directly tested whether naïve Australian English (AusE) and Iberian Spanish (IS) listeners' recognition of spoken Brazilian Portuguese (BP) novel words is related to their individual patterns of L2 discrimination difficulty. The results from our spoken word recognition task indicated that all participants were able to learn the BP novel words and that, at first glance, IS learners had overall higher accuracy than AusE listeners. We also found that minimal word pairs were more difficult to recognise than non-minimal pairs and that overall minimal pairs containing one of the six BP vowel contrasts tested in speech perception were also the most difficult to recognise in newly-learned BP words. The findings of the present study are in line with the models of spoken word recognition such as the Cohort model (Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980), TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994; Norris, Mcqueen, Cutler, & Butterfield, 1997) and its extension to Shortlist-B (Norris & McQueen, 2008), the neighbourhood activation model (Luce & Pisoni, 1998) and its extension to PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000) as well as other studies of spoken word recognition which suggest that difficulty in spoken word recognition is related to multiple word activation and competition between activated words. In particular, Weber and Cutler (2004) showed that Dutch listeners fixated distractor pictures more often and for a longer amount of time when the English names of the target and distractor pictures contained vowels that are auditorily confusable. Although we use a novel word learning task in the present study, listeners' difficulties with the BP vowel contrasts /i/-/e/. /o/-/ ɔ /, /e-ɛ/ and /o/-/u/ are likely due to co-activation of both words in this auditorily confused vowel contrast.

Thus the present results also support the L2LP model's claim that at the initial stage of learning, perceptual and spoken word recognition abilities are closely related. That is, sounds which are confused in speech perception are also difficult to recognise in words. However, we found that unlike the discrimination findings in Elvin et al. (under review a), IS listeners had overall higher spoken word recognition accuracy scores at a group level than AusE listeners. In our analysis of individual differences, we found greater individual variation for AusE listeners than IS listeners in terms of the strength of the relationship between discrimination accuracy and spoken word recognition accuracy. However, most striking were the eight participants (five AusE, three IS) with negative correlations. Interestingly, these eight participants were found to be much worse than the remaining participants at spoken word recognition specifically for the six difficult BP contrasts used in the discrimination task (but not the other contrasts). However, these eight participants' scores for the same six BP contrasts in the discrimination task were not unlike the remaining participants; indeed, some performed much better. We propose that the between-participant differences observed on the spoken word recognition task only could be related to how individuals deal with different tasks that crucially make different demands. Although the two tasks are similar in the fact that they both only offer one competitor per trial, the spoken word recognition task seems to be more demanding than non-native discrimination. Specifically, the discrimination task reported in Elvin et al. (under review a) tested whether listeners are able detect differences between words that had simply been presented to them, and therefore probes speech perception which does not demand higher levels of processing, including that of long term memory. In the spoken word recognition task, on the other hand, listeners were presented with only one newly-learned word at a time and were required to match this to the appropriate referent image. Indeed, Díaz et al. (2012) showed that L2 listeners' performance largely depended on the nature of the task, with high accuracy scores for L2 listeners on an acoustic-phonetic analysis task than on tasks involving lexical processes. In our study the aforementioned eight participants with negative correlations differed from the remaining participants only in spoken word recognition. Thus, it could be that the task demands in spoken word recognition are more challenging for these eight participants than the remainder of the participants.

Recall that the L2LP model's *optimal perception hypothesis* states that learners begin perceiving the sounds in the L2 in a way that is optimal in their L1 according to the acoustic properties of L1 categories. On that basis, the model outlines different learning scenarios. Elvin et al.'s (under review a) study speaks to the optimal perception hypothesis because it provides the particular learning scenarios necessary for successfully acquiring L2 categories and contrasts, which varies for different contrasts such that some will be easier or more difficult to learn than others. Indeed, Elvin et al.'s (under review a) predictions regarding difficulty were mostly borne out in the present study. That is, as in Elvin et al. (under review a) both groups found BP /a/-/ɔ/ and /a/-/ɛ/ easy to recognise and /i/-/e/, /o/-/ɔ/ and /o/-/u/ difficult. The study differs from Elvin et al. (under review a) in the fact that while BP /o/-/u/ is statistically as difficult as /i/-/e/, /o/-/ɔ/ and /e/-/ɛ/, yet as easy as a/-/ɔ/ and /a/-/ɛ/. Furthermore, in the earlier study of speech perception (Elvin et al., under review a), BP /e/-/ɛ/ was not as easy as BP /a/-/ɔ/ and /a/-/ɛ/, yet not as difficult as the other contrasts. However, in spoken word recognition it was just as difficult as /i/-/e/, /o/-/ɔ/ and /o/-/u/. This suggests that these listeners were more sensitive to this contrast in discrimination than in spoken word recognition. The listeners' higher performance for BP /e/-/ɛ/ in discrimination may possibly be evidence of an effect of training. That is, the discrimination task was completed after the spoken word recognition task and this may be a sign of perceptual development for this specific contrast. However, Elvin et al. (2014) also found this BP contrast to be harder to discriminate than BP /i/-/e/ and /o/-/u/, but not as easy as /a/-/ɛ/. Thus it might also be that for this particular contrast, these listeners have difficulty in mapping different pre-lexical representations to their lexical counterparts (Escudero, 2011), however further investigations would be required to determine whether or not this is indeed the case.

For perceptual development to take place, the L2LP model assumes *meaning-driven learning*. The spoken word recognition task, in contrast to the discrimination task from Elvin et al. (under review a), emulates meaning-driven learning because listeners must not only detect differences between the variable auditory forms of different words, but they must also associate them with different meanings (i.e., visual referents). Recall that van Leussen & Escudero (2015) computationally tested the model's claims for meaning-driven learning with two possible ways of processing: *sequential* (strictly bottom-up) and *interactive* (lexical feedback to perception can occur). Although both ways ultimately resulted in successful acquisition of a non-native contrast in L2 words, interactive learners were quicker to successfully acquire the lexical contrast, than sequential learners, i.e., the former required less input than the latter. We propose that the individual differences for the six difficult BP word contrasts in spoken word recognition tasks might also reflect different learning strategies associated with different ways of processing the novel words. That is, the eight listeners who performed poorly on spoken word recognition are likely to be sequential learners who require more input in order to be able to learn words, whereas the majority of listeners are likely to be interactional learners. In line with our discussion of task demands above, it might also be the case that this spoken word recognition task demands more from sequential learners than interactive learners. Future investigations might consider testing whether additional exposure to the target words helps these suspected sequential learners achieve comparable accuracy scores to the remaining suspected interactive learners. It would also be worthwhile investigating whether additional exposure leads to any change in their correlations between discrimination and spoken word recognition.

The findings from the present study would also have important implications for other models of speech perception such as SLM (Flege, 1995) and PAM (Best, 1994; 1995) and its extension to PAM-L2 (Best & Tyler, 2007). In particular, we would expect that L2 speech perception findings explained by these models may also be applicable to findings for spoken word recognition. In particular, PAM-L2 posits that vocabulary size determines L2 sound perception success (Best & Tyler, 2007; see also Bundgaard-Nielsen, et al., 2011). Specifically, PAM-L2 argues that L2 perceptual learning should be most evident early on in L2 learning, before the learner has achieved a sufficiently large vocabulary to adequately follow L2 conversations (estimated to be ~ 6000-7000 words: Nation, 2006) and be focused on higher level aspects of the language (lexical items, morphology, grammar, pragmatics, etc.). Our study has shown that success in speech perception influences success in spoken word recognition, though notably different for some learners than for others. It is likely that the PAM-L2 hypothesis regarding L2 vocabulary size and L2 perceptual learning might also extend to spoken word recognition accuracy. In fact, it is likely that individual differences in L2 learners (even those at similar stages of learning, i.e. beginner, intermediate, advanced) might be explicable according to their differing vocabulary sizes. The findings from the present study have highlighted the relationship between perception and spoken word recognition. Future studies should consider the theoretical development of models such as SLM, PAM and PAM-L2 in linking sound perception to spoken word recognition. In particular, just as SLM explores perceptual difficulties as a means of explaining L2 pronunciation, future studies might look at how these aforementioned theoretical models would reconcile the relationship between perceptual difficulties and spoken word recognition.

In sum, the present study provides further evidence to the growing body of literature (e.g. Escudero et al., 2008, 2013, 2014; Weber & Cutler, 2004) that auditorily or phonetically confusable L2 or non-native vowel contrasts can lead to confusion in spoken word recognition. One of the main claims of the L2LP model is that in addition to speech perception, the model is also applicable to spoken word learning and spoken word recognition as well as speech production. Our study tests the L2LP model as it bases our predictions on data from the same listeners across both tasks. The L2LP model predictions are largely supported in the present study as the same contrasts that were difficult in speech perception were also confused in spoken word recognition. The individual differences approach used in the present study seems to lead to a more accurate comparison of listeners from different language backgrounds thus confirming one of the main proposals of the L2LP model. In particular we may have found evidence that learners use different learning strategies formalised in the L2LP model. It could be that the individuals who show a negative correlation between discrimination and spoken word recognition accuracy are using a sequential learning strategy, who might find the task more demanding because they require longer exposure to the target words before reaching the same level of word learning/recognition as the majority of other participants who we speculate are interactive learners. The fact that these negative correlations were only found when individual variation was investigated highlights Smith and Hayes-Harb’s (2011) claim that one must be careful when reporting group averages only as it may not account for all listeners. That is, individual variation may reveal important information regarding language learning process that cannot be identified at a group level. The findings from the present study emphasise the fact L2 performance should be described, predicted and explained at an individual level as explicitly suggested in the L2LP model.

It is interesting to note that of the eight participants who we predicted to be sequential learners, five were AusE listeners and only three were IS listeners. This leads to the question of whether this different mix of learner types across both groups is principled or accidental. In particular it raises the question of whether or not there is a relationship between the L1 and learning type (especially seeing as it seems that the overall advantage for IS listeners is no longer apparent when these listeners are excluded). It could be that there is something about the language or linguistic environment that leads to a specific language learning type. AusE has a larger vowel inventory than IS and perhaps there is greater potential for languages with larger vowel inventories to have a greater mix of sequential learners because of the multiplicity of available response categories in their L1.

