The Nature of Phonological Representations in a Second Language

by

Erwin La Cruz

A Thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy

Victoria University of Wellington
Te Whare Wānanga o Te Ūpoko o Te Ika a Māui
2015

ήδη θα το κατάλαβες οι Ιθάκες τι σημαίνουν

you will have understood by then what these Ithakas mean

– C.P. Kavafis

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr Paul Warren, for his guidance, assistance and endless patience throughout this process. I would also like to thank my second supervisor, Dr Averil Coxhead, for her constant encouragement and support.

This study would have not been possible without the help of the participants who volunteered to take part in the experiments I carried out in New Zealand and Venezuela.

I am forever indebted to my professors, colleagues and students from my *Alma Mater*, the Universidad de Los Andes, Mérida, Venezuela. In particular, I want to thank my colleagues Dr Medina and Dr Plata for their help in recruiting participants. I would like to acknowledge Prof. Goilo's unconditional support and guidance during all this years.

A big thanks to Myq Larson for the stimulating discussions and for coming to the rescue when my code would not run.

To all of you, ¡que anchos caminos topares!

ABSTRACT

This study explores the nature of phonological representations in a second language (L2). In particular, it explores whether abstract and exemplar phonological representations are available to adult learners of an L2. To do this, the study looks at the ability learners have to perceive and store fine phonetic detail in an L2, and their ability to generalise perceptual learning. Particular attention is paid to the effect of speaker variation in the perception of phonological categories that are difficult to learn in an L2. The population studied is Spanish native speakers who are learning English as a foreign language.

To investigate L2 learners' ability to store acoustically rich representations in their L2, three experiments were conducted. The results from these experiments indicate that L2 learners can store phonetic detail which facilitates lexical access of words previously experienced in the same voice. However, this ability does not guarantee good discrimination of L2 phonological contrasts. On the contrary, acoustic variation due to different speakers can make the perception of non-native contrasts more difficult. Despite this difficulty, this study found that L2 learners can generalise to other voices and lexical items, which suggests that L2 learners can abstract the knowledge gained from exposure to a specific set of exemplars.

This study contributes to the ongoing discussion about the nature of phonological representations. The difficulty of learning new phonological categories in an L2 has been previously presented as evidence of the abstract nature of phonological representations. Nevertheless, the results of this study show that L2 learners can preserve some of the acoustic characteristics of words

they have experienced, which indicates that, for L2 learners, phonological representations cannot be purely abstract. Considering that abstract representations are also available to L2 learners, the results of the thesis support the development of hybrid models of L2 phonology learning.

1 INTRODUCTION

1.1 WHY STUDY PHONOLOGICAL REPRESENTATIONS IN AN L2?

As a learner and a teacher of English as a second language (L2), I have used several pronunciation teaching methods. Some of these methods compared the English and Spanish phonological systems in detail; some used articulatory descriptions of how speech sounds are made, or proposed the use of speech spectrography. Others relied on spelling regularities of the English language or made used of phonetic transcription systems, while others forbade the use of any kind of writing. Although these methods were well structured and the materials well designed, I consider that they failed to address a fundamental aspect of the learning of phonology: the format in which phonological representations are stored. Without this knowledge, it is difficult for teachers to predict how they can best influence the learning process. Understanding how learners of a second language process the speech signal and determining how much phonetic detail they are capable of remembering is a necessary step in creating better teaching methods and materials.

The purpose of this thesis is to better understand what information is stored in the phonological representations and how these develop in the case of a second language. This thesis examines how integrating abstract and exemplar approaches to phonological representations can explain the process of learning new phonological categories in an L2.

Despite the evidence coming from research in the native language (henceforth L1) supporting exemplar models of speech perception (Goldinger, 1998; Hay, Warren, & Drager, 2006; Mullennix, Pisoni, & Martin, 1989; Wedel, 2007), models of L2 phonology continue to propose that phonological representations are abstract (e.g., Best & Tyler, 2007; Escudero, 2005). I argue that such an approach is limiting. The process of learning the phonology of an L2 requires both abstract and episodic phonological representations. As in the L1 (Kuhl et al., 2008), abstract representations are the result of a consolidation process that starts with episodic memories.

In this thesis, I look for evidence that both types of representations are available to L2 learners. I explore whether phonological representations in an L2 are based on phonetically rich episodic memories of the words learners experience. Furthermore, I also explore whether L2 learners are able to generalise perceptual learning acquired with a specific voice and set of words to other voices and lexical items,.

1.2 STRUCTURE OF THE THESIS

Chapter 2 presents an overview of the relevant literature on the general format of phonological representations in the L1 and the L2. The tenets of abstractionist and exemplar approaches are discussed, as well as the limitations of such models in explaining L2 phonology. This chapter concludes with the research questions for the present study.

Chapter 3 focuses on methodological aspects of the experiments carried out in this thesis. First, it presents a review of how voice and word priming have been previously used to determine whether phonological representations contain fine phonetic detail. Second, a review of perceptual assimilation of English phonemes by native speakers of Spanish is presented. This chapter closes with a discussion of the statistical tools used in the analysis of the data collected in the experiments.

Chapter 4 discusses Experiment 1, a mid-term memory task that tested the ability of L1 Spanish learners of English to identify minimal pairs in a second language. This experiment was an exploration of how the acoustic variation resulting from different voices might affect the learners' ability to map acoustic cues onto English phonological categories. Sensitivity to variations in formant structure and in duration were studied. The first type of sensitivity was studied through the use of stimuli contrasting in the vowels FLEECE vs. KIT, a vowel contrast difficult for Spanish speakers[1]. Since duration is not a strong cue for vowel identification in American English (Hillenbrand, Clark, & Houde, 2000), sensitivity to durational cues was evaluated using minimal pairs contrasting between voiced and voiceless plosives.

Chapter 5 presents Experiment 2, in which I investigated whether the difficulty in discriminating between the members of minimal pairs was valid evidence that L2 phonological representations was based on episodic memories. This experiment consisted of a lexical decision task combined with a long-lag repetition, and it looked at the combined effects of lexical priming and voice priming when listening to minimal pairs in an L2. The experiment contrasted the performance of English learners and native English speakers.

Chapter 6 reports Experiment 3. This experiment focused on the ability L2 learners have to generalise perceptual learning to other speakers and to other lexical entries. After training a group of learners of English, I tested their ability to generalise the perceptual knowledge acquired in the training to other voices and lexical items. In this experiment, mouse tracking methodology was used as a tool to analyse the activation of competing lexical entries before a response was given.

In Chapter 7, I present a general discussion of the results obtained in the three experiments. I explain how these results can help better understand the role of episodic memory in L2 phonology learning. I offer some pedagogical considerations to those teaching pronunciation and oral comprehension in an

1 I use Wells's (1982) standard lexical sets to refer to English vowel phonemes due to the large dialectal variation on English vowels and the lack of unified criteria to transcribe them.

L2. Finally, I conclude by discussing the limitations of the present study and proposing future research.

2

EXEMPLAR AND ABSTRACT REPRESENTATIONS IN THE PHONOLOGY OF A SECOND LANGUAGE

Understanding our mother tongue is, for most of us, a natural, effortless activity. We seldom realise that being able to understand up to 230 words per minute in a continuous flow in a noisy environment is an amazing cognitive feat. However, the assumption that speech perception is "effortless" is promptly contradicted when we listen to a foreign language. In order to understand a foreign language, we must set about the task of associating sound patterns contained in the speech signal to the concepts we store in our memory. In broad terms, the information we store in our memories about how a word sounds is its phonological representation.

One of the main questions speech perception theories have to answer is how phonetic detail contained in an ever-changing physical input (sound waves) is mapped onto a lexical entry. More specifically, do we, as proposed by abstract models, carry out a normalisation process of the acoustic signal after which only those aspects that are linguistically relevant are stored and all other details are ignored? Or, as proposed by exemplar models, do we store an episodic trace for each encountered word which preserves detailed acoustic features which encode information such as the sex, age or dialect of speakers?

The present study focuses on the nature of phonological representation in L2. In particular it aims to explore the roles played by abstract and episodic representations in the learning and processing of a second language. This study tries to demonstrate that these two types of representation are not mutually exclusive but that, on the contrary, they complement each other. I propose that episodic representations are needed to learn new phonological categories in the L2, and that as language competence increases, L2 representations become abstract. I will determine the role each type of representation has in the learning of the phonology of an L2. My interest in looking at these issues comes from research in speech perception in L1 and L2 that suggests that both types of representations may be needed to explain how a second language is learnt and processed.

I argue against the postulates of extreme abstract models (e.g., Liberman & Mattingly, 1985) which propose that phonological representations do not retain fine phonetic detail. On the contrary, a model of L2 phonology learning should include episodic representations of words from the target language for the following reasons. Firstly, research has found that listeners can in fact store phonetic detail of particular speakers (e.g., Goldinger, 1996; Hay et al., 2006), and that speakers can even adjust their perception for a particular speaker (Kraljic and Samuel, 2007). Moreover, this capacity to store phonetic detail is necessary for the learning of new phonological categories. That is, in the process of learning an L2 category, an episodic memory of an L2 word –rich in phonetic detail and other indexical information– can be retrieved and compared to new examples of that word produced by other speakers. Without the ability to store fine phonetic detail, learners would be forced to start from zero every time they encountered a word. Secondly, the ability to store episodic memories is necessary not only when listening to others, but it has also been found to be important in self-directed speech during overt rehearsal of the L2. For example, Centeno-Cortés (2003) found that learners of Spanish as a L2 privately rehearse the pronunciation of Spanish words. During this practice, the learners imitated the pronunciation of their teacher. Private repetition can be used not only immediately after an exemplar has been experienced but also later. I suggest that, for this kind of practice, storing phonetic detail is necessary.

However, proposing that phonological representations are solely based on episodic traces is problematic. For example, purely exemplar models (e.g., Bybee, 2001; Goldinger, 1998) argue that the speech signal can be directly mapped onto the lexical representations, and consequently sublexical representations are not needed. In my view, however, abstract phonological representations should also be included in a model of L2 phonology for the following reasons. Firstly, there is extensive evidence that listeners process the speech signal based on sublexical units (e.g, Dupoux, Pallier, Kakehi & Jacques Mehler, 2001; Eulitz & Lahiri, 2004; Nearey, 2001) or that abstract representations are required for proper learning of variants (Peperkamp, Le Calvez, Nadal & Dupoux, 2006).

Secondly, as a result of formal instruction, adults are capable of using abstract symbolic systems such as spelling and mathematical notation. It can be expected that this knowledge will be applied when learning an L2. In fact, phonemic awareness arising from literacy has been shown to have an effect when forming phonological representations in an L1 (e.g., Morais, 2010; Pattamadilok, Morais, De Vylde, Ventura, & Kolinsky, 2009). Thirdly, speech production also seems to require representations to be abstract and discrete. An important characteristic of most speech production models (e.g., Dell, 1986; e.g., Levelt, Roelofs, & Meyer, 1999) is that "their nodes represent whole linguistic units, such as semantic features, syllables or phonological segments. Hence, they are all 'symbolic' models" (Levelt et al., 1999, p. 223). That is, speech production models claim that the phonological information contained in the lexical entries, at least as far as production is concerned, is segmental and abstract.

Finally, it has been found that perceptual learning (i.e., adaptation to a particular speaker) operates at segmental level. Listeners rapidly learn to modify the boundaries of a phonemic category to adapt to deviant pronunciation of a sound by a particular speaker (e.g., Clarke-Davidson, Luce, & Sawusch, 2008; Eisner & McQueen, 2006; Jong, Silbert, & Park, 2009; Norris, McQueen, & Cutler; Kraljic & Samuel, 2007 in L1, and Weber, Boersma, & Aoyagi, 2011, in L2). This perceptual learning can be applied to words containing a deviant sound which has not been heard previously (Eisner & McQueen, 2006; McQueen, Cutler & Norris, 2006), and even in different phonotactic contexts (Jesse & McQueen, 2007). This ability to generalise to other lexical entries indicates that phonological representations contain abstract units.

In this thesis, I propose that the learning of the phonology of a second language requires both abstract and episodic representations. As experience in the L2 increases, the accumulation of L2 exemplars leads to the formation of prototypes that can act as abstract representations of the categories they represent.

2.1 PHONOLOGICAL REPRESENTATIONS IN SPEECH PROCESSING

In this section, I will discuss some of the most influential theories and models of speech processing in the L1 and L2. I will focus in particular on the type of representations these models propose and on some of the limitations of the models.

In order to understand spoken language, one of the tasks presented to the listener is that of mapping the acoustic signal onto a word stored in memory. During this task, it is generally assumed that the listener classifies pieces of the acoustic signal as being representative of a linguistic category. However, a major problem faced in speech perception is the variable nature of speech. A word is never pronounced in the same way even by the same speaker. Factors such as the position in the sentence and number of words in the sentence affect the pitch and duration of words. Likewise, different speakers pronounce words differently, depending on their sex, age, regional or social class. But even parts of an utterance smaller than the word are never the same. For example, plosive consonant /p/ in English is usually aspirated when it precedes a vowel of a stressed syllable, but unaspirated if the syllable starts with /s/, or unreleased in front of another plosive. Moreover, the relative frequency and duration of this consonant will vary according to idiolectal characteristics of each speaker. Despite the lack of invariability in the acoustic signal, listeners are very good at recognising and identifying words. The explanations that have been put forward to explain how listeners achieve this task can be broadly divided into two groups: those that claim that only some information is extracted from the speech signal and consequently only a small part of the information stored in memory, and those that claim that a lot of information is extracted and stored in long term memory.

There is a current debate in the theories of speech perception and spoken word recognition about how much and in what format is phonological information stored. On the one hand, abstract models of representation argue that during word recognition the listener maps aspects of the acoustic signal onto abstract representations, and the rest of the acoustic details are not stored. On the other hand, exemplar models of speech perception claim that people do not represent the speech signal using abstract representations but instead they store detailed memories of the acoustic signal. In other words, the main difference between these approaches is whether they posit that the speech perception mechanism accesses fine phonetic detail.

As mentioned earlier, abstract models claim that in order to deal with speech variability, the speech perception routines map the speech signal onto a series of increasingly abstract representations. For example, in the TRACE Model (McClelland, & Elman, 1986) the acoustic signal is mapped onto phonetic features, these are next mapped onto phonemes, which finally activate words. In other words, the architecture of abstract models posit a level of pre-lexical abstraction. This pre-lexical level contains a limited number of segmental categories. Usually the unit of representation in the pre-lexical level is the phoneme. Moreover, the phoneme is seen as a symbolic unit in that it bears not resemblance to the acoustic signal. For example, Escudero (2005) states that a phoneme is "a completely arbitrary label for a feature combination [...] which has not relation with the auditory world." (p. 48).

