THE RELATIONSHIP BETWEEN SPEECH PERCEPTION AND WORD LEARNING AT THE INITIAL STATE OF SECOND LANGUAGE ACQUISITION

Samra Alispahic

Thesis submitted to Western Sydney University for the degree of Masters of Science (Honours) MARCS

The MARCS Institute for Brain, Behaviour and Development

ARC Centre of Excellence for the Dynamics of Language

Western Sydney University

August, 2017
Principal Supervisor: A/Prof. Paola Escudero

Associate Supervisors: Dr. Karen Mulak
Prof. Denis Burnham
ABSTRACT
The goal of most adult second language (L2) learners is to confidently and efficiently communicate in their target L2. However, this task is not easy. In order to produce new L2 words a learner first needs to perceive the sounds that comprise these words. Considered effortless in one’s native language (L1), distinguishing novel L2 phonemes can be quite difficult for adult learners, and difficulties in L2 speech perception are often attributed to the negative transfer effects of the L1. Research suggests that the size of the L2 vowel inventory relative to the L1 inventory may affect the discrimination and acquisition of L2 vowels. Specifically, if the L1 has a smaller L1 vowel inventory than the L2 this may obstruct L2 vowel perception, while if the L1 has a larger vowel inventory it often facilitates vowel perception. However, the Second Language Linguistic Perception (L2LP) model specifies that it is the L1-L2 acoustic relationships that predict L2 vowel perception, regardless of L1 vowel inventory size. The model further posits that there is continuity between L2 perception and L2 word learning (Escudero, 2005, 2006, 2009; van Leussen & Escudero, 2015). In this view, sounds that are difficult to perceive would yield comparable difficulty in learning L2 words containing the same sounds. The present thesis investigates Australian English (AusE), Peruvian Spanish (PS) and Spanish adult listeners’ non-native vowel perception and word learning of Dutch and the interrelation between these two abilities. In order to establish the initial state of learning the present thesis examines naïve listeners (i.e., AusE, PS and Spanish monolinguals) with no prior knowledge of Dutch.

This thesis comprises two studies, as well as an Introduction and General Discussion. Study 1 investigates the effects of vowel inventory size versus acoustic properties on non-native vowel perception. It compares XAB discrimination and categorization of five Dutch vowel contrasts between monolinguals whose L1 contains more (AusE) or fewer (PS) vowels than Dutch.
It also examines whether cross-language discriminant analyses predict listeners’ categorization patterns and whether these in turn predict their discrimination ability. Study 2 investigates the interrelation between listeners’ initial perceptual and word learning abilities by comparing two sets of previously published data. AusE and PS listeners’ XAB discrimination of five Dutch vowel contrasts, which was presented in Study 1, is compared to AusE and Spanish non-native word learning of minimal pairs containing the same vowel contrasts, which was previously reported in Escudero (2015). This comparison between perception and word learning data sets is done quantitatively, using the same statistical models, and also qualitatively. Results show that the size of participants’ native vowel inventories did not affect performance in either task. Rather it is the L1-L2 acoustic relationships that predicted listeners’ non-native categorisation and discrimination patterns and these in turn predicted their word learning difficulties. Specifically, minimal pair words containing perceptually difficult vowel contrasts were more difficult to discriminate, while word pairs containing perceptually easy contrasts were easier. Results presented in both studies confirm the L2LP proposal that there is continuity between perception and recognition in L2 development.
To Dika and Safet, and to new beginnings.
ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my Principal Supervisor, Associate Professor Paola Escudero for introducing me to the world of research. Her knowledge, experience and support have been invaluable. I thank Dr Karen Mulak for her guidance, patience and training that has taught me so much throughout this thesis. I would also like to extend my thanks to Professor Denis Burnham on his very helpful feedback and suggestions.

I would like to thank MARCS Institute for providing me with such a wonderful work environment. It’s hard to imagine facing technical and administrative upheavals without such an amazing support network. I thank Colin Schoknecht, Ben Binyamin, Johnson Chen, Donovan Govan and Muawiyath Shujau for all of their technical and research equipment support. Likewise, I thank Darlene Williams, Karen McConachie, Jen Armstrong, Craig Purcell and Sonya O'Shanna for all of their remarkable administrative assistance throughout this thesis. I thank my MARCS roommates, Tonya Agostini, Jaydene Elvin and Tina Whyte-Ball, as well as other members of the MARCS student cohort for their friendship, lunch table advice and the wonderful memories we have made along the way. I would like to thank the ARC Centre of Excellence for the Dynamics of Language (project number CE140100041) for supporting my research financially. I would also like to acknowledge and thank all my wonderful participants for their time and effort.

I thank my amazing friends Dani, Vicki, Elizabeth, Marina and Lee for their love and support every step of the way. Last but not the least; I thank my parents for always believing in me. I am eternally grateful for their selflessness, continued support and encouragement.
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.

Submitted: 31/08/2017

Revised: 26/02/2018
# Table of Contents

1 Introduction ......................................................................................................................... 1

1.1 Understanding another language is difficult ................................................................. 1

1.2 Challenges to language learning ...................................................................................... 2

1.2.1 Vocabulary and grammar acquisition .......................................................................... 2

1.2.2 L1-L2 perceptual interference ..................................................................................... 4

1.3 L2 Phoneme Perception ..................................................................................................... 6

1.3.1 Effects of L1 vowel inventory size on L2 sound perception ........................................ 7

1.3.2 Effects of L1-L2 acoustic relationships on L2 sound perception ................................. 10

1.3.3 Is it vowel inventory size and/or acoustic features that drive L2 perception? .............. 11

1.4 Influence of L1 Phonological Space on L2 Development ............................................... 14

1.4.1 Evidence suggesting a link between L2 sound perception and L2 learning .............. 14

1.4.2 Theoretical Models of L2 Speech Perception ............................................................... 21

1.5 Research Aims and Questions ........................................................................................ 29

1.5.1 Research Aims ........................................................................................................... 29

1.5.2 Research Questions ..................................................................................................... 30

2 Study 1 .................................................................................................................................. 33

2.1 Abstract ............................................................................................................................. 33

2.2 Introduction ....................................................................................................................... 34

2.3 Method ............................................................................................................................... 45

2.3.1 Participants ................................................................................................................... 45

2.3.2 Stimuli and Procedure ................................................................................................. 46

2.3.3 Statistical Analysis ....................................................................................................... 48

2.4 Results ............................................................................................................................... 48

2.4.1 Cross-language Discriminant Analyses ...................................................................... 48

2.4.2 Non-native vowel categorization ................................................................................ 49

2.4.3 XAB discrimination task ............................................................................................. 50

2.5 Discussion ........................................................................................................................ 52

3 Study 2 .................................................................................................................................. 59

3.1 Abstract ............................................................................................................................. 59

3.2 Introduction ....................................................................................................................... 60

3.3 Method ............................................................................................................................... 69

3.3.1 Task 1: XAB Discrimination ....................................................................................... 69
LIST OF TABLES

Table 1. Male speakers’ acoustic measures in Hertz of languages of the present study (AusE: Elvin, Williams, & Escudero, 2016; PS: Chládková, Escudero, & Boersma, 2011: Dutch: Adank et al., 2004). .................................................................38

Table 2. Percentage of Dutch vowel token classification to an AusE word and PS vowel based on overall classification patterns of cross-language LDA. .................................................40

Table 3. Percentage of Dutch vowel token classification to an AusE word and PS vowel based on classification patterns of individual dimension cross-language LDAs. .................42

Table 4. Categorization percentages of non-native Dutch vowels to AusE words by AusE listeners tested in the present study and to PS vowels by PS listeners as reported in Escudero and Williams (2011). .................................................................50

Table 5. Difficulty ranking, mean accuracy percentages and standard error (SE) for XAB Discrimination task ranging from most (1) to least difficult Dutch vowel contrast by 22 native AusE and 22 native PS listeners.................................................................52

Table 6. Predicted word recognition learning scenarios for AusE and Spanish listeners.....68

Table 7. Fisher’s LSD pairwise comparisons of significance levels and confidence intervals [CI] for the XAB Discrimination task by AusE and PS listeners.................................77

Table 8. Fisher’s LSD pairwise comparisons of significance levels and 95% confidence intervals [CI] for the Word Learning task by AusE and Spanish listeners. .....................80

Table 9. Percent accuracy and standard error for each contrast for the XAB discrimination and word learning task. .........................................................................................81

Table 10. Collapsed accuracy percentages across the XAB discrimination and word learning task ........................................................................................................82

Table 11. Fisher’s LSD pairwise significance levels for the main effect of contrast in both the XAB discrimination and word learning task.................................................83

Table A 1. AusE and PS participants’ mean accuracy scores, standard deviations and difficulty rankings for the XAB task

Table A 2. AusE and Spanish participants’ mean accuracy scores, standard deviations and difficulty rankings for the word learning task.................................................123
LIST OF FIGURES

Figure 1. Accuracy in discrimination of all non-native Dutch vowel contrasts by 22 native AusE and 22 native PS participants. Standard error bars were treated as Independent Variables. .................................................................51

Figure 2. Illustration of a word learning trial as reported in Escudero et al. (2014) and Escudero (2015). Figure adapted from (Giezen, Escudero, & Baker, 2015). .................72

Figure 3. Discrimination accuracy of all non-native Dutch vowel contrasts by 22 native AusE and 22 native PS participants (Alispahic et al., 2017; Escudero & Wanrooij, 2010). Interaction of vowel contrast and language background is labelled (*). .......................75

Figure 4. Combined overall discrimination accuracy of all non-native Dutch vowel contrasts by 22 native AusE and 22 native PS participants (Alispahic et al., 2017; Escudero & Wanrooij, 2010). Standard error bars were treated as Independent Variables. .................76

Figure 5. Word identification accuracy of minimal pair words containing non-native Dutch vowel contrasts (Escudero, 2015). .................................................................78

Figure 6. Word identification accuracy of non-native Dutch minimal pair vowel contrasts by naive AusE and Spanish listeners (Escudero, 2015). .........................................................79
1 INTRODUCTION

1.1 Understanding another language is difficult

Most people learning a second language (L2) aim to confidently and efficiently converse in their target L2. However, this task is not easy. Naïve learners face not only having to acquire at minimum a substantial amount of new words, a new grammatical system, but also the sound system (i.e. phonetics and phonology) of the new language. Considered effortless in one’s native language (L1), distinguishing novel L2 phonemes and words can be quite difficult for adult learners. Theoretical speech learning models stipulate that perception and learning of an L2 is directly influenced by a person’s L1 and that difficulties in L2 perception primarily arise as a result of how a learner categorizes L2 sounds in relation to their L1 sounds (Best, 1995; Best & Tyler, 2007; Escudero, 2005, 2009; Flege, 1995, 2003). Having a smaller vowel inventory appears to hinder L2 vowel perception (Escudero, Benders, & Lipski, 2009), while having a larger L1 vowel inventory facilitates it (Iverson & Evans, 2007, 2009). Further, L1-L2 difficulties at the phoneme level may extend to recognition and word learning difficulties (Broersma, 2005; Broersma & Cutler, 2008; A. Cutler & Otake, 1994; Anne Cutler & Broersma, 2005; Hayes-Harb & Masuda, 2008; Christophe Pallier, Colomé, & Sebastián-Gallés, 2001; Weber & Cutler, 2004). Specifically, there appears to be a correspondence between phonological and lexical discrimination difficulties faced by learners when distinguishing minimal word pairs (words which differ in a single consonant or vowel) containing non-native contrasts that are difficult to discern in perceptual tasks (Escudero, Hayes-Harb, & Mitterer, 2008).

The present masters thesis is designed to investigate the relationship between listeners’ initial L2 speech perception and L2 word recognition difficulties. Through the lens of The Second Language Linguistic Perception (L2LP) model which makes predictions regarding
listeners’ onset or initial state of L2 learning (Escudero, 2005, 2009; van Leussen & Escudero, 2015), this project examines and compares two speaker groups naïve to the target language – native speakers of Australian English and native speakers of Peruvian Spanish – in their perception of Dutch vowel contrasts and learning of Dutch minimal pair words containing the same contrasts. Specifically, it seeks to answer whether listeners’ initial stages of learning difficulties in L2 vowel perception arise from the effect of native vowel inventory size and/or the relationship between the L1-L2 acoustic features, and whether these difficulties further transpire in the initial stages of L2 word learning.

The current chapter is designed as follows: § 1.1 provides a brief introduction to the present thesis. Section § 1.2 provides an overview of challenges to language learning, § 1.3 will present how L1 perception affects L2 speech perception, while § 1.4 will discuss the influence of L1 phonological space on L2 development of how this affects the recognition of new words in an L2. Then, the most recent theoretical models on non-native and L2 speech perception will be discussed in § 1.4.2 relative to how they predict and explain perceptual and lexical difficulty for L2 learners. Subsequently, a detailed description of the research program and specific research questions will be provided in § 1.5.

1.2 Challenges to language learning

1.2.1 Vocabulary and grammar acquisition

Humans start learning the grammatical components of language very early on. Infants as young as 8 months have been shown to be sensitive to the word order in linguistic sequences of their first language (Benavides-Varela & Gervain, 2017). Word order carries important grammatical functions in sentences, but differs across languages. Similarly, the process of accessing and selecting words, placing them in grammatical utterances and planning speech differs across languages (Benavides-Varela & Gervain, 2017; Gentner & Goldin-Meadow,
2003). For example, Japanese sentence structures follow an Object-Verb (OV) order with postpositions following nouns while the opposite is true for Italian as the verb precedes the object (VO) and prepositions precede nouns (Bernard & Gervain, 2012). Eight-month old infants have been found to prefer the word order of their native language. That is, Japanese and Italian infants mirror the opposite word orders of their respective L1 and this carries over into adulthood (de Jong, 2016; FitzPatrick, 2007). Thus, adult learners approach L2 comprehension and acquisition with an already well-established L1 grammar system. This in turn results in learners’ having to learn how to deconstruct and eventually produce L2 language using the grammatical features of their target language. This is not an easy task and contrary to native speakers, L2 learners’ speech comprehension and production is not very spontaneous or intuitive and is often marked by pauses (de Jong, 2016; FitzPatrick, 2007). This is because L2 learners often pause to identify and segment speech into meaningful units, e.g., phrases or clauses, and to determine their relations, e.g., subject, verb, object (de Jong, 2016; Kim, 2014).

Difficulties for the learner in their acquisition of new L2 structures may occur for a number of reasons. When engaging in the process of L2 comprehension and acquisition listeners use all the available information as the speech signal unfolds (including phonological, metrical, syntactic, morphological and semantic information). However, difficulties in L2 processing may arise as a result of some L2 discourse distinctions not being present in listeners’ L1 (MacWhinney, 2002). In particular, difficulties may arise when comprehension processes at one level (e.g., syntactic encoding) are influenced by processes occurring at other levels (e.g., phonological encoding; Vigliocco, Vinson, Damian, & Levelt, 2002). Native English speakers learning Italian find noun inflections that mark gender and number very confusing (Williams, 1999). For example, in Italian *ragazza* is feminine singular for ‘girl’, while *ragazze* is feminine plural meaning ‘girls’. However, *tigre* is feminine
singular for ‘tiger’ while *tigri* is the plural form. In addition to gender marking not being present in English, the confusion for English learners may stem from the variability in the vowel quality that distinguishes the word endings. In fact, even native speakers of Italian are observed to make frequent lexical errors for e-final feminine words (Bates, Devescovi, Pizzamiglio, D’amico, & Hernandez, 1995). Native Italian speakers appear to treat these words as homophones due to minimal production differences of mid-vowels existing in the Italian language (Renwick & Ladd, 2016). Thus, for an English learner of Italian the difficulties in correctly identifying these lexical items are twofold. If they are unable to perceptually distinguish the difference between two non-native vowels then they will be unable to learn the non-native grammatical rule reliably.

### 1.2.2 L1-L2 perceptual interference

When a listener hears a unit of speech they are required to identify and understand its lexical components. Identifying individual words in running speech can at times be difficult as there are no obvious markers that indicate the start or end of a word boundary (Christophe, Gout, Peperkamp, & Morgan, 2003). Listeners attend to prosodic suprasegmental cues and metrical structures to aid in word segmentation, also referred to as the Metrical Segmentation Strategy (MSS; Cutler & Norris, 1988). However, while listeners are very good at distinguishing metrical word boundaries in their native language (Cutler & Otake, 1994; Cutler & Norris, 1988; Otake, Hatano, Cutler, & Mehler, 1993; Pallier, Christophe, & Mehler, 1997), this does not extend to L2 listening.

Evidence suggests that listeners’ segmentation strategies of their L1, such as the use of language-specific metrical patterns in speech to inform linguistic boundaries, are also applied in L2 segmentation (Otake et al., 1993). For example, French has well defined syllable boundaries, whereas English does not. That is, in French, *balance* is reliably syllabified as
bal'ance, and thus for French listeners ba primes balance more so than balcon, which is syllabified as bal'con. In contrast, English syllable boundaries are less strictly defined, such that ba'lace and bal'ance are equally permissible, and ba or bal prime balance equally. French listeners, but not English listeners, thus attend to syllable boundaries as linguistic boundaries. These skills transfer over to L2 listening, as French listeners will attend to syllable boundaries even when listening to English speech in which they are unreliable, and likewise English listeners ignore the reliable cue when listening to French speech (Cutler, Mehler, Norris, & Segui, 1986).

Language-specific listening has also been shown to transpire in languages that differ in their prosodic suprasegmental features (such as pitch, intensity, and duration). For example, Vietnamese learners of English rely more on the prosodic features of their L1, such as pitch compared to duration, and are able to more accurately perceive words that are produced by an English speaker with a greater pitch range than those with a more restricted pitch range (Nguyễ́n, Ingram, & Pensalfini, 2008).

Likewise, English learners of Spanish produce prosodic prominence within L2 words and phrases more accurately than English learners of French. This is due to the Spanish prosodic typology being more similar to English than French (Colantoni, Marasco, Steele, & Sunara, 2014). Thus, L2 learners classify non-native words in line with the suprasegmental features of their native language. This in turn may cause a number of obstacles in their perception of L2 words, which further results in them being less efficient than L1 listeners when segmenting continuous speech into recognizable words. However, in order to segment continuous L2 speech a listener is first required to perceive the L2 sounds that the speech contains.
1.3 L2 phoneme perception

In the first few months of their life, infants discriminate nearly all speech sounds in the world’s languages (Aslin, Jusczyk, & Pisoni, 1998). However, by 12 months of age infants’ attention to formerly discernible non-native contrasts that do not occur in their language environment is significantly reduced, while attention to contrasts found in the ambient languages persists, with perceptual patterns more closely corresponding to those of adults when listening to their native language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Maye, Werker, & Gerken, 2002; Werker & Tees, 1984). This shift in perceptual attention with particular focus on the acoustic-phonetic contrasts relevant to their native language is considered one of the earliest signs that an infant has begun acquiring their native language (Maye et al., 2002).

This focus on the acoustic-phonetic properties of one’s native language ensures efficient L1 speech processing, but can also result in non-native and L2 perceptual difficulty. Previous studies have suggested that non-native or cross-linguistic speech perception is dependent upon the perceptual filter of the L1 (Best, McRoberts, & Goodell, 2001; Best & Strange, 1992; Boomershine, 2013; Escudero & Boersma, 2002; Guion, Flege, Akahane-Yamada, & Pruitt, 2000). As languages differ on a phonemic level, perceptual difficulty is generally not uniform, and differing phonemic inventories (i.e., number of phonemes), as well as their acoustic characteristics affect L2 perception (Best, Halle, Bohn, & Faber, 2003). Identifying sound contrasts not present in a naïve listener’s L1 appears to be a difficult task, oftentimes resulting in multiple L2 sounds being perceived as a single native category (Kuhl et al., 1992). Theoretical speech learning models (SLM: Flege, 1995; PAM: Best, 1994, 1995; Best & Tyler, 2007; L2LP: Escudero, 2005), posit that non-native sound contrasts mapped to a single native category lead to the greatest discrimination and learning difficulty. For instance,
native speakers\textsuperscript{1} of Dutch rated the American English /θ/\text{-}/s/ and /θ/\text{-}/ʃ/ fricatives more similar to each other compared to native speakers of American English (Johnson & Babel, 2010), which is attributed to the lack of /θ/ in the Dutch sound inventory. Earlier studies have demonstrated that L2 vowel perception patterns, in large part, parallel those of L2 consonants (Best & Tyler, 2007). Nonetheless, compared to consonants vowels contain the most suprasegmental information (Polka & Werker, 1994). Further, vowels are known to be particularly difficult to acquire, as unlike consonants, boundaries between one vowel and another are less distinct (Ladefoged & Johnson, 2011). Therefore, research has endeavoured to compare native and target vowel inventories to establish how the relationship between the two influences learners’ speech perception and L2 development (e.g., Fox, Flege, & Munro, 1995; Gilichinskaya & Strange, 2010; Jia, Kohnert, Collado, & Aquino-Garcia, 2006; Lengeris, 2008). Outlined below are some studies that claim that it is the relationship between L1-L2 vowel inventory sizes that affects L2 perception, as well as those that claim that it is the acoustic L1-L2 differences and similarities that drive L2 perception.

1.3.1 Effects of L1 vowel inventory size on L2 sound perception

Research on non-native and L2 sound perception suggests that if some L2 vowel sounds are not present in a learner’s L1 then they will be more difficult to perceive and acquire (Fox et al., 1995). It is well documented that the English vowel contrast /i/-/ɪ/ (as in the words feel /fɪl/ and fill /fɪl/) poses considerable discrimination difficulties for learners whose L1 does not differentiate these two phonemes (e.g., Greek: Lengeris, 2008; Italian: Flege & MacKay, 2004; Polish: Zaorska, 2015; Portuguese: Escudero, Boersma, Rauber, & Bion, 2009; Russian: Kondaurova & Francis, 2008; Spanish: Escudero & Chládková, 2010). For example, Northern Italian learners of English were asked to perceptually discriminate and categorise

\textsuperscript{1} The term ‘native speaker’ throughout the thesis refers to the native perceiver and producer of a particular L1.
Canadian English vowels (Flege & MacKay, 2004). Discrimination results showed that along with Canadian English /ɪ-/ɪ/, learners found the /ɛ-æ/, /ɒ-ʌ/, and /ɪ-ɛ/ contrasts the most difficult to perceive. As four of the English vowels (/æ/, /ɒ/, /ʌ/, and /ɪ/) are not present in Italian, learners’ discrimination difficulties were consistent with the interpretation that learners perceived the two vowel sounds in each contrast as a single native category. That is, Italian learners with limited English experience categorized Canadian English /ɛ/ and /æ/ as Italian /ɛ/; /ɒ/ and /ʌ/ as Italian /a/; and /ɪ/ and /ɪ/ as Italian /i/ (Flege & MacKay, 2004). On the other hand, perceptual difficulties are not observed for all learners, particularly those whose native inventories do contrast the L2 contrasts, or an analogous contrast. For instance, when presented in non-native words (such as beet /bit/ and bit /bɪt/) German learners of English correctly identify the two English /i/ and /ɪ/ vowels 99% of the time (Flege, Bohn, & Jang, 1997a). Thus, for Italian learners having one /i/ vowel in their native vowel inventory appears to lead to higher discrimination difficulties, compared to German learners who seem to benefit from having two native /i/ and /ɪ/ vowel categories.