For example, Benders et al. (2012) found that listeners who had more response vowel categories available were less sensitive to broad and local acoustic contexts than listeners presented with only two response categories. In addition, they took longer to shift their boundary between Spanish /i/ and /e/ and the authors argued that this delay in the boundary shift was caused by the availability of extra response categories. Thus it could be that there may be more biases toward sequential learners for listeners with more vowel categories in their native language because they need to pay closer attention to the multiplicity of vowel categories activated and the smaller differences between those vowels. Additionally, van Leussen & Escudero (2015) claim that sequential learners have greater difficulties reducing the number of L1 categories employed in perception when learning a language with a smaller vowel inventory. Thus it is more likely that there would be a greater number of sequential learners among listeners with larger vowel inventories because sequential processing arises from the native language having more vowels than the L2 and therefore they have a larger number of categories that they would need to reduce. Further investigation would of course be required to test the claim that different learning types may in fact be language specific. For example, one might investigate whether the number of sequential learners is consistently higher across different dialects of English than the number of sequential learners across different dialects of Spanish. If a relationship between the L1 and learner type is identified it could have important theoretical and pedagogical implications in the areas of L2 language learning and development.

Finally, further research is required for testing the L2LP model's predictions (as well as other models such as PAM-L2) for non-native production, exploring individual variation and again drawing data from the same population. To this end, we are currently investigating non-native production of BP vowels by the same participants from the present study and Elvin et al. (under review a). This ongoing research will allow further understanding of the role of native production in L2 speech perception, L2 spoken word recognition and L2 speech production as well as their interrelation. Future research should also include longitudinal data or data collected from intermediate and experienced listeners across the three L2 abilities (speech perception, spoken word recognition and speech production) in relation to L2 vocabulary learning and learning of other aspects of the L2, in order to further test the L2LP model's (as well as other models of L2 speech) applicability for learners at all stages of L2 development.

5.6. Chapter Summary

This chapter investigated the interrelation between non-native speech perception and spoken word recognition at the initial state of L2 learning at a group and individual level. It also investigated the whether or not individual listeners' follow either a sequential or interactive lexical processing route when learning new words. In line with the results from the previous chapter, AusE and IS listeners found the same BP contrasts equally easy/difficult to recognise, however unlike the findings from the previous study, the IS participants had higher accuracy overall. In terms of an interrelation between non-native perception and spoken word recognition, the findings in this chapter indicated a positive relationship between discrimination accuracy and spoken word recognition for the majority of the participants. Most striking were the individuals with negative correlations. It was proposed that the negative correlations and lower accuracy scores for some participants could be related to task demands or perhaps these participants are sequential learners who require additional input in order to receive comparable accuracy scores to the other word learners. However, further investigation would be required to investigate this claim

The next chapter which is the final chapter of this thesis presents a review of this thesis, specifically addressing the overall research aims and questions. The initial state for AusE and IS non-native speech perception and spoken word recognition will be outlined and predictions for non-native and L2 production will be provided. The chapter will conclude with the overall implications resulting from this thesis and will offer directions for future work.

Chapter 6

Thesis summary, implications and future research

6.1. Thesis summary

The overall aim of this thesis was to investigate how the native language shapes individual differences in Australian English (AusE) and Iberian Spanish (IS) listeners' non-native speech perception and spoken word recognition of Brazilian Portuguese (BP). The thesis also aimed to determine whether or not there was an interrelation between the two communicative abilities. To address the overall aims of the thesis, the following specific research questions were investigated:

1. Do Australian English and Iberian Spanish listeners, whose languages differ in terms of vowel inventory size, have similar problems when perceiving and recognising Brazilian Portuguese novel words?
2. Is acoustic similarity between the native and target language a reliable predictor of non-native perception for listeners at the onset of learning (i.e., the initial state)?
3. Do listeners follow different (individual) developmental paths in non-native speech perception and spoken word recognition?
4. Is there an interrelation between the difficulties identified in individual listeners' non-native vowel categorisation / discrimination patterns and spoken word recognition?

This thesis followed the theoretical framework of the Second Language Linguistic Perception Model (L2LP; Escudero, 2005; 2009; van Leussen, 2015) which considers individual learners at all stages of learning from the onset of language learning right until ultimate attainment and across all communicative abilities (i.e., speech perception, spoken word recognition and speech production).

In Chapter 2, a preliminary investigation of AusE and IS listeners' discrimination of BP vowels was presented (Elvin et al., 2014). The study was developed to address the first and second research questions of this thesis. It therefore investigated vowel inventory size and acoustic similarity as predictors of non-native discrimination difficulty for two groups of listeners with differing vowel inventory sizes. The findings were in line with the L2LP model's claims that discrimination difficulties can be predicted by the detailed acoustic comparisons between the native and target languages. The acoustic predictions in that study were based on previously reported data for IS (Chládková et al., 2011), BP (Escudero et al., 2009) and AusE (Cox, 2006). While the acoustic descriptions for BP and IS were cross-linguistically comparable, the acoustic description for AusE was based on a published data that was produced by adolescent speakers in a different consonantal context and using a different method of data collection and formant analysis. While the acoustic comparison of AusE and BP successfully predicted listeners' discrimination difficulty, the most accurate explanation relies on a thorough analysis of the acoustic properties of AusE vowels with a methodology that could identify the spectral properties of AusE. For this reason, the study in Chapter 3 (Elvin et al., 2016) presented a comprehensive acoustic analysis of AusE vowels. This study can be used as a point of reference for future studies investigating the acquisition of AusE or L2 acquisition by speakers of AusE from Western Sydney where production data from the same listeners intended for testing is not possible to collect.

Specifically, the acoustic analysis of AusE vowels revealed that vowel durations and formant trajectories were affected by the consonantal context in which they were produced, with the hVd context demonstrating some of the most striking differences. Given these differences in duration and formant trajectories differed across the six consonantal contexts, cross-language acoustic comparisons for predicting L2 difficulty may be affected. In particular, the /hVd/ context – which has commonly been used in acoustic analyses of English vowels due to its predictable co-articulatory effects (Cox & Palethorpe, 2004) – could be unsuitable because AusE /hVd/ vowels are acoustically least similar to those in the other consonantal contexts used in this acoustic analysis and also because /h/ is not a word-initial consonant in some languages, including Romance languages such as Spanish and Portuguese.

The study in Chapter 4 (Elvin et al., under review a) was specifically designed to address the first three research questions of this thesis and to build on and improve Elvin et al's (2014) preliminary investigation. This study provided an adequate test of L2LP model as the comprehensive acoustic analyses were based on listeners' own native production data which were then used to predict the listeners' own non-native categorisation and discrimination patterns. The study is a unique contribution to the field of non-native speech perception because, in addition to reporting findings based on groups, it also explored individual variation in the participants' acoustic realisations of vowels, which may predict individual differences in perception performance.

Finally, in Chapter 5, Elvin et al. (under review b) builds on Elvin et al. (under review a) as it addresses the first, third and fourth research questions of this thesis. In particular, it investigated whether individual AusE and IS participants have comparable difficulties when recognising novel BP words and it also explored the role of non-native perception in spoken word recognition. Testing the same participants from Elvin et al. (under review a) and using the same stimuli across the two tasks, the study allowed for an appropriate examination of the interrelation between non-native speech perception and spoken word recognition at both a group and individual level.

The remainder of this chapter will further discuss the specific findings of this thesis which directly relate to the overall research aims, which were to investigate how the native language shapes individual differences in AusE and IS listeners' non-native speech perception and spoken word recognition of BP and whether or not these two communicative abilities are interrelated. Specifically §6.2 will address the role of the native language and §6.3 will address individual variation in non-native speech perception and spoken word recognition. A discussion of whether or not these two communicative abilities are interrelated is then provided in §6.4. The findings from this thesis also indicate which BP contrasts are easy/difficult for naïve AusE and IS listeners to perceive and recognise and these findings can be used to establish their initial state of learning (i.e., the specific difficulties they will have from the onset of learning), which will be presented in §6.5 as well as predictions for L2 development. Predictions for their non-native and L2 production are also made in §6.6 based on their initial state of learning. This chapter concludes with the overall implications of the findings reported in this thesis and with suggestions for future research

6.2. The role of the native language in non-native speech perception and spoken word recognition

Models of speech perception such as the Speech Learning Model (SLM; Flege, 1995), the Perceptual Assimilation Model (PAM; Best 1994; 1995), its extension to PAM-L2 (Best & Tyler, 2007) and the Second Language Linguistic Perception model (L2LP; Escudero, 2005; 2009; van Leussen & Escudero, 2015) all agree that a listeners' native language strongly influences their perception of the sounds in the second language. Previous literature has shown that the size of the native vowel inventory plays an important role in predicting L2 difficulty. In particular, many studies have found that learning a language that has a larger vowel inventory than the native language pose greater difficulties than learning a language with fewer vowels (Iverson & Evans, 2007; 2009). However, studies have also shown that having a larger vowel inventory may also lead to difficulties in perception (e.g., Escudero & Boersma, 2002; Elvin et al., 2014; Vasiliev, 2013).

The role of the native language in non-native perception was investigated in Chapters 2 and 4 of the present thesis which as mentioned above report on Elvin et al. (2014) and Elvin et al. (under review a), respectively. In Chapter 2, Elvin et al. (2014) tested the explanatory power of vowel inventory size and acoustic similarity as predictors of non-native vowel discrimination accuracy. The findings were in line with predictions based on acoustic comparisons. That is, both IS and AusE listeners found BP /a/-/ɛ/ perceptually easy to discriminate and /i/-/e/ and /o/-/u/ perceptually difficult and that IS listeners had overall higher accuracy scores for discrimination. Based on the findings from this study it is evident that the acoustic properties of one's L1 vowel inventory indeed play an important role in determining perceptual success in a new language. Interestingly, the study was in line with findings from Vasiliev (2013) in that cases of the SUBSET scenario or multiple category assimilation may have contributed to the overall lower discrimination accuracy for AusE listeners. In particular, the acoustic comparisons found that the Euclidean Distance between the first and second acoustically closest vowels to BP vowels was much smaller for AusE than for IS, suggesting that both vowels would be a likely choice for AusE listeners but not IS. The findings from the study seem to indicate that when two vowels in a non-native contrast are acoustically similar to the same native vowel(s) listeners are likely to find these sounds difficult to perceive.