Contrasting with this view is the exemplar view. Authors following this theory propose that representations are based on veridical memory traces of the experienced stimuli. The categorisation processed is based on comparing the incoming signal to memory traces of similar stimuli previously experienced. However, exemplar models vary on the levels of representation they allow for. Some authors propose a direct mapping of the input signal onto the word (e.g., Goldinger, 1996, and Johnson, 1998), while others include a pre-lexical, abstract representation level (i.e. the phoneme) in their models. Despite this important difference in the architectures of these models, some authors refer to both these models as exemplar. As I will discuss later, a model that includes a pre-lexical abstract level of representation is in effect a hybrid model. This view is presented by Pierrehumbert (2006) as follows:

> When we consider simultaneously the successes of Varbrul and the strong points of exemplar theory, it is clear that a hybrid model is needed. Hybrid models in which a phonological coding level intervenes between the lexicon and the parametric phonetic description [...] (p. 523)

Due to the differences discussed above, in the following sections, I will make the distinction between purely exemplar and hybrid models, the latter being those that propose a pre-lexical level of representation along with veridical traces of speech input.

In the following sections, I will first look at abstract models and consider the evidence that supports them.

2.1.1 *Abstract models of phonological representation in L1*

In order to explain how listeners can cope with the lack of invariability in the speech signal, abstractionist models (e.g., Boersma, 1998; Browman & Goldstein, 1995, Golstein & Fowler, 2003), postulate that a normalisation process is carried out during speech processing. These models propose that the speech processing system extracts only linguistically relevant features from the acoustic signal and matches them to the abstract phonological representation of words stored in long term memory. A consequence of this normalisation process is that phonetic details related to speaker voice are not stored at the lexical level.

An advantage of these models is that they reduce variation to a finite number of types that can be used to represent an unlimited number of lexical forms. For example, in an abstract model proposing the phoneme as the unit of phonological representation, the words "tip", "tap" and "top", despite the differences in fine phonetic detail caused by coarticulation, share the same first and last phonemes in their phonological representations. Furthermore, a finite number of types can be combined to represent larger language constituents. For example, phonemes are combined into syllables, and syllables are combined into words.

Although abstract models propose phonological representations that are discrete, finite and symbolic, there is no consensus as to what is primitive unit of perception. Different models propose different units, which vary in the degree of abstractness and size, such as the phoneme (e.g., Decoene, 1993; Norris & Cutler, 1988), the syllable (e.g., Mehler et al., 1981; Massaro, 1972;

O'Seaghdha, Chen, & Chen, 2010), demisyllables (e.g. Samuel, 1989) or the phonetic feature (Marslen-Wilson & Warren, 1994).

For the present discussion I will first review models that employ abstract representations based on articulatory features and then those that assume abstract representations based on acoustic features.

Since the phonetic information present in the speech signal is not stable and highly variable, several authors have proposed that articulatory actions offer a invariable unit of perception. As Goldstein and Fowler (2003) propose that:

> [it] is, in fact, possible to decompose vocal tract action during speech production into discrete, re-combinable units. The central idea is that while the articulatory and acoustic products of speech production actions are continuous and context-dependent, the actions themselves that engage the vocal tract and regulate the motions of its articulators are discrete and context-independent. (p. 161)

An example of an abstract model is Liberman and Mattingly's (1985) Motor theory of speech perception. The model proposes that the task of the listener during speech perception is to identify the articulatory gestures involved in the production of a given acoustic signal. For example, the initial sounds of the words "bet" and "get" are perceived as being different because they are produced using different articulatory gestures. On the other hand, the initial sounds of "bet" and "boot" –despite their differences in acoustic cues– are perceived as the same sound as they are produced using the same articulatory gestures. The main tenet of this model is that phonological representations contain the intended articulatory gestures (i.e., neuro-motor commands sent to the speech organs) involved in the pronunciation of a word.

However, not all models that propose that vocal tract actions are the perceptual primitives for speech perception assume that the intended gestures (i.e., mapping onto the listener's own speech motor system) are the object of percept. Instead, other authors (e.g., R. A. Bates, Ostendorf, & Wright, 2007; Best, 1984; Fowler, 1986; Goldstein, & Fowler, 2003) argue that perception of gestures occurs because the acoustic speech signal offers the listener information for its source, that is the object or event that created the sound. This approach has been implemented in computational models that map acoustic parameters into discrete, binary articulatory features that are bundled into phonemes (R. A. Bates, Osterndof, & Wright, 2007; K. N. Stevens, 2002). These gestures are mapped onto the lexical representations, which are specified as segments containing articulatory features. A common characteristic of articulatory specified models is that the phonological representations or invariables they postulate are not acoustic in nature.

Models that assume acoustic representations (e.g., Boersma, 1998; Hillenbrand, Clark, & Nearey, 2001; Escudero & Boersma, 2004; Escudero, 2005; Griffen, 1976) claim that the phonological representations of a word are based on spectral information extracted from the acoustic signal. This information is mapped onto the acoustically-specified segmental units which form the representations of words. Although they are in some respects more concrete than articulatory based representations (as to being less removed from the input sound wave), these representations based on acoustic properties can nevertheless also be considered to be abstract. This is because normalisation of the speech signal is usually implemented, a reduced number of prelexical units is established (proposed to be invariant), and these prelexical units are generally taken to consist of a set of discrete features. Finally, these phonological representations can be taken as symbolic representations since in these models the acoustic signal is not directly stored in the lexical entry.

2.1.2 *Exemplar models of phonology in L1*

The assumptions made by abstract symbolic models dominated research during the second half of the past century. However, the view of abstract systems has recently been challenged (e.g., Goldinger, 1996; Johnson, 2006; Pierrehumbert, Tamati, & Clopper, 2006; Pierrehumbert, 2001b). For example, Port and Leary (2005) claim that the fundamental misconception of the generative paradigm was to assume that language is a discrete and formal system. This led to the view held in these models of "a phonetic space that is closed and contains only static symbolic objects" (p. 982).

In contrast to the paucity of information in phonological representations proposed by abstract models, exemplar models of speech perception (e.g., Goldinger, 1996; Johnson, 2006; Pierrehumbert, 2001b, Välimaa-Blum, 2009) do not consider that during speech perception the input is stripped of all contextual information and decoded into abstract sub-lexical units. Exemplar models propose that much of the phonetic detail is stored in long term memory. These models do not propose that abstractions are formed from the invariant properties of phonemes and words. Instead phonological representations are formed by an accumulation of veridical stored exemplars. These exemplars contained highly detailed acoustic and indexical information, such as gender, social or regional variations.

An important aspect of the exemplar framework is that phonological categories are not given a priori. The tokens of a word which a person encounters are stored in long term memory and eventually categories are formed on a statistical basis, that is, using the accumulation of similarities among the exemplars. The lexicon is not seen to contain just one canonical representation of a word, but a collection of memory traces against which words present in the speech signal can be compared for classification.

An advantage of such models is their intrinsic capacity to deal with variation, such as allophonic, stylistic and social differences observed in words, as well as their capacity to store contextual information about a word. For example, Hay, Warren and Drager (2006) found that social factors can be indexed against lexical items. Likewise, research in the exemplar framework has found that during speech perception people can actually remember paralinguistic information (i.e., speaker specific voice characteristics) when retrieving a word from the lexicon. In a seminal study, Goldinger (1996) showed that participants are faster at recognising a word when it is heard in the same voice that they have previously encountered. Similarly, Nygaard, Sommers and Pisoni (1994) found that word recognition was more accurate when words were presented in a familiar voice. These findings provide evidence that the representations of words are far richer than previously thought.

2.1.3 *Limitations of exemplar approaches*

Although exemplar models can account for how the perception system deals with the lack of invariability in the speech signal, there are several aspects that are not fully explained in these approaches.

One of the challenges they face is establishing the units used for representation. According to episodic models (e.g., Goldinger, 1998; Hawkins, 2003; Johnson, 2006; Klatt, 1979, 1989; Pierrehumbert, 2002), all the stimuli a listener experiences are labelled and stored in category clouds. Categories are formed by overlapping of common characteristics of the tokens stored. The prototype of a category emerges from the area that contains the greatest number of exemplars (Pierrehumbert, 2002). As discussed above some authors in the exemplar field have argued against the phoneme as a valid level of representation. On this respect, Välimaa-Blum (2009) argues that exemplars can only be stored if connected to a meaningful unit, that is the word, and dismisses the possibility of the phoneme. She argues that proposing that the listener creates and updates exemplar clouds of speech sounds is a circular argument. She reasons that as in order to understand an word, the listener must segment a word correctly, but for segmentation to be correct the meaning of the word has be already available. Incidentally, Välimaa-Blum's claim that the word should be the basic unit begs the question of how can the listener establish the boundaries of word episodes in the flow of the speech signal. This difficulty in establishing the word a priori was address by Goldinger (1998). However, other exemplar author have included the phoneme in the architecture of their models (e.g., Hay and Baayen, 2005; Johnson, 2006; Klatt, 1979, 1989).

Moreover, despite the evidence proving that phonetic detail is stored in long term memory and that listeners have the ability to store indexical information about the voice in which they hear a word, there is also compelling evidence that abstract representations are also needed during speech perception.

A major difficulty for exemplar models is catastrophic interference (Ellis, & Humphreys, 1999), that is the loss of knowledge previously acquired due to the learning of new information. In phonemic terms it means the inability of a system to recognise a phonological category due to variation present in new training stimuli. Catastrophic interference is common in models that implement distributed representations, such as are those implemented by purely exemplar models. This occurs because new learning uses the same representational resource (cloud labels), consequently overwriting the previous knowledge. In fact, computational implementations of exemplar models (Cutler, Eisner, McQueen, & Norris, 2010; Goldinger, 2007) have showed catastrophic interference.

Due to catastrophic learning, purely exemplar approaches cannot adequately model perceptual learning. This is an important aspect of speech perception, which refers to the ability listeners have to adapt to different dialects or idiosyncratic pronunciations. Exemplar models propose that adaptation to different dialects or peculiar pronunciations is based on establishing exemplars for that particular dialect or speaker. However, several studies have established (Mitterer & Reinisch, 2013; Norris et al., 2003) that perceptual learning is not limited to the words used in training nor the specific word position a speech sound was presented during training. In fact, these studies have found that listeners could generalise from the few stimuli used in training to the rest of the lexicon, and to different syllable positions. Lexical generalisation means that the perceptual learning must affect representational units smaller than the word. These units must be discrete and abstract to explain how perceptual learning affects phonemes in different words across the lexicon. As Cutler et al. (2010) demonstrated, pure exemplar models cannot predict lexical generalisation, as they lack abstract representations units at the lexical level. To my knowledge, computational implementations trying to demonstrate phonological generalisation have failed. In fact, such attempts have finally recognised the need for an abstract level (e.g. Goldinger, 2007; McLennan & Luce,2005).

Nevertheless, Proponents of exemplar models (e.g. Hay and Baayen, 2005; Hay et al., 2006; Hawkins, 2010) argue that these models can also generalise. For example, Hay et al (2006) propose that generalisation is based on a stochastic comparison of the input to stored exemplars. This is in line with other stochastic models (e.g., Escudero, & Boersma, 2004; Boersma, 1998; K. N. Stevens, 2002). However, a significant difference is that the latter models do use abstract representations at the phoneme level. Although some of the exemplar models include phoneme labels segments (e.g., Hay et al. 2006;

Hawkins, 2003; Pierrehumbert, 2002), these do not solve the problem of generalisation. The first problem with such models is the ontogeny of representations. If generalisation takes place via analogy between labelled categories, this approach begs the question of what is the nature of these labels. Studies on unsupervised learning algorithms, (e.g., Siu, Gish, Chan, Belfield, & Lowe, 2014; Toscano, & McMurray, 2010; Varadarajan, Khudanpur, & Dupoux, 2008) have shown that categories can be formed based on similarities in the input. However, the aim of this algorithms is to acquire an inventory of intermediate symbolic units. That is the categories formed by analogy are not mere labels to collections of similar exemplars, but a series of abstract units that mediate between the acoustical and lexical levels. I would argue that if labels –which are finite, non-continuant, and ultimately abstract– are encoded at the lexical level (as necessary for generalisation), their inclusion in a model actually makes it implicitly hybrid.

Furthermore, there is evidence (e.g., Best, 1984; Kuhl et al., 2008) that during the early stages of phonology acquisition, the infant's lexicon only contains episodic memories of whole, unsegmented words. With more experience in the language, children identify regular patterns (i.e., phonemes) in their L1 words. However, as a result of language attunement to the phonology of the L1, infants become less sensitive to linguistically non-relevant features, limiting their ability to perceive phonetic detail not used in their L1. After this specialisation in L1 takes place, the listener will tend to lexically encode words using the phonemic categories of their L1. Consequently, it has been argued (Pallier, Colomé, & Sebastián-Gallés, 2001) to that adult lexical representations cannot store veridical phonetic information of words they experience.

The limitations of a pure exemplar model have been recognised (Cutler, 2008; Goldinger, 2007; Nguyen, Wauquier, & Tuller, 2009). In fact, Goldinger, who first proposed a purely exemplar model of speech perception, later recognised the need for abstract representations in speech perception. Goldinger (2007) proposes a new model that includes both abstract and episodic representations. His Complementary Learning Systems model is a hybrid model in which two neural networks are involved in speech perception. In this model both abstract and episodic representations are interdependent. The hippocampal network stores traces of stimuli which have been segmented, while the representations in cortical network are formed by overlapping episodes. He presented computer simulations run with this model that satisfactorily predict perceptual learning. However, the Complementary Learning Systems model is not fully developed. For instance, it does not describe the process by which the acoustic signal is mapped onto abstract units, nor does it define the units that are used in lexical or prelexical perception, nor how are these structured. For instance, the model does not stipulate whether the default type of representations are abstract or episodic, nor for which processes is each type of representations dominant. Nevertheless, this attempt to integrate abstract and episodic representations in a model underscores the need for both types of representations in speech perception.

In this regard, McLennan and Luce (2005) propose that both abstract and indexical representations are used during speech processing, and that the kind of representation used is determined by the frequency of occurrence of a particular language feature or "chunk" (i.e., a phonetic feature, an allophone or a word). Features that are highly frequent are more abstract in nature, while low frequency features – such as indexical information – are more of an exemplar nature. High frequency representations resonate faster with the input than low frequency representations. Consequently, abstract representations dominate in the early stages of speech processing, while features that are less frequent because they are related to more specific, detailed surface information (i.e., indexical information) only affect processing later.

In summary, what can be concluded is that voice information and segment information are interdependent, as both are transmitted through the same channel: the acoustic signal. Consequently neither purely exemplar nor purely abstract models can be used to explain speech perception. On the one hand, abstract representations are required for the interpretation of a deviant segment; on the other, episodic memory is required both for generalisation and for delimiting the effect to a particular speaker or dialect.