The findings described above have lead researchers to investigate the relationship between the size of the L1 and L2 vowel inventory and non-native and L2 vowel perception (Bundgaard-Nielsen, Best, & Tyler, 2011; Fox et al., 1995; Lengeris, 2008). In particular, it has been suggested that “the number of vowels in learners’ L1 influences their L2 vowel perception” (Bundgaard-Nielsen et. al., 2011, p.52). That is, having fewer L1 vowels than the target L2 will result in more perceptual difficulties, as it is likely that more than one L2 vowel will be categorized to some L1 categories and the learner will be required to learn new L2 vowels. On the other hand, having a larger L1 vowel inventory than the target language should facilitate L2 perception and eventual learning, since there are sufficient L1 categories available for all L2 sounds to map to without the need for two L2 sounds to map to a single category.
Studies have indeed shown that learners whose L1 vowel inventory contains more sound categories than the target language find it easier to perceive non-native sounds compared to those with fewer first-language sounds. For example, employing a series of perceptual tasks Iverson and Evans (2007) investigated how having a larger, German and Norwegian, or smaller vowel inventory, Spanish and French, compared to English affects L2 vowel perception. While the median years of experience using English varied across the L1 groups (Spanish – 29 years; French – 11 years; German – 28 years; and Norwegian – 9 years), all participants reported having learnt English since childhood. Even with the benefit of experience with English, participants with smaller L1 vowel inventories were less accurate at identifying English vowels. That is, while Spanish and French participants equated multiple English vowels to a single native category (e.g., English vowels /a/, /aɪ/, /aʊ/, and /ɑ/ as Spanish /a/), German and Norwegian participants mapped English vowels to individual L1 categories. These findings led the authors to conclude that there was a clear effect of L1 on L2 perception (Iverson & Evans). The authors further hypothesised that better L2 vowel perception should translate to better L2 vowel learning and that, given their larger L1 vowel inventory, German and Norwegian participants should, at least in theory, find L2 English vowels easier to learn than Spanish and French participants (Iverson & Evans, 2007).

In a follow up study, Iverson and Evans (2009) tested this hypothesis further. A second group of adult German and Spanish participants, who reported learning English since their adolescence, took part in five high-variability auditory training sessions learning vowels embedded in English words. Results show that the German group, who had a larger L1 vowel inventory, was more efficient at learning English vowels. That is, the German groups’ post-training accuracy scores improved twice as much compared to that of the Spanish group. To show the same learning improvement as the German participants, the Spanish participants required an additional ten sessions of training (Iverson & Evans, 2009a). In both studies it
was established that all listener groups appear to rely on primary acoustic cues, such as F1/F2
formant frequencies, formant movement and duration to similar extents in their identification
of English vowels (Iverson & Evans, 2007a, 2009a). These findings are interesting as formant
movement and duration are not present in Spanish and French (Iverson & Evans, 2007a), and
further suggest that although the size of a learner’s L1 vowel inventory may affect their L2
perceptual patterns, inventory size alone is not enough to accurately predict complexities of
L2 perceptual patterns. That is, additional acoustic-phonetic properties are also at play
affecting participants’ perceptual learning of English vowels (Iverson and Evans, 2007,
2009). As a result, it may therefore be useful to take into consideration how the L1-L2
acoustic-phonetic relationship affects listeners’ perceptual patterns of vowels.

1.3.2 Effects of L1-L2 acoustic relationships on L2 sound perception

Distinguishing vowels from one another usually occurs through processing formants
(Ladefoged & Johnson, 2011). Formant values are the product of various speech production
features, such as tongue height (resulting in changes in the first formant, or F1), tongue
backness (F2), and lip roundness (F3), which convey acoustic-phonetic properties to the
listener. Every language has its own unique acoustic-phonetic features. Every language and
accent has its own unique acoustic-phonetic features that native speakers become
perceptually attuned to. This native attunement then colours how subsequent languages and
accents are perceived and produced (PAM: Best, 1994, 1995, 2009). For example, while
Standard Scottish English and Standard Southern British English are closely related varieties
of English, they have different acoustic properties to one another. When listening to Standard
Scottish English and Standard Southern British /i/ and /ɪ/, native listeners' perceptual
behaviours closely align with the production characteristics of /i-ɪ/ in their native dialect
(Escudero & Boersma, 2004). That is, listeners’ perception of English vowels produced in a
different accent reflects their native accent. Research has further demonstrated that this
native accent effect also occurs when listening to L2 speech and determines listeners’ L2 perceptual patterns (Escudero & Williams, 2012). Acoustic differences in Peruvian Spanish and Iberian Spanish vowels lead to varying perceptual patterns between the two groups. That is, the acoustic properties of a listener’s native language determined how novel L2 speech sounds were perceived. Relative to L2 development, as learners’ experience with a target language progresses, their production and perception of L2 acoustic features becomes more nativelike compared to less proficient participants from the same L1 background (Flege et al., 1997). Native Spanish speakers learning English appear to perceive and produce English vowels differently on an acoustic level compared to Spanish speakers who are naïve to English. Specifically, acoustic analyses reveal that native Spanish speakers with more L2 experience may perceive and produce L2 vowels more like English natives, rather than those with the same L1 vowel inventory that have less L2 experience. For experienced L2 learners, this means that as their L2 experience develops so does their L1 perceptual filter.

1.3.3 Is it vowel inventory size and/or acoustic features that drive L2 perception?

As outlined above, differences between acoustic features of L1 and L2 vowels are seemingly an important driving force behind L2 speech perception. Thus, it is essential to address whether at the initial stages of learning difficulties in L2 vowel perception arise from the effect of L1 vowel inventory size and/or the relationship between the L1-L2 acoustic features. Investigating the initial stages of L2 learning will inform what developmental pathways listeners may go through as they progress from inexperienced to experienced L2 learners.

The Second Language Linguistic Perception (L2LP) model (described below; Escudero, 2005, 2006; Escudero et al., 2009; van Leussen & Escudero, 2015) investigates learners’ non-
native and L2 perceptual abilities from the initial to the end state of learning. The model proposes that naïve listeners initially discriminate and perceptually map L2 sounds in line with existing acoustic features of their L1 (Escudero, 2005, 2006, 2009). Thus, at the core of the model is the view that in addition to vowel inventory size, detailed L1-L2 acoustic comparisons are a reliable predictor of listeners’ initial L2 perceptual patterns. The model further proposes that acoustic comparisons should ideally be of quantitative nature so that calculations of acoustic similarities can be made between the native and target language.

One way to quantitatively measure acoustic similarities between languages is through linear discriminant analyses (Klecka, 1980), and a number of studies have in recent times investigated whether these analyses are predictive of listeners’ L2 perceptual patterns. (Strange, Bohn, Nishi, & Trent, 2005) investigated the phonetic similarity between the first three formants of North German and American English vowels using a cross-language discriminant analysis. American English monolinguals were presented with Northern German vowels and were asked to categorize these to native vowel categories. Acoustic comparisons between American English and North German vowels did not always predict listeners’ perceptual patterns (Strange et al., 2005; Strange, Bohn, Trent, & Nishi, 2004). However, a more recent study employed the same analyses as Strange et al. (2004, 2005) and found that perceptual patterns of American English vowels by Russian monolingual listeners were predicted by acoustic similarities between the two languages (Gilichinskaya & Strange, 2010). Similarly, Escudero and Vasiliev (2011) tested monolingual Peruvian Spanish listeners’ identification of the Canadian French and Canadian English /ɛ/ and /æ/ vowels. A linear discriminant analysis showed that acoustic similarities between L1-L2 vowels were a very good predictor of listeners’ perceptual mappings (Escudero & Vasiliev, 2011).

While all four of the above studies tested listeners whose L1 vowel inventories are smaller than that of the L2, recent research has also investigated the explanatory power of
vowel inventory size versus acoustic properties as predictors of L2 perception between naïve Australian English and Iberian Spanish listeners (Elvin, Escudero, & Vasiliev, 2014). While discriminant analyses are useful for plotting vowels of each language acoustically and classifying a probability between L1-L2 classifications they do not provide fine grained acoustic cross-language comparisons. Thus, Elvin et al. (2014) employed Euclidean Distance calculations as a quantitative measure of cross-linguistic similarity by calculating a detailed distance between the L1-L2 vowels in a two-dimensional plane. The study established that acoustic similarity rather than vowel inventory size predicted listeners’ perceptual patterns. That is, Iberian Spanish listeners with a smaller L1 vowel inventory outperformed Australian English listeners when discriminating Brazilian Portuguese vowel contrasts.

There is further evidence that listeners who share a similar L1 vowel inventory appear to perceive non-native vowels in line with their native acoustic properties. Based on vowel inventory size alone, Iberian Spanish and Salento Italian listeners, two languages that share five analogous vowel phonemes, were expected to face similar perceptual patterns of non-native Southern British English vowels (Escudero, Sisinni, & Grimaldi, 2014). However, due to the differing acoustic productions of their respective L1 vowels, listeners’ initial perceptual patterns of Southern British English vowels differed. For example, Salento Italian listeners perceived the Southern British English vowel /ɪ/ vowel as their native /i/ vowel 97% of the time, while Iberian Spanish listeners predominantly map this vowel as Spanish /u/ (42%) and /e/ (36%). These findings are in line with the L2LP model which states that acoustic differences in native vowel production will lead to different L2 vowel perception and that multiple sources of acoustic-phonetic information are employed when perceiving phonological segments (Escudero, 2005).

Thus, the way in which a listener perceives L2 vowels is an important question for L2 development. As outlined above, numerous studies have endeavoured to determine whether it
is the size and/or acoustic properties of a learner’s native vowel inventory that facilitates or impedes L2 perceptual difficulties. The present masters provides further evidence to answer this debate by comparing predictions based on vowel inventory size and on cross-language acoustic comparisons with regard to listeners’ initial perception of non-native vowels.

1.4 Influence of L1 phonological space on L2 development

1.4.1 Evidence suggesting a link between L2 sound perception and L2 learning

Just as L2 learners speak with an accent, they also perceive with an accent. Over the years, second language acquisition research has established that language learning is heavily influenced by the learner’s L1. Therefore, acquiring a second language is not an easy task as a well-established L1 system may interfere with L2 perception and learning. For example, spoken word recognition (SWR) requires the listener to break the incoming speech into recognizable words. However, if a listener is initially unable to accurately perceive the sounds heard they may misinterpret the speech signal. This would disrupt not only their ability to recognise words, but may also disrupt their ability to accurately segment words from the speech stream, setting off a snowball effect. An inability to correctly perceive words would also affect word learning, affecting listeners’ overall L2 language development. Thus, investigating how and to what extent initial L2 sound perception may affect initial L2 word recognition is an important part of psycholinguistic research and L2 development.

1.4.1.1 L1 word recognition

Spoken word recognition refers to the process of segmenting continuous speech into recognizable words. Various models of SWR, while different in their approach, share a similar view in that when a speech signal is heard, multiple sublexical representations of
phonological forms act as mediators between the speech signal and mental lexicon (e.g., Cohort model: Marslen-Wilson, 1987; TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994; Shortlist-B: Norris & McQueen, 2008). The listener’s task is to perceive lower-level acoustic-phonetic probabilities on the prelexical level through a cue-integration approach (McMurray & Jongman 2011), and while these probabilities interact to generate ‘correct’ recognition amongst multiple competing word candidates, select the speaker’s intended word while it is being integrated into its surrounding context. Thus, SWR involves activation of multiple word candidates (also termed ‘cohorts’) in a listener’s mental lexicon and their subsequent competition and selection (e.g., Weber & Scharenborg, 2012). For instance, upon hearing the word initial phoneme, e.g., /k/, words sharing the same onset will also be activated, e.g., cap, cat, cab, catch, and captain (Desroches, Newman, & Joanisse, 2009; Marslen-Wilson, 1987). Lexical competition is then further increased as similar sounding words are also activated such as cop, cape, and clap, as well as rhymes such as map, tap, zap (Desroches et al., 2009a; Luce & Pisoni, 1998). As the speech signal unfolds and the context that the word occurs in becomes clearer, the listener starts to eliminate the low probability candidates until they are left with one ‘winning’ candidate.

Further evidence in L1 word processing has established that due to the temporal nature of speech listeners pay close attention to the acoustic information of the speech signal as it may carry linguistic meaning. For example, when a listener hears a nasalised vowel /a/ (e.g., ‘ran’) cohorts that are acoustically similar are also activated (e.g., ram and rang) compared to those which are not (e.g., rat and rag; Marslen-Wilson & Warren, 1994). Thus, the higher the perceptual similarity between candidates the higher the cohort interference will be. Previous studies have established that an increase in competition does indeed occur for words that share the same phonological onset (e.g., American English: Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Dutch: Dahan, Swingley, Tanenhaus, & Magnuson, 2000; French:
Salverda, Dahan, & McQueen, 2003). For example, American English monolinguals’ eye movements when instructed to move everyday objects on a table, is faster for phonologically dissimilar objects (e.g., apple and candy) than those that share the same phonological onset (e.g., candle and candy; (Tanenhaus et al., 1995).

1.4.1.2 L2 word recognition

As in recognizing a word in one’s L1, it is assumed that non-native speech perception occurs in the same manner (Weber & Cutler, 2004). However, L2 listeners may face a number of additional obstacles in their perception of L2 lexical items compared to native listeners. One of these obstacles for listeners appears to be the activation of both the L1 and L2 lexicon even when in monolingual situations (Spivey & Marian, 1999). In particular, evidence suggests that bilingual listeners appear to activate both the L1 and L2 lexicon for words that share the same phonemic onset in both languages. For example, Russian–English bilinguals upon hearing the English word ‘marker’ also activate the Russian word ‘marku’, meaning ‘stamp’ (Spivey & Marian, 1999). Similarly, proficient Dutch learners of English upon hearing the English word ‘desk’ also activate at the Dutch word ‘deskel’, meaning ‘lid’ (Weber & Cutler, 2004).

Another obstacle for L2 listeners is that they find it more difficult to recognize words that differ in a single phoneme from a high-density neighbourhood (i.e., similar sounding words) than words from a low-density neighbourhood (i.e., non-similar sounding words) resulting in the activation of unintended words (Bradlow & Pisoni, 1999; Marian, Blumenfeld, & Boukrina, 2008). Compared to L1 listeners, L2 learners are observed to be less efficient in deactivating unintended words. For example, Russian learners of German, upon hearing the German word ‘Tisch’ (table) also activate ‘Fisch’ (fish), even though they can hear the difference between the two words (Rüschmeyer, Nojack, & Limbach, 2008). L2 recognition
difficulty is therefore increased and not only limited to similar sounding L2 words but also to words which are phonologically similar between the L1 and L2. From the above examples we see that the influence of the native language also extends to word recognition. However, compared to bilingual and early L2 learners, listeners who started learning their L2 separately from their L1, have difficulties when categorizing and producing phonemes in their later learned language (Rüschemeyer et al., 2008).

1.4.1.3 Influence of L2 speech perception on L2 word recognition

Research suggests that if a listener is unable to auditorily distinguish a non-native sound contrast they will generally have difficulties distinguishing lexical contrasts for L2 words which differ only on the basis of their auditory forms (Escudero, Hayes-Harb, & Mitterer, 2008b). As outlined in § 1.3.1 if some L2 vowel sounds are not present in learners’ L1 then they will be more difficult to perceive. Therefore, it is important to address how the native language, at the initial onset of L2 perception, influences L2 word recognition and subsequent learning. Studies of non-native and L2 word learning have established that discrimination of similar sounding L2 words is amplified for learners whose L1 contains fewer phonemes than the target L2 (e.g., Escudero, 2005; Escudero & Boersma, 2004; Flege, Bohn, & Jang, 1997; McLennan, Luce, & Charles-Luce, 2003). In particular, research suggests that discrimination difficulties of L2 vowel contrasts may influence recognition and the subsequent learning of L2 words contrasted by the same vowels, i.e., minimal pair words. For example, words that differ in a difficult phonetic L2 contrast (e.g., /i/-/ɪ/ as in the words sheep and ship for native speakers of Spanish) seem to be treated as homophones in L2 spoken-word recognition (Weber, Broersma, & Aoyagi, 2011).
1.4.1.4 Effects of perceptually difficult and perceptually easy L2 vowel contrasts on L2 word recognition

Evidence suggests a correspondence between phonological and lexical difficulties faced by learners in their discrimination of words that contain perceptually difficult non-native contrasts (Escudero et al., 2008). Dutch learners of English appear to have difficulty perceiving the difference between the novel English /æ–ɛ/ vowel contrast (Broersma, 2005; Cutler & Broersma, 2005). This contrast is particularly difficult for Dutch learners as their native vowel inventory contains the phonetic category /ɛ/ but not /æ/ (Cutler & Broersma, 2005). Using an eye-tracking paradigm, Weber and Cutler (2004) investigated lexical competition in non-native spoken word recognition. Highly proficient Dutch learners of English and native British English listeners participated in the study. On a computer screen, participants were presented with four black and white line-drawing pictures and four geometric shapes contained within a cell grid, while listening to auditory stimuli produced by a male native speaker of British English. The high confusibility target-competitor sets were chosen on the basis that they contained English vowels often confused by Dutch listeners (e.g., /æ/ as in panda and /ɛ/ as in pencil). The English word *panda* /pændə/ has the same meaning in Dutch, while *pencil* /pɛnsl/ is *potlood* /pɔtloʊd/ in Dutch. The low-confusability target-competitor sets contained English vowels that Dutch participants would not likely confuse (e.g., /ɒ/ as in *bottle* and /i/ as in *beetle*). Alongside the target-competitor sets, two phonologically unrelated line-drawing pictures of distractors, such as a dress or strawberry, were presented on the screen. The participants were instructed to click on a certain line-drawing picture and move it to a particular geometric shape (e.g., ‘Click on the panda. Now put it on top of the circle.’). Dutch participants fixated their gaze more and longer at distractor pictures containing perceptually difficult L2 vowel contrasts compared to perceptually easy L2 vowel contrasts (i.e., /æ-ɛ/ vs. /ɔ-ɪ/). This effect was not found for the
native English participants. Upon hearing the phoneme /æ/, English participants reject pencil as a likely word candidate, while Dutch participants do not do this until the second syllable of the word is heard. However, Dutch participants’ looking time to the screen was not prolonged when a word contained a phoneme present in their native language (e.g. /ɛ/ as in the word pencil). That is, when Dutch participants heard the word pencil, panda did not distract their eye fixations. No group difference was observed for target-competitor sets that contained vowel contrasts such as /u/ and /i/ as these were easy for Dutch listeners to discern.

In a follow up study, Escudero and colleagues (2008) investigated whether perceptually difficult L2 vowel contrasts account for the asymmetric perceptual patterns found by Weber and Cutler (2004) even when learning novel L2 words. In the auditory condition, highly proficient Dutch learners of English were first trained to associate 20 English non-words to line-drawing pictures. Ten of the non-words were target words containing the perceptually difficult English /æ-ɛ/ vowel contrast, such as tenzer /tɛnzə/ and tandek /tændək/. Another 10 non-words acted as distractors and contained perceptually easy English vowels, such as /ɛ-u/ and /æ-u/. A distractor was paired with an identical target word, that only differed in the first syllable vowel, which was always /u/ (e.g., /tɛnzə-tunzə/ and /tændək-tundək/). Findings show that, Dutch learners were unable to discriminate the first syllable of non-words that contained English /æ–ɛ/ (as in /pæn/ in panda and /pɛn/ in pencil; Escudero et al., 2008). That is, Dutch participants’ fixations to pictures of English non-words containing /æ/ or /ɛ/ was equal (i.e. symmetric). These findings suggest that, even if participants are proficient in their L2, difficulties in L2 vowel perception will also affect novel L2 learning for words contrasted by the same vowel sounds.

Pallier and colleagues (2001) were interested to investigate whether bilinguals store L2 words containing word pairs distinguished by perceptually difficult vowel contrasts as homophones or as distinct lexical items. Spanish-dominant bilinguals have difficulty
perceiving the Catalan /e/-/ɛ/ and /o/-/ɔ/ contrasts (e.g., ‘néta’ /nɛtə/ meaning ‘granddaughter’ and ‘neta’ /nɛtə/ the feminine form meaning ‘clean’), as they are not phonemically differentiated in Spanish (Pallier, Bosch, & Sebastian-galles, 1997; Sebastián-Gallés & Soto-Faraco, 1999). Spanish-dominant and Catalan-dominant Spanish-Catalan bilinguals were presented with auditory forms of Catalan minimal pairs and pseudowords that were comprised of either common contrasts or Catalan-specific contrasts (e.g., dóna ‘s/he gives’ and dona ‘woman’) in a repetition priming task. Compared to Catalan-dominant bilinguals, Spanish-dominant bilinguals were unable to differentiate minimal pairs that were lexically contrasted by /e/-/ɛ/ and /o/-/ɔ/, and did indeed treat them as homophones. That is, as these contrasts are not differentiated in Spanish, lexical items containing the same sounds were phonologically indistinguishable to listeners.

Further evidence shows that learners are readily able to recognize L2 minimal pairs when they are lexically contrasted by perceptually easy phonemes (Broersma, 2005; Escudero, Simon, & Mulak, 2014; Weber & Cutler, 2004). Spanish listeners find the Dutch contrasts /i-a/ and /u-a/ easy to discriminate, as these vowels are also contrasted in their native language (Escudero & Wanrooij, 2010a). Using the same non-object picture line drawings as those reported in Escudero and colleagues (2008), Spanish native speakers, with varying proficiencies in L2 English and L3 Dutch, and Dutch monolinguals took part in a Dutch word-learning task involving pseudo-words contrasted by minimal and non-minimal word pairs. The minimal word pairs were monosyllabic and contrasted by Dutch vowel sounds known to be either perceptually difficulty or easy for naïve Spanish listeners. Both participant groups were trained to associate a pseudo-word to a non-object. Their perceptual mapping and reaction time was subsequently tested. The results reveal that Spanish participants’ recognition performance was on par with that of native Dutch listeners when presented with novel minimal pair words containing perceptually easy vowel contrasts (e.g., /piχ/ and /pαχ/;
93% for Spanish learners, 95% for Dutch native listeners; (Escudero, Broersma, & Simon, 2013). Likewise, native speakers of Dutch are more accurate at distinguishing perceptually difficult Dutch minimal pairs than naïve Spanish listeners (Escudero, Simon, et al., 2014). These findings are in line with those described above that show that Dutch learners of English are able to distinguish words contrasted by the English /ɒ-i/ (e.g., ‘bottle’ and ‘beetle’), as these vowels are similar to distinct phonemes in their language (Weber & Cutler, 2004).