Chapter 4 (Elvin et al., under review a) was designed to build on the findings reported in Elvin et al. (2014) with the inclusion of native vowel production and non-native categorisation data in addition to non-native discrimination data collected from the same participants. In light of the fact that the acoustic analysis in this study was based on actual production data rather than reported means, acoustic similarity between the native and target languages could be determined by cross-linguistic discriminant analyses. The cross-linguistic discriminant analyses may be more appropriate for measuring acoustic similarity than Euclidean Distances because, while the latter relies on acoustic similarity between averaged F1 and F2 values only, the former can include individual measurements of F1, F3, F3, formant trajectory and duration.

In line with Elvin et al. (2014) the study found that despite the differing vowel inventory sizes and in line with acoustic predictions, both AusE and IS listeners found the same BP contrasts equally easy/difficult to discriminate. The study further found that acoustic similarity positively correlated with these listeners' own non-native categorisation patterns. Acoustic and perceptual overlap scores were then calculated to test whether acoustic similarity and non-native categorisation patterns are reliable predictors of discrimination difficulty. As predicted by acoustic similarity and non-native categorisation patterns, BP /a/-/ɛ/ was perceptually easy to discriminate. Interestingly, the results for BP /i/-/e/ and /o/-/u/ were in line with predictions based on acoustic similarity, whereas the results for BP /a- ɔ/, /e/-/ɛ/ and /o/-/ɔ/ were more in line with predictions based on the perceptual overlap scores of non-native categorisation patterns. These findings seem to suggest that for the most accurate predictions of discrimination difficulty, a combination of acoustic similarity and non-native categorisation patterns would be most effective. Importantly, when examining the correlation between discrimination difficulty and acoustic similarity and non-native categorisation, the findings indicate that both predictors positively correlated with discrimination difficulties. However the strength of this correlation varied across language groups, with a stronger correlation between acoustic similarity and discrimination difficulty for AusE listeners and the opposite was the case for IS.

The results from Elvin et al. (2014) and Elvin et al. (under review a) clearly show that the native language strongly influence non-native perception and in particular the discrimination of non-native contrasts. In fact, patterns of discrimination difficulty were dependent on the L2LP learning scenario identified for each vowel contrast, which were similar across the two language groups. That is, contrasts which contained evidence of L2LP's NEW or SUBSET scenarios (containing an acoustic or perceptual overlap where the two vowels in a given BP contrast are acoustically similar or categorised to two or more of the same native categories) resulted in similar discrimination difficulties for IS and AusE listeners as both scenarios resulted in a neutralisation of the non-native contrast. In addition, both of these learning scenarios are likely to be more difficult than contrasts where only the SIMILAR scenario is present.

While Elvin et al. (2014) and the study reported in Chapter 4 (Elvin et al., under review a) show the importance of the native language in determining success in non-native perception, Elvin et al., (under review b) which was reported in Chapter 5 reveals that the influence of the native language extends to spoken word recognition. In line with the two studies that investigate non-native perception, in spoken word recognition, the BP contrasts that were easy and/or difficult to recognise were the same across the AusE and IS language groups. Although the participants were the same across Elvin et al. (under review a) and Elvin et al. (under review b), the IS listeners had higher accuracy overall in word recognition but not in perception. This suggests that word learning and word recognition might be easier for learners whose vowel inventory is smaller than the target language. However, some of the overall difference in accuracy scores could also be explained by individual variation in language groups and this will be further discussed below.

In sum, the results from the present thesis indicate that at the initial state of learning, non-native contrasts are perceived and recognised according to the way in which vowel categories are perceived and recognised in the listeners' native language. Importantly, the size and subsequent learning scenarios identified for each language group was predictive of discrimination difficulties. These studies reveal that despite their differing vowel inventory sizes, AusE and IS have the same difficulties with the same BP contrasts (albeit for differing reasons, namely SUBSET scenario for AusE and NEW scenario for IS). One striking result in this thesis is that in contrast to many studies of L2 speech perception, AusE listeners' who have a larger vowel inventory than the target language did not have higher accuracy than IS listeners whose vowel inventory is smaller than the target language.

6.3. Individual variation in non-native speech perception and spoken word recognition

In the process of L2 speech acquisition, different learners ultimately achieve different levels of attainment even when they share the same language background. Individual variation in AusE and IS listeners' non-native perception and spoken word recognition was investigated in Chapter 4 (Elvin et al., under review a) and Chapter 5 (Elvin et al., under review b). In Elvin et al. (under review a) each individual's acoustic classification patterns (i.e. the predicted probabilities) were compared against their own non-native categorisation patterns. The results for individuals are consistent with those at a group level where a positive relationship between acoustic similarity and non-native categorisation was observed.

Individual listeners' discrimination patterns were also predicted by their individual acoustic and perceptual overlap scores, although the strength of this relationship varied across listeners. In fact there were a couple of instances where a negative correlation was observed in which a higher degree of acoustic and/or perceptual overlap did not correspond with lower accuracy scores. Overall, it appeared that IS individuals had stronger correlations between perceptual overlapping and discrimination difficulty, whereas acoustic overlapping was a better predictor for AusE. The fact that each language group's discrimination difficulty was better predicted by a different method (acoustic vs perceptual overlap) may have important implications for the field of L2 speech perception. However, in order to explore these implications, we would first need to rule out the possibilities that these small discrepancies between how well acoustic overlap and perceptual overlap predict discrimination difficulty are not related to the limitations of the statistical tool used to measure acoustic similarity and the design of the non-native categorisation task.

For example, the cross-language discriminant analyses do not consider auditory and acoustic cue-weighting and listeners might be weighing certain cues more than others, which cannot be observed in a discriminant analysis model. Furthermore, it could be that perceptual overlap scores do not predict as well for AusE listeners because they have many more response options and they may be selecting some of these options at random. In the case of IS, listeners had a very small number of response options in the discriminant analyses. Perhaps the probabilities of categorisation were higher for IS because the model forces categorisation even if none of the categories are a good match. If AusE listeners are choosing some options at random and the discriminant analysis model is indeed forcing categorisation for IS this might result in reduced perceptual overlap scores for AusE and inflated acoustic scores for IS. Additionally, the discrimination analyses and non-native categorisation task were based on one repetition of each vowel per speaker (7 vowels x 10 speakers). To rule out the possibility of this finding being task or design dependent, one would need to determine whether listeners (and in particular AusE listeners) would be more consistent in their responses if they heard multiple repetitions by the same speaker. Perhaps including multiple repetitions of each vowel by the same speakers in the discriminant analysis, the probabilities might not be as forced for IS listeners.

Individual variation was also observed in the results of the spoken word recognition task. In particular, the strength of the variation was stronger for AusE than IS listeners. Most striking were the individual negative correlations between discrimination and spoken word recognition for some of the participants. Specifically, in Elvin et al. (under review b) there were eight participants (5 AusE, 3 IS) with a negative correlation between discrimination and word recognition accuracy. Interestingly, word recognition scores of the six BP contrasts were on average 10.5 (AusE) and 10% (IS) lower than the remaining participants with positive slopes. However, in discrimination the accuracy scores of these eight participants were comparable to the remaining participants with positive slopes and in the case of AusE they were higher. Their lower performance in word recognition was not due to a failure to perform the task because they scored above 90% on non-minimal word pairs and with the exception of one, above 75% on minimal-easy pairs, which is in line with group averages. It is clear that these eight listeners had a disproportionately harder time with these six BP contrasts in word recognition than in discrimination. Because five of the eight participants with negative correlations were AusE, this could possibly explain why their overall accuracy scores for these six contrasts were lower than IS. A re-examination of word recognition accuracy across the six BP contrasts revealed that although the overall advantage for IS participants across the six BP contrasts was still present when these 8 participants were included, the effect was not as strong. Indeed, some of the variation between the two groups' accuracy scores in the *fafe-fɛfe* contrast could be explained by the performance of these listeners with negative correlations. This highlights the importance of analysing individual accuracy results in order to confirm that any effects of language identified are indeed language-specific and not driven by the performance of a small number of individuals.

When further investigating the negative correlations for the eight participants (5 AusE, 3 IS) and overall lower performance in word recognition than in discrimination, Elvin et al. (under review b) proposed that the between-participant differences could be related to how individuals deal with the different tasks that were presented to them, which imposed different cognitive demands on participants. Although the two tasks were similar in the fact that they both only offer one competitor per trial, the spoken word recognition task seems to be more demanding than non-native discrimination. Specifically, the discrimination task reported in Elvin et al. (under review a), is an auditory and phonetic task in which listeners were presented with three words one after the other and required to match the first word to the second or third, blocked by contrast for half the participants and randomised for the other half. No effect of condition was found suggesting that in discrimination, listeners perform the same regardless of the way in which the stimuli is presented. In contrast, the spoken word recognition involves meaning-driven learning, that is, learners attempt to improve their perception and recognition of the L2 whenever the current state of the grammar leads to misunderstandings (van Leussen & Escudero, 2015). The spoken word recognition task therefore comprises higher-order levels of processing, in which listeners were presented with a single newly-learned word at a time and were required to match this to the appropriate referent image. Previous studies in L1 acquisition by infants (e.g., Escudero, et al., 2014; Werker & Curtin, 2005) have shown that differing task demands play an important role in lexical development. In L2 acquisition, Díaz et al. (2012) showed that L2 listeners' had higher accuracy scores on an acoustic-phonetic analysis task than on tasks that involved lexical processing. Thus, it could be that the task demands in spoken word recognition, which require lexical processing rather than purely acoustic-phonetic processing, are more challenging for these eight participants than the remainder of the participants.

Another explanation for the performance by these eight participants could be that they use different learning strategies. Recall that for perceptual development to take place, the L2LP model assumes meaning-driven learning. The spoken word recognition task, in contrast to the discrimination task from Elvin et al. (under review a), emulates meaning-driven learning because listeners must not only detect differences between the variable auditory forms of different words, but they must also associate them with different meanings (i.e., visual referents). The L2LP model accounts for two possible types of learners: those who are *sequential* learners (i.e., those who use bottom-up processing) and those who are *interactive* learners (those who allow lexical feedback to perception). Results from computer simulations of these two processing types reported in van Leussen & Escudero (2015) indicate that learners who applied either processing route were shown to succeed in acquiring the L2 contrast. However, sequential learners were slower than interactive learners as they required a greater quantity of L2 input. It is possible that the eight listeners who performed poorly in spoken word recognition are sequential learners, who require additional exposure to the target words in order to achieve the same level of achievement as the other listeners who are likely to be interactive learners. Also, in line with the explanation related to task demands, it might be the case that this spoken word recognition task demands more from sequential than interactive learners.