The studies reviewed above show how challenging it is to determine the nature of phonological representations. Furthermore, the importance of defining the nature of phonological representations is crucial not only in modelling speech perception in general, but also in modelling how the phonology of a second language is learned. In the next section I review how phonological representations have been seen in several models of L2 learning.

2.1.4 *Phonological representations in models of L2 learning*

The performance of a person in a second language is influenced by their experience of their first language. Language specificity for the first language emerges very early in life. Kuhl, Tsao and Liu (2003) showed that infants are sensitive to small phonetic differences present both in their native language and in foreign languages. However, as they become more tuned to their L1 categories, their ability to perceive contrasts in other languages decreases. Kuhl's Native Language Magnet theory (NLM) of speech acquisition (Kuhl et al., 2008) claims that as L1 phonetic categories are formed, the perceptual space around L1 categories is warped, causing loss of the capacity to perceive fine phonetic details of other sounds close in that perceptual space. The consequence of this warping is the loss of the ability to discriminate between L1 sounds and perceptually highly similar non-L1 sounds. The phonological representations in this model are equivalent to sound segments and the information they contain is of an acoustic nature.

This model of L1 perception is compatible with Flege's Speech Learning Model (Flege, 1995), in that L2 learners show greater difficulty in learning categories that are phonetically close to categories in their L1 than they do in learning L2 categories that are more distant from their L1. Both the Native Language Magnet (NLM) and the Speech Learning Model (SLM) share the view that speech perception is based on the same auditory processes that are employed for general acoustic perception. In both models, sound categories are only represented at the phonetic level. In other words, speech sound categories are formed based only on their acoustic characteristics, and the models do not include phonological representations: a finite set of discrete units made up of distinctive features in which lexical items are spelled out.

However, the lack of a phonological (i.e., symbolic, abstract) level of representation has implications for the notion of what learning is and for the type of knowledge a L2 speaker may have about the target language. SLM's view is that L2 speakers are considered to have correctly and completely learnt a phonetic category if they are able to produce speech sounds that are identical to those of a native speaker. However, failure to imitate the production of a native speaker does not mean that the learner has not created a new phonological category in the target language. A learner could have the same number of phonological categories as a native speaker, even though his pronunciation varies greatly from that of a native speaker. For example, even if a Japanese L1 learner of English pronounces the word "light" with an alveolar tap and the word "right" with an alveolar approximant, it can be said that he can distinguish between the phonological categories /l/ and /r/.

Contrary to SLM, the Perceptual Assimilation Model L2 (PAM-L2) (Best & Tyler, 2007) recognises that learning the phonology of a second language does not just involve the phonetic level. As Best and Tyler suggest:

> [the] perceptual objects/events that are relevant to L2 speech learning are not merely phonetic. Language-relevant speech properties are differentiated not at the phonetic level but also at the higher-order phonological level, as well as at the lower-order gestural level. (Best & Tyler, 2007, p 25).

Consequently, an important difference between SLM and PAM-L2 is the inclusion in the latter of a phonological level of representation. The phonological level comprises information about articulatory gestures. PAM-L2 proposes that during speech perception the listener can recover from the acoustic signal the articulatory gestures that were employed in pronouncing speech sounds. By including a module for identifying articulatory gestures, PAM-L2 is able to approach speech perception at two levels: a phonetic level and a phonological level. Since articulatory gestures are considered to be the relevant information contained in the speech signal and since they are relatively small in number, much of the variation observed at the phonetic level can be minimized. PAM-L2 is based on the Perceptual Assimilation Model, developed by Best and her colleagues (Best & Hallé, 2010; Best, McRoberts, & Goodell, 2001; Best, 1984). PAM is similar to the Motor Theory of speech perception (Liberman & Mattingly, 1985) in proposing articulatory features as the primitive of speech perception. However, contrary to the Motor Theory, PAM proposes that the listener perceives articulatory information in the acoustic signal, not the articulatory motor commands sent to the articulatory organs.

Another model of L2 perception is the Automatic Selective Perception model (Strange, 2011). As in PAM-L2, the phonological units proposed in this model are articulatory gestures. These representations are common to both perception and production routines. An important aspect this model shares with PAM-L2 is that speech perception is seen as an information-seeking activity. This means that the way listeners process the acoustic signal depends on the task they have to carry out. The model proposes that information from the acoustic signal can be extracted using two perception modes: phonological and phonetic. The phonological mode is used by adults when listening to L1 speech in dialects similar to their own. In this mode, variations due to speech rate or minor variations in the speech of familiar speakers are ignored. In the phonetic mode, listeners attend to context-dependent allophonic detail. This mode of perception is used when adjusting to a different dialect and when learning L2 sequences that contain segments that are different from those of the L1, that is when attention is focused on fine phonetic detail. For this model L2 phonology learning consist on acquiring selective perception routines for the perception of L2 segments.

There are several advantages in acknowledging both a phonetic and a phonological level in a model. For example, including phonetic representations can account for findings that suggest that perception of phonetic categories (allophones) changes during the life span, while not necessarily implying a change in phonological categories. Likewise, phonological assimilation (i.e., fusing two L2 phonological categories into one category) need not imply that the associated phones are perceived as identical at the phonetic level. The inclusion of a phonological level, then, makes PAM-L2 and the Automatic Selective Perception models better at predicting L2 speech perception than SLM. However, proposing that articulatory gestures are the basic unit of speech perception is problematic.

There is strong evidence that articulatory features are needed to explain findings in speech production. Also, some authors argue that the perception and production systems should share the same representations, or a "common currency". Goldstein and Fowler (2003) claim articulatory features can serve this role. However phonological representations based on articulatory features are not fully satisfactory in perception models. Phonological categories often group together sounds which are articulatory quite different. For example, in Caribbean dialects of Spanish syllable final /s/ is pronounced as a voiceless alveolar fricative, or as a voiceless glottal fricative. What these sound have in common is their acoustic properties. As Ohala (1996) argued, phonological patterns across languages are better explained based on acoustic-auditory properties of speech sounds rather than on their articulations. Finally, the representations proposed in PAM-L2 cannot account for the storage of indexical information, as they cannot contain fine acoustic detail that is not related to articulatory gestures, such as the timbre of vowels determined by differences in length of the vocal track or of the nasal cavities across speakers. Consequently, no voice facilitation effect would be predicted for acoustic variation not related to articulatory gestures, as no acoustic details could be retrieved about the timbre of a particular voice.

Another model in second language perception that postulates that phonological representations are completely abstract is the Second Language Linguistic Perception (L2LP) model (Escudero, 2005). L2LP builds on the Functional Phonology model (Boersma, 1998). Like PAM-L2, L2PL proposes that learners' initial L2 perception is based on their L1 sound system. So initial difficulties L2 learners face in discriminating L2 phonological categories are determined by how similar the target categories are to those that already exist in the L1. However, L2PL does not propose that the acoustic signal is mapped onto articulatory gestures. Instead, the acoustic signal is first mapped onto a series of acoustic parameters, which in turn activate abstract phonological representations. L2PL proposes that two types of tasks are carried out during language comprehension; a perceptual task and a representational task. The perceptual task consists of learning how to map acoustic cues onto L2 phonological categories, while the representational task involves establishing correct representations for lexical items containing the new learnt category. In the perceptual task, the acoustic signal is mapped onto phonological representations by means of a perception grammar. This grammar establishes the perceptual boundaries for phonetic categories. For example, an English perception grammar contains phonological cue constraints such as "[...]'an F1 of 300 Hz is not /i/' or 'a duration of 120 ms is not /i/' as well as similar constraints that map F1 and duration values onto /ɪ/" (Escudero, 2005, p. 48). L2PL maintains that the learning of an L2 phonology entails a retuning of the perceptual boundaries in the cue constraints used in the perceptual task. Once this perceptual change has taken place, the representational task can correctly store lexical items in the L2. Finally, L2PL says that phonological representations are constituted of segments (i.e., vowels and consonants) but these are highly abstract, symbolic units which contain no articulatory or phonetic information. The perceptual grammars set out in the model, makes L2PL capable of using with fine phonetic detail to create new phonological categories. However, it is not explicit about the use of indexical information contain in the speech signal. Also, it makes no claim about the possibility of storage of fine phonetic detail in long term memory.

The models of speech perception discussed above –arguably the most influential models in L2 research in this area– vary as to how concrete or abstract their phonological representations are. Nevertheless, they share the view that the information needed to process an L2 is discrete, categorical and abstract.

2.2 JUSTIFICATION FOR THE CURRENT STUDY

The effect of L1 interference and transfer on L2 learning has been satisfactorily described by the L2 learning models discussed above. However, in order to better understand how L2 phonology is acquired, a model should account for how learners store and retrieve the information available to them. The previous discussion of the models of how phonology is learned in a second language underscores the fact that L2 models propose abstract representations only. The assumption that only abstract information is stored cannot account for how a person forms and uses the phonological representations of two languages.

There is substantial evidence coming from research in L1 showing that listeners can store episodic traces of the words they experience. However, none of the L2 models focuses on the role of episodic memories in the learning of the phonology of a second language. I argue that this is a limitation on these models. Without the ability to store fine phonetic detail, L2 phonology learning would not occur. For example, without episodic representations, it would impossible for the learner to compare the fine phonetic detail contained in new input to that of previously experienced exemplars.

However, neither a purely exemplar nor a purely abstract model can fully explain how the phonology of an L2 is learned. I argue that both types of representations are needed. Episodic representations have a predominant role at the early stages of learning new L2 phonological categories. However, as exposure to the target language and competence in it increases, abstract phonological representations for L2 emerge. Also I propose that abstract representations are a crucial part of L2 phonology. Although the abstract representations consolidated in the learner's L1 have an impact on their ability to discriminate

L2 phonemes, abstract representations are needed in order to generalise the ability to discriminate a new phonological category across the lexicon.

2.3 research questions

The literature review presented above underscores the need to investigate the role phonetic detail has in the learning of the phonology of a L2. The research questions that arise from the literature review are as follows:

1. Are L2 learners capable of storing acoustically rich representations in their L2?

2. Are the L2 phonological representations of L2 learners predominantly episodic in nature?

3. Can L2 learners generalise the perceptual learning acquired on a limited set of stimuli to the rest of their lexicon?

Answers to these questions provide an insight into the nature of phonological representations in an L2. Furthermore, a better understanding of the capacity and limitations L2 learners have for storing phonetic detail will certainly guide better teaching practice in the L2 classroom. Moreover, although the present study is focused on L2 perception, its findings are also pertinent to the field of speech perception in general. The importance of the present study is that it tries to disentangle the complex relationship between abstract and episodic representations in the learning of L2 phonology. Moreover, given that the focus of this thesis is the nature of phonological representations, the results obtained from this study can contribute to the ongoing research into the nature of phonological representations in general.

3 METHODOLOGICAL CONSIDERATIONS

Some of the disagreements seen in the previous chapter concerning the nature of phonological representations may be the result of the different methodologies employed. As discussed in the review of speech perception models, one of the strongest arguments in favour of exemplar models of speech perception is the evidence that listeners can store indexical information about the voices in which stimuli are presented. In turn, the ability to generalise the perceptual learning resulting from exposure to a few exemplars to the rest of the lexicon has been held as one of the main reasons to propose that the phonological representations are abstract in nature. In this study, I try to reconcile these two views, at least for the L2 learning context, by looking for evidence that both abstract and episodic representations are available to L2 learners. If the early stages of learning the phonology of an L2 rely on episodic representations, learners should be less capable of generalizing phonological contrasts to other lexical items and to other speakers. In this case, beginning learners should be able to discriminate phonological contrasts in the L2 better when the words are heard in a familiar voice, but will be slower and produce a higher error rate when they are asked to detect these contrasts in new words or words pronounced in a new voice. On the other hand, if learners use abstract representations, the knowledge they acquire from a limited number of exemplars could be generalised across the lexicon. Generalisation would allow learners to accurately discriminate words uttered by different speakers. Moreover, phonological discrimination should not be affected by variability in voice.

In order to establish how much phonetic detail L2 learners are capable of storing, I carried out two experiments. The first experiment (Chapter 4) looked at the effects of speaker variation on the ability to discriminate phonological contrast in a L2. In this experiment, I expected to obtain information about how speaker variation affected the sensitivity of L1 Spanish learners of English to temporal and spectral acoustic cues not used in their L1. Two contrasts identified in previous studies as problematic for L1 Spanish learners of English (see section 3.3 and 4.2) were selected. These were the contrasts between the FLEECE/KIT vowels and between voiced and voiceless plosive consonants. This experiment also aimed to find out if these contrasts were suitable to investigate whether learners stored phonetic detail in their L2 phonological representations.

The second experiment (Chapter 5) looked at whether L2 learners were able to store fine phonetic detail (as related to indexical information) despite the difficulty they had to discriminate between L2 categories. Finally, to determine whether L2 learners are able to generalise phonological knowledge, I carried out a perceptual training experiment (Chapter 6) in which the effect of voice variation was investigated.

In particular, the designs of the first two experiments presented in this thesis are influenced by the methodology used by Pallier, Colomé and Sebastián-Gallés (2001) and Dufour, Nguyen, & Frauenfelder (2010), in that the effect of repetition priming is used to determine the ability L2 learners have to store phonetic detail.

3.1 PRIMING

One way to study the information stored in a phonological representation is to look at the ease with which a lexical entry can be accessed. Priming is frequently used in psycholinguistic studies as it offers a tool for the investigation of word recognition. Facilitatory effects can be observed between stimuli that are related in meaning or in form. Faster and more accurate responses to a stimulus are observed when it is preceded by a word to which it is related semantically. For example, the word "chair" is accessed more quickly if preceded by the word "table" than by "dog". Similarly, priming can happen when two stimuli are similar phonologically, like the words "cap" and "captain".

3.1.1 *Repetition priming*

Repetition priming is the facilitatory effect that occurs when a word is processed that has previously been experienced recently. The effects of repetition priming can be observed in several tasks such as perceptual identification, oral repetition (naming), and lexical decision. The effect is behaviourally observed as a decrease in reaction time between the first and second presentation of an experimental item, or in a faster reaction time in a primed condition than in an unprimed condition. Repetition priming has been extensively used in several lexical decision and naming experiments in the first language (e.g., Forbach, Stanners, & Hochhaus, 1974; Goldinger, 1996; Kirsner & Smith, 1974; Scarborough, Cortese, & Scarborough, 1977). More recently, repetition priming has also been used in various studies on L2 perception (Colomé & Miozzo, 2010; Costa, Santesteban, & Ivanova, 2006; Dufour & Nguyen, 2008; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008).

Most theories of repetition priming (see Tenpenny (1995) for an overview) postulate that word recognition involves the simultaneous activation of several lexical entries. The level of activation of a lexical entry depends on how compatible it is with the speech signal. The lexical entry with the highest level of activation is the one that is eventually selected. A lexical entry that has been selected recently retains a high level of activation from this prior exposure, and this facilitates subsequent selection. In studies of speech perception this effect has been proposed as evidence both for detailed exemplar representations and for abstract representations. As an example of the first of these, it has been observed that repetition priming produces faster reaction times when the stimuli are presented in the same voice than when they are presented in different voices (e.g., Goldinger, 1996; Mullennix et al., 1989). Exemplar models view this effect as evidence that word recognition involves direct comparison of the acoustic signal to previous episodes of a word.