Together, these findings suggest that L2 word recognition uses language-specific phonological representations, and that there is a relation between native vowel inventories and non-native word learning. Failure to distinguish the speech signal within a foreign word at the initial stages of learning may indicate an unintended word, or a mispronunciation, resulting in further L2 development difficulties (e.g., L2 grammatical learning difficulties). Thus, it is vital to investigate how and to what extent initial L2 vowel perception affects listeners’ initial non-native word learning, specifically, of minimal pair words contrasted by the same sounds.

However, it is essential to first address the theoretical viewpoint on the continuity between non-native speech perception, word learning and spoken word recognition (Elvin, 2016; Pallier et al., 1997). Outlined below are summaries of the most recent theoretical models of non-native L2 speech perception with their either potential or explicit proposal on how speech perception and word recognition interact in non-native and L2 development.

1.4.2 Theoretical Models of L2 Speech Perception

1.4.2.1 The Speech Learning Model (SLM)

The Speech Learning Model (SLM: Flege, 1995) was developed to define experience and age-related limitations on listeners’ perception and consequent acquisition of L2 vowels and
consonants. Designed to address ultimate attainment of L2 production by experienced learners, as opposed to beginners, it explores changes that may occur over the lifespan. Further, the model explores the relation between production and perception of speech sounds, suggesting that learners’ L2 production is influenced by their existing L1 phonemes. It posits that learners’ ease or difficulty in learning new L2 sounds is influenced by where they occur in the acoustic space relative to existing L1 sound categories. That is, when a new L2 phoneme is very similar to an existing L1 phoneme, the learner will assimilate the L2 phoneme to the L1 category. Thus, it suggests that predictions of perception and production of L2 sounds can be made through acoustic comparisons between two languages.

Furthermore, SLM states that more native-like perception will occur if a learner is able to detect phonetic differences between L1 and L2 phonemes. According to SLM, learners are able to create new L2 phonetic categories if perceptible phonetic differences between L1 and L2 phonemes exist. That is, an adult L2 learner is more likely to create a new phonetic category for L2 vowel sounds that are more phonetically distinct from their closest L1 sounds, particularly when the listener’s native vowel inventory has fewer sounds than the target language (Flege, 1995). As a result, production of L2 sounds depends on learners’ phonetic category representations. Moreover, if sounds of the L1 and L2 are perceived to be similar, learners will find it more difficult to establish these L2 sounds as distinct phonemes. This in turn will limit their attainment of new sound categories (Flege, 1995). If the model’s predictions extend to word learning, learners should more easily learn and recognize L2 words containing more dissimilar than similar phonemes.

1.4.2.2 The Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM; Best, 1994, 1995) identifies variations in non-native discrimination about assimilation and discrimination differences for diverse types of
non-native contrasts (Best et al., 2001). It addresses naïve listeners’ perception of non-native phonological contrasts as influenced by their L1 (Best 1994, 1995). PAM stipulates that through the process of perceptual assimilation, non-native phonemes will be categorized in accordance with similar articulatory gestures that exist between non-native and native phonemes. As a means to incorporate L2 acquisition into the model, PAM was further extended into PAM-L2 (Best & Tyler, 2007). PAM-L2 predicts that perceptual learning, while likely at all ages, is strongly influenced by an individual’s entire language learning experience (Best & Tyler, 2007; Bundgaard-Nielsen et al., 2011).

PAM and PAM-L2 posit three perceptual assimilation patterns of non-native segments and their categorization in accordance with the features of native categories. The first pattern of assimilation occurs when a non-native segment is perceived to exhibit a similarity to a native category, with its goodness of fit within that category ranging from excellent to poor. The second pattern arises when a speech sound is unrecognizable and cannot be assimilated to any particular native category. The third pattern results when a sound is heard as non-speech as it bears no similarity to any native phoneme (Best, 1994; 1995). As well, PAM postulates six perceptual assimilation patterns by which non-native contrasts can be categorized relative to the native phonological space, and their predicted degree of discrimination. If each item of a non-native contrast is perceived by a listener to exhibit similarity to a separate native category, Two Category assimilation occurs. This pattern predicts good to excellent discrimination of non-native vowels (Best, 1994). When both items in a contrast are assimilated as equally good exemplars of a single native category, Single Category assimilation occurs, leading to poor discrimination. Category Goodness also occurs when both items are assimilated to a single native phoneme, but when one non-native token is judged as a better exemplar of that phoneme than the other. As a result of this pattern, the level of discrimination is predicted to vary from moderate to very good in accordance with
the disparity of goodness fit to the native phoneme between the two non-native tokens (Best, 1995). An Uncategorised-Categorised pattern occurs when one non-native phoneme is assimilated to a native category while the other falls in phonetic space outside of native categories, leading to very good discrimination (Best, 1995). However, if both sounds fall within a phonetic space but are unable to be categorized to a native category this is referred to as Uncategorised-Uncategorised. Discrimination is expected to vary from poor to very good depending on the non-native sounds’ proximity to each other and to L1 categories within the L1 phonological space (Best, 1995). If a listener were to identify non-native contrasts as non-speech sounds, a Non-Assimilable type would occur, predicting good to very good discrimination due to their degree of discrepancy from one another (Best, 1995). Previous studies have supported these predictions. For example, in their identification and discrimination of three synthetic American English approximant contrasts, Japanese learners had difficulty identifying the English /r/-/l/ contrast, assimilating it according to the SC pattern, resulting in poor discrimination (Best & Strange, 1992). In sum, PAM and PAM-L2 posit that perceptual assimilation and learning of L2 sounds occur at a gestural, phonetic and phonological level (Best & Tyler, 2007).

Relative to the native language having an effect on L2 word learning, PAM-L2 predicts that perceptual learning is affected by a listener’s entire learning experience. Thus, the way that a listener learns L2 words will not only depend on how L2 sounds are perceived, but also at what stage of their L2 development this occurs at. It is proposed that early L2 lexical development expedites the forming of L2 phonological categories (Bundgaard-Nielsen et al., 2011). Nevertheless, as the listener’s L2 development progresses and a higher L2 vocabulary is attained a curtailing of perceptual learning and phonological fossilization will likely occur (Bundgaard-Nielsen, Best, Kroos, & Tyler, 2012). That is, re-attunement and rephonologization of L2 phonemes will decrease as the listener settles on an accented version
of the target language (Bundgaard-Nielsen et al., 2011). While there is a suggestion that L2 perception and word learning interact, within their framework neither PAM nor PAM-L2 do not propose specific calculations that can measure and/or predict the interplay of these two abilities.

1.4.2.3 The Second Language Linguistic Perception Model (L2LP)

The Second Language Linguistic Perception model of cross-language speech perception (L2LP: Escudero, 2005, 2006, 2009; van Leussen & Escudero, 2015) is unique in that it looks at learners from the initial to the end state of learning and encompasses both speech perception and word learning. The model proposes that non-native sounds are categorized according to the acoustic features of naïve listeners’ L1. This leads to naïve listeners and L2 learners discriminating and perceptually mapping non-native sounds in line with existing acoustic features of their L1 (Escudero, 2005, 2006, 2009; van Leussen & Escudero, 2015). The model proposes that learners whose native language has fewer sounds than the target language are required to learn new sounds (Escudero & Chládková, 2010), while learners with more sounds than the target language are required to unlearn sounds (Escudero & Boersma, 2002). According to L2LP, L2 learners will categorize new target language sounds through a type of learning mechanism referred to as distributional learning, which is a mechanism based on tracking the statistical distributions of auditory acoustic-phonetic information (Escudero, 2005).

Auditory-driven learning performed by the L2LP’s Gradual Learning Algorithm (GLA), guided by L1 lexical representations, enables listeners to adjust their perceptual mappings to match their target L2. Previous studies have shown improvement in discrimination of difficult L2 contrasts by listeners who heard a phonetic distribution of speech sounds along an acoustic continuum that encompass the two vowel categories in a bimodal distribution.
(Escudero, Benders, & Wanrooij, 2011; Maye & Gerken, 2001; Maye et al., 2002). As a result, the model posits that comparisons of acoustic properties between the L2 and L1 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 quantitative measures of cross-linguistic similarity, and the L2LP model makes these comparisons through the *optimal perception hypothesis* - that the perceptual mappings of speech signals depend on particular characteristics of the listener’s production environment (Escudero, 2005). It is hypothesized that a listener, as an optimal perceiver with experience in sound production, identifies auditory inputs as sounds that a speaker most likely intended by making perceptual decisions based on the acoustic cues that are most reliable in their perceptual environment (Escudero, 2005, 2006, 2009; van Leussen & Escudero, 2015). Computing the optimal L1 perception allows for predictions of the listener’s initial state of L2 learning process, which is the perceptual system that learners will initially use in their L2 learning (Escudero, 2005). The model further hypothesizes that computing listeners’ optimal L1 and L2 perception (i.e., acoustic-phonetic differences between the native and target language), will determine the particular nature of L2 learning tasks required to be undertaken by listeners in order to achieve optimal target L2 perception (Escudero, 2005).

These computational analyses between L1 and L2 acoustic values can be used to predict perceptual mapping patterns and discrimination difficulties. For instance, L2LP proposes that the most difficult perceptual pattern to discern occurs when listeners perceive two target language sounds in line with a single phonemic category of the L1. This is referred to as *New Scenario* (Escudero, 2005). This learning scenario usually occurs for listeners with fewer L1 categories compared to the L2 and requires them to split their existing L1 category and/or create new categories and perceptual mappings (Escudero, 2005). It is predicted to pose
immense difficulties for the listener in both speech perception and recognition of words that are distinguished by the same sounds.

The second scenario that is predicted to pose medium difficulties for the listener is referred to as *Subset Scenario* and occurs when L2 phonological categories are perceived in line with more than one counterpart of their L1 (Escudero, 2005). In this scenario the listener usually has more native sound categories than the target language and is therefore required to adjust their L1 category boundaries. Provided that there is no boundary overlap between multiple native categories Subset Scenario is predicted to result in good discrimination, however, discrimination difficulties are predicted to occur if boundary overlap occurs (van Leussen & Escudero, 2015). The L2LP model outlines that perception and recognition difficulties at the initial stage will be easier for *Subset Scenario* compared to *New Scenario*, as the former only involves L1 category boundary shifting while the latter requires additional auditory-driven category formation (Escudero, 2005). As a result, it is posed that a listeners’ L2 perceptual and recognition developmental paths will differ based on the learning scenario they initially encounter. The L2LP model states that, in the *New Scenario* perceptual learning will precede lexical learning while in the *Subset Scenario* lexical development will occur before perceptual development (Escudero, 2005). The final scenario that the model addresses is referred to as *Similar Scenario* and occurs when two non-native sounds are categorized across two native categories (Escudero, 2005). Due to an equal number between L1/L2 categories this learning scenario is predicted to be the easiest for listeners. However, even if category numbers are equal between listeners’ native and target language, no two languages have the exact same acoustic properties. Therefore, the learning task in this scenario requires the listener to slightly adjust or shift their L1 perceptual category boundaries and perceptual cue weightings as they progress to optimal L2 perception and lexicon-driven learning (Escudero, 2005).
While L2LP stipulates that predictions about non-native vowel discrimination can be made through detailed acoustic comparison between sound categories of a native and target language, evidence shows that these predictions are not always correct (Strange, 2007). For instance, discriminant analysis did not predict perceptual assimilation patterns of Northern German vowels by American English listeners (Strange, Bohn, Trent, & Nishi, 2004). However, many studies have indeed shown acoustic comparisons to be successful predictors of non-native vowel discrimination (e.g., Escudero & Boersma, 2004; Escudero et al., 2013; Escudero & Chládková, 2010; Escudero & Vasiliev, 2011; Escudero & Williams, 2011, 2012; Gilichinskaya & Strange, 2010). Spanish monolingual listeners’ categorization of vowels in two English varieties differed as a result of the acoustic relationship between the non-native vowels and their L1 (Escudero & Chládková, 2010). While these studies have involved perception of vowel sounds, the L2LP model (Escudero, 2005, 2009) specifies that perception of L2 sounds should translate to L2 word learning. Therefore, sounds that are easy for L2 learners to perceive and acquire should also be reflected in learners’ abilities to learn L2 words containing the same, while sounds that are difficult to perceive should also be difficult recognize.

1.4.2.4 Comparison of Theoretical Models

As outlined above, both the SLM and L2LP models predict L2 perception based on acoustic-phonetic differences between the L1 and L2. However, SLM’s focus lies on advanced learners, while L2LP focuses on naïve listeners. Furthermore, the SLM focuses on perception of individual phonemes, and does not make predictions regarding perception of non-native contrasts (Best et al., 2001). For this reason, SLM is unable to make predictions regarding the questions put forth by the present thesis, which investigates the relationship between naïve listeners’ initial L2 vowel perception and initial L2 word learning. While PAM, PAM-L2, and the L2LP model make predictions about how unfamiliar L2 categories
will be categorised to the native phonological space, PAM and PAM-L2 do not propose a specific basis by which these categorizations occur and how they can be predicted. In contrast, L2LP uses specific acoustic measurements that outline why particular perceptual patterns occur. For this reason, PAM and PAM-L2 are not entirely suited for the present thesis, which in addition to establishing the relationship between initial L2 perception and learning of L2 words, also endeavours to answer whether a measurable relationship between the L1 and L2 predicts performance in the L2. The present Master's thesis examines data from listeners naïve to the target language to investigate how the native language, at the initial onset of L2 perception, influences L2 perceptual patterns. Therefore, to establish the initial state of L2 learning it will test the hypotheses postulated by the L2LP model for both speech L2 perception and word learning.

1.5 Research Aims and Questions

1.5.1 Research Aims

As outlined in § 1.3.3, an important aspect of L2 research is to establish whether listeners’ initial L2 perceptual and initial L2 word learning patterns coincide. Previous research suggests that exposure to and experience with an L2 has a direct bearing on L2 sound perception (e.g., Best & Strange, 1992) and that perceptual learning of difficult L2 contrasts is possible (e.g., Best & Tyler, 2007). For instance, the perceptual vowel space of native Spanish learners of English has been found to differ compared to that of native Spanish speakers naïve to English (Flege et al., 1994, 1995). This is because Spanish learners with greater L2 experience appear to perceive and produce L2 vowels more like English natives, compared with those with the same L1 vowel inventory who have less L2 experience. Similarly, research has shown that intermediate and advanced English learners of French discriminate French vowels presented in non-words that are not present in their native
language while English monolinguals are not (Darcy et al., 2012). Therefore, the present thesis is concerned with establishing listeners’ initial state or onset of L2 learning as specified by the L2LP model and investigating the interaction between the L1-L2 vowel inventories and their effect on non-native speech perception, and, in turn, the influence of initial non-native speech perception on initial L2 spoken word recognition. To establish listeners’ initial state of L2 learning accurately, perceptual and word recognition data from listeners who are naïve to the target language are examined.

1.5.2 Research Questions

In this thesis, we examine and compare the role of two aspects of the listeners’ native language: whether it is the size of listeners’ native vowel inventory, the acoustic properties of their L1 vowels, or both that predict initial perceptual difficulties and how and to what extent these affect listeners’ initial word learning abilities.

To investigate how these two properties of the listeners’ native language may affect their non-native perception and spoken word recognition the following questions are addressed:

1) Does the relative size of the L1 versus L2 vowel inventory affect vowel perception and word learning performance? If so, does having a larger rather than smaller L1 vowel inventory aid listeners in their initial L2 perceptual and initial L2 word learning abilities?

2) Does initial L2 vowel perception affect initial L2 word learning such that minimal pairs contrasted by perceptually difficult sounds are also difficult to learn?

3) Which is the better predictor of perception performance and word learning performance: L1-L2 acoustic comparisons or relative vowel inventory size between the L1 and L2?
These research questions are addressed across two studies that have been written in preparation for submission to journals.

1.5.2.1 Study 1: How do acoustic properties predict perception of unfamiliar Dutch vowels by adult monolingual Australian English and Peruvian Spanish listeners?

As outlined in the literature review, research relating to L2 vowel perception has suggested that it is the size of listeners’ native vowel inventories compared to the target language that may affect their L2 perceptual patterns (e.g., Iverson and Evans, 2007, 2009). Other research suggests that it is the L1-L2 acoustic relationship that plays a larger role in listeners’ L2 perceptual patterns. However, most of this research was conducted with learners who already had experience with the L2, which is not ideal when investigating listeners’ initial state of L2 learning. Accordingly, Study 1 was designed to address the first and third Research Questions by exploring how native vowel inventory size and acoustic L1-L2 relationships may affect naïve listeners’ initial perception of L2 vowels. In particular, listeners with more versus fewer native vowel categories relative to the target language were examined. First, Australian English (AusE) monolingual listeners’ categorization and discrimination of five non-native Dutch vowels was tested. Second, AusE results were then compared to those of Peruvian Spanish (PS) listeners reported in Escudero and Wanrooij (2010) and Escudero & Wanrooij (2011). Finally, the study employed detailed cross-language discriminant analyses as a means to test whether listeners’ initial L2 perceptual patterns could be predicted based on L1-L2 acoustic relationships, as predicted by the L2LP model.
1.5.2.2 Study 2: Does initial L2 vowel perception influence initial L2 word learning

With respect to the second Research Question, the literature review has highlighted the importance of investigating how initial L2 speech perception affects initial L2 word learning. Research on L2 speech perception has established that accurate perception of some vowel contrasts is more difficult than others and that these difficulties may carry over to L2 word learning. Specifically, L2 word learning studies suggest that the most difficult minimal pairs to learn are those that are differentiated by perceptually difficult contrasts, while those that are differentiated by perceptually easy contrasts are easy to attain. Using previously published word learning data (Escudero, 2015), Study 2 compares AusE and PS listeners’ initial L2 word recognition patterns to initial vowel XAB discrimination patterns reported in Study 1. Specifically, Study 2 seeks to establish whether the same difficulties encountered in non-native speech perception also transpire in non-native word learning at the initial stages of L2 learning. Further, we test the L2LP model’s tenet that L1-L2 acoustic relationships predict both non-native vowel perception and non-native word learning, thereby implying continuity between both abilities (Escudero, 2005, 2006, 2009; van Leussen & Escudero, 2015).
2 STUDY 1

ACOUSTIC PROPERTIES PREDICT PERCEPTION OF UNFAMILIAR DUTCH VOWELS BY ADULT AUSTRALIAN ENGLISH AND PERUVIAN SPANISH LISTENERS (ALISPAHIC, MULAK & ESCUDERO, 2017)²

2.1 Abstract

Research suggests that the size of the second language (L2) vowel inventory relative to the native (L1) inventory may affect the discrimination and acquisition of L2 vowels. Models of non-native and L2 vowel perception stipulate that naïve listeners’ non-native and L2 perceptual patterns may be predicted by the relationship in vowel inventory size between the L1 and the L2. Specifically, having a smaller L1 vowel inventory than the L2 impedes L2 vowel perception, while having a larger one often facilitates it. However, the Second Language Linguistic Perception (L2LP) model specifies that it is the L1-L2 acoustic relationships that predict non-native and L2 vowel perception, regardless of L1 vowel inventory. To test the effects of vowel inventory size versus acoustic properties on non-native vowel perception, we compared XAB discrimination and categorization of five Dutch vowel contrasts between monolinguals whose L1 contains more (Australian English) or fewer (Peruvian Spanish) vowels than Dutch. No effect of language background was found, suggesting that L1 inventory size alone did not account for performance. Instead, participants in both language groups were more accurate in discriminating contrasts that were predicted to

be perceptually easy based on L1-L2 acoustic relationships, and were less accurate for contrasts likewise predicted to be difficult. Further, cross-language discriminant analyses predicted listeners’ categorization patterns which in turn predicted listeners’ discrimination difficulty. Our results show that listeners with larger vowel inventories appear to activate multiple native categories as reflected in lower accuracy scores for some Dutch vowels, while listeners with a smaller vowel inventory seem to have higher accuracy scores for those same vowels. In line with the L2LP model, these findings demonstrate that L1-L2 acoustic relationships better predict non-native and L2 perceptual performance and that inventory size alone is not a good predictor for cross-language perceptual difficulties.

2.2 Introduction

In adulthood, perception of sound categories in a second language (L2) is broadly thought to occur through the lens of the native language (L1). That is, L2 sound categories are mapped to categories of the L1 (Best, 1995; Best & Tyler, 2007; Escudero, 2005, 2006, 2009, Flege, 1995, 2003). L2 perception difficulties are thus thought to arise from a lack of one-to-one mappings of categories between the L2 and the L1 – for example, when two L2 sound categories map to a single L1 category, as in Japanese listeners’ mapping of English /r/ and /l/ to the single Japanese category, /ɺ/. As difficulty in the perception of certain L2 sounds can extend to difficulties in recognizing words containing the same sounds, it is important to consider how and to what extent L1 and L2 sound inventories interact in L2 perception. The relationship between the size of the L1 and L2 vowel inventory may predict non-native and L2 vowel perception (Bundgaard-Nielsen, Best, & Tyler, 2011; Fox, Flege, & Munro, 1995; Lengeris, 2009). In this view, having fewer L1 vowels than the target L2 will result in more perceptual difficulties, as more than one L2 vowel will be categorized to some L1 categories. That is, a consequence of a smaller vowel inventory is the fact that two vowels in a non-native category will be perceived as one single sound. By extension, having more L1 vowel
categories than the L2 should facilitate L2 perception, since there are sufficient L1 categories for all L2 sounds to map to without the need for two L2 sounds to map to a single category. There is ample evidence demonstrating that L2 learners frequently struggle with sounds not present in their L1 (Escudero, 2005; Escudero & Boersma, 2002a; Flege et al., 1997; Fox et al., 1995; Morrison, 2003). For instance, Mexican Spanish listeners, having a small five-vowel inventory, categorized Canadian English /i/ and /ɪ/ vowels to their single /i/ native category (Morrison, 2002). By the same token, individuals whose L1 vowel inventory contains more sound categories than the target language have been shown to outperform listeners with fewer first-language sounds. For example, native speakers of German and Norwegian – two languages that have a larger and more complex vowel system than English – identified English vowels more accurately than French and Spanish native speakers, whose L1 vowel inventories are smaller than that of English (Iverson & Evans, 2007, 2009). However, in this case, native speakers of all four languages relied on primary acoustic cues, such as F1/F2 formant frequencies, formant movement and duration in their perception of the English vowels, despite formant movement and duration not being present in Spanish and French, suggesting that in addition L1 vowel inventory size affecting perceptual accuracy, other acoustic-phonetic properties are also at play (Iverson & Evans, 2007, 2009). Together, these findings further suggest that while the scope of a learner’s L1 vowel inventory may affect their L2 perceptual patterns, inventory size alone is not enough to accurately predict complexities of L2 perceptual patterns.