In sum, the analyses of individual differences in Elvin et al. (under review a) and Elvin et al. (under review b) are in line with previous studies (e.g., Escudero & Boersma, 2004; Escudero et al., 2009; Mayr & Escudero, 2010; Wanrooij et al., 2013) that test the L2LP model's claims that individual learners follow different developmental paths when learning to perceive and recognise words in a new language. In both studies the individual variation seems to be greater for AusE participants than for IS listeners, which may suggest that there is more potential for individual variation for speakers with a larger vowel inventory than the target L2.

6.4. The interrelation between non-native perception and spoken word recognition

The final research question in this thesis was whether or not there is an interrelation between non-native perception and spoken word recognition. A number of studies that test non-native and L2 word learning and word recognition have found that minimal-word pairs which consist of a perceptually difficult contrast are also difficult to learn and recognise (Broersma, 2002; Escudero, et al., 2012; Escudero, et al., 2008; Pallier et al., 2001; Weber & Cutler, 2004). Identifying links between non-native perception and spoken word recognition often poses many challenges for investigators due to methodological reasons. For instance, Levy and Law (2010) explain that not only do different tasks have different demands, but most studies use different techniques for assessing these different communicative abilities. Furthermore, in order to reliably assess the relationship between these abilities, analyses should be based on individual performance using the same participants in both tasks and ensuring that the methodology in each task is comparable and controlled.

In Chapter 5, Elvin et al. (under review b) adheres to the aforementioned guidelines by investigating the same AusE and IS listeners across the non-native categorisation, discrimination and spoken word recognition tasks. Furthermore, to ensure the comparability of perceptual and word recognition abilities, the same stimuli were used across both experimental tasks. When comparing the patterns of difficulty for the six BP contrasts investigated in discrimination and spoken word recognition, the patterns are almost identical. The exception was BP /e/-/ɛ/ which was not as easy as BP /a/-/ɔ/ and /a/-/ɛ/, yet not as difficult as the other contrasts in discrimination, but just as difficult as /i/-/e/, /o/-/ɔ/ and /o/-/u/ in spoken word recognition. A possible explanation for this finding could be that the discrimination task was influenced by training effects. Many studies have shown that speech perception improves after exposure to or training on non-native or L2 sounds and contrasts (e.g., Cenoz & Lecumberri, 1999; Schwab et al., 1985; Strange & Dittmann, 1984). Perceptual training has also been shown to influence L2 production (e.g., Bradlow et al., 1997). When comparing the interrelation between discrimination in Elvin et al. (under review a) and spoken word recognition in Elvin et al. (under review b) it is important to note that the spoken word learning and word recognition task was always conducted prior to non-native discrimination task. Thus, the listeners were more sensitive to the BP /e/-/ɛ/ contrast in discrimination could be some evidence of training effects and perceptual development. Interestingly, the listeners reported in Elvin et al. (2014) also found this contrast to be easier to discriminate than BP /i/-/e/ and /o/-/u/, but not as easy as /a/-/ɛ/. Therefore it could be that for this particular contrast, the listeners in Elvin et al. (under review b) have difficulty in mapping the different pre-lexical representations to their lexical counterparts (Escudero, 2011), however further investigations would be required to determine why this was the only contrast to differ across the two tasks.

The L2LP model posits that a link between speech perception and spoken word recognition exists, however it does not predict that this specific link will be straightforward. This is because factors such as training effects (described above), task demands and learning strategies (which were both described in §6.4 to account for individual variation) could influence the nature of the correlation between speech perception and spoken word recognition. A series of Spearman's rank correlations were conducted at both a group and individual level to statistically test whether accuracy in perception was linked to accuracy in word recognition. At a group level, a positive correlation was found between the percentage of errors in discrimination and word recognition for IS listeners. A positive relationship was also observed for AusE listeners; however, this relationship was less reliable and did not reach significance. Yet at an individual level, reliable positive correlations for both AusE and IS listeners confirmed the L2LP claim that non-native contrasts which are perceptually easy/difficult to discriminate were also be easy/difficult to recognise at the lexical level. Although the negative correlations that were observed for the eight participants discussed above do indicate that the link between speech perception and spoken word recognition is not always straightforward, the overall findings from this study suggest that for most individuals a positive link between speech perception and spoken word recognition exists.

In sum, in line with previous studies (Broersma, 2002; Escudero, et al., 2012; Escudero, et al., 2008; Pallier et al., 2001; Weber & Cutler, 2004), the findings from Elvin et al. (under review b) provide some evidence that at the initial state of learning, there is an interrelation between non-native perception and spoken word recognition. More specifically, the discrimination difficulties reported in Elvin et al. (under review a) are in line with the spoken word recognition difficulties reported in Elvin et al. (under review b). Importantly, this interrelation is not only present in group findings, but in fact this correlation is stronger when investigating the relationship for individual learners.

6.5. The initial state for Iberian Spanish and Australian English learners of Brazilian Portuguese

This thesis has examined non-native perception and spoken word recognition of Brazilian Portuguese (BP) by individual naïve Australian English (AusE) and Iberian Spanish (IS) participants. With the findings presented in this thesis, the initial state learning for AusE and IS learners of BP can now be established, specifically in non-native speech perception and spoken word recognition. Interestingly, the findings from the two perception studies in Chapters 2 and 4 (Elvin et al., 2014 and Elvin et al, under review a) and the spoken word recognition study in Chapter 5 (Elvin, et al., under review b) indicate that in line with predictions based on acoustic similarity, AusE and IS listeners do not differ in terms of the BP contrasts that they find easy or difficult to perceive and learn. Furthermore, the despite variation in terms of accuracy scores, discrimination and word recognition difficulties at an individual level were comparable to those at a group level. Thus the initial state of learning described below is applicable to individuals in both language groups tested in this thesis.

At the onset or initial state of learning, AusE or IS learners of BP had little difficulty perceiving and recognising the BP /a/-/ɔ/ and /a/-/ɛ/ contrasts. This is likely due to the fact that each vowel in each of these BP contrasts is perceived as two separate native categories. On the other hand, AusE and IS individuals did find BP /i/-/e/, /o/-/u/ and /o/-/ɔ/ difficult due to both of the vowels in these contrasts being acoustically similar to and perceptually overlapping with the same native vowel(s). Finally, at the initial state BP /e/-/ɛ/ is not as difficult as BP /i-/e/, /o/-/u/ and /o- ɔ/ to perceive, but it is difficult to recognise.

One would expect that when learning BP, some contrasts may be easier and might take less time to attain than others. Perceptual development will ultimately take place as learners attempt to improve recognition by updating their lexical representations (Escudero, 2005; van Leussen & Escudero, 2015). The L2LP model predicts that the level of difficulty will be higher for cases of the SUBSET and NEW learning scenarios than for the SIMILAR scenario. Thus, for AusE and IS learners of BP, it is likely that development will first occur for /a/-/ɔ/ and /a/-/ɛ/ as the SIMILAR scenario was identified for these listeners. This will likely be followed by development in BP /e/-/ɛ/ which was not easy as the first two BP contrasts, but it was not as difficult as the remaining contrasts. Finally, it is likely that the development of the BP contrasts /i/-/e/, /o/-/u/ and /o/-/ɔ/ will take place last. This is because cases of the NEW and SUBSET scenarios were identified for these contrast, in which both vowels in the BP contrast were perceived as the same native vowel(s) thus making them difficult to perceive and recognise.

6.6. Predictions for non-native and L2 production

Recall that the L2LP model claims that non-native speech perception, spoken word recognition and speech production are all interrelated. Specifically, the model states that learners will initially perceive and produce L2 sounds in the same way as they perceive and produce sounds in their L1. The findings in Chapter 2 (Elvin et al., 2014) and Chapter 4 (Elvin et al., under review a) have shown that listeners' discrimination difficulty is influenced by the acoustic similarity between their own L1 vowel categories and those of the L2. Discrimination difficulty was also predicted by AusE and IS' listeners' categorisation of non-native vowels to native vowel categories. It is therefore expected that naïve AusE and IS listeners' vowel productions of the same non-native BP novel words used as stimuli in this thesis would match the acoustic properties of their L1 vowel categories. Below I make predictions for non-native and L2 vowel production for the same listeners tested in Elvin et al. (under review a) and Elvin et al. (under review b). In order to test whether these predictions are met, the acoustic similarity between these individuals' non-native vowel production, their own native vowel categories as well as native BP speakers' vowel productions should be measured. Ideally, the analyses of non-native vowel production should be similar to those reported in Elvin et al. (2014) and Elvin et al. (under review a), namely measurements of Euclidean Distances and cross-linguistic discriminant analyses. Additionally, AusE and IS speakers' non-native vowel productions could also be judged by native speakers to determine how native-like their vowel productions are. It is important to note that the predictions made below may be subject to individual variation and that some of these predictions may not be borne out due to any additional difficulties (e.g., established motor commands) that the learner may have beyond perception.

When producing BP vowels in a /fVfe/ context, AusE and IS listeners are expected to produce separate vowel categories for those BP contrasts that were perceptually easy to discriminate and recognise (i.e., /a- ɛ/ and /a- ɔ/). That is, it is likely that an AusE and an IS speaker will produce two distinct vowels with similar acoustic properties to those native vowels in which they were categorised and are in close acoustic proximity. In the BP /a/-/ɛ/ contrast, AusE listeners are likely to produce BP /a/ with acoustic properties that are similar to AusE /ɐ/, /ɐː/ and /æ/. For BP /ɛ/ AusE listeners are expected to produce a vowel that is distinct from BP /a/ and acoustically similar to AusE /e/ and/or /eː/. Likewise for the BP /a/-/ɔ/ contrast AusE speakers are likely to produce BP /ɔ/ with similar acoustic properties to /ɔ/ or /oː/. IS speakers are also expected to produce two separate categories for the BP /a- ɛ/ and /a/-/ɔ/ vowel contrasts. In particular they are expected to produce a BP /a/ that is acoustically similar to IS /a/ and BP /ɛ/ with similar acoustic properties to IS /e/. According to IS listeners' non-native categorisation of BP /ɔ/, they should produce this vowel in the same way as IS /o/, however the acoustic analyses in Elvin et al. (under review a) also indicated that this vowel was acoustically similar to IS /e/ and /o/ and there may be a chance some IS listeners will produce BP /ɔ/ in a similar manner.