Nevertheless, other studies propose that repetition priming is evidence for abstract representations. For example, in a study comparing Catalan dominant and Spanish dominant bilinguals, Pallier et al. (2001) found that L2 Catalan speakers could experience repetition priming between the members of minimal pairs. In the lexical decision task, Spanish dominant bilinguals showed the same level of priming between the repeated words and between members of minimal pairs which involved Catalan specific contrasts. However, no repetition priming was found when the minimal pair members contrasted in phonemes that were common to Spanish and Catalan. Pallier et al. (2001) concluded that minimal-pair priming was due to L2 listeners' inability to store fine phonetic detail in the lexicon. They argued that the fact that the L2 speakers in their study, who had a high proficiency in Catalan and had more than 15 years of exposure to this language, had failed to learn Catalan-specific contrasts was evidence that acoustic information was not stored in the lexicon.

The same effect of priming between members of minimal pairs was found by Dufour et al. (2007). Their study looked at the perception of standard French by native speakers of Southern French, a dialect that does not have the contrast between close-mid and open-mid vowels standard French has (e.g., épée "sword" /epe/ vs épais /epɛ/ "thick"). In a follow-up study, Dufour, Nguyen and Frauenfelder (2010) trained a group of speakers of Southern French to discriminate between the /e/ and /ɛ/ vowels of standard French. After successful training, the Southern listeners showed priming within these minimal pairs. The authors concluded that this effect is evidence for the existence of abstract representations.

One of the limitations of the method used by Pallier et al. (2001) and Dufour, Nguyen and Frauenfelder (2007, 2010) is that these studies used one voice only to record their stimuli. As a consequence, these studies overlooked the effect of speaker variation. However, the effect of voice variation on repetition priming has been investigated in L1. McLennan and Luce (2005) carried out a study which focused on the effects of the difficulty of a task on the access of indexical information. In a lexical decision task, they found that there was no voice facilitation effect when the task was easy, that is when the pseudowords were quite different from real words. However, a same voice facilitation effect was found when the pseudowords were very similar to real words. McLennan and Luce concluded that these findings were consistent with a hybrid model of spoken word recognition which employs both abstract and episodic phonological representations. They propose that abstract representations are activated first, but when the task requires more time to be completed, then episodic representations are also activated.

Follow up studies by Dufour and Nguyen (2010) and by Orfanidou, Davis, Ford and Marslen-Wilson (2011), have failed to find a strong priming effect due to voice repetition. However, Orfanidou et al. found a voice repetition effect for the pseudowords. Since pseudowords did not have a representation before the experiment took place, the priming effect between pseudowords suggests that the listeners had stored episodic traces of the stimuli. Orfanidou et al. propose that hybrid models are better equipped to explain the findings of their study (i.e., no voice effect for real words, but an effect for pseudowords) than a purely abstract one.

Despite the value of using different voices in repetition priming, this type of paradigm is seldom used in the field of L2 spoken word recognition. To the best of my knowledge, no study to date has explored repetition priming in relation to the effects of voice variation on minimal pair discrimination. However as I discuss next, speaker variation has a strong effect on speech perception and in L2 perceptual learning.

3.2 USING SEVERAL VOICES IN TRAINING

It has been found that a high-variability phonetic training method (Logan, Lively, & Pisoni, 1991) promotes more robust perceptual learning in an L2. In this method, the contrast to be learned is presented in several phonotactic contexts and in several voices. Several studies (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Iverson & Evans, 2009; Iverson, Hazan, & Bannister, 2005; Iverson, Pinet, & Evans, 2012; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994) have found that learning with this method has long term effects and generalizes better to words and speakers not used during training. Similarly, L2 vocabulary learning seems to benefit from voice variation (Barcroft & Sommers, 2005; Trofimovich, 2005).

Nevertheless, although the previous studies indicate that robust learning happens when several voices are used in training, other studies have found that voice variation makes certain experimental tasks more difficult. For example, in their study of L1 perception of foreign accented speech, Sidaras, Alexander, and Nygaard (2009) noticed that lexical recognition was made extremely difficult by voice variability. A similar finding is reported by Creel and Dahan (2010) for L1 vocabulary learning by preschool children. The difficulty probably arises from the speaker variability. In case of L2 learners this means that in order to discriminate new phonological categories in the L2, they must be able to identify the general acoustic cues for each category, while ignoring the variation introduced by differences between talkers. The challenge is that although the outcome of high-variability training is more robust than training with low-variability, the former is more demanding cognitively, as attention has to be directed to the correct acoustic cues. In fact, Hazan and Simpson (2000) found that enhancing the acoustic cues relevant for the distinction of English consonants helped L2 English learners to attend to those cues. Given the effects that voice variation has on L2 perception, I aimed for a balance between voice variation and perceptual difficulty for L2 categories when designing the experiments reported in the present study.

3.3 L1 SPANISH PERCEPTION OF ENGLISH VOWEL CONTRASTS

The PAM-L2 model (Best & Tyler, 2007) predicts that the ease of perception of L2 contrasts depends on the degree of assimilation of these L2 contrasts to L1 categories. As seen in Figure 1, Spanish has far fewer monophthongal vowels than English. Consequently, it is expected that L1 Spanish learners of English perceptually assimilate several English vowels into a single Spanish category. For example, since Spanish has only one vowel in the close front vowel space, the contrast between the FLEECE and KIT vowels should be difficult for Spanish speakers.

In fact, the perceptual assimilation of the English vowels FLEECE and KIT is a well-known difficulty for Spanish L1 learners of English and has been reported in several studies (Escudero & Chládková, 2010; e.g., Escudero, 2005; Flege, Bohn, & Jang, 1997; Iverson & Evans, 2009; Konopka, 2011; Morrison, 2008).

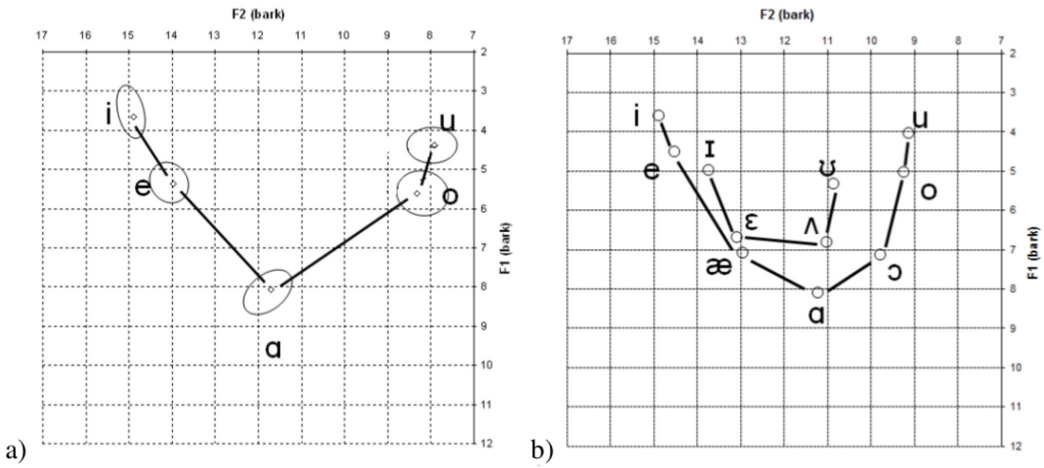

Figure 1 a) Formant values for Spanish, b) formant values for American English vowels. The values are taken from seven adult female native speakers of Spanish born and educated in Mexico and 12 adult female native speakers of English born and educated in Chicago. (Taken from Konopka, 2011, p. 80).

I decided to study the effects of voice variation using minimal pairs contrasting the FLEECE and KIT vowels for the following reasons. First there is perceptual assimilation of these vowels in learners of all dialects of Spanish. Second, the vowels FLEECE and KIT are high frequency phonemes. Cruttenden (2001) reports KIT and FLEECE as the second (8.33%) and seventh (1.65%) most frequent vowels in English. Moreover, there are several high frequency minimal pairs in English contrasting these vowels, and this greatly facilitated the selection of materials for the experiments. As a consequence, the results of the study will have practical implications for Spanish native speakers learning English as a second language.

In both General American and Standard Southern British English, FLEECE and KIT differ in spectral quality and duration. Several studies (e.g., Bohn & Flege, 1990; Flege, 1991; Kondaurova & Francis, 2008) have found that native English speakers rely primarily on spectral information to discriminate between these vowels, but that they can rely on duration when the spectral information is ambiguous (Flege et al., 1997).

The studies carried out with L1 Spanish learners of English do not show agreement on the acoustic cues preferred by this group of learners. For example, Escudero (2005) proposed that, since Spanish does not use length as a distinctive feature, L1 Spanish learners of English would be more sensitive to durational than to spectral differences between these vowels. She argued that in the initial stages of learning this contrasts, L1 Spanish learners would use only length to differenciate between FLEECE and KIT, and that only later they would start using spectral differences. However, Morrison (2008) found that this was not the case. In his study, L1 Spanish learners of Canadian English showed no preference for length over spectral differences. On the contrary, they carried out a multidimensional goodness-fit to determine the category of these English vowels, and with more experience in the language L1 Spanish learners used durational cues. Similarly, Kondaurova and Francis (2010) found that while L1 speakers of American English rely primarily on spectral information to discriminate between these vowels, L1 Spanish listeners were not consistent on the acoustic cues they attended, with some participants relying more on spectral information and others relying more on duration. Moreover, Kondaurava and Francis found that the type of training used would determine which acoustic cue was attended by their participants. The studies previously discussed indicate that although duration is an important cue, successful learning of the FLEECE and KIT contrast should be based on spectral quality.

The design of Experiment 2 needed to include a vowel contrast that served as a control for the perception of FLEECE-KIT contrast. The spectral and durational differences between the two vowels serving as controls should be equivalent to those between FLEECE-KIT. Moreover, these vowels should produce several high frequency minimal pairs in English. However, the control vowels should not be as likely to be perceptually assimilated by L1 Spanish learners as FLEECE and KIT are.

In my teaching experience I have noticed that the DRESS and TRAP is not as difficult to learn by L1 Spanish speakers as the FLEECE and KIT contrast. The DRESS vowel is commonly assimilated to Spanish /e/ while the TRAP vowel is assimilated to Spanish /a/. This pattern has been observed in several perceptual studies. My observations are consistent with studies carried out on the perception of English vowels by L1 Spanish speakers of English. For example, Flege (1991) reports this pattern of misidentification: KIT for Spanish /i/, DRESS for Spanish /e/ and TRAP for Spanish /a/. Similarly, Flege, Bohn and Jang (1997) found that although L1 Spanish L2 English learners did not produce significant differences in the production of FLEECE and KIT, they produced an even larger spectral difference between English DRESS and TRAP than did the Native English speakers. Some studies have found that L1 Spanish listeners can perceptually assimilate English vowels DRESS and TRAP vowels to Spanish /e/. For example, Escudero and Chládková (2010) found that this was one of the perceptual assimilation patterns they found in a study using synthetic stimuli and monolingual L1 Spanish listeners. However, the assimilation pattern was observed for General American stimuli but not for Southern British English stimuli. They argued that the reason for this was that DRESS and TRAP have lower F2 values in Southern British than in General American. Consequently in the former dialect, TRAP lies outside of the Spanish /e/ category boundary.

These findings indicate that although this contrast is potentially assimilable by L1 Spanish learners, it is not as problematic as the FLEECE/KIT contrast. For this reason, I included DRESS and TRAP as a control a set of the stimuli in Experiment 2.

3.4 statistical tools

The main purpose of the statistical analysis carried out in the present study was to determine differences in performance between groups of participants in different experimental conditions. The most common statistical analysis used in this type of study has been the Analysis of Variance (ANOVA). However, mixed-effects analysis offers a superior approach to ANOVA in situations where the data is correlated (Ieno, Walker, Zuur, Savelieve, & Smith, 2009).

The data analysed in the present study were obtained from repeated measures of participants. Data obtained from repeated measures are generally correlated by participant and stimulus. These correlations violate the assumptions of standard ANOVA. However, the variance generated by these random factors can be explicitly modelled in mixed-effect models. Mixed-effect models include a fixed effect structure and a random effect structure. As recommended by Baayen (2008), repeated measure data analysis using mixed-effects models should include a fixed structure with those variables that are constant in all conditions (e.g., the treatments or conditions included in the experiment) and a random structure with those variable that are random (e.g., participants). In the present study, the use of mixed-effect models allowed me to model the error term arising from difference in performance between participants and also the difference caused by different words.

Another advantage of mixed-effects models is that more participants can be kept in case of missing data. When using ANOVA to analyse repeated measures data, it is necessary to discard participants if they are missing even a single measurement. In contrast, mixed-effects models are more robust to missing data, and participants with missing measurements can still be used (on the condition that missing data is missing at random). This robustness with respect to unbalanced data makes mixed models better at handling uneven spacing of repeated measurements.

Mixed models can be extended to generalized mixed models. Mixed-effect logistic regression can be used to predict the outcome of a categorical dependent variable. This statistical tool can be used in the analysis of accuracy of responses. Finally, I find that results of mixed model analysis are easier to interpret than the results of classical ANOVA. For example, the effect of each variable can be estimated without carrying out other statistical tests, such as a Tukey's test after an ANOVA. Moreover, the polarity of the estimate of a variable indicates whether this variable has an incremental or decremental effect.

A challenging aspect in mixed-effects modelling is selecting the best model. In the analysis of the data, I followed the protocol proposed by Ieno et al. (2009). This protocol prescribes a series of steps to find the best model. The selection process starts by fitting a complex model that contains all the explanatory variables and as many interactions as possible (i.e., the maximal or beyond optimal model). Based on this model, simpler models are subsequently fitted. First, keeping the fixed structure of the maximal model, the optimal random structure is found by comparing a series of models with simpler random structures. The random structure of the model that best fits the data (as determine by the lowest Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC)) is kept. Next, the optimal fixed structure is selected among a series of models fitted following a stepwise backward process. For the comparison of fixed effects, models are fitted using the Maximum Likelihood (ML) method. An ANOVA test is used to compare how well these models fit the data. The model with the lowest logistic likelihood or BIC is selected. This model is then refitted using the Restricted Maximum

Likelihood (REML) method. Finally, the fit of the optimal model is checked by visually inspecting plots of residuals vs. fitted values.

Given the large number of models that can be evaluated during the selection process, in the discussion of the results of the experiments, I only present the results of the optimal models. For reference purposes, I have also included the output of beyond-optimal models in the appendixes.