While several theories have proposed that L1-L2 relationships affect perception, they differ in some ways. The Speech Learning Model (SLM: Flege, 1995, 2003) proposes that non-native phonemes are perceived in accordance with learners’ L1 acoustic properties. However, its focus lies predominantly on advanced learners’ perception of individual phonemes, rather than naïve learners and vowel contrasts. The Perceptual Assimilation
Model (PAM: (Best, 1995), its extension to L2 learning (PAM-L2: Best & Tyler, 2007) and the Second Language Linguistic Perception model (L2LP: Escudero, 2005, 2006, 2009) focus on naïve listeners’ perception of non-native and L2 contrasts, and propose that the features of listeners’ native phonemes predict whether and to what extent contrasts will be discriminated and learned during L2 acquisition. However, PAM and PAM-L2 propose that it is the articulatory similarity/dissimilarity between L1-L2 sounds that influence and predict naïve listeners’ non-native sound perception and later L2 development. The L2LP model (Escudero, 2005, 2006, 2009) is a computational model that takes into account listeners’ learning trajectory from the initial state to ultimate attainment. It proposes that listeners will initially perceive non-native and L2 sounds in line with the acoustic features of their L1 sound system (Escudero & Chládková, 2010; Escudero, Sisinni, & Grimaldi, 2014). The model further specifies that apart from the number of vowels in a listener’s L1 relative to the L2, detailed acoustic-phonetic comparisons between the L1 and L2 determine listeners’ perceptual mapping and discrimination of non-native sounds.

L2LP posits that acoustic comparisons should ideally be quantitative measures of cross-linguistic similarity as this will allow for predictions of listeners’ initial state of the overall L2 learning process, as this is the perceptual system that learners will initially use (Escudero, 2005). One method of quantifying cross-linguistic acoustic similarity is through linear discriminant analyses (LDA) models. In order to make initial L2 perceptual difficulty predictions, LDA models allow for cross-language similarity to be established independent of listeners’ identification or discrimination performance (e.g., Gillichinskaya & Strange, 2010; Strange, Bohn, Nishi, & Trent, 2005; Strange, Bohn, Trent, & Nishi, 2004). However, some studies that have used LDA models have claimed that acoustic comparisons are not always good predictors of cross-language speech perception. For instance, in an examination of phonetic similarity between the first three formants of North German and American English
vowels (Klecka, 1980), acoustic similarities between American English and North German vowels did not always predict perceptual similarity (Strange et al., 2005, 2004). In contrast, using the same discriminant analyses as Strange and colleagues (2004, 2005), a more recent study established that acoustic similarities were a good predictor of categorization patterns of American English vowels by Russian listeners (Gilichinskaya & Strange, 2010). Likewise, recent research has indeed shown that the L1/L2 acoustic relationship affects sound perception (e.g., Elvin, Escudero, & Vasiliev, 2014; Escudero et al., 2014; Vasiliev, 2013). For example, despite the fact that Iberian Spanish listeners have a smaller vowel inventory in comparison to Australian English, they outperformed AusE listeners in their discrimination of six Brazilian Portuguese vowel contrasts (Elvin et al., 2014).

To first establish the effects of vowel inventory size versus acoustic properties in non-native vowel perception, AusE listeners’ XAB discrimination of five Dutch vowel contrasts (/a-ɑ/, /ɪ-i/, /y-y/, /i-y/, and /ɪ-ʏ/) were compared to Peruvian Spanish (PS) listeners who took part in the same XAB task as reported in Escudero and Wanrooij (2010). As listeners’ discrimination patterns should be predicted by their categorization patterns, listeners’ categorization of the same target vowels was then compared to those of PS listeners reported in Escudero and Williams (2011). As shown in

Table 1, AusE and PS vary not only in the number of phonemes present in each vowel inventory, but also in their F1, F2 and F3 acoustic properties.
Table 1. Male speakers’ acoustic measures in Hertz of languages of the present study (AusE: Elvin, Williams, & Escudero, 2016; PS: Chládková, Escudero, & Boersma, 2011; Dutch: Adank et al., 2004).

<table>
<thead>
<tr>
<th>Language</th>
<th>Vowel</th>
<th>Measure</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F1 (Hz)</td>
<td>F2 (Hz)</td>
<td>F3 (Hz)</td>
</tr>
<tr>
<td>Australian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/i:/</td>
<td>320</td>
<td>2339</td>
<td>2948</td>
</tr>
<tr>
<td></td>
<td>/u/</td>
<td>332</td>
<td>2336</td>
<td>2968</td>
</tr>
<tr>
<td></td>
<td>/ε/</td>
<td>467</td>
<td>2085</td>
<td>2799</td>
</tr>
<tr>
<td></td>
<td>/æ/</td>
<td>695</td>
<td>1763</td>
<td>2669</td>
</tr>
<tr>
<td></td>
<td>/e:/</td>
<td>757</td>
<td>1349</td>
<td>2582</td>
</tr>
<tr>
<td></td>
<td>/ɜ/</td>
<td>743</td>
<td>1386</td>
<td>2581</td>
</tr>
<tr>
<td></td>
<td>/ɔ/</td>
<td>584</td>
<td>1040</td>
<td>2540</td>
</tr>
<tr>
<td></td>
<td>/ɔː/</td>
<td>439</td>
<td>846</td>
<td>2575</td>
</tr>
<tr>
<td></td>
<td>/ʊ/</td>
<td>378</td>
<td>948</td>
<td>2490</td>
</tr>
<tr>
<td></td>
<td>/uː/</td>
<td>341</td>
<td>1796</td>
<td>2427</td>
</tr>
<tr>
<td></td>
<td>/ ø/</td>
<td>468</td>
<td>1637</td>
<td>2581</td>
</tr>
<tr>
<td></td>
<td>/aː/</td>
<td>329</td>
<td>2343</td>
<td>2980</td>
</tr>
<tr>
<td></td>
<td>/ɛː/</td>
<td>452</td>
<td>2092</td>
<td>2792</td>
</tr>
<tr>
<td></td>
<td>/æː/</td>
<td>660</td>
<td>1099</td>
<td>2557</td>
</tr>
<tr>
<td></td>
<td>/œ:/</td>
<td>745</td>
<td>1613</td>
<td>2617</td>
</tr>
<tr>
<td></td>
<td>/ʊː/</td>
<td>480</td>
<td>956</td>
<td>2530</td>
</tr>
<tr>
<td></td>
<td>/æː/</td>
<td>698</td>
<td>1844</td>
<td>2676</td>
</tr>
<tr>
<td></td>
<td>/əː/</td>
<td>636</td>
<td>1442</td>
<td>2527</td>
</tr>
<tr>
<td>Peruvian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/a/</td>
<td>612</td>
<td>1356</td>
<td>2337</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>455</td>
<td>1929</td>
<td>2532</td>
</tr>
<tr>
<td></td>
<td>/i/</td>
<td>323</td>
<td>2186</td>
<td>2789</td>
</tr>
<tr>
<td></td>
<td>/o/</td>
<td>483</td>
<td>942</td>
<td>2315</td>
</tr>
<tr>
<td></td>
<td>/u/</td>
<td>371</td>
<td>824</td>
<td>2356</td>
</tr>
<tr>
<td>Dutch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/i/</td>
<td>278</td>
<td>2162</td>
<td>2665</td>
</tr>
<tr>
<td></td>
<td>/u/</td>
<td>361</td>
<td>1919</td>
<td>2536</td>
</tr>
<tr>
<td></td>
<td>/a/</td>
<td>670</td>
<td>1425</td>
<td>2485</td>
</tr>
<tr>
<td></td>
<td>/ɑ/</td>
<td>578</td>
<td>1172</td>
<td>2435</td>
</tr>
<tr>
<td></td>
<td>/y/</td>
<td>259</td>
<td>1734</td>
<td>2205</td>
</tr>
</tbody>
</table>
If vowel inventory size is indeed a reliable predictor of non-native vowel perception, AusE listeners, whose vowel inventory is larger than that of PS, should outperform PS listeners in their discrimination of the five Dutch contrasts (/a-ɑ/, /i-ɪ/, /y-ʏ/, /i-ʏ/, and /ɪ-ʏ/). However, the L2LP model states that acoustic-phonetic similarities between the native and target language predict perceptual mapping patterns and outlines different learning scenarios that predict discrimination difficulties prior to testing. For instance, a difficult scenario of L2 learning is the New Scenario whereby listeners perceive two target language sounds in line with a single native category, (Escudero, 2005). Given that four of the Dutch vowels presented, namely /a/, /i/, /y/, and /ʏ/, are not part of the Spanish vowel inventory, it is expected that Spanish listeners will find contrasts containing these sounds (e.g., /a-ɑ/, /i-ɪ/, and /ɣ-ɣ/) relatively difficult to discriminate, and are likely to categorize these across single native categories, namely /a/, /i/ and /e/, resulting in New Scenario. An easier pattern of discrimination occurs when listeners equate two L2 sounds with two L1 categories, referred to as Similar Scenario (Escudero, 2005). AusE listeners should find at least two Dutch contrasts less difficult than PS listeners, as AusE contains two /i-ɪ/ and two /e, ɣ/ vowels. Additionally, as Dutch /ɣ/ and /ɣ/ are not present in the AusE inventory and it is further predicted that AusE will encounter Subset Scenario by equating each of these sounds to two or more native categories (Escudero & Boersma, 2002). This scenario often occurs for listeners with larger vowel inventories and difficulty is predicted to be higher than for New Scenario. That is, if perceptual overlap between the non-native and native categories occurs, then listeners are predicted to perceive a non-native contrast as the same multiple native categories. However, if no perceptual overlap occurs then the learning scenario should be

---

3 PAM and PAM-L2 describe these patterns as Single Category assimilation, Two Category assimilation, and Uncategorized assimilation, respectively (Best, 1995; Best & Tyler, 2007).
easier than New Scenario to discern, but should not be easier to discern compared to Similar Scenario.

If listeners’ L1 vowel inventory size affects non-native discrimination difficulty, AusE listeners are predicted to outperform PS listeners overall in their discrimination of the Dutch vowel contrasts. This is in contrast to acoustic comparisons, where comparable perceptual difficulties across both listener groups would be expected. All of the aforementioned studies that used LDAs as a means of testing the predictive nature of listeners’ L2 perception only tested listeners whose L1 vowel inventory is smaller than that of the L2. Thus, we further used LDA models to test whether acoustic similarities are predictive of categorization patterns by listeners with a smaller and larger vowel inventory compared to the target language (Escudero & Vasiliev, 2011; Gilichinskaya & Strange, 2010; Strange et al., 2005, 2004). As described in the method section, these analyses model AusE and PS listeners’ likely classification patterns of Dutch vowels, and in turn predict their likely discrimination difficulties (Escudero & Vasiliev, 2011; Gilichinskaya & Strange, 2010; Strange et al., 2005, 2004). Table 2 presents the AusE and PS cross-language classification data percentages of the most frequent Dutch vowel classification to an AusE word and PS vowel.

**Table 2. Percentage of Dutch vowel token classification to an AusE word and PS vowel based on overall classification patterns of cross-language LDA.**

<table>
<thead>
<tr>
<th>Dutch Vowel</th>
<th>AusE classifications</th>
<th>PS classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>dress: 5 fleece: 35  foot: 45 goose: 5 kit: 5 lot: 10 nurse:</td>
<td>a: 90 e: 10 i: 5 o: 45 u: 5</td>
</tr>
<tr>
<td>/æ/</td>
<td>5 15 60 15 5 10</td>
<td>5 95</td>
</tr>
<tr>
<td>/i/</td>
<td>10 90</td>
<td>100</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>25 10 65</td>
<td>70 30</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>5 75 20</td>
<td>100</td>
</tr>
<tr>
<td>/y/</td>
<td>75 25</td>
<td></td>
</tr>
</tbody>
</table>
Furthermore, the L2LP model posits that when distinguishing between L2 categories, listeners employ multiple sources of acoustic-phonetic information in their perception of phonological segments (Escudero, 2005). Previous research has indeed demonstrated that close attention is paid to the most salient acoustic cue of a particular sound (see Curtin, Fennell, & Escudero, 2009; Escudero, Sisinni, & Grimaldi, 2014; Mayr & Escudero, 2010). For instance, Salento Italian listeners’ perceptual patterns of standard Southern British English vowels were tested to establish their initial state in the acquisition of the Southern British English vowel system (Escudero, et al., 2014). The results suggest that Southern British English vowels were initially mapped relative to the acoustic properties of the listeners’ native vowel system. For example, the first two formants of Southern British English /ɪ/ and /ɔː/ fall between Salento Italian /i/-e/ and /o-u/, respectively. However, Salento Italian listeners perceived these sounds as corresponding to their native /i/ and /o/ categories, displaying the use of single acoustic dimensions in their categorization. That is, F2 was the defining acoustic measure for English /ɪ/, and F3 for /ɔː/ (Escudero, et al., 2014). Thus, to test classification power of each individual acoustic measure in our study, we conducted additional stepwise discriminant analyses in each language based on F1, F2, F3 as well as duration. Table 3 presents the AusE and PS cross-language classification data based on individual acoustic dimensions.
Table 3. Percentage of Dutch vowel token classification to an AusE word and PS vowel based on classification patterns of individual dimension cross-language LDAs.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Dutch Vowel</th>
<th>AusE classifications</th>
<th>PS classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>dress</td>
<td>fleece</td>
</tr>
<tr>
<td>F1 Bark and duration</td>
<td>/æ/</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>/a/</td>
<td>35</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>/ʏ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1, F2 Bark and duration</td>
<td>/æ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/a/</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ʏ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1, F2, F3 Bark and duration</td>
<td>/æ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/a/</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/ɪ/</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>/ʏ/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the cross-validation classification sets in Table 2 and Table 3 our predicted perceptual patterns for each Dutch contrast by AusE and PS listeners are as follows:

- **Dutch /ɪ-ɪ/**

It is expected that both listener groups should face comparable difficulties in their discrimination of the Dutch /ɪ-ɪ/ contrast. Listeners are predicted to predominantly perceive these two non-native vowels in line with a single native category, namely AusE /ɪ/ and PS /i/. Additionally, based on the stepwise classifications AusE /i/ exhibits F2 and F3 acoustic similarity to Dutch /i/ more than AusE /ɪ/. Therefore, AusE listeners are further predicted to
exploit these differences and are expected to categorize Dutch /i/ as AusE /i/ some of the time, in addition to AusE /i/. PS listeners are predicted to also exploit F2 and F3 by categorizing Dutch /i/ as PS /e/ and /u/.

- Dutch /ʏ-/y/

Based on the overall LDA classifications, AusE listeners are expected to encounter New Scenario in their discrimination of the Dutch /ʏ-/y/, as they are likely to perceive both vowels as AusE /ʉ:/ . However, based on the stepwise DAs, these vowels are acoustically similar to AusE /ʉ:/, /u/ and /o/ (F1), /ɛ/, /u:/ and /i/ (F2) and /ɛ/, /u/ and /u:/ (F3). Thus, in line with these parameters, AusE listeners are predicted to encounter the Subset Scenario by categorizing Dutch /ʏ-/y/ across multiple native categories, namely AusE /ɛ/, /i/, /o/ and /u:/.

PS listeners are predicted to not face difficulties in their discrimination of Dutch /ʏ-/y/ and are expected to encounter Similar Scenario by perceiving both sounds in line with a distinct native phoneme namely PS /e/ and /i/.

- Dutch /i-/y/

In line with overall perceptual similarity, PS listeners are predicted to categorize Dutch /i-/y/ as PS /i/. However, based on the stepwise classifications, Dutch /y/ also exhibits some F2 and F3 similarity to PS /u/. While minimal, these differences between backness and rounding are predicted to aid PS listeners when discerning these two sounds. Similarly, AusE listeners are predicted to exploit all three acoustic parameters in their perception of Dutch /i-/y/. That is, based on height (F1), Dutch /i/ is acoustically most similar to AusE /u/ and /u:/, while Dutch /y/ is acoustically most similar to AusE /u:/.

However, based on the first two formants Dutch /i/ is closest to AusE /i/, while Dutch /y/ is most similar to AusE /u:/.

Therefore, AusE listeners are predicted to face Similar Scenario by categorizing Dutch /i-/y/ as AusE /i/ and /u:/.

This learning scenario is also predicted by the cross-linguistic DA classification patterns.
Based on overall LDA classifications, both listener groups are expected to encounter Similar Scenario when differentiating Dutch /ʏ-ɪ/, as these vowels are acoustically similar to distinct native phonemes. However, based on the stepwise DAs both Dutch /ʏ/ and /ɪ/ bear F1 similarity to AusE /ɪ/, F2 similarity to AusE /ʉ:/ and /ɪ/, and F3 similarity to AusE /ɪ/, /ʉ:/ and /ʊ/. However, based on classification percentages (e.g., F2: Dutch /ʏ/ → AusE /ʉ:/, 80% and /ɪ/ 15%, Dutch /ɪ/ → AusE /ɪ/, 95% and /ʉ:/, 5%), it is predicted that AusE listeners, facing Subset Scenario, should differentiate these two phonemes due to the low acoustic overlap between the F2 and F3 cues. In contrast, PS listeners are predicted to exhibit lower discrimination accuracy compared to AusE listeners due to a higher acoustic overlap across all three acoustic dimensions and are further expected to classify both vowels as PS /i/, /e/ and /u/.

Dutch /a-ɑ/

In line with the overall DA and stepwise models, AusE listeners are predicted to find the Dutch /a-ɑ/ somewhat challenging to discern as these vowels were classified across two or more AusE vowel categories, while PS listeners are predicted to encounter New Scenario by predominantly mapping the sounds in line with PS /a/.

In sum, if predictions based on listeners’ L1 vowel inventories size are borne out, AusE listeners, whose vowel inventory is larger vowel than that of PS, are expected to have higher discrimination accuracy than PS listeners for all five Dutch contrasts. Cases of New Scenario are predicted for PS and that of Similar Scenario for AusE listeners. Alternatively and following L2LP’s acoustic hypothesis, if acoustic differences between L1 and L2 influence non-native sound perception, both listener groups’ discrimination difficulties should yield comparable results. That is, both listener groups are expected to face the New, Similar and
Subset Scenarios. To test these contrastive hypotheses, naïve AusE listeners’ XAB discrimination and categorization of five Dutch vowel contrasts (/a-ɑ/, /i-ɪ/, /y-ʏ/, /i-ʏ/, and /i-ε/) was compared to those of naïve PS listeners reported in previous studies (Escudero & Wanrooij, 2010; Escudero & Williams, 2011).

2.3 Method

2.3.1 Participants

Twenty-two monolingual AusE students aged 18-45 years ($M_{age} = 24.1$ years; 11 females) participated for course credit at Western Sydney University. Participants were born and raised in Greater Western Sydney, and reported no experience with Dutch or any hearing impairment.

Non-native vowel categorization data from the same AusE listeners were compared to non-native vowel categorization data from 40 PS monolinguals (20 females) from Lima, Peru reported in Escudero and Williams (2011). Participants ranged in age from 18-30 years\(^4\), and reported no knowledge of Dutch or hearing impairment. XAB discrimination data from our AusE listeners was then compared to discrimination data of 22 PS listeners reported in Escudero and Wanrooij (2010). Listeners were monolinguals aged 17-28 years ($M_{age} = 20.95$; 10 females) born and raised in Lima, Peru their entire life and reported no knowledge of Dutch.

Participant data collection for the present study was carried out in accordance with the Human Research Ethics Committee (HREC), Western Sydney University, approval number H9373.

\(^4\) The mean age for these participants was not reported by the previous authors.
2.3.2 Stimuli and Procedure

Both groups of participants first completed a two-alternative forced choice XAB discrimination task followed by a non-native categorization task. The auditory stimuli for the XAB discrimination task were 20 naturally produced tokens of each of the 5 Dutch vowels /a/, /ɑ/, /ι/, /i/, and /ʏ/, extracted from recordings produced by 20 native Standard Northern Dutch speakers (10 females) in monosyllabic utterances in a neutral non-word /sVs/ consonantal context embedded within a carrier sentence (Adank, van Hout, & Smits, 2004). In the XAB task, listeners heard three sounds in a row and were then asked to indicate whether the first sound (X) sounded more like the second (A) or third (B) sound by clicking on one of two yellow squares (viz. “2” and “3”) presented on a computer screen. There was an inter-stimulus interval of 1.2 s, which was selected because it is long enough to trigger phonological activation (Escudero & Wanrooij, 2010a; Werker & Logan, 1985), and an inter-trial interval of 0.5 s following the participant’s selection. The experiment was conducted in Praat and consisted of five blocks (one for each contrast – /a-ɑ/, /i-ɪ/, /y-ʏ/ and /i-ʏ/) containing 80 trials each.

Stimuli for the non-native vowel categorization task were 20 naturally produced tokens of each of the 12 Dutch monophthongal vowels, /ɑ, a, e, ɛ, i, ɪ, o, ɔ, u, ʊ, ʏ, y/, extracted from the same speakers and context as in the XAB task. As the present study compares non-native discrimination and categorization, we report categorization results only for the same vowels presented in the XAB task. The task consisted of 240 randomized test trials and participants completed 12 practice trials prior to beginning. In each trial, PS listeners were asked to categorize one Dutch vowel token to one of the nine PS (/a, e, i, o, u, ei, eu, ue, ou/) and 12 AusE (/i:, ɪ, e, e:, æ, ɐ:, ɐ, ɔ, o:, ʊ, ʉ:, ɜ:/) vowels presented orthographically on the screen. According to the Orthographic Depth Hypothesis (ODH; Katz & Frost, 1992), Spanish has a very straightforward correspondence between phonemes and their graphemic representations.
(Escudero & Wanrooij, 2010). That is, each grapheme tends to represent one phoneme only. However, English is not orthographically transparent and vowels can’t reliably be presented using orthography unless they are embedded in words. Therefore, AusE listeners were asked to categorize the vowel to native words that each contained an AusE vowel (heed, hid, hood, who’d, hair, head, heard, hall, had, hut, hot, hard). There was a between-trial interval of 1 s and listeners could take a short break after every 24 trials.

Stimuli for both tasks were presented through headphones at a comfortable hearing level. Testing of AusE participants took place in a quiet room at the MARCS Institute, Western Sydney University. PS participants were tested in a quiet room at the Pontificia Universidad Católica del Perú, in Lima (Escudero & Wanrooij, 2010; Escudero & Williams, 2011). Before starting each task, listeners completed a practice session to familiarize themselves with the testing procedure. Each listener took approximately one hour to complete both tasks.