It is expected that both AusE and IS speakers will have difficulties producing two distinct vowel categories in the BP contrasts /i/-/e/, /o/-/u/ and /o/-/ɔ/. This is because evidence of L2LP's SUBSET scenario for AusE and the NEW scenario for IS was found in the acoustic analyses and the non-native categorisation patterns reported in Elvin et al. (under review a). For each of the BP vowels in the contrast /i/-/e/, AusE listeners are expected to produce both vowels as one single vowel category that is acoustically similar to AusE /iː/, /ɪ/ and /ɪə/. In non-native categorisation, BP /e/ was also categorised as AusE /e/ and /eː/ and it is possible that some non-native vowel productions would be similar to these two vowels. Although both BP vowels are also acoustically similar to AusE /ʉː/, they were never categorised to this category and it is unlikely that AusE speakers will produce a BP /i/ or /e/ vowel similar to AusE /ʉː/. IS listeners are likely to produce both BP /i/ and /e/ as one single vowel category with similar acoustic properties to IS /i/. However, a large percentage of BP /e/ vowel tokens were categorised as IS /e/ and it is likely that some productions of BP /e/ could fall in the acoustic space for IS /e/. The vowels in the BP /o-/u/ contrast are likely to be produced as one single vowel category with similar acoustic properties as AusE /ʊ/ and /oː/. Likewise, given the acoustic similarity and non-native categorisation patterns reported in Elvin et al. (under review a), it is likely that IS listeners will produce a single vowel category for both BP /o/ and /u/ that falls somewhere in the IS /o/-/u/ vowel space. In the case of BP /o/-/ɔ/ although there was less acoustic overlap for AusE listeners than IS listeners, there was a large amount of perceptual overlap. The results from discrimination and word recognition tasks also indicated that this contrast was difficult to perceive and recognise. It is therefore expected that AusE listeners' productions of BP /o/-/ɔ/ will be in line with their non-native categorisation patterns. In particular, both vowels should be produced similarly to AusE /oː/. Likewise IS listeners should produce both vowels as IS /o/.

Finally, in the case of BP /e/-/ɛ/, this vowel contrast was easier to discriminate than it was to recognise in word recognition for both AusE and IS listeners. According to acoustic similarity between both vowels should be produced as two separate categories. However in their non-native vowel categorisation there was a partial perceptual overlap. Therefore AusE listeners may produce BP /e/-/ɛ/ as distinct vowel categories. As mentioned above, BP /e/ should be produced similarly to AusE /iː/, /ɪ/ and /ɪə/ for and BP /ɛ/ should be produced as AusE /e/ and /eː/. As previously mentioned, some tokens of BP /e/ were also perceived as AusE /e/ and /eː/ and AusE speakers may produce this BP vowel with similar acoustic properties. A similar case is expected for IS listeners, while BP /e/ should predominately be produced as IS /i/ and BP /ɛ/ as IS /e/, there may be some cases where BP /e/ is also produced as IS /e/.

6.7. Implications and future research

The present study has shown how the native language influences Australian English and Iberian Spanish listeners' non-native perception and spoken word recognition and how this shapes individual differences within groups. Specifically, the findings from this thesis have shown that acoustic similarity and non-native categorisation patterns influence discrimination difficulty. The findings have also indicated that in most cases, individual listeners' word recognition difficulties are related to their difficulties in discrimination. As a result there are a number of implications for the field of non-native and L2 speech perception and spoken word recognition.

Firstly, in Chapter 4 (Elvin et al., under review a), when investigating individual variation in the relationship between acoustic similarity and non-native discrimination difficulty and between non-native categorisation patterns and non-native discrimination difficulty, non-native categorisation was a better predictor of discrimination difficulty for IS listeners and acoustic similarity was a better predictor for AusE. Furthermore, within groups, some speakers showed a stronger effect of acoustic overlap, while others showed a stronger effect of non-native categorisation. PAM predicts that perceivers detect and use such articulatory properties in determining cross-language phonetic similarities (i.e., non-native categorisation or perceptual assimilation patterns). On the other hand L2LP claims that acoustic similarity can predict both non-native categorisation and discrimination difficult. Interestingly, the results from Elvin et al. (under review a) indicate that a combination of predictions based on non-native categorisation patterns and acoustic similarity is necessary for the most accurate predictions of non-native discrimination difficulty. Thus, future work might further aim to examine correlations between articulatory and acoustic properties of the target vowels and evaluate their contributions to non-native/L2 categorisation and discrimination patterns.

The analyses of individual differences suggested that different task demands and different learner types among individuals might have caused the negative correlations for some of the AusE and IS participants. In regards to the different learner types, it may be that the listeners with lower word recognition accuracy scores are sequential learners (i.e., those who rely on strictly bottom-up processing) who require more input in order to be able to learn words with difficult non-native contrasts, whereas the majority of listeners are likely to be interactive learners (i.e., those who allow lexical feedback to perception). Future studies would need to confirm the proposal made in this thesis that these listeners with lower accuracy scores and negative correlations between non-native speech perception and spoken word recognition are in fact sequential learners. For example, studies might consider including two learning phases (with different exposure amounts) and two identical testing phases to find out if there is indeed a mix of sequential vs. interactive learners in both language groups and to determine whether sequential learners do in fact benefit from a greater amount of exposure to target words.

It is interesting to note that there was a greater amount of individual variation for AusE than for IS listeners and of the eight possible sequential learners, five were AusE. This leads to an unanswered question of whether differences in mixes of learner types are principled or accidental. In other words, is there anything about the native language or linguistic environment that favours one particular learning type over the other? It could be that there is greater potential for languages with larger vowel inventories to have a greater mix of sequential learners because of the multiplicity of available response categories in their L1. In that respect and as discussed in Chapter 5 (Elvin et al. under review b), Benders et al. (2012) found that listeners who had more response vowel categories available were less sensitive to broad and local acoustic contexts than listeners presented with only two response categories. In addition, they took longer to shift their boundary between Spanish /i/ and /e/

and the authors argued that this delay in the boundary shift was caused by the availability of extra response categories. Thus it could be that there may be more biases toward sequential learners for listeners with more vowel categories in their native language because they need to pay closer attention to the multiplicity of vowel categories activated and the smaller differences between those vowels. Additionally, van Leussen & Escudero (2015) claim that sequential learners have greater difficulties reducing the number of L1 categories employed in perception when learning a language with a smaller vowel inventory. Thus it is more likely that there would be a greater number of sequential learners among listeners with larger vowel inventories because the sequential processing arises from the native language having more vowels than the L2 because they have a larger number of categories that they would need to reduce. A finding of this kind would have important theoretical questions and should be investigated in future empirical studies.

One of the strengths of the L2LP model that has been reliably tested in this thesis is that it considers individual variation and individual development at all stages of learning and across all communicative abilities (non-native perception, spoken word recognition and non-native production). Considering that the findings for speech perception were also in line with the PAM theoretical framework, it is likely that this would extend to spoken word recognition. This thesis has empirically established a relationship between native vowel production and non-native perception and spoken word recognition at an individual level, it seems appropriate that other models of L2 perception (e.g., SLM, PAM, PAM-L2) should be expanded to consider the same scope as L2LP in linking individual learners' sound perception to spoken word recognition at all stages of learning.

Given that the present thesis is the first to investigate the individual differences in non-native perception and spoken word recognition of Brazilian Portuguese by two groups of listeners whose native languages differ in vowel inventory size, limitations and possibilities for future research are not surprising. For instance, this thesis was limited in the fact that it investigates non-native perception and spoken word recognition of Brazilian Portuguese monophthongs produced in the /fVfe/ context only. Future investigations could extend this study to both BP monophthongs and diphthongs produced in a variety of consonantal contexts. Because the focus was on novel words, future studies should also consider the application of the present study in the learning of real Portuguese words.

The experiments reported in Chapters 4 (Elvin et al., under review a) and 5 (Elvin et al., under review b) were conducted during the same experimental session. Consequently, all participants completed the spoken word recognition task prior to the non-native perception task, as the emphasis was to replicate a more natural L2 learning setting. However, as mentioned in Chapter 5, it might be that training effects (arising from the discrimination task being performed after exposure to the novel words) influenced listeners' discrimination accuracy. Specifically, listeners may have had poorer discrimination performance had they not first learned the words. Future studies might consider counterbalancing these tasks to control for training effects or testing discrimination prior to word learning and word recognition and again after, or administer tasks in different sessions or on different days. Studies have also shown the influence of orthography in non-native speech perception and spoken word recognition, so additional conditions with and without orthography may provide additional insight into individual learners' developmental patterns.

In sum, this thesis has presented a reliable test of the L2LP model which claims that acoustic similarity between the listeners' own native production and those of the target language can predict non-native categorisation and discrimination patterns. It also provides additional evidence that non-native speech perception is linked to spoken word recognition. The findings are consistent with previous studies (e.g. Escudero et al., 2008, 2013, 2014; Weber & Cutler, 2004) which find that auditory confusable L2 or non-native vowel contrasts are also confused in word recognition. The individual differences approach used in this thesis seems to lead to a more accurate comparison of listeners from different language backgrounds thus confirming one of the main proposals of the L2LP model. Importantly the findings from the studies reported in this thesis are not consistent with studies (e.g., Iverson & Evans, 2007; 2009) which show that learners with a larger vowel inventory than the target language are likely to better learners than those with a smaller vowel inventory. In fact, the performance of AusE listeners was comparable to that of IS and in the case of Elvin et al. (under review b) and Elvin et al. (2014), worse than IS. This finding suggests that while it is still very important to consider vowel inventory size together with acoustic similarity, having a larger vowel inventory than the target language is not always a guarantee for success, at least at the initial state of learning. Finally, the findings from the present study emphasise the fact L2 performance should be described, predicted and explained at an individual level as explicitly stated in the L2LP model. Further development of the L2LP model might include investigations that confirm individual performance at various stages of L2 development and longitudinal studies that track learners' development in non-native speech perception, spoken word recognition and production from the initial state to ultimate attainment.

6.8. Concluding summary

This thesis has investigated the influence of the native language on Australian English and Iberian Spanish listeners' non-native perception and spoken word recognition of Brazilian Portuguese. The thesis presented an adequate test of some of the L2LP model's core assumptions, including the appropriate scope of L2 perception models. Specifically, acoustic similarity was predictive of listeners' non-native categorisation and discrimination patterns, and that difficulty in discrimination positively correlated with word recognition difficulty. Finally, the analysis of individual differences in both abilities was essential for understanding group findings, particularly in word recognition. These findings of individual differences also provide support for the L2LP claims that there are different developmental patterns for individual learners at the initial state of L2 learning.