4

VOICE REPETITION AND SPOKEN WORD RECOGNITION IN AN L2

As discussed in Chapter 1, one of the pressing topics in spoken word recognition is the general format of phonological representations. Although most influential models are abstractionist, there is evidence suggesting that phonological representations contain fine phonetic detail. As a consequence, exemplar and hybrid models have been proposed. However, it is still unclear if or how exemplar memories are accessed during spoken word recognition.

Crucial to this discussion is to find out whether listeners are able to store fine phonetic detail in their phonological representations. Since phonetic detail is necessary to differentiate speakers and phonological categories, the experiment reported in this chapter was designed to explore how voice variation affected the perception of the acoustic cues used in a second language.

The experiment reported in this chapter looked at the effect of form priming in a second language. As discussed in Chapter 3, the study of form priming in an L2 is very useful tool in trying to better understand the nature of the phonological representations in a second language. Two aspects of form priming were considered, phonological priming and speaker priming.

Two contrasts that are difficult to learn by L1 Spanish speakers were used. The experiment studied the contrast between the FLEECE/KIT vowels and the contrast between voiced/voiceless plosives. The experiment aimed to establish the suitability of the combination of speaker variation and minimal pair discrimination as a viable method to study the ability to store phonetic detail in L2 phonological representations. The information obtained in this experiment was important to answer the first question set out in this thesis: Are L2 learners capable of storing acoustically rich representations in their L2?

4.1 INTRODUCTION

Anecdotal and empirical evidence on the learning of the phonology of a second language by adult learners indicates that achieving a level of competence identical to that of a native speaker is extremely rare if not impossible. This difficulty has been attributed to neural maturational processes that take place in early life (e.g., Kuhl et al., 2008). As discussed in Chapter 2, the dominant phonological theories during the last half of the twentieth century explain this inability to form new categories in terms of models of speech perception which posit that some form of normalisation of the acoustic signal is carried out before the relevant phonological information can be mapped onto the lexicon. This normalisation process is seen as necessary so that listeners can cope with the high variability typically observed in speech. The normalisation process is supposed to eliminate variability and convert the continuous acoustic flow into discrete units (for example, phonemes, codas, syllables) which are finally stored in the lexicon.

For the purpose of the present experiment, the most important claim of such models is that the phonological representation contains very little phonetic detail. There are at least two reasons why this claim is important and relevant to the current study. First, the difficulty that L2 learners have in creating new phonological categories in their second language can be explained by the lack of phonetic detail retained in phonological representations after the normalisation process mentioned above. Second, the paucity of phonetic detail presents challenges to adult speakers of a second language that would have implications beyond lexical processing. For example, the lack of phonetic detail in the phonological representations might not only make the learning of new phonological categories difficult, but it might also make speaker discrimination difficult. After all, linguistic and indexical details are transmitted simultaneously through speech.

Recent research has found that indexical and lexical information are tightly related. There are findings that show that participants recognise words faster and more accurately when they are presented in the same voice (e.g., Goldinger, 1996; Mullennix et al., 1989). These findings indicate that phonological representations are richer than previously thought. Nevertheless, the processing of indexical and linguistic information contained in the acoustic signal is complex.

Although voice discrimination and lexical discrimination are carried out by different areas of the brain (A. A. Stevens, 2004), indexical and linguistic information are closely interdependent. For example, Remez, Fellowes and Rubin (1997) found that individuals could recognise a familiar voice based on phonetic information alone. In this study, participants were presented with a natural sentence (i.e., recorded normally) followed by two synthetic sentences, one of which was created based on the natural recording presented previously and the other was based on a recording of the same sentence but by a different speaker. The participants were asked to report which of these two sentences was based on the natural sentence presented at the beginning of the trial. Remez and his colleagues found that the participants were able to match synthetic stimuli to natural recordings. Since the sinewave stimuli cannot carry much indexical information about the speaker in terms of voice quality, this shows that idiosyncratic phonetic information at the segmental level can be used in talker recognition.

Another important aspect of the processing of indexical information is that, although it can be language independent, talker identification is easier if the listener knows the language of the talker (Goggin, Thompson, Strube, & Simental, 1991; Nygaard et al., 1994; Thompson, 1987). In the case of L2 learners, Sullivan and Schlichting (2007) found that the ability of English L2 beginner learners to recognise a familiar voice improved as their knowledge of English increased. More recently, Winters, Levi, and Pisoni (2008) investigated what type of indexical information was processed in a foreign language. They found that both language independent and language specific acoustic cues were used in voice identification. They noticed, however, that listeners used phonetic information to identify a talker if they knew the language of the talker. If the listeners did not know the language, they would rely on language independent information, that is to say, information related to the physical characteristics of the talkers, such as voice quality, and characteristic fundamental frequency (F0) range. In a follow-up study, Levi, Winters, and Pisoni (2011) studied the facilitatory effect of voice familiarity in lexical recognition. They found that familiarity with a talker created an advantage in word recognition. However, no advantage was observed if the talker used a language unknown to the listeners. Levi et al. (2011) concluded that listeners may only integrate talker and linguistic information in representations in a language that they know.

Previous studies on L2 phonology acquisition have looked at the role of voice variation in new category learning (e.g., Bradlow et al., 1997; Iverson et al., 2012; Lively et al., 1994). These studies have focused on training L2 learners to identify L2 phonemes. However, there is still a need to consider the effects of voice variation on lexical access in a second language. As discussed above, the processing of phonological information in an L2 is affected by voice familiarity. In other words, the granularity of phonological representations in an L2 seems to be determined by the knowledge of the target language. To the best of my knowledge, no study to date has looked at voice priming effects in the perception of minimal pairs in an L2.

4.2 GRANULARITY IN L2 REPRESENTATIONS

In order to study how much acoustic detail learners stored in their L2, two phonemic contrasts were studied. One was the contrast between a series of voiced and voiceless plosive consonants; the other was the contrast between the close front vowels FLEET and KIT. English learners need to be sensitive to fine phonetic detail in order to discriminate these contrasts, and correctly access lexical entries in the L2.

The reason for looking at the contrast between voiced and voiceless consonants in this experiment was to study how sensitive L1 Spanish learners of English were to acoustic cues not used in their L1. Spanish and English have the same number of plosive consonants. Hence, the challenge for Spanish L1 learners is not learning new categories, but learning to map new acoustic cues onto the correct voicing categories. Spanish and English word initial plosives differ in the acoustic cues they use. For example, in English voiced plosives have a short positive voice onset time (VOT) and voiceless plosives are aspirated. In contrast, Spanish voiced plosives are fully voiced (i.e., have negative or zero VOT) and voiceless plosives are unaspirated (Obediente, 1998) and typically have shorter VOTs than voiceless plosives in English.

The ability to notice differences in VOT between plosives has been previously used in studies in L2 perception (e.g., Abramson & Lisker, 1972; Flege & Eefting, 1987; V. L. Hazan & Boulakia, 1993). In the case of L1 Spanish speakers, Ju and Luce (2006) found that high proficiency L2 English speakers were very sensitive to differences in VOT. However, this may not be the case for beginner/intermediate learners. Furthermore, the variation introduced by different talkers may affect the ability learners have to notice these differences. Differences in VOT have been found for sex (Whiteside & Irving, 1998; Whiteside & Marshall, 2001), age (Ryalls, Simon, & Thomason, 2004), and even menstrual cycle (Wadnerkar, Cowell, & Whiteside, 2006).

The vowel contrast studied poses a different challenge to L1 Spanish learners. In this case, the learners need to notice acoustic differences between the English close front vowels in order to learn a new phonological category. As discussed in Chapter 3, L1 Spanish speakers have difficulty discriminating between the vowels FLEECE and KIT (e.g., Escudero & Chládková, 2010; Escudero, 2005; Morrison, 2008). L1 Spanish learners of English tend to perceptually assimilate both FLEECE and KIT vowels to Spanish /i/, the only close front vowel in this language. Morrison (2008) found that L1 Spanish beginner learners of Canadian English were sensitive to both the spectral and duration cues for English vowels. He proposed that at early stages of learning L1 Spanish speakers distinguished English FLEECE and KIT based on a goodness-of-fit between English vowels and Spanish . The tokens that were good examples of Spanish were categorised as KIT and poor examples labelled as FLEECE. Again this comparison process could be made much more complex if talker differences are considered. Studies involving acoustic comparisons of English and Spanish vowels have found there is a great deal of variation across categories in individual speakers (e.g., Bradlow, 1995; Escudero & Chládková, 2010; Escudero, Simon, & Mitterer, 2012).

4.3 VOICE PRIMING

The present experiment consists of a mid-term lexical priming recognition task, based on the study by Palmeri, Goldinger and Pisoni (1993), who examined the effects of voice variability on word recognition in native speakers of American English. Palmeri et al. (1993) found that listeners were able to recognise a word more quickly when it was presented for a second time in the same voice than when the second presentation was in a different voice. This priming effect was taken as evidence that the participants stored particular details about the several voices (20 in total) in long term memory. The purpose of the current experiment was to determine if a similar effect would be found in a second language.

The current experiment included two voices, one female and one male. The decision to include a contrast between a male and a female voice was taken to maximise the particular indices for the voice variation variable. The decision to include only two voices was to allow the use of a smaller number of trials in the experiment. It would have been very difficult to produce counterbalanced experiment lists since the experiment included two main factors (same word vs. minimal pair) and voice repetition (same voice vs different voice). Another difference from the study of Palmeri et al. (1993) was the inclusion of minimal pairs in the current experiment. Minimal pairs were used to find out the extent to which the memories of stimuli presented in the experiment are acoustically fine-grained.

While both English and Spanish share the same number of phonological categories for plosives, Spanish only has one vowel category in the close front vowel space. Including the latter contrast allowed me to determine whether the difficulty in perceiving the vowel contrast was due to not being able to perceive fine phonetic difference per se, or whether this difficulty arises from the mismatch in phonological categories between the L1 (Spanish) and the L2 (English). Differences in perception and production in VOT between L1 and L2 have previously been studied (e.g., Antoniou, Best, Tyler, & Kroos, 2011; Flege & Eefting, 1987; López, 2012) as a tool to look into the process of phonological acquisition in an L2. Concerning fine phonetic detail, McMurray, Tanenhaus, and Aslin (2009) found that sensitivity to differences in VOT within a category can facilitate lexical recovery in the L1. By including the minimal pairs for the contrast discussed above, the experiment could better study whether learners could store acoustic details that help them discriminate not only different voices, but also different lexical items.

4.4 EXPERIMENTAL DESIGN

The experiment followed a one-group repeated measures design. All participants were presented with the same number of stimuli and trials. The factors considered in the experiment were as follows:

- *Presentation*: This variable indicated whether a form was presented for the first time or for the second time in the experiment. It had two levels, "first" and "second".
- *Voice*: refers to the voice that recorded a given stimulus. This factor had two levels: male and female.
- *Word pair*: The stimuli were grouped into pairs. The pairs consisted of repeated words (e.g., *queen/queen*), minimal pairs (e.g., *ship/sheep*) and pseudo minimal pairs (*brick/breek*). Repeated words in the same voice condition consisted of two different recordings of the same word by the same speaker.
- *Voice pair*: refers to whether the second item in a pair (i.e., in a minimal pair or in a repeated word pair) was presented by the same voice (SV) or by a different voice (DV). This factor is obviously only relevant for the second item in a pair.
- *Word type*: There were two levels for this factor: real words (rw) and pseudo-words (pw).

- *Sound type*: The experiment considered the difference between consonants and vowels.

- *Contrast*: This variable was nested in the previous variable. One vowel and two consonant contrasts were included. The vowel contrast was the high front vowels (FLEECE and KIT), while the consonant contrast was voicing. Only labials (/b/ vs. /p/) and velars (/g/ vs. /k/) were included. The reason for excluding alveolar consonants was that the Spanish counterparts have a different place of articulation (i.e., dental), which might have affected the way Spanish learners of English perceived them.

4.5 PARTICIPANTS

Sixty five undergraduate students majoring in English at the University of Los Andes in Mérida, Venezuela, took part in the experiment. All participants were native Spanish speakers who had started learning English after adolescence. Approval to carry out the experiment was obtained from the Human Ethics Committee of Victoria University Wellington (see Appendix 3). The participants were majoring in Modern Languages and had already taken three courses in English. Their average age was 22 years (min = 17, max = 43, *SD* = 5.8). All participants reported having normal hearing. The participants were recruited by inviting students taking third year and fourth year courses to take part in the experiment. All volunteered to take part in the experiment for no reward.

4.6 METHOD

The participants were asked to decide whether words were presented for the first or for a second time during the experiment. The software used to present stimuli and record responses was Perceval (André, Ghio, Cavé, & Teston, 2003). This program presented the stimuli and registered the response choice and the reaction time measured from the beginning of each stimulus. The participants entered their responses via a game pad. According to Perceval's documentation, the use of a game pad can offer a time resolution close to one millisecond.

4.7 MATERIALS

4.7.1 *Selection of stimuli*

The real word stimuli were selected from the 3000 most frequent words in the Corpus of Contemporary American English (COCA, Davies, 2008)[1]. In total, 80 words and 20 pseudowords were included. Ten of the 80 words were presented only once. The rest of the stimuli were grouped in pairs. The word pairs consisted of 20 repeated word pairs, 12 minimal pairs (e.g. "ship" and "sheep") and 20 pseudo minimal pairs (e.g. "court" and "gort"). The pseudo minimal pairs consisted of a real word and a pseudoword. Eight pseudowords were included for the vowel contrast and 12 for the consonant contrasts. These pseudowords were included to determine how sensitive English learners were to the acoustic features that distinguish minimal pairs in the absence of lexical representations. Repetition priming between members

1 I decided to American English in this thesis as it is the variety most used in Latin America. It is the variety usually taught in ESL classrooms and the one to which learners are more frequently exposed.

of minimal pairs should happen if both lexical entries share the same phonological representation. However, this does not prove that L2 learners are not sensitive to fine phonetic detail, as they could have wrongly assumed that the acoustic differences are simply alternative productions of the same phoneme. However, pseudowords have no lexical representation. Consequently, repetition priming in the case of pseudo minimal pairs would certainly indicate that the listeners are not sensitive to the acoustic cues used to discriminate the L2 categories.

Since it was expected that the participants would perceive many of the minimal pairs as actual repetitions, there was a chance that the participants expected all the words to be repeated (after the first few unique presentations). To minimise this kind of strategy, 10 words were presented only once in the experiment. Neither the words included in the repeated pairs nor those presented only once contained vowels or consonants featuring in the contrasts included in the minimal and pseudo minimal pairs.