We implemented a vowel-intrinsic normalization procedure where the first three formant values for each language were converted from the Hertz to the Bark scale using Traunmüller’s (1990) critical band rate equation (1; see Syrdal & Gopal, 1986). This procedure is typically used for modeling human vowel perception, compared to a vowel-extrinsic procedure, which is traditionally used for automatic speech recognition purposes (Adank, Smits, & van Hout, 2004; Gerstman, 1968; Lobanov, 1971; McMurray & Jongman, 2011; Neary, 1978).

\[
F_i^B = 26.81 \times \left( \frac{F_i}{1960 + F_i} \right) - 0.53.
\]

(1)

Two separate LDA models were first trained using the cross-validation method reported in Strange et al. (2004, 2005). Each LDA included F1/F2/F3 bark values and vocalic duration as an additional parameter as well as the six (/a/, /a/, /i/, /i/, /y/, /y/) target L2 Dutch vowels.
(AusE: Elvin, Williams, & Escudero, under review; PS: Chládková, Escudero, & Boersma, 2011; Dutch: Adank et al., 2004).

2.3.3 Statistical Analysis

A mixed-effects logistic model examining listeners’ correct and incorrect responses was used to establish any effects of vowel inventory size and acoustic properties on L2 perception of all non-native Dutch vowel contrasts presented in the XAB task. In particular, we analyzed participants’ correct responses, with participant, speaker and XAB trial as random effects, and vowel contrast and language background as fixed effects. As a means of establishing discrimination ranking of contrasts both within and between participant groups, we then conducted further post-hoc pairwise comparisons. The statistical model was chosen as it is appropriate for evaluating data of categorical nature (see: Arnon, 2010; Baayen, Davidson, & Bates, 2008; Jaeger, 2008).

2.4 Results

2.4.1 Cross-language Discriminant Analyses

LDA models yielded 84.2% overall correct classification for AusE and 96.7% for PS. Percentages of the most frequent Dutch vowel classification to an AusE word and PS vowel are presented in Table 4. To inform the contribution of duration, we additionally ran two LDA models that did not include duration as a factor. While the classification parameters remained the same the models yielded slightly lower correct classification percentages when duration was removed; 72.2% overall correct classification for AusE and 94.05% for PS.

Two additional (one per language group) stepwise classification models were then trained and tested. Each step in the model contained the same acoustic parameters, vocalic duration as well as the same six target L2 Dutch vowels as the LDA models. The AusE stepwise DA
yielded 33.9% for F1 and duration, 71.3% for F1, F2 and duration, as well as 73.3% for F1, F2, F3 and duration correct classification. Whereas the PS model yielded 55.1 for F1, 87.7% for F1, F2 and duration, in addition to 90.1% for F1, F2 and F3 correct classification.

2.4.2 Non-native vowel categorization

Table 4 presents the percentage of times (>5%) a Dutch vowel token was classified to an AusE and PS vowel. Instances of the New Scenario were observed for both groups, whereby two non-native Dutch sounds were mapped to a single native category: PS participants categorized both Dutch /i/ (94%) and /y/ (59%) to the single PS /i/ and both Dutch /ɤ/ (53%) and /ʊ/ (49%) sounds to PS /ɛ/. AusE listeners mainly classified Dutch /u/ (40%) and /i/ (48%) to AusE /i/, while PS participants classified Dutch /i/ as their native PS /i/ (94%). PS participants classified Dutch /i/ across two native categories, namely /i/ (39%) and /e/ (49%). AusE listeners mostly mapped Dutch /a/ to an acoustically similar AusE counterpart, /ɐː/ (47%), while Dutch /a/ was mapped most frequently to AusE /ɔ/ (40%). PS listeners categorized Dutch /ɑ/ as PS /a/ (59%) and /o/ (33%), while Dutch /a/ was mapped to PS /a/ (96%). Furthermore, instances of the Subset Scenario, which involves non-native vowels being categorized as more than two native categories, was observed for AusE listeners e.g., /ɤ/ → /ɛ/ (19%), /ɔ/ (19%), /uː/ (14%) and /y/ → /ʊː/ (28%), /ʊ/ (17%), /i/ (17%). PS listeners categorized these sounds mainly across two acoustically distinct native categories, PS /e/ (53%) and /i/ (59%), encountering Similar Scenario.
Table 4. Categorization percentages of non-native Dutch vowels to AusE words by AusE listeners tested in the present study and to PS vowels by PS listeners as reported in Escudero and Williams (2011).

<table>
<thead>
<tr>
<th>Dutch Stims</th>
<th>AusE Responses</th>
<th>PS Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/ɪ/ /i/ /a/ /ɑ/ /u/</td>
<td>/ɪ/ /e/ /ei/ /eu/ /a/ /o/ /ou/ /u/ /ue/</td>
</tr>
<tr>
<td>heed hid head heard hair had hard hall hat hood who'd</td>
<td>6 10 15 47 6</td>
<td>96</td>
</tr>
<tr>
<td>/a/</td>
<td>9 13 7 13 40</td>
<td>39 49 7</td>
</tr>
<tr>
<td>/ɪ/</td>
<td>8 40 20</td>
<td>6 6</td>
</tr>
<tr>
<td>/u/</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>/e/</td>
<td>10 53 25</td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>19 14</td>
<td></td>
</tr>
<tr>
<td>/ei/</td>
<td>13 19 10 6</td>
<td></td>
</tr>
<tr>
<td>/eu/</td>
<td>17</td>
<td>27 28 59 32</td>
</tr>
</tbody>
</table>

2.4.3 XAB discrimination task

To determine whether discrimination differed between participants whose native language had more (AusE) or fewer (PS) vowels compared to Dutch, we compared performance between AusE and PS listeners. A mixed-effects binary logistic model analyzing participants’ correct responses, with participant, speaker and XAB trial as random effects, and vowel contrast and language background as fixed effects revealed a main effect of contrast ($\chi^2 (4, N = 17600) = 38.7, p < .001$). While there was no main effect of language background ($\chi^2 (1, N = 17600) = .112, p = .738$), there was an interaction of vowel contrast and language background ($\chi^2 (4, N = 17600) = 16.5, p = .002$). Fishers’ LSD-corrected post-hoc pairwise comparisons revealed that PS listeners had marginally more correct responses than AusE listeners for /i-i/ (p = .053, 95% CI [-0.52, -0.003]), whereas AusE participants were marginally more correct for /u-i/ than PS listeners (p = .086 [-0.44, 0.03]). Figure 1 presents listeners’ discrimination accuracy of the five non-native Dutch vowel contrasts.
Figure 1. Accuracy in discrimination of all non-native Dutch vowel contrasts by 22 native AusE and 22 native PS participants. Standard error bars were treated as Independent Variables.

Fisher’s LSD-corrected post-hoc pairwise comparisons revealed that AusE participants had more correct responses for /i-y/ than /i-i/ (p < .001, 95% CI [-0.09, -0.18]), /a-ɑ/ (p = .002, [-0.14, 0.03]), /γ-y/ (p < .001, [-0.16, -0.05]), and /γ-ɪ/ (p < .001, [-0.12, -0.06]); for /γ-ɪ/ than /i-i/ (p = .043, [-0.09, 0.00]); and for /a-ɑ/ than /i-i/ (p = .046, [-0.10, 0.00]). PS participants had more correct responses for /γ-y/ than /a-ɑ/ (p = .002, 95% CI [-0.16, -0.04]) and /γ-i/ (p = .002, [-0.15, -0.03]); for /i-y/ than /a-ɑ/ (p = .003, [-0.16, -0.03]) and /γ-ɪ/ (p < .001, [-0.13, -0.04]); and for /i-i/ than /γ-ɪ/ (p = .043, [-0.11, 0.00]), and trended towards more correct responses for /i-i/ than /a-ɑ/ (p = .081, [-0.11, 0.00]). Table 5 presents listeners’ discrimination ranking from most to least difficult contrast, along with mean accuracy percentages.
Table 5. Difficulty ranking, mean accuracy percentages and standard error (SE) for XAB Discrimination task ranging from most (1) to least difficult Dutch vowel contrast by 22 native AusE and 22 native PS listeners.

<table>
<thead>
<tr>
<th>L1 listener group:</th>
<th>AusE</th>
<th>PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty ranking</td>
<td>Dutch vowel contrast (% accuracy)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>/i-/  (64.1%) SE 1.4</td>
<td>/a-a/  (63.1%) SE 2.1</td>
</tr>
<tr>
<td>2</td>
<td>/a-a/  (64%) SE 2.1</td>
<td>/i-/  (64%) SE 1.5</td>
</tr>
<tr>
<td>3</td>
<td>/i-/  (66.8%) SE 2.3</td>
<td>/i-/  (69.5%) SE 2.2</td>
</tr>
<tr>
<td>4</td>
<td>/i-/  (68.4%) SE 1.5</td>
<td>/i-/  (72.2%) SE 1.7</td>
</tr>
<tr>
<td></td>
<td>/a-a/  (68.8%) SE 2.1</td>
<td>/y-y/  (72.6%) SE 2.0</td>
</tr>
<tr>
<td></td>
<td>/i-/  (76.9%) SE 2.2</td>
<td></td>
</tr>
</tbody>
</table>

2.5 Discussion

We examined whether the size and/or acoustic properties of native vowel inventories relative to the target language aid or impede L2 perceptual difficulties by directly comparing L2 vowel discrimination and categorization patterns by two listener groups with varying vowel inventory sizes. The study also tested whether cross-linguistic LDA and stepwise models were predictive of AusE and PS listeners’ vowel classification patterns, which in turn should predict their discrimination patterns. Based on a larger and more complex native vowel inventory, AusE listeners were predicted to perform better overall than PS listeners due to their vowel inventory being larger than PS. While there was no effect of language background, an interaction of language background and contrast was observed, with results suggesting that vowel inventory size does not fully explain non-native vowel discrimination. In fact, PS listeners were marginally better than AusE listeners in their discrimination of Dutch /i-/i/, while there was a trend for AusE listeners having a higher accuracy in only one contrast, namely /y-/i/.
Our findings are in line with our acoustic predictions that further support L2LP’s tenet that L1-L2 acoustic proximities predict listeners’ initial perception and discrimination patterns. That is, both listener groups appear to employ perceptual cues from their L1 when perceiving non-native sounds. Specifically, AusE and PS listeners’ perceptual patterns were influenced by L1-L2 acoustic differences and both listener groups faced comparable difficulty in their perception of non-native Dutch vowels. As predicted by our cross-language LDA models, AusE listeners found Dutch /ɪ-Y/ their second most challenging contrast. AusE listeners mapped this contrast across multiple native categories, Dutch /ɪ-/ as AusE /ɛ-u-ʊ:/ and Dutch /y/ as AusE /u:-ɔ-ʊ/, leading to an overall lower discrimination performance compared to PS listeners who mapped these vowels across two acoustically distinct native categories, PS /ɛ/ and /ʊ/ respectively. Further evidence for acoustic L1-L2 overlap across multiple native categories affecting listeners’ perceptual patterns, irrelevant of their L1 vowel inventory size, can be observed for Dutch /ɪ-/i/. Even though AusE has two and PS one /ʊ/ vowel, both listener groups had difficulty when discriminating this contrast. As predicted by the stepwise DAs, PS listeners employed F2 and F3 to classify Dutch /ɪ/ across two native categories, namely /ʊ/ and /ɛ/ while Dutch /ʊ/ was classified solely as PS /ʊ/. On the other hand, AusE listeners mapped Dutch /ɪ-/i/ predominantly as AusE /ɪ/ but also /ɛ/, /ʊ/, /u:/ and /ʊ/. These results are in line with those of earlier studies that show English listeners’ initial perceptual patterns are primarily influenced by spectral cues when perceiving the Dutch tense-lax /ɪ-/i/ contrast, providing further evidence that acoustic properties influence listeners’ perceptual patterns of non-native sounds (e.g., Lengeris, 2008). Research has established that in their perception of high-front vowels listeners are more sensitive to vowel-intrinsic formant movement than duration (e.g., Bennett, 1968; Stevens & House, 1963; Tiffany, 1953). Specifically, English listeners are almost entirely unaffected by changes in duration
for vowel contrasts such as /i-ɪ/, /e-ɛ/, and /u-ʊ/, even though a large and noticeable difference exists in the production of these vowels (Hillenbrand & Nearey, 1999). Thus, perceptual evidence suggests that even though “vowel duration varies substantially across individual vowel categories the degree to which a given vowel can be distinguished from its neighbors is based on spectral characteristics” (Hillenbrand, 2013, p.25). AusE listeners’ low discrimination performance and categorization of Dutch /i-i/ across multiple native categories was therefore due to a higher acoustic overlap between the AusE and Dutch categories, compared to PS. As a result, PS listeners outperformed AusE listeners in their discrimination of Dutch /i-i/ who found this their most challenging contrast to discern.

AusE listeners’ perceptual patterns are also reflective of acoustic overlap between the number of referents available to AusE listeners. In the present study, AusE listeners were given more response categories compared to PS listeners as AusE has a larger vowel inventory. Earlier studies have shown that vowel categorization is affected by the number of mental representations available to a listener (e.g., Benders, Escudero, & Sjerps, 2012; Elvin et al., 2014). For instance, PS listeners were less accurate in their categorization of Spanish /i-e/ when given more response categories, /a-e-i-o-u/, compared to fewer response categories, /i/ and /e/ (Benders et al., 2012). Listeners who were given two response categories were found to be more sensitive to F1 changes allowing for an early boundary shift, while those with five options were found to constrain their sensitivity to acoustic context effects resulting in a slower boundary shift. Specifically, the authors argue that listeners who were given the response option /a/ activated more mental representations and were implicitly expecting to hear /a/, thus delaying the boundary shift between /i/ and /e/. Further evidence of the number of mental representations and acoustic overlap affecting perceptual performance is suggested by Elvin et al. (2014). The authors suggest that AusE listeners’ overall lower accuracy of BP vowels may be due to that a larger number of mental representations are activated for AusE
than IS listeners. Moreover, AusE listener’s discrimination accuracy was lower for the non-native vowel contrasts that bear complete or partial acoustic overlap to native categories. As a result, AusE listeners mapped non-native sounds across two or more native categories (e.g., e-i and /o-u/). However, listeners’ accuracy was not affected for contrasts that were mapped across multiple native categories but involved no acoustic neutralization (e.g., /a-e/).

Our results further support findings by Elvin et al. (2014) as AusE listeners showed higher discrimination accuracy for contrasts involving Subset Scenario, in which listeners equate a non-native sound across two or more native categories, with minimal to no acoustic overlap. As predicted, AusE listeners had low discrimination difficulty for Dutch /i-y/ and found this their least challenging contrast. Based on the LDA model predications listeners were predicted to classify this contrast across two native categories, namely /ɪ-ʉː/. Non-native vowel categorization results show that AusE listeners appear to utilize all acoustically close native vowel categories in their perceptual differentiation of Dutch /i-y/. That is, listeners categorized Dutch /i/ as AusE /ɪ/, /i/ and /ɛ/, while /y/ was categorized as AusE /uː/, /ʊ/ and /u/. As presented in Table 3, while there was some acoustic overlap to AusE /uː/ based on F1 listeners appear to exploit backness (F2) and rounding (F3) differences to distinguish this contrast. Similarly, PS listeners’ high discrimination accuracy shows that listeners’ also exploit F2 and F3 differences in their perception of Dutch /i-y/. That is, PS listeners categorized Dutch /i/ to PS /i/ and Dutch /y/ as PS /i/ and /u/. These results indicate that while there is acoustic overlap between Dutch /i/ and PS /i/ across all three acoustic dimensions, F2 and F3 appear to be the defining cues for PS listeners’ perception of Dutch /y/.

It is well known that PS listeners face New Scenario in their perception of Dutch /a-ɑ/ and equate this contrast as PS /a/, resulting in low discrimination accuracy (e.g., Escudero & Williams, 2012). While PS listeners did face New Scenario for this contrast, AusE listeners encountered Subset Scenario with low perceptual overlap, leading to an overall higher
accuracy percentage compared to PS listeners. Our findings are in line with L2LP which stipulates that Subset Scenario should be difficult for non-native listeners, but less difficult than the New Scenario. A similar pattern can be observed for Dutch /ɣ-ɪ/. AusE listeners faced Subset Scenario by perceptually mapping Dutch /ɣ-ɪ/ across multiple native categories. However, as predicted by the stepwise classifications, AusE listeners made use of F2 and F3 differences between Dutch /ɣ-ɪ/ to discern the contrast. Conversely, PS listeners were predicted to have lower discrimination accuracy for this contrast as both Dutch vowels exhibit acoustic overlap across all their acoustic dimensions. Our predictions were borne out, as listeners found this one of their most challenging contrasts to discern facing Subset Scenario by predominantly mapping Dutch /ɣ-ɪ/ as PS /i, e, u/ leading to a lower mean accuracy compared to AusE listeners. Furthermore, AusE listeners mapped Dutch /ɣ-ʏ/ across multiple native categories, namely /ɛ/, /ɪ/, /ʊ/ and /新陈/. Due to an acoustic overlap across all three acoustic cues, AusE listeners found this their second most challenging contrast. PS listeners categorized Dutch /ɣ-ʏ/ predominantly to native /e/ and /i/, encountering Similar Scenario.

In line with L2LP, our perceptual results suggest that for both listener groups non-native phonemes are easier to discern when they are in acoustic proximity to distinct native categories and are categorized across acoustically similar native counterparts. In addition, listeners with larger vowel inventories seem to activate multiple native categories reflected in the perceptual patterns of some L2 vowels. This demonstrates that for the most part, having a larger and more complex first-language vowel inventory is not a good predictor for L2 perceptual difficulties as reported in previous literature (e.g., Iverson & Evans, 2007, 2009). Furthermore, activation of multiple native categories for non-native contrasts involving acoustic or perceptual overlap results in lower discrimination accuracy, such as categorization of Dutch /a-ɑ/ and /ɣ-ɪ/ for PS listeners, while medium to good discrimination is seen for
contrasts that are not completely neutralised across native categories, such as categorization of Dutch /i-yl/, /a-α/ and /ʏ-ɪ/ for AusE listeners. Findings further suggest that both listener groups transfer perceptual cues from their native language when discriminating non-native contrasts. Moreover, rather than overall LDA classifications, our findings suggest that individual cues offer a more detailed insight into naïve listeners’ perceptual patterns. That is, as established in earlier studies, listeners with varying L1 vowel inventories appear to access the complex interaction of spectral and temporal information in their perception of L2 sounds (Lengeris, 2008). Thus, our findings are in line with L2LP model’s acoustic hypothesis that stipulates that L1-L2 acoustic relationships are predictive of listeners’ initial perceptual patterns, as well as previous research that demonstrates acoustic proximities rather than vowel inventory size offering more detailed non-native/L2 perceptual pattern predictions (e.g., Elvin et al., 2014; Escudero et al., 2014; Escudero & Williams, 2011).

Nevertheless, further analyses should be undertaken to generate more accurate perceptual predictions based on quantitative measures of cross-linguistic similarity between the target language and listeners’ L1, such as Euclidean Distances. While the present study uses F1, F2 and F3 measurements reported in earlier studies, vowel-extrinsic speaker normalization procedures (e.g., Lobanov, 1971) require fundamental frequency (F0) in addition to the first three formant measurement as a means of computing average formant values across speakers. Since F0 values were not available in the AusE corpora used in the present study, we were unable to compute formant means across genders as F0 values, which are part of a detailed ED comparison between L1/L2 languages. This will allow for a more detailed analysis of the present data and comparison of acoustic overlap for contrasts that exhibit Subset to that of earlier research (e.g. Elvin et al., 2014). In line with the present findings, and as suggested by one of our reviewers, an interesting avenue for future research may be to also include
perception tasks that simply require listeners to write down the perceived non-native sound instead of categorizing it to a native category option.

In sum, our findings demonstrate that regardless of AusE and PS listeners’ varying native vowel inventories it is the L1-L2 acoustic relationships that predict their non-native vowel perception. The findings also show that cross-linguistic LDA and stepwise models were predictive of AusE and PS listeners’ vowel classification patterns, which in turn predicted their discrimination patterns. Ongoing research will further examine whether our results extend to L2 word recognition abilities in words that differ in the same Dutch vowel contrasts. Findings may inform possible future language learning programs which could include customizing individual L2 learning according to native language.
3 STUDY 2

ADULT LISTENERS’ NON-NATIVE PERCEPTUAL PATTERNS PREDICT DIFFICULTIES IN NOVEL WORD RECOGNITION

3.1 Abstract

This study investigated how and to what extent non-native perceptual patterns of five Dutch vowel contrasts affect listeners’ non-native Dutch word learning patterns of minimal pair words containing the same sounds. Specifically, we tested the L2LP model’s tenet that non-native vowel discrimination and acquisition are predicted by L1-L2 acoustic relationships, and that non-native perceptual abilities will translate to non-native word learning abilities (Escudero, 2005, 2006, 2009; van Leussen & Escudero, 2015). With respect to non-native word learning, the model predicts that words containing contrasts that are either difficult or easy for non-native listeners to discriminate should correspondingly be more difficult or easy to learn. To test these claims, we compared data from Alispahic et al. (2017) on Australian English (AusE) and Peruvian Spanish (PS) listeners’ XAB discrimination of five Dutch vowel contrasts with data from Escudero (2015) showing AusE and PS listeners’ learning of minimal pair words containing the same Dutch vowel contrasts. Data sets from both published studies were examined using the same statistical models, and results were further compared qualitatively. In line with the L2LP model, if vowel contrast discrimination relates to minimal pair learning, results from both experiments should show a similar pattern of vowel contrast difficulty. Indeed, minimal pairs containing perceptually difficult vowel contrasts were difficult to discriminate, while word pairs containing perceptually easy contrasts were better discriminated. Participants’ native language did not affect performance in either task. These results suggest that L1/L2 acoustic relationships explain both perception
and word learning, confirming the L2LP proposal that there is continuity between perception and recognition in L2 development.

3.2 Introduction

To learn and recognize novel words, naïve L2 listeners must discriminate any non-native speech sounds contained within those words. This study concerns the extent to which non-native sound perception affects foreign word recognition, more specifically, whether there is a positive relationship between discrimination difficulty of particular L2 speech contrasts and learning non-native words containing those contrasts.