References

Alispahic, S., Escudero, P., & Mulak, K. E. (2014). Difficulty in discriminating non-native vowels: Are Dutch vowels easier for Australian English than Spanish listeners? *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 1293–1296. http://doi.org/10.13140/2.1.2572.6726

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*(2), 233–250. http://doi.org/10.1016/S0095-4470(03)00036-6

Auer, E. T. (2009). Spoken word recognition by eye. *Scandinavian Journal of Psychology, 50*(5), 419–25. http://doi.org/10.1111/j.1467-9450.2009.00751.x

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels : The role of individual differences. *IRAL, International Review of Applied Linguistics in Language Teaching, 44*(3), 231–250. http://doi.org/10.1515/IRAL.2006.010

Benders, T., Escudero, P., & Sjerps, M. J. (2012). The interrelation between acoustic context effects and available response categories in speech sound categorization. *The Journal of the Acoustical Society of America, 131*(4), 3079. http://doi.org/10.1121/1.3688512

Best, C. T. (1994). The Emergence of Native-Language Phonological Influences in Infants : A Perceptual Assimilation Model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words, 167*, 167–224.

Best, C. T. (1995). A direct realist perspective on 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. T., Faber, A., & Levitt, A. G. (1996). Assimilation of non-native vowel contrasts to the American English vowel system. *Journal of the Acoustical Society of America,* *99*(4), 2602–2603.

Best, C. T., Halle, P., Bohn, O., & Faber, A. (2003). Cross-language perception of nonnative vowels : Phonological and phonetic effects of listeners' native languages. *In Proceedings of the 15th international congress of phonetic sciences* (pp. 2889–2892).

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109*(2), 775–794.

Best, C. T., & Strange, W. (1992). Effects of Phonological and Phonetic Factors on Cross-Language Perception of Approximants. *Journal of Phonetics, 20*(3), 89–108.

Best, C. T., & Tyler, M. D. (2007). Non-native and second-language speech perception: commonalities and complementarities. In O. Bohn & M. J. Munro (Eds.), *Language Experience in Second-Language Speech Learning: In Honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Billington, R. (2011). Location, location, location! Regional characteristics and national patterns of change in the vowels of Melbourne adolescents. *Australian Journal of Linguistics, 31*(3), 275-303.

Boersma, P., and Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 6.0.05, retrieved 8 November 2015 from http://www.praat.org/

Bohn, O., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics 11*(03), 303–328.

Bohn, O. S., Best, C. T., Avesani, C., & Vayra, M. (2011). Perceiving through the lens of native phonetics: Italian and Danish listeners perception of English consonant contrasts. In *Proceedings of the XVII International Congress of Phonetic Sciences* (pp. 336–339).

Bohn, O.S., & Flege, J. E. (1992). The Production of New and Similar Vowels by Adult German Learners of German. *Studies in Second Language Acquisition, 14*(02), 131–158.

http://doi.org/10.1017/S0272263100010792

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*(4), 447–465.

Bradley, D. (2008). Regional characteristics of Australian English: phonology. In *Varieties of English: The Pacific and Australasia* (pp. 111–123).

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America, 97*(3), 1916–1924.

Bradlow, A. R., Pisoni, D. B., Akahane-yamada, R., & Tohkura, Y. (1997). Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*(4), 2299–2310.

Broersma, M. (2002). Comprehension of non-native speech: inaccurate phoneme processing and activation of lexical competitors. In *INTERSPEECH* (pp. 261–264).

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*(01), 51–67. http://doi.org/10.1017/S0142716410000287

Cenoz, J., & Lecumberri, L. G. (1999). The effect of training on the discrimination of English vowels. *IRAL-International Review of Applied Linguistics in Language Teaching*, *37*(4), 261–276.

Chládková, K., Escudero, P., & Boersma, P. (2011). Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *The Journal of the Acoustical Society of America, 130*(1), 416–28. http://doi.org/10.1121/1.3592242

Clopper, C. G., & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. In *Language Variation and Change*,18, pp. 193–

221). http://doi.org/10.1017/S0954394506060091.

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second Language Speech*. Cambridge University Press.

Cox, F. (2006). The Acoustic Characteristics of /hVd/ Vowels in the Speech of some Australian Teenagers. *Australian Journal of Linguistics*, *26*(2), 147–179. http://doi.org/10.1080/07268600600885494

Cox, F., & Palethorpe, S. (1998). Regional variation in the vowels of female adolescents from Sydney. *Proceedings of the 5th International Conference on Spoken Language Processing*, *6*, 2359–2362.

Cox, F., & Palethorpe, S. (2004). The Border Effect: Vowel Differences across the NSW - Victorian Border. *Proceedings of the 2003 Conference of the Australian Linguistics Society*.

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word – object associative learning. *Developmental Science, 12*(5), 725–731. http://doi.org/10.1111/j.1467-7687.2009.00814.x

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research, 14*(4), 389-405.

Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle & J. M. Beck (Eds.), *A Figure of Speech: A Festschrift for John Laver* (pp. 63–91). London: Lawrence Erlbaum Associates.

Cutler, A., & Pasveer, D. (2006). Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. In *3rd International Conference on Speech Prosody,* TUD press.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*, 269–284. http://doi.org/10.1016/j.wocn.2005.06.002

Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences, 105*(42), 16083–16088.

Díaz, B., Mitterer, H., Broersma, M., Escera, C., & Sebastián-Gallés, N. (2015). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognition*, 1-16.

Díaz, B., Mitterer, H., Broersma, M., & Sebastian-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences, 22*(6), 680–689. http://doi.org/10.1016/j.lindif.2012.05.005

Diaz Granado, M. (2011). L2 and L3 acquisition of the Portuguese stressed vowel inventory by native speakers of English.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On Being Blinded by Your Other Language: Effects of Task Demands on Interlingual Homograph Recognition. *Journal of Memory and Language, 42*(4), 445–464. http://doi.org/10.1006/jmla.1999.2697

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On Being Blinded by Your Other Language: Effects of Task Demands on Interlingual Homograph Recognition. *Journal of Memory and Language, 42*(4), 445–464. http://doi.org/10.1006/jmla.1999.2697

Dijkstra, T., Van Jaarsveld, H., & Brinke, S. Ten. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition, 1*(1), 51–66. http://doi.org/10.1017/S1366728998000121

Dörnyei, Z., & Shekhan, P. (2003). Individual differences in second language learning. In D.

J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 589–601). Malden, MA: Blackwell.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*, 315–30.

Elliott, A. R. (1995). Foreign Language Phonology : Field Independence , Attitude , and the Success of Formal Instruction in Spanish Pronunciation. *The Modern Language Journal, 79*, 356–371.

Ellis, R. (2004). Individual Differences in Second Language Learning. In *The handbook of applied linguistics* (p. 525).

Elvin, J., & Escudero, P. (2014). Perception of Brazilian Portuguese vowels by Australian English and Spanish Listeners. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech. Concordia Working Papers in Applied Linguistics*, *5*, 145–156.

Elvin, J., & Escudero, P. (2015). Predicting discrimination accuracy through cross-linguistic acoustic analyses', in *The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

Elvin, J., Escudero, P., & Vasiliev, P. (2014). Spanish is better than English for discriminating Portuguese vowels: Acoustic similarity versus vowel inventory. *Frontiers in Psychology*, *5*, 1188. http://doi.org/10.3389/fpsyg.2014.01188

Elvin, J., Escudero, P., Williams, D., Shaw, J. A., & Best, C. T. (under review a). The role of acoustic similarity and non-native categorisation in predicting non-native discrimination: Brazilian Portuguese vowels by English vs. Spanish listeners.

Elvin, J., Escudero, P., Williams, D., Shaw, J. A., & Best, C. T. (under review b). The role of speech perception and individual differences when learning novel words in a non-native language: English vs Spanish learners of Portuguese

Elvin, J., Vasiliev, P. & Escudero, P. (in press). Production and Perception in the Acquisition of Spanish and Portuguese. In M. Gibson & J. Gil (eds) *Romance Phonetics and Phonology*, Oxford University Press

Elvin, J., Williams, D., & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *The Journal of the Acoustical Society of America, 140*, 576–581. http://doi.org/10.1121/1.4952387

Escudero, P. (2001). The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. *Proceedings of the 25th Annual Boston University Conference on Language Development*, 250–261.

Escudero, P. (2005). Linguistic Perception and Second Language Acquisition. *PhD Dissertation*, Utrecht University.

Escudero, P. (2006). Second language phonology: The role of perception. In *Phonology in context* (pp. 109-134). Palgrave Macmillan UK.

Escudero, P. (2009). The linguistic perception of similar L2 sounds. *Phonology in Perception, 15*, 152–190.

Escudero, P. (2011). Speech processing in bilingual and multilingual listeners. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Handbook of Laboratory Phonology* (pp. 407–417). Oxford University Press.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37*(4), 452–465. http://doi.org/http://dx.doi.org/10.1016/j.wocn.2009.07.006

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of*

America, 130(4), EL206–12. http://doi.org/10.1121/1.3629144

Escudero, P., Best, C. T., Kitamura, C., & Mulak, K. E. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Frontiers in Psychology, 5*. http://doi.org/10.3389/fpsyg.2014.01059

Escudero, P., & Boersma, P. (2002). The Subset Problem in L2 Perceptual Development : Multiple-Category Assimilation by Dutch Learners of Spanish. In *Proceedings of the 26th annual Boston University conference on language development* (pp. 208–219). Somerville, MA: Cascadilla Press.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(4), 551–585.

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. a H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America, 126*(3), 1379–93. http://doi.org/10.1121/1.3180321

Escudero, P., Broersma, M., & Simon, E. (2013). Learning words in a third language: Effects of vowel inventory and language proficiency. *Language and Cognitive Processes, 28*(6), 746-761. http://doi.org/10.1080/01690965.2012.662279

Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America, 128*(5), EL254–9. http://doi.org/10.1121/1.3488794

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

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics, 40*(2), 280–288.

http://doi.org/http://dx.doi.org/10.1016/j.wocn.2011.11.004

Escudero, P., Simon, E., & Mulak, K. (2014). Learning words in a new language: orthography doesn't always help. *Bilingualism: Language and Cognition, 17*(02), 384–395.