4.7.2 *Creation of pseudowords*

The pseudo minimal pairs consisted of a real word and a pseudoword. Care was taken to select words that did not form a minimal pair with any of the other stimuli. The pseudowords were created by replacing one segment in a real word with its contrasting segment. That is, in the case of the close front vowels, the FLEECE vowel was replaced with the KIT vowel, and vice-versa. In the case of the consonants, voiced plosives were replaced with their homorganic voiceless plosives. For example, the pseudoword "grick" /ɡrɪk/ was created by replacing the FLEECE vowel in the word "greek" /ɡriːk/ with the KIT vowel. Table 1 presents examples of the different word pairs. There were four pseudowords containing the FLEECE vowel, and four containing the

KIT vowel. In addition, three pseudo minimal pairs were set up for each of the consonants /p, b, k, g/. Pseudowords were presented either before or after their corresponding real words. In total 22 pseudowords were included (see Appendix 1 for the complete list of stimuli used in this experiment).

Table 1 Example of word pairs used in Experiment 1.

Item	Stimulus	Word pair	Contrast
11	sheep	Minimal pair	Vowel
11	ship	Minimal pair	Vowel
4	brick	Pseudo minimal	Vowel
4	breek	Pseudo minimal	Vowel
37	church	Repeated word	Vowel
37	church	Repeated word	Vowel

4.7.3 *Organization of stimulus set*

The stimuli were presented in pseudo-random order. The average distance between primes and targets was 10.3 stimuli (max=27, min=2). No breaks were included in the test. Only one list was used in the experiment, so the same stimulus order was used for all participants.

4.7.4 *Recording and preparation of stimuli*

The stimuli were recorded by one male and one female speaker. They were both native speakers of American English with mid-Western accents. Although both knew some Spanish at the time of the recording, they had spoken only English until early adulthood. The speakers were in their late twenties at the time of the recording. The speakers were recorded in a quiet room, using a Behringer C-1U USB microphone connected to a laptop. The digital recordings were made using a sampling rate of 44.1 kHz and their amplitude was normalised using CoolEdit 2000 (Johnstone, 1999). Next, the words in the recordings were tagged and segmented into individual wav files using Praat (Boersma & Weenink, 2009). A 10 ms period of silence was added at the beginning of all stimuli in order to avoid any possible audio clipping.

4.8 PROCEDURE

The participants took the test individually. Each participant was provided with a pair of headphones and a USB gamepad controller, connected to a laptop computer. The use of a gamepad offered greater time accuracy than using a keyboard. The test was carried out in a quiet room. The participants were instructed to indicate for each stimulus if they were hearing it for the first time in the experiment (gamepad button 1) or if it was a repetition (gamepad button 2). The participants used their thumbs to press the buttons. They were asked to respond as quickly and accurately as they could to each stimulus. There was a 3000 ms timeout for each trial. The experiment began with a training phase that contained 10 trials. The test took approximately 15 minutes to complete.

4.9 RESULTS

In the course of the experiment, two variables were collected, response type ('first' or 'repeated') and reaction time. Although the participants were instructed to report their responses as first or repeated, the responses were actually coded as *new* and *old*. This was to avoid confusion between the levels of the factor order of *presentation* (first, second) and the *response type* (first, repeated). In the case of the second items of the minimal pairs, a correct response means that they were reported as *new*. The analysis of response type will be presented first, followed by the analysis of reaction time.

The results from three participants were discarded because they did not complete the test, and responses for two more participants were not recorded due to a technical problem. In addition, data for three further participants were excluded from the analysis based on a high level of incorrect answers and missing responses. These participants had more than 35% incorrect responses and more than 10% missing responses. In total, data from 57 participants were retained for the analysis. Minimal data cleaning was carried out for these participants, leading to the exclusion of reaction times less than 200 ms from the beginning of the presentation of the audio file (four responses in total). These responses were considered to have been given before the cognitive processes needed to carry out the task were completed. A total of 5815 data points were included in the final analysis.

4.9.1 *Analysis of response choice*

As discussed in the introduction to this chapter, one of the aims of this experiment was to find out how difficult it was for Spanish speakers to discriminate minimal pairs in English that employ acoustic features that are used differently in their native language. Of interest in this experiment was sensitivity to differences in critical absolute VOTs in the voiced/voiceless plosive contrasts and differences in spectral quality in the vowel contrast (as discussed in section 3.3, for L1 Spanish speakers, the main challenge is noticing spectral cues used to discriminate between FLEECE and KIT). If Spanish native speakers had difficulties discriminating minimal pairs involving the contrasts used in this experiment, it was expected that they would report the second presentation from a minimal pair set as a repeated word, even though it is a different word in English (e.g., reporting "cheap" as repeated after having heard"chip"). In this case, the number of responses "repeated" should be similar for both the second presentation of a repeated word as for the presentation of the second member of a minimal pair.

The first step was to look at the accuracy rate of responses to the first item of a pair. The differences in accuracy rate to the first item would help to establish a baseline with which to compare the accuracy rate for the second item of the pairs. As seen in Table 2, there was variation in accuracy rate across different word pair contrasts, with first items of repeated words having the highest accuracy rate and first members of bilabial pairs having the lowest. To find out if these differences in the proportion of correct reponses were significant, a multiple proportions χ^2 test was carried out (King, 2014; Schwarz, 2011). [2]

For this test the variations in the observed proportions are compared to an expected value, which can be the total of cases or a theoretically motivated value. In the case of responses to the first item, the observed value was the accuracy rate and the expected value was 100, which is the maximum value corresponding to all responses being correct. The result of the multiple proportion test indicated that the differences in accuracy rate for the first item across the four conditions (including repeated words) were not significant ($\chi^2(3) = 5.469, p = 0.147$). Consequently, it can be concluded that identifying the first member of the repeated word pairs (i.e., the control items) was not easier than the other contrasts. In addition, a multiple proportions χ^2 test was carried for the first member of the minimal pairs only. This test showed that the differences in accuracy rate among the different contrasts was not significant ($\chi^2(2) = 2.048, p = 0.359$), which indicates that none of first member of the contrasts was prone to produce more errors.

2 I chose to begin the analysis of the response choice with a multiple proportion test as this presents a clearer overview of the base line performance. However, mixed-effects logistic regression models were fitted to the response choice data to analyse the effect of the different factors studied in this experiment. These analysis are presented in the next section.

Table 2 Percentage of correct responses for the first and second presentation of the items in the various word pairs.

Presentation	Repeated words	bilabial	velar	vowel
1st	93.17	83.03	86.61	89.93
2nd	82.42	41.75	52.32	27.45

The next step was to compare the accuracy rates for the second items of the pairs. The accuracy rate for repeated words drops more than 10%, but it is still well above chance, which would be 50%. For comparison, this accuracy rate (82.42%) for second items is not much lower than the 84% Dufour et al. (2010) found in their study on perception of French vowels by native speakers. These results for the control items indicate that although some items were not reported as repeated, the learners did not have difficulties identifying the repeated control words.

Compared to the accuracy rate for the second items of the repeated word pairs, the accuracy rate for their second items of the minimal pairs was much lower, which indicates that the learners had difficulty discriminating between the two items of the minimal pairs, and tended overall to report the second items as *old*. Moreover, there was great variation among the contrasts, with the contrast vowel showing the lowest accuracy rate. A multiple proportion (using an expected proportion of 100) showed that these differences were significant ($\chi^2(2) = 12.928, p = 0.001$). These results indicate that vowel minimal pairs pose the greatest difficulty for the learners. This finding was confirmed by values of the standardised residuals of the χ^2 which showed that the vowel minimal pairs had the largest deviation (-2.051 vs. 1.85 for velars).

Response choice and voice repetition

One of the aims of the experiment was to explore the role that speaker variation has in the ability to discriminate minimal pairs. For this reason, the experiment included *voice repetition* as a factor. To measure the effect of *voice repetition*, the analysis now focuses on the responses given to the second item of each pair. As discussed on Chapter 3, mixed-effects models were considered the best tool to analyse the data obtained in experiments. To study the relationship between *voice repetition* and *contrast* multi-factor mixed-effects logistic regression models were fitted. The R package lme4 (D. Bates & Bolker, 2013) was used to carry out the mixed-effect logistic regressions. First, a beyond optimal model was fitted (see Appendix 1). This model included the variable *correct response* (correct = 1, incorrect =0) as the dependent variable and predictors *voice repetition* (same vs. different), and *contrast* (bilabial vs. velar vs. vowel) as well as an interaction between these factors. Also the control variables inter-item *distance*, *word frequency* (coded according to family frequency lists , Nation, 2012), *word type* (real word vs. pseudoword) and *trial number* were included. This model included the variable *participant* and *word* as random factors, as well as random slopes for *trial number* by participant. The random structure of this model included the factors participant and word, as well as random slope for trial number by participant. These two factors are commonly included in repeated sample studies to account for variation due to individual participants and stimuli. Also a slope for trial number by participant is included to account for fatigue and task experience effects along a experiment (Baayen, 2008). However, the pertinence of these factors in the random structure is ultimately determined in the model selection process.

As explained on Chapter 3, I followed the protocol proposed by Ieno et al. (2009). In this protocol a stepwise backwards selection process was carried out, in which progressively simpler models were fitted. From these, the best random structure was selected based on the lowest AIC, while the best fix structure was chosen best on the lowest ANOVA. Using this protocol, we can obtain the model that best fits the data. The optimal model obtained (see Appendix 2) returned a significant effect for the variable *distance* ($\beta = -0.028, SE = 0.008, z = -3.52, p < 0.001$). This means that the larger the numbers of trials between the members of a pair, the less likely the participants would score a correct response. In other words, the larger the number of trials the more like they would report the second item as a repetition. The factor *word type* was also significant ($\beta = 0.45, SE = 0.103, z = 4.36, p < 0.001$), which showed that real words received more correct responses than pseudowords. That is, pseudowords were more likely to be taken as repetitions than real words. The optimal model also returned a significant interaction between *voice repetition* and *contrast*. As can be seen in Figure 2, for the repeated words, the stimuli presented in the same voice had a higher change of receiving a correct response ($\beta = 0.569, SE = 0.164, z = 3.48, p < 0.001$). However, *voice repetition* had a negative effect when the second member of a minimal pair was presented. The second minimal pair members for vowels ($\beta = -0.886, SE = 0.242, z = -3.66, p < 0.001$) and bilabials ($\beta = -0.990, SE = 0.234, z = -4.22, p < 0.001$) were more likely to be perceived as repeated words when both members of a minimal pair were presented in the same voice. The effect of voice was not significant in the case of velar minimal pairs ($\beta = 0.011, SE = 0.255, z = 0.04, p < 0.96$).

4.9.2 *Analysis of reaction time*

For the analysis of reaction time, the responses to the second item of the pairs were compared to those of the first item. The reason for this was that the effects of voice repetition would be observed as differences in response times to the second members of the word pairs. The reaction times were recorded in milliseconds, however the distribution of the data was skewed towards high values. This distribution required the data to be transformed in order to fit

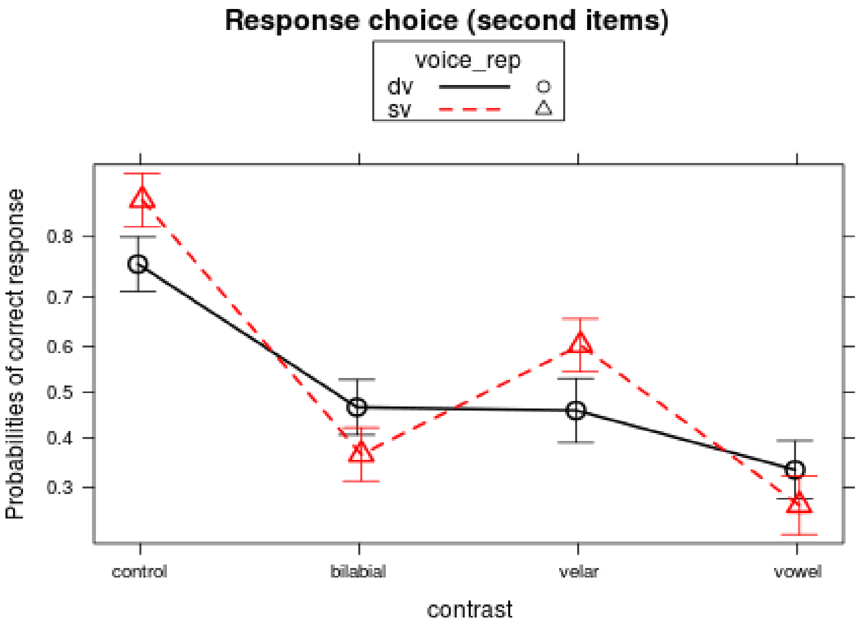

Figure 2 Effect of voice repetition on response choice for second items of word pairs (actual values).

the linear models discussed below. Following the ladder of powers (Velleman & Hoaglin, 1981), the reciprocal square root function was found to be the best method to transform the reaction time data.

As to the regression model applied, given the differences in average reaction times across the participants, using mixed-effects models was found to be the best tool. Accordingly, all the models fitted for the analysis of reaction time included participants as a random effect. Random slopes for participants against trial number were also included. This variable was present in all the models fitted. Only correct responses were included in the analysis. As in the analysis of response choice, the R package lme4 (D. Bates & Bolker, 2013) was used.

The first model fitted to the reaction times of all word pairs included the factors *voice repetition*, *presentation*, inter-item *distance*, stimulus *duration*, word *frequency* and *trial number*. Also the model included a three way interaction for *voice repetition*, *presentation* and *contrast*. The model included *participant* as a random factor and a random slope for *trial number* (centered) by *par-*

ticipant. The output of this model is included in Appendix 3. A step-wise backward selection process was followed to obtain an optimal model (see Appendix 4). This model showed that words with lower frequencies were responded to more slowly than words with high frequencies ($\beta = 0.111, SE = 0.051, t = 2.16$)[3]. Also the duration of the stimuli had a significant effect ($\beta = 0.003, SE = 0.0003, t = 10.59$), with longer stimuli receiving slower responses. There was a fatigue effect as indicated by the effect of trial number ($\beta = 0.016, SE = 0.003, t = 6.3$). Finally the interaction between *voice repetition*, *presentation* and *contrast* was significant for all contrasts. A plot of the coefficients of the variables taking part in the interaction (see Figure 3) indicates that the effects of *voice repetition* were different for each contrast. Consequently, separate analyses for each contrast were carried out.