A number of lexical models propose that word recognition occurs once a speech signal is heard and information is extracted at a pre-lexical level allowing access of meaningful lexical representations (e.g., Cohort model: Marslen-Wilson, 1987; TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994; Shortlist-B: Norris & McQueen, 2008). That is, multiple sublexical representations of phonemes act as mediators between the speech signal and the mental lexicon. Thus, spoken word recognition involves activation of multiple word candidates in a listener’s mental lexicon and subsequent competition across candidates (e.g., Weber & Scharenborg, 2012). For example, once a listener hears the initial phoneme, e.g., /k/, words sharing the same word-initial phoneme (termed ‘cohorts’) will also be activated (e.g., cap, cat, cab, catch, and captain) resulting in cohort interference (Desroches, Newman, & Joanisse, 2009; Marslen-Wilson, 1987). The set of competitors is further increased by activation of lexical items that share a similar neighborhood (e.g., words that differ from cap by only one phoneme, like cop, cape, and clap, which also includes rhymes, map, tap, zap; Desroches et al., 2009; Luce & Pisoni, 1998). The greater the phonological overlap between the perceived word and the activated candidate words, the stronger a competitor a word candidate will be.
Acoustic-phonetic variation between activated lexical competitors is an additional factor that accounts for difficulties in word recognition according to continuous mapping models (e.g., TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994; Shortlist-B: Norris & McQueen, 2008). Due to the temporal nature of speech, listeners must integrate multiple sources of information, including acoustic features, when perceiving phonemes in the speech stream (McClelland & Elman, 1986). If a listener hears a nasalised vowel /a/ such as in the word ‘ran’ then competitor sets that are acoustically similar will be more strongly activated (e.g., ram and rang) compared to those which are not (e.g., rat and rag; Marslen-Wilson & Warren, 1994).

During L1 word processing, competitor activation is increased for words that are phonologically similar. In an eye-tracking study that investigated the effects of visual and linguistic information in spoken language comprehension, monolingual American English listeners were instructed to fixate their eyes to everyday objects presented to them (Tanenhaus et al., 1995). It was established that when visually presented with multiple items, the mean time to initiate eye-movement to the correct object (e.g., candy) presented in the display was slower for items that shared a phonologically similar onset with another object (e.g., candle), compared to those that did not (e.g., apple). Increased competition between candidates sharing the same word onset was also found in other studies (e.g., Dutch: Dahan, Swingley, Tanenhaus, & Magnuson, 2000; French: Salverda, Dahan, & McQueen, 2003).

It’s possible that activation of competitor sets is increased in an L2 environment. To test whether bilinguals are able to deactivate one language while using the other Spivey and Marian (1999) tested Russian-English late bilinguals using the same eye-tracking paradigm as Tanenhaus et al. (1995). During separate Russian and English sessions, participants were instructed to move everyday objects presented to them on a table (Spivey & Marian, 1999). There were two conditions in each language session: (i) an interlingual-distractor-present
condition, in which the objects shared the same phonological onset in both languages (e.g., Russian target marku, ‘stamp,’ and English competitor marker), and (ii) an interlingual-distractor-absent condition in which the distractor object was replaced by an object that did not share the same phonological onset (e.g., a ruler, lineika in Russian). Participants made more eye movements to the distractor in the interlingual-distractor-present condition than the interlingual-distractor-absent condition (Spivey & Marian, 1999). The authors concluded that bilingual listeners appear to activate lexicons of both of their languages even when in a monolingual situation (Spivey & Marian, 1999). As a result, for L2 learners recognition difficulties are increased as acoustic features, phonemes and word candidates from both of their languages are activated leading to an increase in word competition and slower overall word recognition (Vroomen & de Gelder, 1995; Weber & Cutler, 2004).

If non-native listeners are unable to correctly perceive the speech signal, then this may result in word recognition and word learning difficulties. Therefore, it is important to address the theoretical issue of the continuity between non-native speech perception, word learning and spoken word recognition (Pallier et al., 2001; Elvin, 2016). Current models of non-native and L2 speech perception such as the Speech Learning Model (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 (Escudero, 2005, 2009; van Leussen & Escudero, 2015) share a common assumption that it is the L1-L2 relationships that affect listeners’ speech perception. However, L2LP is the only model that explicitly links perceptual and lexical difficulties and functions as a computational model and that specifies the relationship between perception and word recognition in non-native and L2 acquisition (van Leussen & Escudero, 2015). Uniquely, L2LP considers learners from the initial to end state of learning and hypothesizes that comparisons of acoustic properties between the L2 and L1 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). Set within the L2LP framework are learning scenarios that detail listeners’ learning stages from naïve to experienced L2 learners. L2LP proposes that if a listener is unable to discriminate two non-native sounds and maps them onto a single native category then they will also have difficulty discriminating words that are distinguished by these same sounds. This is known as NEW scenario (L2LP; Escudero, 2005, 2009). The NEW scenario requires listeners either to create a new L2 category or to split their existing single L1 category (van Leussen & Escudero, 2015). Alternatively, good discrimination would be predicted to occur when a contrast in the target language is perceived as two separate native categories. This pattern requires the listener simply to replicate and adjust their native boundaries to match those of the L2 contrast. This is referred to as the SIMILAR scenario (van Leussen & Escudero, 2015). A third pattern, predicting poor-to-good discrimination, is the SUBSET scenario (Escudero & Boersma, 2002), which refers to listeners mapping an L2 sound to two or more native categories. This pattern is predicted to result in good discrimination provided that there is no boundary overlap between multiple native categories. Alternatively, if boundary overlap occurs, it is predicted that this scenario will lead to poorer discrimination.

Taking into account listeners’ L1 sound inventories, speech perception studies have extensively documented that most L2 listeners struggle to perceive and produce L2 phoneme contrasts not present in their native language. Specifically, research has uncovered perceptual difficulties faced by learners whose L1 contains fewer phonemes than the target L2 (e.g., Escudero, 2005; Escudero & Boersma, 2004; Flege, Bohn, & Jang, 1997; McLennan, Luce, & Charles-Luce, 2003). Vowel inventories differ in size across languages, estimated to vary from 3 to 24 distinct vowels (Maddieson, 1984; Vallée, 1994). Research has extensively shown that vowels are particularly difficult to master. That is, if some vowel sounds are not present in a listener’s native language, then perception is deemed to be more difficult,
affecting listeners’ recognition accuracy of novel vowel sounds. For instance, Southern British English has a large vowel inventory, containing 11 monophthongal vowels, compared to Greek and Japanese, which each have only 5 vowels. Greek and Japanese listeners categorize multiple Southern British English vowels as belonging to a single native vowel category; Greek listeners categorized English /aː/, /ɔ/ and /ɔː/ as Greek /ɔ/, and Japanese listeners categorized English /ɜː/ and /aː/ as Japanese /aː/ (Lengeris, 2009). This difficulty is presumably due to the continuous dimensions by which vowels differ within a language (e.g., Beddor & Strange, 1982; Polka & Bohn, 1996). That is, boundaries between one vowel and another are less distinct than those of consonants and are further influenced by phonetic variation across contexts (Keating & Huffman, 1984; Ladefoged & Johnson, 2011).

The most difficult L2 vowel contrasts to distinguish are those that are perceived as belonging to a single native category. Failure to discriminate L2 vowel contrasts may influence recognition and the subsequent learning of L2 words contrasted by the same vowels, i.e., minimal pairs (e.g., sheep /ʃip/ and ship /ʃɪp/; Weber, Broersma, & Aoyagi, 2011). Earlier studies have shown that L2 listeners appear to have lower recognition accuracy of word pairs containing vowel contrasts believed to be perceptually difficult compared to words containing vowel contrasts believed to be easy to discriminate (Mirijam Broersma, 2002; Escudero et al., 2013, 2008a; C. Pallier et al., 1997; Weber & Cutler, 2004). For example, Dutch learners of English more frequently judged non-words containing difficult L2 contrasts as real words relative to native English listeners (Broersma, 2002). A novel and perceptually difficult vowel contrast for Dutch listeners is English /æ–ɛ/ (Broersma, 2005; Cutler & Broersma, 2005). This contrast has been shown to be particularly difficult as the Dutch sound system contains the phonetic category /ɛ/ but not /æ/ (Cutler & Broersma, 2005). Studies have shown that L2 phonological difficulties may translate to word recognition difficulties. For instance, when highly proficient Dutch learners of English were
taught bisyllabic English nonwords using an eyetracking word recognition paradigm (Escudero et al., 2008), these Dutch learners were unable to discriminate the first syllable of nonwords that contained English /æ–ɛ/ (as in /pæn/ in *panda* and /pɛn/ in *pencil*; Escudero et al., 2008). Spanish bilinguals appear to follow a similar pattern in their perception of minimal pair words that contain perceptually difficult vowel contrasts. Spanish-dominant bilinguals have difficulty perceiving the Catalan /e–ɛ/ and /o–ɔ/ contrasts, as they are not phonemically differentiated in Spanish (Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés & Soto-Faraco, 1999). In a follow up study, the authors tested whether Spanish-dominant bilinguals store Catalan words containing these pairs as homophones or as distinct lexical items (Pallier, Colomé, & Sebastián-Gallés, 2001). Using a repetition priming task, Spanish-dominant and Catalan-dominant Spanish-Catalan bilinguals were presented with auditory forms of Catalan minimal pairs and pseudowords that were comprised of either common contrasts or Catalan-specific contrasts (e.g., *dóna* “s/he gives” and *dona* “woman”). Compared to Catalan-dominant bilinguals, Spanish-dominant bilinguals were unable to differentiate minimal pairs that were lexically contrasted by /e–ɛ/ and /o–ɔ/, and did indeed treat them as homophones. That is, as these contrasts are not differentiated in Spanish, lexical items containing the same sounds were phonologically indistinguishable to listeners. Spanish-dominant bilinguals were able to differentiate Catalan words that were not lexically contrasted by /e–ɛ/ and /o–ɔ/ (e.g., ‘néta’ /nɛtə/ meaning ‘granddaughter’ and ‘neta’ /nɛtə/ the feminine form meaning ‘clean’) just as well as the Catalan-dominant bilinguals. The authors concluded that while Spanish-dominant bilinguals can master the Catalan lexicon, their lexical representations differ from those of native Catalan listeners (Pallier et al., 2001). Together, these findings suggest that if a listener is unable to auditorily distinguish a non-native sound contrast they will have difficulties distinguishing between words differentiated by those contrasts (Escudero et al., 2008).
Further research has also shown an advantage for listeners whose L1 contains more vowels than the L2 (e.g., Iverson & Evans, 2007, 2009). Listeners of German and Norwegian, whose L1 contains more vowels than English, were more accurate in their identification of English vowels compared to native speakers of French and Spanish, whose native language contains fewer vowels than English (Iverson & Evans, 2007, 2009). Previous studies have also shown that listeners’ recognition accuracy for L2 minimal pairs is higher when the minimal pairs are lexically contrasted by perceptually easy phonemes (Broersma, 2005; Escudero, Simon, & Mulak, 2014; Weber & Cutler, 2004). For instance, Spanish listeners find the Dutch contrasts /i-a/ and /u-a/ easy to discern, as these vowels are also differentiated in their native language (Escudero & Wanrooij, 2010). Spanish listeners’ recognition performance was on par with that of native Dutch listeners when presented with novel minimal pair words containing the same sound contrasts (e.g., /piχ/ and /paχ/; 93% for Spanish learners, 95% for Dutch native listeners; Escudero, Broersma & Simon, 2012). Similarly, Dutch learners of English were found to readily distinguish lexical contrasts containing the English /ɒ-i/ (e.g., “bottle” and “beetle”), as these vowels are similar to distinct phonemes in their language (Weber & Cutler, 2004). These findings suggest that L2 word recognition uses language-specific phonological representations, particularly between native vowel inventories and non-native word learning. However, all of the aforementioned studies tested bilinguals’ L2 perceptual patterns and word recognition patterns separately. Moreover, compared to unfamiliar non-native speech perception, research suggests that exposure to and experience with an L2 has a direct bearing on L2 perception (Best & Strange, 1992), and that perceptual learning of difficult non-native and L2 contrasts is possible (e.g., Best & Tyler, 2007).

The present study explicitly compares how and to what extent initial phoneme discrimination patterns relate to word recognition patterns in minimal word pairs.
Accordingly, this study is the first that compares naïve AusE and PS listeners’ L2 perceptual and word recognition patterns of Dutch. As a means of investigating the continuity between non-native phonological and lexical difficulties, we compare Australian English (AusE) monolinguals’ (Alispahic, Mulak, & Escudero, 2017) and Peruvian Spanish (PS) monolinguals’ (Escudero & Wanrooij, 2010) perceptual patterns for five Dutch vowel contrasts, namely /a-ɑ/, /i-i/, /y-y/, /i-y/, and /ɪ-y/ (Adank, van Hout, et al., 2004) to word recognition data (Escudero, 2015) of minimal pair words containing the same sounds. This comparison is particularly interesting as AusE and PS have very different native vowel inventories: AusE contains thirteen monophthongs (/i:, ɪ, ɪə, e, e:, æ, ɐ:, ɐ, ɔ, o:, ʊ, ʉ:, ɜ:/; Cox, 2005), whereas PS contains only five, /a, e, i, o, u/ (Chládková et al., 2011).

If non-native vowel contrast discrimination relates to minimal pair learning, results from both experiments should show similar perceptual patterns. In addition, if L2 perceptual patterns are linked to L2 word learning, in line with the tenet of the L2LP model that difficulties encountered in vowel perception should carry over to word learning, it is predicted that both listener groups should show comparable difficulty across both tasks. Specifically, word recognition difficulties are predicted to transpire for minimal pairs that are differentiated by perceptually difficult vowel contrasts, while higher recognition accuracy should be reflected for minimal pairs that contain perceptually easy vowel contrasts.

Alispahic et al. (2017) established that for AusE and PS naïve listeners the Dutch /i-i/ vowels are difficult to differentiate. Both listener groups face a NEW scenario when categorizing these sounds by equating them as a single native phoneme. Therefore, in line with the L2LP model, it is expected that both AusE and Spanish listeners should face difficulties in their recognition of minimal pairs containing this contrast. Moreover, contrasts that were perceived across two distinct native categories (e.g., Dutch /a-ɑ/ → AusE /v:/ and /ɔ/:; Dutch /i-y/ → PS /i/ and /u/) are predicted to result in a SIMILAR scenario and should
result in good recognition of minimal pairs containing the same sounds. Furthermore, a SUBSET scenario which involves non-native vowels being mapped across more than two native categories should pose discrimination and word recognition difficulty in instances where acoustic overlap occurs, such as Dutch /ɣ-ɪ/ for PS listeners (Alispahic et al., 2017). Alternatively, medium to good recognition is expected for SUBSET contrasts that are not completely neutralised across native categories, such as Dutch /i-ɣ/ and /ɣ-ɪ/ for AusE listeners (Alispahic et al., 2017). However, if vowel inventory size is indeed a reliable predictor of non-native word recognition patterns, then AusE listeners, having a larger vowel inventory compared to PS listeners, should simply outperform PS listeners in their recognition of all minimal pair words containing the five Dutch contrasts (/a-ɑ/, /ɪ-ɪ/, /ɣ-ɣ/, /i-y/, and /i-ɣ/). AusE and PS listeners’ predicted word recognition learning scenarios are set out in Table 6.

**Table 6. Predicted word recognition learning scenarios for AusE and Spanish listeners.**

<table>
<thead>
<tr>
<th>Dutch vowel contrast</th>
<th>AusE</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i-ɪ/</td>
<td>New</td>
<td>New</td>
</tr>
<tr>
<td>/ɣ-ɣ/</td>
<td>Subset</td>
<td>Subset</td>
</tr>
<tr>
<td>/ɣ-ɪ/</td>
<td>Subset</td>
<td>Subset</td>
</tr>
<tr>
<td>/a-ɑ/</td>
<td>Similar</td>
<td>New</td>
</tr>
<tr>
<td>/i-y/</td>
<td>Subset</td>
<td>Similar</td>
</tr>
</tbody>
</table>

In sum, L2 contrasts and minimal pairs involving New Scenario are predicted to elicit poor discrimination and word learning, while the opposite should occur for those that involve Similar Scenario. L2 vowels that map across more than two native categories are predicted to pose more difficulties in the Subset Scenario, compared to those that are not completely neutralised across native categories.
3.3 Method

The procedural details are split between the XAB task, in which data from AusE participants (from Alispahic, Mulak, & Escudero, 2017) and PE participants (from Escudero & Wanrooij, 2010) are analysed; and a word recognition task in which data from AusE and Spanish participants (from Escudero, 2015) are analysed.

3.3.1 Task 1: XAB Discrimination

Task 1 is the same task previously reported by Alispahic et al. (2017). We provide a summary of the participants, stimuli and procedure below.

3.3.1.1 Participants

Alispahic and colleagues (2017) tested 22 monolingual AusE students aged 18-45 years ($M_{age} = 24.1$; 11 males) who participated for course credit and were recruited via the Western Sydney University SONA Research Participation System. Participants were born and raised in Greater Western Sydney, and reported little to no experience with languages other than English, and no prior experience with Dutch. Testing was conducted in a quiet room at Western Sydney University. Alispahic et al. (2017) compared the results of AusE participants to those of 22 PS listeners from a dataset previously reported in Escudero and Wanrooij (2010). The PS participants were monolinguals aged between 17 and 28 years ($M_{age} = 21.0$; 10 females), who reported no prior knowledge of Dutch, and who lived in Lima, Peru their entire life. PS participants were tested in a quiet room at the Pontificia Universidad Católica del Perú, in Lima. Neither group of participants reported any hearing impairment.

3.3.1.2 Stimuli and Procedure

Stimuli were 20 naturally produced tokens of each of the 5 Dutch vowels (/a/, /ɑ/, /ɪ/, /i/, /y/) presented as five vowel contrasts, /a-ɑ/, /ɪ-ɪ/, /y-ʏ/, /i-ʏ/ and /i-ɣ/. Vowels were extracted
from recordings produced by 10 male and 10 female native Standard Northern Dutch speakers in a non-word /sVs/ context embedded within a carrier sentence (Adank et. al, 2004). During the task, listeners heard three sounds which were played at a comfortable volume through headphones and were presented using the Praat computer program (Boersma & Weenink, 2005).

Listeners were asked to decide whether the first heard sound (X) was more like the second (A) or more like the third (B) heard sound. On the computer screen listeners were shown two yellow squares, viz. “2” and “3,” and were asked to click on a square that corresponded to the sound that best matched the first heard sound. There were five XAB tasks, one for each of the five contrasts (/a-/a/, /i-/i/, /y-/y/, /i-/y/, /i-/i/) and each included 80 trials, a total of 400 trials. The inter-stimulus interval between the X and A and the A and B stimuli was 1.2 s, which was selected because it is long enough to trigger phonological activation (Escudero & Wanrooij, 2010; Werker & Logan, 1985; Van Hessen & Schouten, 1999). There was an inter-trial interval of 0.5 s between the participant’s response click and presentation of the next trial. Participants were presented with four practice trials consisting of synthetic vowel tokens to familiarize themselves with the testing procedure. The practice trials could be repeated once if requested by the participant.

3.3.2 Task 2: Word learning

3.3.2.1 Participants

Escudero (2015) reports on AusE and Spanish listeners’ ability to associate 12 Dutch pseudowords with 12 visual referents consisting of line drawings of nonsense objects. While the study tested both monolinguals and bilinguals/multilinguals, for the present study we only use the data of English and Spanish monolingual listeners. The AusE monolingual participants were 43 undergraduate students from Western Sydney University, aged between
18 and 35 years\textsuperscript{5} who reported no knowledge of any other language, and no exposure to Dutch. Escudero (2015) compared the results of AusE listeners to those of 30 Spanish monolingual listeners from a previous dataset reported in Escudero, Simon, et al., (2014). Spanish participants were undergraduate students from Pontificia Universidad Católica del Perú, in Lima originally from Spain or Latin America\textsuperscript{6}, who reported only basic knowledge of English and no exposure to Dutch, and who were between 18 and 25\textsuperscript{7} years of age. Both language groups reported normal hearing and normal or corrected-to-normal vision.

3.3.2.2 Stimuli

Stimuli comprised 12 Dutch pseudowords presented in a carrier sentence. Each pseudoword was randomly paired with a line drawing picture from Shatzman and McQueen (2006), previously used with L2 learners (Escudero et al., 2008, 2013, 2014). Half of the pseudowords contained one of the six Dutch vowels /ɪ, i, a, ɑ, y, y/, and were monosyllables produced in /pVχ/ context. The remaining six pseudowords were disyllabic and contained different consonants and vowels. Three had a long vowel or diphthong (/ˈbeːptːuː/, /ˈfoːmpəl/, /ˈtœykfɔm/) and three a short vowel in the stressed syllable (/ˈjɔmtoː/, /ˈkɛsta/, /ˈsurkɛt/). Two Dutch carrier sentences were produced by a native female speaker of Dutch (“Dit is een X” [“This is an X’] and “Klik op de X’” [“Click on the X’]). Sentences and pseudowords were recorded at the Institute of Phonetic Sciences at the University of Amsterdam in a soundproof booth, and were stored at a sample rate of 44.1 kHz. Carrier sentences and pseudowords were read one by one in clear citation style. To keep the context of the carrier sentence constant,

\textsuperscript{5} The mean age for these participants and gender balance was not reported by the previous authors.

\textsuperscript{6} The authors of the study state that a recent acoustic comparison of Spanish vowels spoken by speakers from Madrid and Lima (Chladkova, Escudero & Boersma, 2011) shows only small differences in the acoustic properties of vowels across dialects, especially in the pVpV (V = vowel) context, which is the context closest to the one used for the Dutch vowels included in the present study. Therefore, the authors did not expect any variation in the learning of non-native minimal word pairs between Spanish speakers from Spain and different countries in Latin America.

\textsuperscript{7} The mean age for these participants and gender balance was not reported by the previous authors.
pseudowords were cross-spliced from the end of one carrier sentence onto the end of another carrier sentence that originally contained the same word.

3.3.2.3 Procedure: Word Learning Phase

The word learning phase consisted of two parts. As shown in Figure 2, the sound file and visual image started simultaneously. During the first part, individual pseudowords were presented to listeners auditorily in the sentence “Dit is een X” (“This is an X”) while a referent image was displayed on the computer screen. The sound file and image started simultaneously, while the latter stayed in the middle of the screen for 2 s. Once the referent image disappeared and 1.5 s after the offset of the previous sound file, the same pseudoword was presented in the sentence “Klik op de X” (“Click on the X”) and the target image was presented side-by-side with another image (the distractor) at the offset of the sentence.

![Figure 2](image-url)

_Figure 2._ Illustration of a word learning trial as reported in Escudero et al. (2014) and Escudero (2015). Figure adapted from (Giezen, Escudero, & Baker, 2015).

The position of targets and distractors on the screen (left vs. right) was counterbalanced and presented in a pseudorandom order – the same target could appear only twice in succession, and targets could appear in the same place on the screen a maximum of five times in succession. Participants were required to press on a computer key that corresponded to the left-right location of the named image. The two images were displayed on the screen until a key was pressed. There was a time-out of 10 s, with the next trial starting 1 s after the key
press or the end of the time out period. The word learning phase consisted of 72 learning trials with each of the 12 pseudowords presented 6 times.