Escudero, P., Sisinni, B., & Grimaldi, M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *The Journal of the Acoustical Society of America, 135*(3), 1577–1584. http://doi.org/10.1121/1.4864477

Escudero, P., & Vasiliev, P. (2011). Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the Acoustical Society of America, 130*(5), EL277–EL283.

Escudero, P., & Wanrooij, K. (2010). The Effect of L1 Orthography on Non-native Vowel Perception. *Language and Speech, 53*(3), 343–365. http://doi.org/10.1177/0023830910371447

Escudero, P., & Williams, D. (2011). Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *The Journal of the Acoustical Society of America, 129*(1), EL1–EL7.

Escudero, P., & Williams, D. (2012). Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *The Journal of the Acoustical Society of America, 131*(5), EL406–12. http://doi.org/10.1121/1.3701708

Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition, 133*(2), 408–413. http://doi.org/10.1016/j.cognition.2014.07.002

Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America, 139*(1), EL1-EL5. http://doi.org/10.1121/1.4939608

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America, 84*(1), 70-79.

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

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). Timonium, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437–470. http://doi.org/http://dx.doi.org/10.1006/jpho.1997.0052

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25*(2), 169–186.

Flege, J. E., Mackay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America, 106*(5), 2973-2987.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*, 3125-3134.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Halle, P. (2008). Cross language phonetic influences on the speech of French – English bilinguals. *Journal of Phonetics, 36*(4), 649–663. http://doi.org/10.1016/j.wocn.2008.04.001

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. http://doi.org/10.1121/1.411974

Gilichinskaya, Y. D., & Strange, W. (2010). Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the Acoustical Society of America, 128*(2), EL80–5. http://doi.org/10.1121/1.3462988

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language, 109*(2), 55–67. http://doi.org/10.1016/j.bandl.2008.01.005

Gordon, L. S. (2008). Factors affecting English speakers' perception of L2 vowels. ProQuest. Georgetown University.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia, 9*(3), 317-323.

Grabe, E., & Low, E. L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. *Papers in Laboratory Phonology, 7*, 515–546.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107*(5), 2711–2724.

Harrington, J., & Cassidy, S. (1994). Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English. Language and Speech, 37(4), 357-373.

Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general , and cultivated Australian English vowels. *Australian Journal of Linguistics, 17*(2), 155-184.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research, 1*, 5–33.

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the Phonological Forms of New Words: Effects of Orthographic and Auditory Input. *Language and Speech, 53*(3), 367–

381. http://doi.org/10.1177/0023830910371460

Herrmann, F., Cunningham, S. P., & Whiteside, S. P. (2014). Speaker sex effects on temporal and spectro-temporal measures of speech. *Journal of the International Phonetic Association, 44*(01), 59–74. http://doi.org/10.1017/S0025100313000315

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*, 3099–3111.

Hockett, C. F. (1955). A manual of phonology. *IJAL, Memoir 11*, Indiana University Publications in Anthropology and Linguistics. Waverly Press: Baltimore, Maryland

Honikman, B. (1964). Articulatory Settings. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, & J. L. M. Trim (Eds.), In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday 12 September 1961 (pp. 73–84). London: Longmans, Green & Co.

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), 2842–54. http://doi.org/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. http://doi.org/10.1121/1.3148196

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*, 47–57. http://doi.org/10.1016/S0

Jacewicz, E., & Fox, R. A. (2013). Cross-Dialectal Differences in Dynamic Formant Patterns in American English Vowels. In *Vowel inherent spectral change* (pp. 177-198). Springer Berlin Heidelberg.

Jin, S., & Liu, C. (2013). The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *The Journal of the Acoustical Society of America, 133*(5), EL363-EL369. http://doi.org/10.1121/1.4798620

Jin, Y., Díaz, B., Colomer, M., & Sebastián-Gallés, N. (2014). Oscillation encoding of individual differences in speech perception. *PLoS ONE, 9*(7). http://doi.org/10.1371/journal.pone.0100901

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: the status of subjacency in the acquisition of a second language. *Cognition, 39*(3), 215–58.

Kendall, R. R. (2004). The Perception and Production of Portuguese Mid- Vowels by Native Speakers of American English. *BYU Dissertation*

Kisler, T., Schiel, F., & Sloetjes H. (2012). Signal processing via web services: the use case WebMAUS. In *Proceedings of Digital Humanities* (pp. 30–34). Hamburg.

Kluge, D. C., Rauber, A. S., Reis, M. S., Bion, R. A. H., & Program, G. (2007). The Relationship between the Perception and Production of English Nasal Codas by Brazilian Learners of English. In *Interspeech* (pp. 2297–2300).

Kondaurova, M. V, & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America, 124*(6), 3959. http://doi.org/10.1121/1.2999341

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97*(22), 11850–11857.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech

Perception and Later Language Development: Implications for the “Critical Period .” *Language Learning and Development*, *1*(3-4), 237–264.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21.

Kuhl, P. K., Williams, K. A, Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606–8.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Boston: Wadsworth Cengage Learning.

Landick, M. (1995). The Mid-Vowels in Figures : Hard Facts. *The French Review, 69*(1), 88–102.

Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults’ French vowel learning. *The Journal of the Acoustical Society of America, 126*(5), 2670–2682.

Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America, 128*(3), 1290–305. http://doi.org/10.1121/1.3466879

Llisterri , J . (1995). Relationships between speech production and speech perception in a second language. In *Proceedings of the XIIIth International Congress of Phonetic Sciences, 4,* 92-99

Luce, P. A, Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62*(3), 615–625. http://doi.org/10.3758/BF03212113

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. http://doi.org/10.1097/00003446-199802000-00001

Mackay, I. R. A., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America*, *110*(1), 516–528. http://doi.org/10.1121/1.1377287

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Routledge.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71–102.

Marslen-Wilson, W., Marslen-Wilson, W., Tyler, L. K., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), b101–b111.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(03), 279–297. http://doi.org/10.1017/S1366728909990022

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. http://doi.org/10.1016/0010-0285(86)90015-0

McQueen, J. M. (2007). Eight questions about spoken-word recognition. *The Oxford Handbook of Psycholinguistics*, 37–53.

Mitchell, A. G. (1946). The pronunciation of English in Australia. Sydney: Angus and Robertson.

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

Mora, J. C., Keidel, J. L., & Flege, J. E. (2010). Why are the Catalan contrasts between /e/-/ɛ/ and /o/-/ɔ/ so difficult for even early Spanish-Catalan bilinguals to perceive? *New Sounds 2010. Sixth International Symposium on the Acquisition of Second Language Speech*, (DeKeyser 2000), 325–330.

Morrison, G. S. (2003). Perception and Production of Spanish Vowels by English Speakers. In *Proceedings of the 15th International Congress of Phonetic Sciences* (Vol. 2003, pp. 1533–1536). Barcelona.

Morrison, G. S. (2006). L1 & L2 Production and Perception of English and Spanish Vowels: A Statistical Modelling Approach. *Doctoral dissertation,* University of Alberta.

Morrison, G. S. (2009). L1-Spanish Speakers' Acquisition of the English/i/-/I/Contrast II: Perception of Vowel Inherent Spectral Change. *Language and speech, 52*(4), 437-462. http://doi.org/10.1177/0023830909336583

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, *21*(1), 81–108.

Nation, I. S. P. (2006). How Large a Vocabulary Is Needed For Reading and Listening? *Canadian Modern Language Review*, 63(1), 59–82.

Norris, D. (1994). Shortlist - a Connectionist Model of Continuous Speech Recognition. *Cognition, 52*(3), 189–234. http://doi.org/10.1016/0010-0277(94)90043-4

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. http://doi.org/10.1037/0033-295X.115.2.357

Norris, D., Mcqueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*(3), 191–243.

O’Brien, M. G., & Smith, L. C. (2010). Role of first language dialect in the production of second language German vowels. *IRAL - International Review of Applied Linguistics in Language Teaching*, *48*(4), 297–330.

Oliviera, I. G. (2006). A aquisição do sistema vocálico por falantes de português europeu como língua não materna. Universidade de Lisboa.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64*, B9–B17. http://doi.org/10.1016/S0010-0277(97)00030-9

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access : exemplar-based vs. abstract lexical entries. *Psychological Science*, *12*(6), 445–449.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*(1), 461–472. http://doi.org/10.1121/1.3593366

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2 : a review. *Journal of Phonetics*, *29*(2), 191–215. http://doi.org/10.006/jpho.2001.0134

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39*(4), 467–478. http://doi.org/10.1016/j.wocn.2010.08.007

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners’ perception and production of English vowels. *Journal of Phonetics, 40*(3), 491–508. http://doi.org/10.1016/j.wocn.2012.01.001

Rauber, A. S., Escudero, P., Bion, R. A. H., & Baptista, B. O. (2005). The Interrelation between the Perception and Production of English Vowels by Native Speakers of

Brazilian Portuguese Graduate Program in Applied Linguistics. *INTERSPEECH, 2,* 2913–2916.

Recasens, D., & Espinosa, A. (2009). Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Communication, 51*(3), 240–258. http://doi.org/10.1016/j.specom.2008.09.002

Scholes, R. J. (1968). Phonemic interference as a perceptual phenomenon. *Language and Speech, 11*, 86–103. http://doi.org/10.1177/002383096801100202

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some Effects of Training on the Perception of Synthetic Speech. *The Journal of the Human Factors and Ergonomics Society, 27*(4), 395–408.

Sebastián-Gallés, N., & Díaz, B. (2012). First and Second Language Speech Perception: Graded Learning. *Language Learning, 62*, 131–147. http://doi.org/10.1111/j.1467-9922.2012.00709.x

Sebastián-Gallés, N., Echeverrı, S., & Bosch, L. (2005). Memory and Language The influence of initial exposure on lexical representation : Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*(2), 240–255. http://doi.org/10.1016/j.jml.2004.11.001

Sebastián-gallés, N., Soriano-mas, C., & Baus, C. (2012). Neuroanatomical markers of individual differences in native and non-native vowel perception. *Journal of Neurolinguistics, 25*(3), 150–162. http://doi.org/10.1016/j.jneuroling.2011.11.001

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition, 72*(2), 111–123. http://doi.org/10.1016/S0010-0277(99)00024-4

Shatzman, K., & McQueen, J. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68*(1), 1–16.

http://doi.org/10.3758/BF03193651

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.