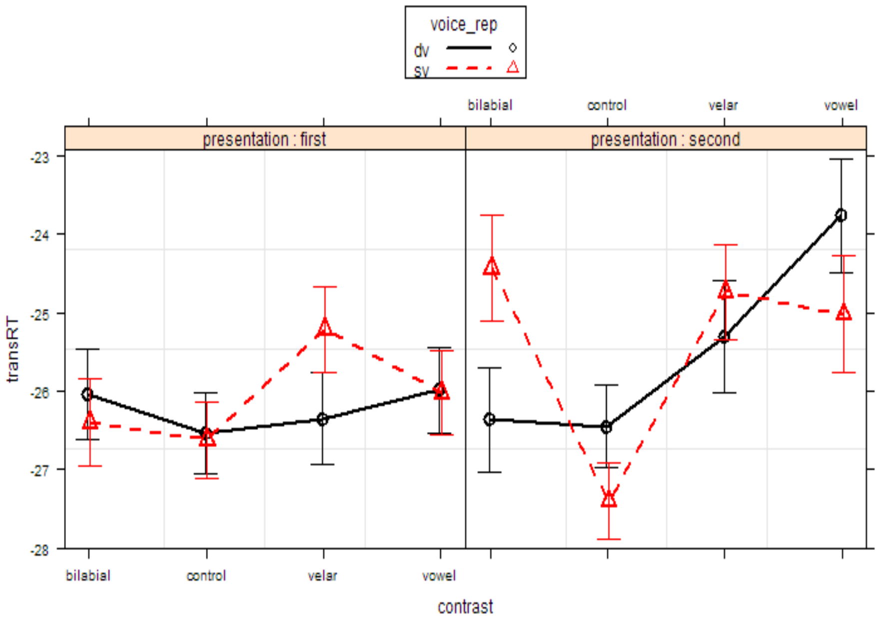

Figure 3 Effects of voice repetition on reaction time of second items of pairs (all pairs included).

The responses for the repeated words (i.e., the items serving as controls) were analysed first. The first model fitted for these pairs included the factors *voice repetition*, *presentation*, inter-item *distance*, stimulus *duration*, word

3 In linear mixed models, *t*-values above 2 or below -2 are considered to be significant.

frequency and *trial number*, as well as an interaction between *voice repetition* by *presentation*. In the selection process, the word *frequency* and inter-item *distance* variables had no significant effect and were discarded. The optimal model (see Appendix 5) showed that stimulus duration had a significant effect ($\beta = 0.005, SE = 0.0004, t = 12.69$) with responses to longer stimuli taking more time. The interaction between voice repetition and presentation was significant ($\beta = -1.123, SE = 0.254, t = -4.41$). This means that correct ('old') responses to repeated words were faster than correct ('new') responses to the first presentation of those words, but only when the first and repeated presentations were in the same voice (Table 3). This in turn suggests that the listeners had the capacity to store indexical information about the stimuli they heard.

Table 3 Summary of reaction time values (in milliseconds) for repeated words (actual values).

Presentation	**Voice repetition**	**N**	**mean**	**SD**
first	dv	414	1503.24	418.05
first	sv	623	1492.69	437.33
second	dv	335	1505.31	402.45
second	sv	575	1403.48	330.24

Next the analysis of real minimal pairs was carried out. The model fitted to this data (Appendix 6) included a three way interaction for *voice repetition*, *presentation* and *contrast*, as well as the control variables *duration*, *frequency* and *distance*. The interaction was significant ($\beta = -3.706, SE = 0.959, t = -3.86$). As seen in Figure 4, velar minimal pairs did not differ according to presentation and voice repetition. In order to further analyse the effects of voice repetition on presentation, separate analyses were carried out for each contrast.

First, a model was fitted for the minimal pairs contrasting vowels. The optimal model obtained (Appendix 7) included only the factors *voice repetition*,

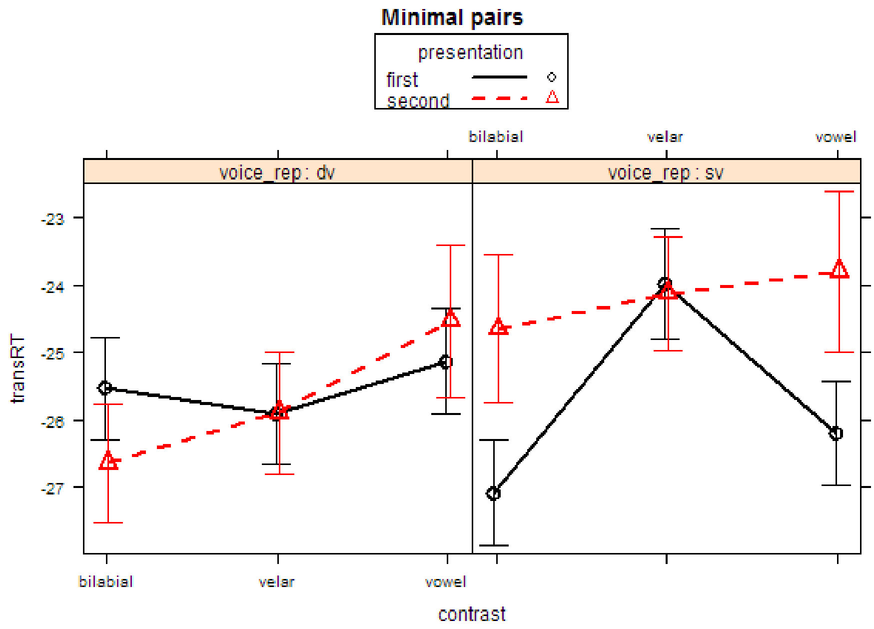

Figure 4 Effects of voice repetition on reaction time to second items of minimal pairs.

presentation, and an interaction between these variables. This interaction was significant ($\beta = 1.811, SE = 0.76, t = 2.38$). This means that it took longer for the participants to report the second member of a FLEECE-KIT minimal pair as 'new' when both members were presented in the same voice. The slower responses and the low number of correct responses for the second member of the pairs (see Table 4) indicate that it was very difficult for the learners to discriminate between the FLEECE and KIT vowels, and that the difficulty was exacerbated by experiencing a minimal pair in the same voice. This finding suggests that the participants might have focused on the indexical characteristics of the speech signal, but failed to focus on the acoustic cues that are used to contrast the English vowels used in the minimal pairs.

For the analysis of bilabial plosives, the optimal model included *voice repetition* and *presentation* as well as an interaction between these factors (see Appendix 8). The interaction showed that responses to the second item of minimal pairs were slower in the same voice condition ($\beta = 3.355, SE = 0.733, t = 4.58$). This finding is similar to the results obtained for the vowel minimal pairs, which in turn suggests that as with vowel minimal pairs, voice repetition made it more difficult for the learners to notice the differences in the acoustic cues used to discriminate these minimal pairs. As seen in Table 5, reaction times to second members of a minimal pair were much slower in the same voice condition.

However, the model fitted to the minimal pairs containing velar plosives (Appendix 9) showed no significant interaction between voice repetition and presentation ($\beta = 1.066, SE = 1.008, t = 1.06$). As seen in Table 6, the number of correct responses to second items was higher than in the case of vowel minimal pairs. However, the reaction time means for these items did not vary greatly in the same voice condition.

Finally the pseudo minimal pairs were analysed. This group of word pairs was interesting because it could give some indication of the effect that the lack of a lexical representation can have in L2 perception of phonological contrasts. As in the case of minimal pairs, first a model was fitted including the variables *voice repetition*, *presentation* and *contrast*, as well as a three-way interaction between these factors. The control variables *word duration*, and inter-item *distance* and *trial number* were also included. The optimal model (Appendix 10) excluded the factors *distance* and *duration*, but showed that the three-way interaction was significant ($\beta = -3.539, SE = 0.865, t = -4.09$).

Table 4 Reaction times for minimal pairs contrasting vowels.

Contrast	**Voice repet.**	**Presentation**	**Correct responses** (total responses)	**mean**	**SD**
vowel	dv	First	91 (113)	1707.08	403.59
vowel	dv	Second	32 (107)	1836.00	459.95
vowel	sv	First	102 (112)	1407.08	298.99
vowel	sv	second	26 (107)	1793.88	485.75

Figure 5 presents the interactions between presentation and contrast for each voice condition.

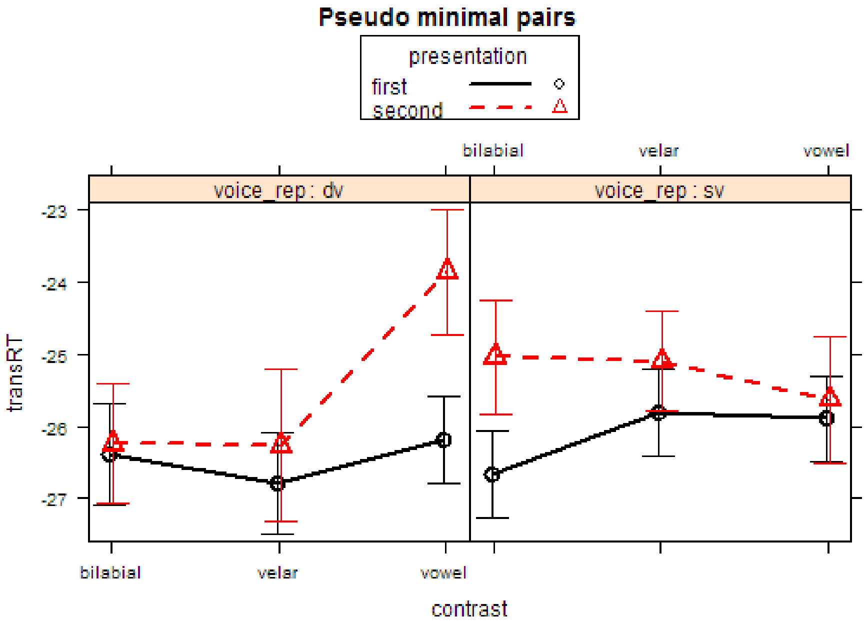

Figure 5 Effects of voice repetition on reaction time to second items of pseudo minimal pairs.

To facilitate the analysis, models were fitted separately for each contrast. The results from the vowel pseudo minimal pairs (Appendix 11) revealed a significant interaction between voice *repetition* and *presentation* ($\beta = -2.066, SE = 0.596, t = -3.47$), which means that the second stimuli presented in the same voice were responded to faster. In the case of bilabial pseudo minimal pairs (Appendix 12), there was also a significant interaction between these factors

Table 5 Reaction times for minimal pairs contrasting bilabial plosives.

Contrast	**Voice repet.**	**Presentation**	**Correct responses** (N total)	**mean**	**SD**
bilabial	dv	First	98 (110)	1545.12	447.53
bilabial	dv	second	60 (107)	1457.53	475.17
bilabial	sv	First	98 (110)	1503.33	396.59
bilabial	sv	second	37 (107)	1831.51	483.40

($\beta = 1.489, SE = 0.635, t = 2.34$). However, *voice repetition* had an inhibitory effect on the second items rather than the facilitatory effect observed for vowels pseudo minimal pairs. The model for bilabial pseudo minimal pairs also returned a significant effect for duration ($\beta = 0.005, SE = 0.001, t = 5.01$), which indicates that reaction times were slow for longer stimuli. Finally, the model fitted for the pseudo minimal pair contrasting velars (see Appendix 13) returned a non-significant effect for the interaction between voice repetition and presentation ($\beta = 0.91, SE = 0.666, t = 1.37$). Table 7 presents a summary of the mean reaction times for pseudo minimal pairs.

The most noticeable difference in reaction time between minimal pairs and pseudo minimal pairs was observed for vowels. Contrary to minimal pairs, having the pseudo minimal pairs presented in the same voice facilitated the recognition of the members as different words. One reason for this finding could have been a lack of a lexical representation for the pseudo words. In the case of minimal pairs, the students might have assumed that they were homophones, and that the acoustic differences between these words were just alternative pronunciations. Consequently, both lexical entries shared their phonological representations. On the contrary, when the learner heard a pseudoword, no representation was activated, and so it was faster for them to report as a new word.

Table 6 Reaction times for minimal pairs contrasting velar plosives.

Contrast	**Voice repet.**	**Presentation**	**Correct responses** (N total)	**mean**	**SD**
velar	dv	First	98 (112)	1643.64	475.36
velar	dv	Second	63 (106)	1673.30	459.95
velar	sv	First	84 (110)	1709.08	453.72
velar	sv	Second	63 (106)	1708.30	458.24

Table 7 Reaction times for pseudo minimal pairs (actual values).

Contrast	**Voice repet.**	**Presentation**	**Correct responses** (N total)	**mean**	**SD**
bilabial	dv	First	131 (165)	1476.47	477.74
bilabial	dv	second	67 (167)	1447.01	467.15
bilabial	sv	First	177 (222)	1548.65	469.71
bilabial	sv	second	84 (213)	1657.54	477.89
velar	dv	First	106 (112)	1519.61	442.56
velar	dv	second	37 (110)	1567.33	450.65
velar	sv	First	191 (219)	1598.84	450.54
velar	sv	second	118 (215)	1651.72	488.74
vowel	dv	First	199 (221)	1545.63	382.67
vowel	dv	second	61 (210)	1688.49	478.43
vowel	sv	First	207 (220)	1555.24	429.11
vowel	sv	second	57 (217)	1556.47	441.99

4.10 discussion

The results showed that voice repetition had an effect on L2 speech perception. Listening to a word for a second time in the same voice did facilitate lexical access. This is similar to what previous studies had found in an L1 (Goldinger, 1998; Palmeri et al., 1993). The same voice facilitation effect reveals that, even in a second language, indexical information is readily stored. Hence, the results of this experiment show that phonological memory is richer than would be predicted by extreme abstract models. The voice repetition effect supports the argument that indexical information is not discarded after the acoustic signal is mapped onto the phonological representations.

However, the effect of voice repetition poses a challenge in the L2. The difficulty is in separating indexical variation and linguistic variation. As seen in the greater number of incorrect responses and longer reaction times for second items of minimal pairs presented in the same voice, voice repetition can cause L2 learner to perceive members of minimal pairs as though they were homophones. This could indicate that although L2 representations can store fine acoustic details, these details are not fine grained enough, or not stored in the right way, to enable L2 learners to correctly categorise L2 words. Not surprisingly, the results show that some contrasts not present in the L1 are difficult to discriminate in the L2. In particular, the results indicate that for these Spanish speakers, the contrast between the FLEECE and KIT vowels is more difficult than the contrast between voiced and voiceless consonants.

The difficulty L2 learners experience in discriminating minimal pairs could be caused by their attention being directed to voice variation as a strategy to identify if a word had been repeated, which could arise from the design of the experiment. In other words, explicitly asking the participants to identify repeated words could have influenced them to disregard acoustic details used to discriminate English phonemes. This seems to be the case for pseudo minimal pairs, where the majority of the participants reported that the pseudowords were repetitions of a real word. However, the fact that in some cases the participants reported pseudowords as new words indicates that they were sensitive to the acoustic cues that distinguish the English front vowels. In other words, they were able to perceive enough phonetic detail to allow them compare a given exemplar to the prototypes in their phonological representations. This finding suggests that the fine phonetic detail can be noticed by L2 learners.

The results of this experiment helped my study in two ways. Firstly, it helped to narrow the focus of phonological contrasts to be studied. Given the high error rate in the case of front vowels, this contrast proved to be worth to be studied in the following experiments. Secondly, it showed how intermediate learners of English reacted to pseudowords that need not differ greatly from real monosyllabic words, and so provided an insight into the type of pseudowords that can be used with this population. This knowledge was integrated in the design of the second experiment, which looked more closely at lexical activation and voice variation. In the next chapter, I present and discuss Experiment 2, which compared the performance of L1 Spanish learners of English and native speakers of English. The task used in that Experiment 2 allowed me to better understand whether fine phonetic detail is accessed during spoken word recognition.