### 3.3.2.4 Procedure: Test Phase

The ensuing test phase was the same as the second part of the word learning trial, with the exception that the next trial started 0.5 s after the key press. Of the 264 test trials, 204 comprised 51 non-minimal word pairs each presented four times (e.g., ‘beep toe’ – ‘fompel’ or ‘beep toe’ – ‘pag’). Thirty-two of the trials consisted of eight perceptually easy minimal word pairs (e.g., ‘pag’–‘pieg’) presented four times each. The other 28 trials were of seven perceptually difficult minimal word pairs (e.g., ‘pag’–‘paag’) presented four times each.

Participants were tested individually in a soundproof booth and both the training and test phases took approximately 30 minutes to complete. Escudero, Simon and Mulak (2014) divided these fifteen minimal pairs into two groups based on Spanish listeners’ vowel discrimination performance reported in Escudero and Wanrooij (2010), and the likelihood that listeners would be able to discriminate between contrasts: (1) Perceptually difficult pairs: /ɪ–i/, /ɪ–y/, /ɪ–ʏ/, /ɪ–ɪ/, /ɪ–ɪ/ and; (2) Perceptually easy pairs: /ɪ–ɑ/, /ɪ–a/, /i–a/, /i–i/, /i–a, /a–ɪ/, /a–a/, /a–y/, /a–y/, /a–y/. To determine the relationship between non-native sound perception and non-native word learning, the present study focused its analyses on the same five perceptually difficult contrasts presented to listeners in the XAB task, namely /a–ɑ/, /i–i/, /y–ɪ/, /i–ɪ/ and /i–ɪ/.

### 3.3.3 Analysis

To establish whether XAB discrimination related to L2 word recognition performance, the present study reanalysed results reported in Escudero (2015). Escudero (2015) used a repeated-measures analysis of variance (ANOVA), whereas Alispahic, Mulak and Escudero (2017) used a mixed effects logistics model for their analysis, as is appropriate when
evaluating categorical responses (see Arnon, 2010; Baayen, Davidson, & Bates, 2008; Jaeger, 2008). The XAB discrimination and L2 word recognition analyses will be compared by employing the same statistical model across both tasks, as in Alispahic and colleagues (2017). Qualitative comparisons of L2 vowel perception and L2 word learning data of listeners’ accuracy percentages across both experiments will also be made to determine whether speech perception patterns translate to word learning patterns.

3.3.4 Ethics Approval

This study was conducted in accordance with the Human Research Ethics Committee (HREC), Western Sydney University, approval number H9373.

3.4 Results

3.4.1 XAB Discrimination Task

The overarching aim of this study was to examine the interrelation between non-native vowel perception and non-native word learning, employing the same statistical analyses, so only the findings of the XAB discrimination task from Alispahic et al. (2017) are reported here. We thus compare listeners’ accuracy performance across the XAB and word learning tasks. For visual reference, we include figures and tables adapted from those reported by Alispahic and colleagues.

Alispahic et al. (2017) compared XAB discrimination between naïve AusE and PS listeners to determine whether discrimination differed between participants whose native vowel inventory was smaller (PS) or larger (AusE) than in Dutch. Listeners’ discrimination accuracy of five non-native Dutch vowel contrasts is presented in Figure 3. In their analysis the authors employed a mixed-effects logistic model, that included the participants’ correct responses, with participant, speaker and XAB trial as random effects, and vowel contrast and
language background as fixed effects revealed a main effect of contrast ($\chi^2 (4, N = 17600) = 38.7, p = <.001$). While there was no main effect of language background ($\chi^2 (1, N = 17600) = .112, p = .738$), there was an interaction of vowel contrast and language background ($\chi^2 (4, N = 17600) = 16.5, p = .002$). Fishers’ LSD post-hoc pairwise comparisons revealed that PS listeners had marginally more correct responses than AusE listeners for /ɪ-i/ ($p = .053, 95\%$ CI [-0.52, -0.003]), whereas AusE participants made marginally more correct responses for /ʏ-ɪ/ than PS listeners ($p = .086 [-0.44, 0.03]$).

**Figure 3.** Discrimination accuracy of all non-native Dutch vowel contrasts by 22 native AusE and 22 native PS participants (Alispahic et al., 2017; Escudero & Wanrooij, 2010). Interaction of vowel contrast and language background is labelled (*).

Furthermore, overall pairwise comparisons of the main effect of contrast revealed that compared to Dutch /ʏ-y/ ($p = .007, [-.09, -.01]$), /ɪ-i/ ($p = .001, [-.12, -.05]$), /a-ɑ/ ($p = .001, [-.13, -.05]$) and /ʏ-ɪ/ ($p = .001, [-.06, -.12]$), Dutch /i-y/ was the least challenging contrast to discriminate. There was a trend of more correct responses for Dutch /ʏ-y/ compared to /ɪ-i/ ($p = .090, [-.07, .01]$), /a-ɑ/ ($p = .054, [.08, .00]$) and /ʏ-ɪ/ ($p = .060, [.08, .00]$), while accuracy
was lower for Dutch /i-ɪ/ compared to /a-ɑ/ (p = .730, [-.05, .04]) and /y-ʏ/ (p = .784, [-.04, .03]). Thus, the combined overall difficulty ranking from least to most challenging contrast to discern is: (1) /i-ɪ/ >> (2) /y-ʏ/ ~ /i-ɪ/ ~ /a-ɑ/ ~ /i-ʏ/. Figure 4 presents discrimination accuracy of the five non-native Dutch vowel contrasts combined over listener groups.

**Figure 4.** Combined overall discrimination accuracy of all non-native Dutch vowel contrasts by 22 native AusE and 22 native PS participants (Alispahic et al., 2017; Escudero & Wanrooij, 2010). Standard error bars were treated as Independent Variables.

Using Fisher’s LSD pairwise comparisons, significance levels and 95% confidence intervals for participants’ discrimination accuracy across the two language groups are set out in Table 7.
Table 7. Fisher’s LSD pairwise comparisons of significance levels and confidence intervals [CI] for the XAB Discrimination task by AusE and PS listeners.

<table>
<thead>
<tr>
<th>XAB Discrimination Task</th>
<th>/a-α/</th>
<th>/i-i/</th>
<th>/γ-y/</th>
<th>/γ-ɪ/</th>
<th>/i-y/</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Listener Group:</strong></td>
<td>AusE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a-α/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i-i/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/γ-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/γ-ɪ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Listener Group:</strong></td>
<td>PS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a-α/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i-i/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/γ-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/γ-ɪ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As can be seen, AusE participants had more correct responses for /i-y/ than /i-i/, /a-α/, /γ-y/, and /γ-ɪ; for /γ-y/ than /i-i/; and for /a-α/ than /i-i/. PS participants had more correct responses for /γ-y/ than /a-α/ and /i-γ; for /i-y/ than /a-α/ and /γ-ɪ/; and for /i-i/ than /γ-ɪ/, with a trend towards more correct responses for /i-i/ than /a-α/.
3.4.2 Word Recognition Task

For L2 word recognition accuracy the same statistical analysis as in the XAB discrimination task was used (Alispahić et al., 2017). A mixed-effects logistic model with participant and picture trial as random effects, vowel contrast and language background as fixed effects and participants’ correct response as dependent variable revealed a main effect of contrast ($\chi^2 (4, N = 660) = 20.43, p < .001$). There was no main effect of language background ($\chi^2 (1, N = 660) = 1.04, p = .308$) or interaction of vowel contrast and language background ($\chi^2 (4, N = 660) = 3.84, p = .428$). To test for the main effect of contrast, Fishers’ LSD pairwise comparisons revealed that there was a trend toward more correct responses by Spanish listeners compared to AusE listeners for /i/-/y/ ($p = .081, [-1.80, 0.15]$). Listeners’ accuracy for the novel words containing the five non-native Dutch vowel contrasts is presented in Figure 5.

![Figure 5](image.png)

**Figure 5.** Word identification accuracy of minimal pair words containing non-native Dutch vowel contrasts (Escudero, 2015).
Moreover, pairwise comparisons for the main effect of contrast revealed that Dutch /i-y/ was the least challenging minimal pair to discern compared to Dutch /i-i/ (\(p = .001, [-.34, -.10]\)), /a-a/ (\(p = .001, [-.27, -.08]\)), /y-y/ (\(p = .001, [-.38, -.10]\)), and /y-i/ (\(p = .049, [-.01, -.26]\)). There was no difference in accuracy between /i-i/ and /a-a/ (\(p = .563, [-.10, .19]\)), /y-y/ (\(p = .805, [-.14, -.11]\)) and /y-i/ (\(p = .168, [-.04, .23]\)), nor Dutch /a-a/ and /y-y/ (\(p = .314, [-.17, .06]\)) and /y-i/ (\(p = .369, [-.06, .16]\)). Therefore, the combined overall difficulty ranking from least to most challenging minimal pair vowel contrast is (1) /i-y/ >> (2) /y-y/ ~ /i-i/ ~ /a-a/ ~ /i-y/. Presented in Figure 6 are the combined mean L2 word recognition accuracy percentages for the two listener groups.

![Graph](image)

**Figure 6.** Word identification accuracy of non-native Dutch minimal pair vowel contrasts by naive AusE and Spanish listeners (Escudero, 2015).

While there was no interaction between vowel contrast and language background we were interested to compare Fisher’s LSD pairwise comparisons between the XAB and word learning task. Therefore, we employed an additional post-hoc Fisher’s LSD pairwise comparison. Significance levels and 95% confidence intervals for participants’ word learning
accuracy involving the five Dutch vowel contrasts across the two language groups are set out in Table 8.

**Table 8.** Fisher's LSD pairwise comparisons of significance levels and 95% confidence intervals [CI] for the Word Learning task by AusE and Spanish listeners.

<table>
<thead>
<tr>
<th>Dutch vowel contrast</th>
<th>/a-a/</th>
<th>/i-i/</th>
<th>/y-y/</th>
<th>/y-i/</th>
<th>/i-y/</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Listener Group:</strong></td>
<td>AusE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a-a/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i-i/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-i/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Listener Group:</strong></td>
<td>Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a-a/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i-i/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-i/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AusE participants had more correct responses for /i-y/ than /a-a/ and /y-y/, with a trend toward more correct responses than /i-i/; more correct responses for /y-i/ than /a-a/; and a trend toward more correct responses for /i-i/ than /a-a/. PS participants had more correct
responses for /ɪ-y/ than /ɪ-i/ and /ʏ-y/, with a trend for more correct responses compared to /ʏ-ɪ/.

3.4.3 Summary of Results

For visual comparison, listeners’ mean accuracy percentages of the five Dutch vowel contrasts in both the XAB discrimination and word learning task, along with the learning scenario elicited, is set out in Table 9.

Table 9. Percent accuracy and standard error for each contrast for the XAB discrimination and word learning task.

<table>
<thead>
<tr>
<th>L1 listener group:</th>
<th>AusE</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɑ-ɑ/</td>
<td>64.8%</td>
<td>41.7%</td>
</tr>
<tr>
<td>/ɪ-ɪ/</td>
<td>67.9%</td>
<td>56.9%</td>
</tr>
<tr>
<td>/ʏ-ʏ/</td>
<td>69.4%</td>
<td>48.6%</td>
</tr>
<tr>
<td>/ʏ-ɪ/</td>
<td>69.8%</td>
<td>63.9%</td>
</tr>
<tr>
<td>/ɪ-y/</td>
<td>78.4%</td>
<td>70.8%</td>
</tr>
</tbody>
</table>

While there was no main effect of language background across either task, our statistical analyses revealed that there was an effect of contrast. Accuracy percentages across the two tasks collapsed between participant language groups are in Table 10.
Table 10. Collapsed accuracy percentages across the XAB discrimination and word learning task.

<table>
<thead>
<tr>
<th>Task: Dutch vowel contrast</th>
<th>Overall accuracy percentages</th>
<th>XAB</th>
<th>Word Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent accuracy</td>
<td>Standard Error</td>
<td>Percent accuracy</td>
<td>Standard Error</td>
</tr>
<tr>
<td>/a-a/</td>
<td>67.6</td>
<td>1.8</td>
<td>56.0</td>
</tr>
<tr>
<td>/i-i/</td>
<td>66.8</td>
<td>2.1</td>
<td>51.7</td>
</tr>
<tr>
<td>/y-y/</td>
<td>70.9</td>
<td>2.1</td>
<td>50.1</td>
</tr>
<tr>
<td>/y-i/</td>
<td>67.2</td>
<td>1.5</td>
<td>61.1</td>
</tr>
<tr>
<td>/i-y/</td>
<td>75.9</td>
<td>2.0</td>
<td>73.8</td>
</tr>
</tbody>
</table>

Fisher’s LSD pairwise comparisons revealed that the overall difficulty ranking from least to most challenging vowel contrast was the same across both tasks: (1) /i-y/ >> (2) /y-y/ ~ /i-i/ ~ /a-a/ ~ /i-y/. Significance levels for the main effect of contrast across both tasks are set out in Table 11.
Table 11. Fisher’s LSD pairwise significance levels for the main effect of contrast in both the XAB discrimination and word learning task.

<table>
<thead>
<tr>
<th>Combined</th>
<th>/a-ɑ/</th>
<th>/ɪ-ɪ/</th>
<th>/y-y/</th>
<th>/y-i/</th>
<th>/ɪ-y/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a-ɑ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɪ-ɪ/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-i/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɪ-y/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Task:             |       |       |       |       |       |
| Word Learning     |       |       |       |       |       |
| /a-ɑ/             |       |       |       |       |       |
| /ɪ-ɪ/             |       |       |       |       |       |
| /y-y/             |       |       |       |       |       |
| /y-i/             |       |       |       |       |       |
| /ɪ-y/             |       |       |       |       |       |

3.5 Discussion

The aims of this study were twofold: (i) to examine the relationship between non-native perceptual and lexical difficulties by comparing non-native vowel discrimination and word recognition of minimal pairs containing the same vowels, and (ii) to investigate whether (a) the size of vowel inventories relative to the target language, or (b) the relationship between acoustic properties in the native and non-native language better explain non-native word recognition. To this end, discrimination patterns of five Dutch vowel contrasts, /a-ɑ/, /ɪ-ɪ/, /y-y/,
ɣ/, /i-ɣ/ and /ɪɣ/ by AusE and PS listeners reported in Alispahic et al. (2017) were compared to non-native word recognition patterns of AusE and PS listeners reported in Escudero (2015). To test the continuity between non-native perception and word recognition we employed the same statistical analyses across both tasks. Our findings suggest that vowel inventory size does not provide a sufficient explanation of listeners’ non-native perceptual patterns and that, in line with the L2LP model, naïve listeners’ discrimination difficulties translate to word recognition difficulties. Justification for this conclusion follows.

Consider Table 11. Results show that there was no main effect of language background or interaction between vowel contrast and language background in listeners’ recognition patterns. As predicted, word recognition difficulties were found for words differentiated by perceptually difficult contrasts, while higher word learning accuracy was reflected in contrasts that were easily discerned. For instance, AusE listeners found Dutch /i-ɪ/ the most difficult to distinguish likely because this contrast elicited a NEW scenario in their categorization patterns (Alispahic et al., 2017). That is, AusE listeners categorized Dutch /i/ and /ɪ/ in line with a single native vowel, namely AusE /ɪ/. This perceptual difficulty was further reflected in their mean recognition accuracy of minimal pair words containing Dutch /i-ɪ/. These results show that AusE listeners’ perceptual difficulty has a direct bearing on their recognition of minimal pair words containing the same phonemes.

PS listeners appear to follow a similar pattern as AusE listeners. PS listeners found minimal pairs containing perceptually difficult contrasts (e.g., Dutch /a-ɑ/ and /ɪ-i/) equally difficult to differentiate. That is, PS listeners likely face NEW scenario in their categorization of these sounds by perceptually mapping them across a single native category, namely PS /a/ and /ɪ/ (Alispahic et al., 2017). As a result, high discrimination difficulty of these contrasts leads to word learning and subsequent word identification difficulties for minimal pairs containing the same sounds. Our results further support earlier findings that suggest Spanish
learners’ discrimination accuracy of novel Dutch minimal pairs is lower for perceptually
difficult contrasts (e.g., /a-ɑ/ or /i-ɪ/; Escudero et al., 2013).

Furthermore, Alispahic and colleagues (2017) established that listeners show lower
discrimination accuracy for non-native contrasts involving the SUBSET scenario, a scenario
that involves an acoustic or perceptual overlap of non-native contrasts with multiple native
sound categories. The results show that discrimination difficulties involving SUBSET with
acoustic overlap on the phoneme level extends to word recognition difficulties. For instance,
PS listeners were shown to categorize Dutch /ʏ-ɪ/ across multiple native categories. Dutch /ɣ/
was categorized as PS /i/, /ɛ/ and /u/, while Dutch /ɣ/ was also categorized as PS /i/, /ɛ/ and
/u/, resulting in PS listeners’ low discrimination accuracy (Alispahic et al., 2017). These
discrimination difficulties were also reflected in PS listeners’ word recognition patterns. As
shown in Table 9, recognition accuracy for minimal pairs containing Dutch /ɣ-ɪ/ ranked as the
most difficult for PS listeners. Likewise, AusE listeners’ recognition of minimal pairs
involving SUBSET that contain acoustic overlap with native sound categories, such as Dutch
/ɣ-ɣ/, ranked as one of the most difficult across both tasks. Conversely, low recognition
difficulty was seen for SUBSET contrasts that were not completely neutralized across native
categories, such as Dutch /i-ɣ/ and /ɣ-ɪ/ for AusE listeners. As set out in Table 9, minimal
pairs containing these contrasts ranked as the least difficult for AusE listeners to discern.
Similarly, PS listeners found words containing Dutch /i-ɣ/ easy to differentiate as was evident
by their high accuracy for this contrast across both tasks. Our findings suggest that word-
recognition difficulties reflect those of vowel discrimination patterns.

Our findings also show that contrasts are not treated equally across the two tasks.
Discrimination accuracy was much higher compared to word-recognition. A likely
explanation for this is that the nature of the task affects listeners’ perceptual performance.
These findings are in line with those of earlier studies showing that L2 listeners’ accuracy is
higher for acoustic-phonetic tasks compared to tasks involving lexical processing (Amengual, 2016; Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012; Núria Sebastian-Galles & Baus, 2005).

It was not possible to test what effect individual learners’ perceptual paths might have on these results as the tasks were not completed by the same listeners. Nevertheless, the statistical comparisons show there was no main effect of language background in either task, so there is a definite suggestion of continuity between non-native vowel perception and non-native word-learning. XAB discrimination revealed a marginal interaction of language background and two of the five vowel contrasts; for /i-i/ more correct responses by PS than AusE listeners, and for /y-y/ more correct responses by AusE than PS listeners. These perceptual leads dissipate for both listener groups as the task demands increase – there were no such language background by vowel contrast interactions in the word recognition results. However, both tasks reveal a main effect of contrast. Pairwise comparisons of the main effect of contrast show that difficulty rankings, from least to most difficult, were identical across both tasks (1) /i-y/ >> (2) /y-y/ ~ /i-i/ ~ /a-a/ ~ /i-y/; see Table 11).

In sum, the relationship between non-native phonological and non-native word learning was investigated. AusE listeners, having a larger L1 vowel inventory, did not outperform Spanish listeners in their initial L2 perceptual and initial L2 word learning. These results suggest that the size of listeners’ native vowel system alone did not account for listeners’ vowel discrimination and word learning patterns. Rather, it is proposed that the acoustic-phonetic relationship between the native and non-native language that determines non-native perceptual and word recognition patterns. These results support one of the main tenets of the L2LP model that a direct link exists between L2 sound perception and L2 word learning. This adds to the growing body of research (e.g., Elvin, 2016; Escudero et al.2008, 2013, 2014; Weber & Cutler, 2004) showing that perceptually difficult contrasts affect and have an
negative impact on non-native word recognition. Future research should focus on directly comparing naïve AusE and Spanish listeners who complete the same perception and word-recognition tasks. Findings may inform possible future language learning programs which could include customizing individual L2 learning according to native language.
4 GENERAL DISCUSSION

4.1 Thesis aims summary

This thesis investigated the existence and nature of the relationship between non-native vowel perception and non-native word learning. Specifically, its focus was to investigate whether it is the size of a learner’s native vowel inventory, the acoustic properties of the L1 vowels, or both that predict non-native perceptual patterns. In doing so, this investigation extends the comparison of vowel discrimination patterns and word recognition patterns by AusE and PS naïve listeners of Dutch.

The following questions were addressed:

1) Does the relative size of the L1 versus L2 vowel inventory affect vowel perception and word learning performance? If so, does having a larger rather than smaller L1 vowel inventory aid listeners in their initial L2 perceptual and initial L2 word learning abilities?

2) Does initial L2 vowel perception affect initial L2 word learning such that minimal pairs contrasted by perceptually difficult sounds are also difficult to learn?

3) Which is the better predictor of perception performance and word learning performance: L1-L2 acoustic comparisons or relative vowel inventory size between the L1 and L2?

Study 1 addressed these questions by first establishing the effects of listeners’ native language on their initial L2 vowel perception. It presented an investigation of naïve AusE and PS listeners’ discrimination and assimilation patterns of five Dutch vowel contrasts. Specifically, it investigated the effects of vowel inventory size versus acoustic L1-L2 relationships on listeners’ non-native perceptual performance. Using Linear Discriminant Analyses models it also tested whether acoustic similarities were predictive of listeners’
cross-language categorization patterns and whether these, in turn, were predictive of listeners’ discrimination patterns.

**Study 2** built on the investigation presented in Chapter 2, by comparing listeners’ initial perceptual patterns of five Dutch contrasts to word learning patterns of minimal pair words containing the same sounds. Specifically, it tested whether minimal pair words containing perceptually difficult sound contrasts were easily recognized or whether listener’s perceptual difficulties were also reflected in their word learning abilities.

Thus, this thesis aims were to investigate the effects of vowel inventory size and L1/L2 acoustic relationships on listeners’ non-native vowel perception. Furthermore, it sought to uncover if and to what extent perceptual patterns also extend into listeners’ non-native word recognition abilities. A summary of the studies and discussion of their findings is presented below. This is followed by the implications to current theoretical models and future directions sections.