Smith, B. L., & Hayes-Harb, R. (2011). Individual differences in the perception of final consonant voicing among native and non-native speakers of English. *Journal of Phonetics, 39*(1), 115–120. http://doi.org/10.1016/j.wocn.2010.11.005

Soto-Faraco, S., Sebastián, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*(3), 412–432. http://doi.org/10.1006/jmla.2000.2783

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In *Language Experience in Second-Language Speech Learning: In Honor of James Emil Flege* (pp. 35–55).

Strange, W., Bohn, O.-S., Nishi, K., & Trent, S. a. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America, 118*(3), 1751. http://doi.org/10.1121/1.1992688

Strange, W., Bohn, O.-S., Trent, S. a., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America, 115*(4), 1791. http://doi.org/10.1121/1.1687832

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.

Strange, W., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America, 109*(4), 1691–1704. http://doi.org/10.1121/1.1353594

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language*, *58*(2), 366–379. http://doi.org/http://dx.doi.org/10.1016/j.jml.2007.11.002

Temple, L. (2000). Second Language learner speech production. *Studia Linguistica*, *54*(2), 288–297.

Tsao, F., Liu, H., & Kuhl, P. K. (2004). Speech Perception in Infancy Predicts Language Development in the Second Year of Life : A Longitudinal Study. *Child Development*, *75*(4), 1067–1084.

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), 4–21. http://doi.org/10.1159/000356237.

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, *6*, 1–12. http://doi.org/10.3389/fpsyg.2015.01000

Vasiliev, P. (2013). The Initial State for Californian English Learners of Spanish and Portuguese Vowels. University of California, Los Angeles.

Vasiliev, P., & Escudero, P. (2011). Cross-language perception of Brazilian Portuguese vowels by Californian English listeners. *The Journal of the Acoustical Society of America*, *130*(4), 2574.

Vroomen, J., & de Gelder, B. (1995). Metrical Segmentation and Lexical Inhibition in Spoken Word Recognition. *Journal of Experimental Psychology*, *21*(1), 98–108.

Wanrooij, K., Escudero, P., & Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel distribution ? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, *41*(5), 307–319.

http://doi.org/10.1016/j.wocn.2013.03.005

Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, 106(1), 458-468.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1–25.

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(3), 387–401. http://doi.org/10.1002/wcs.1178

Weiand, K. (2007). Implementing Escudero's model for the subset problem. Rutgers Optimality Archive, 913.

Werker, J. F., & Curtin, S. (2005). PRIMIR : A Developmental Framework of Infant Speech Processing. *Language learning and development, 1*(2), 197–234.

White, L., & Mattys, S. L. (2007). Calibrating rhythm : First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522.

Williams, D., & Escudero, P. (2014). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the Acoustical Society of America, 136*(5), 2751–2761. http://doi.org/10.1121/1.4896471

Williams, D., Leussen, J. Van, & Escudero, P. (2015). Beyond North American English: Modelling vowel inherent spectral change in British English and Dutch," in *The Scottish Consortium for ICPhS 2015 (ed.) Proceedings of the 18th International Congress of Phonetic Sciences,* August 10–14, University of Glasgow, Glasgow, UK.

Zahorian, S. A., & Jagharghi, A. J. (1993). Spectral shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America, 94*, 1966–1982.

Appendix A: Participant Background Questionnaire - English

Date: Participant ID:

Participant Background Information

1. Date of Birth (e.g., 12-Jan-1984): ____-______-________

day month year

2. Did/do you have any specific *hearing* difficulties, or *reading* difficulties (e.g., difficulties learning to read as a child), or *language development or speaking* difficulties (e.g., delayed language onset, stuttering, lisping, more-than-average difficulties in learning new words or remembering the names of objects)? Yes/No

If so, what is/was the nature of the problem? At what age did it occur? Did/does it require special assistance (e.g., hearing aids, a reading tutor, a language/speech therapist)?

3. Please tell us every place you have lived (or stayed for more than a month), in order, starting with the place you were born.

Country	**City or Region**	**Dates lived there?**	**How old were you?**
Australia	*Milperra*	*5/03 - 8/05*	*birth - present*

4. Where did your mother grow up (town/region/country/ age of arrival in Australia)?

Where did your father grow up (town/region/country/ age of arrival in Australia)?

5. What is your native language? ____________________________

What is your mother's native language? ____________________________

What is your father's native language? ____________________________

Do your parents generally speak with a particular regional accent of Australian English?

Yes / No

If so, which accent? ____________________________

6. Was there any other adult from a different country, *or* from a different city/region, who spent a large amount of time with you when you were growing up (for example, a grandmother, a live-in housekeeper)? If so, who, and during what ages in your childhood?

Where did this person grow up? (town/country) ____________________________

What is his or her native language? ____________________________

7. Please tell us what languages you speak, how long you have spoken them, and how well you speak and understand them.

(1=hardly at all; 7=highly fluent)

Language	**Years Used**	**Where did you use this language?**	**How well do you speak it?**	**How well do you understand it?**
Spanish	*2006-2008*	*high school*	*2*	*3*

7. Please provide your e-mail address if you are willing to be contacted for participation in future experiments related to this study

Appendix B: Participant Background Questionnaire - Spanish

Fecha: _______________ Código del Participante: ______________

INFORMACIÓN SOBRE EL IDIOMA DEL PARTICIPANTE

1. Fecha de nacimiento (por ejemplo, 12/01/1984): ______/______/______

día mes año

2. ¿Alguna vez has tenido dificultades *auditivas*, o dificultades *para la lectura* (por ejemplo, problemas de aprendizaje de lectura durante la niñez), o dificultades *en el desarrollo del lenguaje o en el habla* (por ejemplo, retraso en el inicio del habla, tartamudez, balbuceo, dificultades mayores que la media para aprender palabras nuevas o para recordar el nombre de los objetos)?

☐ SI ☐ NO

En caso afirmativo, ¿cuál es/fue la naturaleza del problema? ¿A qué edad ocurrió? ¿Necesitaste o necesitas alguna ayuda especial (por ejemplo, dispositivos de ayuda auditiva, tutor de lectura, terapeuta del lenguaje o del habla)?

3. Por favor señala en orden los lugares en que has vivido (o visitado durante más de un mes), comenzando por el lugar en que naciste.

	País	Ciudad o Región	Entre qué fechas	A qué edad
Ej: Nacimiento	*España*	*Madrid*	*5/1/1990 – 5/5/2015*	*Del nacimiento hasta la actualidad*

4. ¿Dónde creció tu madre (pueblo/ciudad/país)? ______________________________

¿Dónde creció tu padre (pueblo/ciudad/país)? ______________________________

5. ¿Cuál es tu idioma nativo? ______________________________

¿Y el de tu madre? ______________________________

¿Y el de tu padre? ______________________________

6. ¿Ha habido algún otro adulto de otro país, ciudad o región diferente con el que conviviste durante tu infancia (por ejemplo, una abuela, una asistenta del hogar, etc.)? De ser así, ¿quién?, ¿durante cuánto tiempo?

____

____

¿Dónde creció esta persona? (pueblo/país) ______________________________

¿Cuál era el idioma nativo de esta persona? ______________________________

7. Por favor señala qué idioma(s) hablas, por cuanto tiempo los has hablado y cómo de bien los hablas y entiendes.

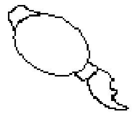
Idioma	Años de aprendizaje	¿Dónde empleaste este idioma?	1 = muy poco 7 = altamente fluido ¿Cómo de bien lo hablas?	¿Cómo de bien lo entiendes?
Ingles	*2006-2008*	*Secundaria*	*2*	*3*

Appendix C: English word prompts used in Elvin, Williams & Escudero (2016)

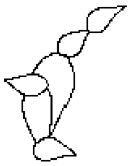
Vowel	dVt	fVf	sVs	bVp	gVk	dVk	hVd	example
iː	deet	feef	seess	beep	geek	deet	heed	see
ɪ	dit	fif	sis	bip	gik	dik	hid	bib
ɪə	deert	feerf	seerss	beerp	geerk	deerk	heerd	beer
e	det	fef	sess	bep	gek	dek	hed	web
eː	dairt	fairf	sairss	bairp	gairk	dairk	haird	hair
ɜː	durt	furf	surss	burp	gurk	durk	hurd	fur
ɐ	dut	fuf	suss	bup	guk	duk	hud	tub
ɐː	dart	farf	sarss	barp	gark	dark	hard	car
æ	dat	faf	sass	bap	gak	dak	had	cat
oː	dawt	fawf	sawss	bawp	gawk	dawk	hawd	saw
ɔ	dot	fof	soss	bop	gok	dok	hod	sob
ʊ	doott	fooff	sooss	boopp	gookk	dookk	hood	book
ʉː	doot	foof	soos	boop	gook	dook	hood	food
æɪ	dayt	fayf	sayss	bayp	gayk	dayk	hayd	ray
ɑe	dight	fighf	sighss	bighp	gighk	dighk	highd	high
æɔ	dowt	fowf	sowss	bowp	gowk	dowk	howd	cow
əʉ	doat	foaf	soass	boap	goak	doak	hoad	coat
oɪ	doyt	foyf	soyss	boyp	goyk	doyk	hoyd	boy

Appendix D: Word learning and word recognition stimuli

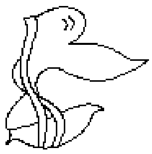
The following line drawings from Elvin, Escudero, Williams, Shaw & Best (under review c) were selected from the Shatzman and McQueen corpus and paired with one of the 14 Brazilian Portuguese words.



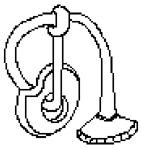
[fafe]



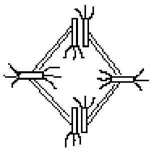
[fɛfe]



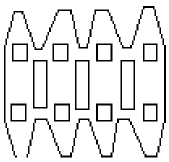
[fefe]



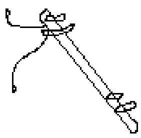
[fife]



[fofe]



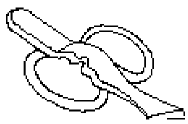
[fɔfe]



[fufe]

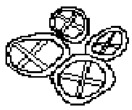
[kɔko]

[kuke]

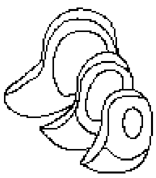


[pipe]

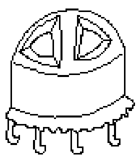
[popo]



[sase]



[sɛso]



[teko]