### 4.2 Summary of Findings

#### 4.2.1 Study 1: Effects of native language on non-native vowel perception

Theoretical models of non-native speech perception share a common assumption that listeners will perceive non-native and L2 sounds through the lens of their native sound inventory (Best, 1994, 1995; Best & Tyler, 2007; Escudero, 2005, 2006, 2009; Flege, 1995). In this view, listeners will initially map non-native speech sounds in line with their native sound categories. As vowel inventories vary across languages, previous literature has established that there appears to be a link between the difficulty and/or ease of perceiving non-native speech sounds relative to listeners’ vowel inventory size. For instance, if a listener’s native vowel inventory has fewer phonemes than the target language then this may
pose perceptual difficulties as their native vowel inventory does not contain enough phoneme categories for them to map new sounds to (Escudero & Chládková, 2010; Iverson & Evans, 2007). Conversely, having a larger L1 vowel inventory compared to the target language has been shown to benefit listeners’ non-native and L2 perceptual abilities (Iverson & Evans, 2007, 2009). However, in addition to listeners’ native vowel inventory size affecting non-native vowel perception, previous studies have also shown that L1-L2 acoustic relationships may affect non-native and L2 perceptual patterns (Elvin, Williams, & Escudero, 2016; Escudero & Williams, 2012). That is, irrelevant of the vowel inventory size between the native and target language, listeners will map non-native and L2 sounds to the acoustically most similar native counterpart. To investigate these claims, Study 1 tested non-native perceptual patterns of listeners whose native vowel inventory contains more and fewer phonemes than the target language. Specifically, it compared AusE and PS listeners’ categorization and discrimination of non-native Dutch vowel sounds. Study 1 also investigated whether cross-linguistic LDA and stepwise models predicted listeners’ categorization patterns and whether these were indicative of their discrimination difficulties.

In a perceptual categorization task, AusE and PS listeners were asked to categorize six Dutch vowels (/a/, /ɑ/, /ɪ/, /i/, /ʏ/, and /y/) to their native counterparts. Categorization predictions based on LDA and stepwise models were borne out. Results show that L1/L2 acoustic proximities and perceptual cues influenced AusE and PS listeners’ categorization patterns. To determine the effects of non-native categorization on non-native discrimination, both listener groups took part in a discrimination task presented in the XAB format. To test discrimination accuracy between listeners whose native vowel inventory differed in size relative to the target language, a mixed-effects binary logistic model analysing participants’ correct responses, with participant, speaker and XAB trial as random effects, and vowel contrast and language background as fixed effects was employed. This revealed a main effect
of contrast, no main effect of language background, and an interaction of vowel contrast and language background. Fishers’ LSD-corrected post-hoc pairwise comparisons revealed that PS listeners had a marginally higher accuracy for /ɪ-i/, compared to AusE listeners who were marginally more accurate in their discrimination of /ʏ-ɪ/ than PS.

Results presented in Study 1 suggest that having a larger vowel inventory compared to the target language does not always offer an advantage to listeners when perceiving some non-native sounds, and that it is the L1-L2 proximities which affect listeners’ categorization patterns which in turn are indicative of listeners’ discrimination patterns. For instance, even though AusE contains a lax and tense /i/ vowel, PS listeners were marginally better than AusE listeners in their discrimination of Dutch /ɪ-ɪ/. As predicted and due to the acoustic L1/L2 overlap that exists between the AusE and Dutch categories, AusE listeners categorized this contrast mainly across a single native /ɪ/ counterpart. In their categorization of the same contrast, PS listeners were able to employ the F2 and F3 perceptual cues to classify Dutch /ɪ/ across two native categories namely /i/ and /e/, while Dutch /i/ was classified solely as PS /i/, leading to a higher discrimination performance compared to AusE listeners. Earlier research has established that English listeners are more sensitive to vowel-intrinsic movement than duration when listening to some English vowel contrasts (e.g., /i-ɪ/, /e-ɛ-ɛ/, and /u-ʊ/; (Hillenbrand & Nearey, 1999), and these results establish that this also carries of over into their non-native perceptual patterns.

AusE listeners categorized Dutch vowels across multiple native categories more frequently than PS listeners. AusE listeners’ discrimination accuracy was lower for contrasts involving multiple native categories that had partial or complete acoustic overlap (e.g., /ʏ-ʏ/) compared to those with minimal to no acoustic overlap (e.g., /ɪ-ʏ/, /a-a/ and /ʏ-ɪ/). These findings further support those of previous studies that have established AusE listeners’ lower discrimination accuracy for non-native vowel contrasts that are mapped across multiple
native categories while involving acoustic overlap (Elvin et al., 2014). While PS has a smaller vowel inventory compared to AusE, PS listeners also encountered similar perceptual patterns for Dutch contrasts that were categorized across multiple PS categories. For instance, PS listeners’ discrimination accuracy was lower for Dutch /y-/ due to a higher acoustic overlap across more than one native category compared to Dutch /i-/.

Theoretical models on non-native and L2 speech perception stipulate that two non-native sounds should be easy to differentiate if they are mapped across two distinct native categories. PS listeners categorized Dutch /y-/ as PS /e/ and /i/, which lead to their highest discrimination accuracy across all Dutch vowel contrasts presented. Overall, PS listeners had comparable discrimination accuracy compared to AusE listeners. In particular, these findings are interesting as four of the six presented Dutch vowels are not found in PS. Therefore, results presented in Study 1 further support the notion that the native acoustic vowel space plays an integral part in listeners’ initial perception of non-native sounds.

4.2.2 Study 2: The relationship between non-native speech perception on non-native word learning

Previous studies suggest that, similarly to L2 speech perception, the L1-L2 relationship may affect listeners’ L2 word recognition accuracy. A number of studies have established that sounds which are not present in learners’ native inventories will be difficult to recognize and learn in L2 words (e.g., Broersma, 2002; Pallier et al., 2001; Weber & Cutler, 2004). While very few studies have compared the initial link between non-native speech perception and non-native word learning, none have compared AusE and PS listeners’ vowel discrimination and word learning difficulties. This is an important area in psycholinguistic research as it has been shown that listeners’ perceptual trajectory changes as they gain experience with their L2 (Best & Strange, 1992; Best & Tyler, 2007). To answer the first
research question and establish any effects of vowel inventory size and acoustic properties on both non-native perception and non-native word learning. Study 2 compared listeners with fewer and more vowels than the target language with the aim of determining the relationship between these two abilities. Specifically, AusE listeners’ XAB discrimination data reported by Alispahic et al. (2017) and presented in Study 1 was compared to word learning data reported by Escudero (2015) analysed using the same method. To this end, this is the first study that compared AusE and PS listeners’ non-native discrimination patterns to non-native word recognition of Dutch minimal pair words.

As outlined in section 3.3.2, in the word learning task, AusE and PS monolingual listeners were presented with minimal pair words that were differentiated by the same five Dutch vowel contrasts, /a-ɑ/, /i-ɪ/, /y-ʏ/, /i-ɛ/, and /i-ɨ/, as presented in the XAB task. To determine the relationship between the two abilities and considering that the data reported in Escudero (2015) was originally analysed using a repeated measure of variance (ANOVA), the results were re-analysed using the same statistical model as for the XAB task. A mixed-effects logistic model with participant and picture trial as random effects, vowel contrast and language background as fixed effects and participants’ correct response as dependent variable revealed a main effect of contrast, no main effect of language background, or interaction of vowel contrast and language background. Fishers’ LSD-corrected post-hoc pairwise comparisons revealed that there was trend toward more correct responses by PS listeners compared to AusE listeners for Dutch /i-ɛ/.

Overall, results showed that contrasts which were perceptually difficult to discriminate also posed difficulties in listeners’ word recognition patterns. For instance, mean accuracy percentages showed that AusE listeners found the Dutch /a-ɑ/, /y-ʏ/ and /i-ɪ/ minimal pairs the most challenging to learn. In the XAB discrimination task, these contrasts also ranked as the top three most challenging for AusE listeners to discriminate. Conversely, perceptually
easy contrasts, e.g., /y-ɪ/ and /i-y/, had higher recognition accuracy. Similarly, PS listeners showed low word learning accuracy for perceptually difficult contrasts (e.g., Dutch /a-ɑ/, /i-ɪ/ and /y-ɪ/), while words containing Dutch /i-y/ were easier to recognise as evident by PS listeners’ high percentage accuracy for this contrast across both tasks. However, while PS listeners were able to discriminate the Dutch /y-y/ contrast they found it challenging to differentiate when presented in a word learning task. However, these comparisons were based simply in order of difficulty ranking of mean accuracy percentages. To statistically determine whether or not difficulties were the same for speech perception and word recognition further comparisons of the main effect of contrast reveal that combined overall difficulty ranking from least to most challenging minimal pair vowel contrast was the same across both tasks:

1) /i-ʏ/ >> (2) /y-ʏ/ ~ /i-i/ ~ /a-ɑ/ ~ /ɪ-ʏ/. These findings establish that irrelevant to vowel inventory size both listener groups followed a similar path in non-native vowel perception and non-native word learning. Importantly, the present findings contribute to an emerging area in psycholinguistic research that compares the effects of cross-linguistic similarity on listeners’ initial perception of non-native vowels and how this translates into their non-native word learning. In sum, findings outlined in Study 2 demonstrate that listeners’ initial non-native discrimination difficulties also carry over into their non-native word recognition abilities.

### 4.3 Implications for Theoretical models

Similarly to how listeners employ the sound system of their L1 to initially perceive non-native phonemes the effects of lexical competition are transferred from listeners’ native to non-native word recognition. If a listener is unable to match the acoustic-phonetic variation that exists between activated lexical competitors then this will result in lower word recognition accuracy (TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994; Shortlist-B: Norris & McQueen, 2008). Previous research has indeed shown that increased activation
of L1-L2 competitor sets leads to higher difficulty for lexical entries which in turn corresponds to accuracy percentages varying across perceptual and word recognition tasks (Amengual, 2016; Díaz et al., 2012; Nuria Sebastian-Galles, 2005). The present findings are in line with earlier research and show that acoustic-phonetic similarity has bearing on listeners’ initial sound perception and that once a sound is heard activation of competing lexical items will ensue. Due to lexical processing eliciting higher cognitive demands, compared to phonological processing, the mean accuracy percentages across the perceptual and world learning tasks presented differed.

These findings thus have implications for non-native and L2 speech perception models. Within their framework PAM and PAM-L2 do not propose specific calculations that can measure and/or predict how these two abilities interact. While the SLM model does not make predictions on learners’ spoken word recognition and word learning patterns if the model’s predictions extend to word learning, learners should more easily learn words containing more dissimilar than similar phonemes. However, the SLM predictions are not borne out for either L2 perception or L2 word recognition of the present thesis. Out of the present models, L2LP is the only model that explicitly links listeners’ speech perception and word recognition abilities. L2LP posits that non-native and L2 vowel discrimination and acquisition is predicted by L1-L2 acoustic relationships, and that non-native perceptual abilities should translate to non-native and L2 word learning abilities (L2LP; Escudero, 2005; 2006; 2009; van Leussen, 2015).

The present findings support one of L2LP’s main claims that listeners will initially perceive non-native sounds in line with the acoustic properties of their native language and that difficulties encountered in perception will further transpire into their non-native word learning. The results of the present thesis support previous findings that suggest that adult listeners are sensitive to and employ spectral and temporal perceptual cues in their perception
of foreign language vowels (Lengeris, 2009), and that these vowel distinctions affect their non-native word learning abilities as has previously been shown in the case of infants (Curtin et al., 2009). For example, infants as young as 15-months are able to learn words differentiated by the /i-ɪ/ contrast because they exploit the F1 differences that exist between these two vowels. That is, as outlined in section 2.2, close attention is paid to the most salient acoustic cue of a particular sound and this salience of acoustic information informs not only phonemic but also lexical learning. Therefore, in line with the present findings it is proposed that future studies should consider the theoretical development of current models that do not directly test how non-native and L2 speech perception affects non-native and L2 word learning. As a result, these studies and theoretical developments may further inform on the relationship between perceptual difficulties, spoken word recognition and production of non-native L2 speech.

4.4 Future Directions

The present thesis endeavoured to establish the interrelation of non-native speech perception and non-native word learning by Australian English and Peruvian Spanish listeners. Specifically, it demonstrated that acoustic relationships between listeners’ native language and the target language have bearing on their non-native categorization patterns and that these in turn are indicative of their discrimination patterns. It also established that vowel contrasts that are perceptually difficult to discriminate are also challenging to acquire when presented in novel minimal pair words. However, a number of future directions should be undertaken to build upon and strengthen the current findings.

4.4.1 Participants

Findings presented in Study 1 tested the perception of Dutch vowels by naïve AusE and their results were compared to those of PS listeners reported in Escudero and Wanrooij
Participants’ non-native XAB perceptual patterns from Study 1 were then compared to non-native word recognition patterns of listeners reported in Escudero (2015), as outlined in Study 2. As a result, future research (e.g., in the form of a PhD) should endeavour to have the same participants take part in both the perceptual and word learning tasks. This is an important future avenue as recent findings show that, while analysing overall group effects, it is also essential to consider individual listeners’ perceptual and word learning patterns (Elvin, 2016). Moreover, having the same participants take part in both the non-native perceptual and non-native word-learning task will allow for a direct statistical comparison between the listener groups that was unable to be included in the statistical models of the present thesis.

The current findings offer an insight into the initial perceptual patterns of Dutch by monolingual AusE and PS listeners. However, previous studies have also shown that perception shifts once a listener has been exposed to their L2. Future studies should also pursue investigating listeners’ perceptual development from their initial non-native stage to one where they have been exposed to or actively learning their target language for a period of time. These developmental comparisons will not only offer greater insight into language specific development but can also further investigate the learning scenarios and stages put forth by the L2LP model.

4.4.2 Stimuli

In Study 2, listeners were presented with non-words. In ongoing research it would be interesting to test non-native word learning patterns of real words in an unknown language so as to more closely approximate real L2 learning contexts. A further possibility is to then compare the initial perceptual patterns to those of listeners that have been exposed to the unknown language for a period of time. Moreover, just as vowel inventories differ across
languages so do consonant inventories. Thus, future research may also include stimuli that, in addition to vowels, also investigates the effects of consonants on listeners’ perceptual patterns. This would allow for a direct comparison of non-native vowel and consonantal perception and may inform future research on how L2 perceptual trajectories change once a non-native listener advances through their L2 development.

4.4.3 Tasks

Recall that in the categorization task presented in Study 1, AusE listeners were given more response options compared to PS listeners due to a larger AusE vowel inventory. That is, the number of response options were controlled to match the native vowel inventory size of each group. However, previous research has documented that vowel categorization is affected by the number of mental representations available to a listener and that categorization is affected by the amount of response categories offered to listeners (e.g., Benders et al., 2012; Elvin et al., 2014). As a result, future research should seek to include an equal number of response options across all listener groups. For instance, if listeners are given fewer response options than those reported in the current thesis, these results could then be compared to the present findings that suggest listeners’ perceptual patterns being reflective of acoustic overlap between the number of referents available to them. Moreover, this will further test how and to what degree the number of response options available affect listeners’ perceptual categorization results.

4.5 Practical Implications

In today’s global climate, foreign language learning not only facilitates economic cross-border mobility, cultural exchange and easier integration of migrants, but also offers cognitive and long term health benefits to those who speak more than one language. For instance, children’s concept formation, classification, creativity, analogical reasoning, and
visual-spatial skills have all been positively related to bilingualism in comparison to monolingualism (Bialystok, 1991). Adult bilingual listeners are shown to have stronger motor executive functions, perform better on conflict management tasks and encode the fundamental frequency of sounds better in noisy environments compared to monolingual listeners (Abutalebi et al., 2012; Bialystok, Craik, & Luk, 2012; Krizman, Marian, Shook, Skoe, & Kraus, 2012). These bilingual benefits also further extend to age-related cognitive concerns such as Alzheimer’s and memory function in older adults. That is, apart from exhibiting better episodic memory older bilingual adults report the onset of Alzheimer symptoms 5.1 years later than their monolinguals peers (Craik, Bialystok, & Freedman, 2010; Schroeder & Marian, 2012).

However, while more than half the world’s population is bilingual and many European, Asian and African countries have more than one official language, learning a foreign language is not on the educational forefront when it comes to English speaking countries, specifically Australia. For instance, learning a language other than English (LOTE) is at an all-time historical low across Australian secondary schools (EducationHQ Australia, 2016). Out of 77,433 enrolled students that completed their Higher School Certificate in NSW during 2015 only about 6% studied a second language (Board of Studies NSW, 2016). This number is even lower for students that study a novel foreign language irrelevant of their cultural background. For example, with approximately 1.2 billion speakers Chinese is currently the most used language in the world (Ethnologue, 2017), yet out of the enrolled 935 students studying Chinese for their Higher School Certificate only 153 were not native speakers (Board of Studies NSW, 2016). These stark numbers also extend to Australian Universities which has seen a decline of language programmes being offered. But, why is that?
It has been well documented that learning to successfully master a new language can be a very stressful experience for many people. Oftentimes, it can take years of rigorous classroom repetition of sounds, words and sentences before a person is confident enough to verbally communicate in their second language. Literature in language teaching suggests that apart from concentrating on teaching students knowledge and skills it is important to take into consideration that language learning is an individual process that cannot be covered by a ‘one size fits all’ approach (Abhakorn, 2008; Ehrman, Leaver, & Oxford, 2003). However, the way foreign languages are taught in many classrooms around the world is precisely through a universal approach. That is, teaching and learning-method ideas have generally been independently developed from second language acquisition research (Cook, 2013). This is not good, as these ideas do not take into consideration how, for example, second language perception and acquisition research has, for a number of years, demonstrated that a listener’s L1 has bearing on their L2 perception, learning and development. Therefore, information on L2 language learning based on scientific research is crucial in order to help shape and tailor L2 language teaching programmes.

The findings of the present thesis show that, for an AusE monolingual learning Dutch words it would be beneficial to know that, initially, the student will find discriminating the /i/-/ɪ/ Dutch vowels challenging. This discrimination difficulty at the phoneme level will then further affect their recognition and learning of words that are differentiated by those sounds. In this view, if a student is partaking in a spelling dictation task and is asked to transcribe words they are presented with, it is highly likely that they will be unable to correctly identify and transcribe words that are differentiate by those sounds. As a result, this may lower their grades and discourage the student from wanting to continue learning the language. Therefore, the practical implications of the present thesis are that it sheds light on how listeners with specific language backgrounds hear novel sounds and words. Specifically, it exhibits the
interrelation of sound perception and word recognition at the initial stages of learning and shows that perceptual difficulties in L2 sound perception translate to L2 word learning difficulties. The present findings may benefit future language teaching techniques and help shape tailored language teaching programmes that factor in listeners’ native language backgrounds. If research is able to exemplify exactly what learning stages a student will go through when learning an L2 then this may aid teachers to develop pathways (i.e., tasks) that will make language learning a more proficient and enjoyable experience for everyone involved.

4.6 Final Conclusion

The present thesis tested the initial interrelation of non-native speech perception and non-native word learning by Australian English and Peruvian Spanish listeners. Along with a literature overview, it presented two experiments that established how and to what degree acoustic cross-linguistic similarity compared to vowel inventory size, affects listeners’ perception of non-native vowel contrasts. It then compared these results to listeners’ word learning of minimal pairs that were differentiated by the same vowel contrasts. The findings presented establish that acoustic relationships between listeners’ native language and the target language have a bearing on their non-native categorization patterns and that these in turn are indicative of their discrimination patterns. Importantly, the present thesis established that perceptually difficulties encountered in vowel discrimination translate to novel minimal pair learning. As the present findings show real word implications, future directions should aim to investigate listeners’ perceptual development from their initial non-native stage to one where they have been exposed to or actively learning their target language for a period of time.
5 REFERENCES


6 APPENDICES

6.1 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.

<table>
<thead>
<tr>
<th>Country</th>
<th>City or Region</th>
<th>Dates lived there?</th>
<th>How old were you?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Milperra</td>
<td>5/03 - 8/05</td>
<td>birth - present</td>
</tr>
</tbody>
</table>

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

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

6. What is your native language? ___________________________

7. What is your mother’s native language? ___________________________
8. What is your father’s native language? __________________________

9. Do your parents generally speak with a particular regional accent of Australian English?
   
   Yes / No
   
   If so, which accent? __________________________

10. 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 language? __________________________

11. 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)

<table>
<thead>
<tr>
<th>Language</th>
<th>Years Used</th>
<th>Where did you use this language?</th>
<th>How well do you speak it?</th>
<th>How well do you understand it?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>2006-2008</td>
<td>high school</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

12. Please provide your e-mail address if you are willing to be contacted for participation in future experiments related to this study.
6.2 Appendix B: Additional Tables

Additional tables that include participants’ mean accuracy scores, standard deviations and difficulty rankings for both the XAB and word learning task are set out below.

**Table A 1.** *AusE and PS participants’ mean accuracy scores, standard deviations and difficulty rankings for the XAB task.*

<table>
<thead>
<tr>
<th>Dutch vowel contrast</th>
<th>XAB Discrimination Task</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/a-ɑ/</td>
<td>64.8%</td>
<td>10.1</td>
<td>3</td>
<td></td>
<td>1</td>
<td>63.9%</td>
<td>9.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɪ-ɨ/</td>
<td>67.9%</td>
<td>6.8</td>
<td>1</td>
<td></td>
<td>2</td>
<td>70.5%</td>
<td>10.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-ʏ/</td>
<td>69.4%</td>
<td>10.7</td>
<td>2</td>
<td></td>
<td>3</td>
<td>73.8%</td>
<td>9.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-ɨ/</td>
<td>69.8%</td>
<td>7.2</td>
<td>3</td>
<td></td>
<td>1</td>
<td>64.9%</td>
<td>7.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɨ-ʏ/</td>
<td>78.4%</td>
<td>10.4</td>
<td>4</td>
<td></td>
<td>3</td>
<td>73.4%</td>
<td>8.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table A 2.** *AusE and Spanish participants’ mean accuracy scores, standard deviations and difficulty rankings for the word learning task.*

<table>
<thead>
<tr>
<th>Dutch vowel contrast</th>
<th>Word Learning Task</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/a-ɑ/</td>
<td>41.7%</td>
<td>28.4</td>
<td>1</td>
<td></td>
<td>1</td>
<td>61.7%</td>
<td>28.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɪ-ɨ/</td>
<td>56.9%</td>
<td>20.7</td>
<td>1</td>
<td></td>
<td>1</td>
<td>55.0%</td>
<td>25.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-ʏ/</td>
<td>48.6%</td>
<td>23.4</td>
<td>1</td>
<td></td>
<td>1</td>
<td>51.7%</td>
<td>24.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/y-ɨ/</td>
<td>63.9%</td>
<td>27.4</td>
<td>2</td>
<td></td>
<td>1</td>
<td>58.3%</td>
<td>29.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ɨ-ʏ/</td>
<td>70.8%</td>
<td>24.6</td>
<td>2</td>
<td></td>
<td>2</td>
<td>76.7%</td>
<td>30.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>