5 VOICE PRIMING IN A SECOND LANGUAGE

5.1 INTRODUCTION

The previous chapter explored the effects of voice repetition on L2 perception. The results obtained in Experiment 1 showed that L1 Spanish learners of English had difficulties perceiving the acoustic cues that are used in English to discriminate between voiced and voiceless plosives and between the two close front vowels in General American English. The greatest difficulties, as revealed by the error rates, were observed in the set of minimal pairs contrasting the KIT and FLEECE vowels. Furthermore, the results obtained in Experiment 1 showed that L2 learners were faster at identifying a word presented for a second time when the same voice was used in the first and second presentation. The facilitation of voice repetition observed in Experiment 1 suggests that L2 listeners were able to store fine phonetic detail about the stimuli they heard.

However, the findings of Experiment 1 have a limitation. Given the task of the experiment, which was to report repeated words, L2 listeners might have focused their attention on the idiosyncratic features of the voices as a strategy to identify repeated items, while ignoring the acoustic cues needed to discriminate the members of a minimal pair. Moreover, the task did not ensure that the participants had to activate their lexical representations. But even if they did activate a lexical entry, the task involved participants having to decide if the stimuli they had just heard was new or repeated. This kind of task also involves access to an episodic memory. As Morais (2010, p. 65) argues, "one should not accept the claim that lexical access involves non-linguistic episodic traces when the task, besides or beyond lexical access, requires access to these episodic traces." Considering the limitations of Experiment 1, I designed a new experiment that allowed me to investigate the ability to store fine phonetic detail without encouraging participants to remember the stimuli.

The experiment discussed in this chapter (henceforward, Experiment 2) aims to find out how rich the acoustic detail is in the phonological representations developed by learners of a second language. The approach taken to find this out is twofold. On the one hand, I investigate whether L2 learners can store acoustic detail that is related to a particular speaker (e.g., pitch, loudness, and differences in the power spectrum related to the timber of the voice), which facilitates the recognition of words presented in the same voice. On the other hand, I explore whether L2 learners can store acoustic detail that helps them learn new phonological categories in the target language. For example, in the case of English vowels, this means that the learners are able to store acoustic detail, such as formant structure and vowel duration, that allows them to discriminate between the minimal pairs such as *sheep* and *ship*.

Previous studies have found that bilingual speakers experience priming between the members of a minimal pair if this contains phonemes that are difficult for them to distinguish in the target language. For example, in a priming experiment using Catalan stimuli containing phonological contrasts difficult for L1 Spanish speakers, Pallier et al. (2001) found that Spanish dominant speakers responded more quickly to the word *neta*, /ˈnetə/ "clean", when this word was preceded by the word *néta* /ˈnɛtə/ "granddaughter". However, these studies – which concluded that L2 speakers do not store acoustic detail – employed only one voice to present the stimuli. I revisit these conclusions by including two different speakers for the presentation of the stimuli in my experiment. The inclusion of different speakers allowed me to find out more about the granularity (i.e., the amount of acoustic detail) of the phonological representations in a second language. If a facilitation effect was observed using the same voice in the case of L2 learners, this would be an indication that acoustic information does permeate into the lexicon.

The experiment described below consists of a lexical decision experiment involving mid-term memory lexical priming in which minimal pairs and repeated words were used as stimuli. The stimuli were recorded by two speakers of General American English. The experiment focused on the perception of the English close front vowels (FLEECE and KIT) and the mid and open-mid front vowels (DRESS and TRAP) by native speakers of Spanish who were learning English as a second language.

5.2 EXPERIMENTAL DESIGN

The experiment used a repeated measures design and included two groups: a group of native speakers of English and a group of English learners who were native speakers of Spanish. The variables included in the experiment were:

- *Presentation*: This variable indicates whether a word was presented for the *first* time or for the *second* time in the experiment.
- *Word pair*: Two kinds of word pairs were included in the experiment: minimal pairs and repeated words.
- *Voice pair*: refers to whether the words in a pair were presented in the same voice (sv) or in two different voices (dv).
- *Vowel height*: The experiment considered the contrasts between the close front vowels FLEECE and KIT on the one hand, and the mid front vowel DRESS and the open-mid front vowel TRAP on the other. In the analysis the first two vowels are coded as *close* and the last two as *open*.

5.3 PARTICIPANTS

Twenty five native speakers of American English and 23 Spanish speakers who were learning English as a foreign language took part in this experiment. The learners were required to be native speakers of Spanish who started learning English after adolescence. They were recruited among students majoring in Modern Languages at the University of Los Andes in Venezuela. The students had taken at least three semesters of English at university level, which represents 300 hours of English instruction. Two of the English learners were discarded due to their high number of missing responses (more than 70%). Since the stimuli were recorded by speakers of General American English, the participants recruited for the English native group were also required to be familiar with this accent. Therefore, the participants for this group were recruited among exchange students from American universities visiting Victoria University of Wellington. All the participants reported normal audition. There were 15 female and 10 male English native speakers, and 15 female and 8 male Spanish native speakers. The average age was 22 years (min = 17, max = 34, *SD* = 3.22).

5.4 METHOD

The participants were asked to carry out a lexical decision task. They were presented with a series of spoken words, and they had to decide whether each word was a real word or a non-word. As in Experiment 1, a gamepad connected via the USB port to a PC was used to collect the participants' responses. The software used to manage the presentation of the experiment was a multi-platform and open source program called Open Sesame (Mathôt, Schreij, & Theeuwes, 2012). This program offers good time resolution (down to one millisecond) and more flexibility than Perceval (André et al., 2003). For example, one of the advantages of Open Sesame's flexibility was that a script could be set up so that demographic information could be collected as part of the experiment session. Also different types of feedback (image, sound or both) could be implemented for individual blocks. This was important for the practise block, which included visual feedback after each trial (the word *correct* printed in green in the middle of the screen). Likewise, the two different lists used in the experiment were automatically loaded without the researcher having to manually load different scripts during the course of the experiment. Anonther advantage of Open Sesame is that it is possible to include scripts written in Python in an experiment. I used this flexibility to write a plugin for Open Sesame that allowed me to collect responses using a gamepad (see Appendix 1).

5.5 MATERIALS

One hundred and twenty English monosyllabic words were selected for this experiment. The stimuli consisted of 40 minimal pairs and 40 pairs of repeated words. There were 20 minimal pairs for the lexical sets KIT and FLEECE and 20 minimal pairs for the lexical sets DRESS and TRAP. The repeated pairs included 10 words for each of the lexical sets KIT, FLEECE, DRESS and TRAP. The repeated words did not have a minimal pair involving the relevant contrast.

The frequency of the words was calculated from the spoken part of the Corpus of American English (Davies, 2008) which contains 90 million words as to July 2012. The minimum frequency of the words selected (i.e., "telly") was 20 (0.22 per million) and the maximum was 1707000 (1896 per million), with a mean of 7195 (78 per million). Since the Spanish learners of English would have a smaller vocabulary in English, the stimuli were selected among

Table 8 Real words use in Experiment 2, classified according to BNC/COCA word frequency lists.

BNC-COCA list	Percentage of total	Number of words
1st 1000	50.86	59
2nd 1000	26.72	31
3rd 1000	3.45	4
4th 1000	8.62	10
5th 1000	3.45	4
6th 1000	4.31	5
7th 1000	2.59	3

the seven thousand most frequent words in English according to the BNC/-COCA word families frequency lists (Nation, 2012). Also, care was taken to avoid having minimal pairs whose members differed greatly in frequency. For example, both members of the minimal pair "feel - fill" are among the first thousand most frequent words in English. Table 8 presents a breakdown of number of the words selected by frequency rank according to the BNC/COCA frequency lists.

For the lexical decision task, 120 pseudowords were created. The pseudowords were monosyllabic and had legal syllable structures in English. None of the pseudowords formed a minimal pair with a real word using the other vowel from the contrast being studied. For example, the pseudowords "sprim" and "spreem" do not form a minimal pair with real English words containing the FLEECE or KIT vowels. This condition was included since the results of Experiment 1 showed that L2 speakers could be primed by pseudowords if the real word and the pseudoword contrasted in vowels which were difficult to discriminate. The pseudowords were grouped similarly to the real words. That is, there were pairs of repeated pseudowords and of pseudo minimal pairs.

The minimum number of trials between the members of a given pair was seven and the maximum 15 (*M* = 11, *SD*= 2.05). The number of trials between items is similar to that used previous in experiments using a similar paradigm (e.g., Dufour et al., 2010; Pallier et al., 2001). Two lists were created to counterbalance the effect of voice repetition. This means that the pairs that were presented in the same voice in one list were presented using two different voices in the other list. Ten additional words and pseudowords having the characteristics of the stimuli described above were used in the practice block and as the first five trials of the test blocks. A list of the stimuli used in this experiment can be seen in Appendix 2.

One male and one female native speaker of General American English recorded the stimuli. These speakers were different from those used in Experiment 1. The recordings were digitised at a sampling rate of 44.1 kHz and their amplitudes normalised. The stimuli were segmented and tagged following the same procedure described for Experiment 1.

5.6 procedure

The participants took the test individually in a quiet room. The stimuli were presented binaurally over headphones. The participants sat in front of a computer and were given the instructions on the computer screen. They were instructed to hold a gamepad with both hands and to use their index fingers to press the buttons on the gamepad. They were told to press the right button (button 8) if they heard a real word, or to press the left button (button 7) if they heard a nonword. This button assignment was reversed for left-handers. They were asked to respond as quickly and accurately as possible. When I piloted the experiment, I found that the term pseudoword difficult for the participants to understand, and that the term "nonword" was much easier. For this reason, the latter was used in the instructions. However, the term "pseudoword" is used in this analysis.

Table 9 Percentage of responses by native language. (Number of responses between parenthesis).

Language	Time out	Button 5	Button 6	pw	rw
English	0.32 (25)	0.00 (0)	0.00 (0)	48.80 (3831)	50.88 (3994)
Spanish	0.94 (62)	0.58 (38)	0.49 (32)	38.29 (2525)	59.71(3937)

The experiment started with a practice block of 20 trials in which the participants received feedback after each trial. The practice block was followed by a block of 190 trials and a block of 160 trials. The participants could take a short break (less than 5 minutes) between blocks. There was a 3000 ms timeout for each trial. The test took approximately 25 minutes to complete.

5.7 RESULTS

5.7.1 *Response accuracy*

In the course of the experiment two variables were collected, response accuracy and reaction time. As explained in the procedure section, the participants had to decide whether each stimulus they heard was a real word or a pseudoword by pressing the relevant button on the game-pad. Missing responses occurred when the participants did not press a button at all. Table 9 shows that both groups of participants had a very low percentage of missing responses, which indicates that they had enough time to complete the task within the 3 seconds allowed for each trial. However, as the other buttons on the gamepad were not disabled during the experiment, some of the learners pressed other buttons beside 7 and 8. Further inspection showed that these responses were given by two participants only. It is possible that they pressed these buttons inadvertently since the rest of their responses followed the pattern observed in the other learners.

Table 10 Number of correct responses.

Language	Word type	Total	N correct	Percentage
English	pw	3930	3502	89.11
English	rw	3887	3563	91.66
Spanish	pw	3276	2006	61.23
Spanish	rw	3255	2699	82.92

In the analysis of accuracy, only real word (rw) and pseudo-word (pw) responses were included. Logistic regression mixed-effects models were used to carry out the analysis of response accuracy. The statistical software used was R and the package lme4 (D. Bates & Bolker, 2013). The protocol recommended by Ieno et al. (2009) was followed to fit the models.

It was expected that the accuracy of the learners would differ significantly from that of the native speakers, considering that it would be more difficult for the former to distinguish between real words and pseudowords given their more limited vocabulary.

As seen in Table 10, which presents the total number of correct responses, the probability of correct responses appears to be related to the native language of the participants and the type of word. To confirm this, a model was fitted which included the dependent variable *accuracy* (coded as 0 for incorrect responses and 1 for correct responses) and *native langua*ge and *word type* as factors, together with an interaction between these factors. Also the factors *trial* and word *frequency* were included. For the analysis of accuracy, all the models included the random effect *participant* and *stimulus*. The results of the optimal model (see Appendix 1.) showed that the learners performed less well than the native speakers, with the learners being less likely to give correct responses ($\beta = -1.886, SE = 0.132, z = -14.259, p < 0.001$). Secondly, real words were more likely to receive a correct response ($\beta = 0.277, SE = 0.133, z = 2.08, p < 0.038$). The coefficient for the interaction ($\beta = 0.983, SE = 0.101, z = 9.77, p < 0.001$) showed that the difference between real word and pseudoword accuracy was more marked for the learners, primarily because they were more likely than the native speakers to give an incorrect "rw" response when the stimulus was a pseudo word. In other words, the learners were more likely to produce false positives.

The next step was to investigate if the other factors studied in this experiment had an effect on response accuracy. These factors were *presentation* (first or second presentation of any member of a pair), *vowel height* (close or open), *voice repetition* (same or different voice), *word pair* (real or pseudoword) and, in the case of real words, *word frequency*. In order to simplify the analysis of the effects these factors had on each the native speakers and the learners, both groups were analysed separately. This decision was justified considering the significant differences in overall accuracy reported above.

The analysis of these factors started with an additive model for the responses of the native speakers to real words. The variables included were: *vowel height*, *word frequency* (logarithmically transformed), *presentation*, *word pair* and *voice repetition*. Through a backward model selection, a minimal model was obtained which showed that the variables *voice repetition* and *vowel height* did not contribute significantly and so were excluded from the final model (Appendix 2.). The additive model showed that, compared to the minimal pairs, the possibility of correct responses was higher for the repeated word pairs (β = 0.606, SE = 0.237, z = 2.56, p < 0.01). Also, the second items of the pairs received more correct responses ($\beta = 0.410, SE = 0.156, z = 2.63, p < 0.01$). The results of this model showed that *frequency* nearly reached significance ($\beta = 0.269, SE = 0.157, z = 1.716, p < 0.086$). The lack of effect was probably due to the high frequency of the words used in the experiment. In order to find out if response accuracy increased the second time an item was presented, an additive model was compared to a model including an interaction between the factors *word pair* and *presentation*. However, a model comparison determined that the interaction was not justified.

In the case of the responses to pseudowords, an additive model was fitted with the same variables analysed for the real words, except *word frequency*. This model was compared to a model with an interaction between *presentation* and *word pair*. An ANOVA of the deviance of both models indicated that the model with the interaction was more adequate ($\chi^2(1) = 8.997, p < 0.01$) and so this model was kept (see Appendix 3. for complete output). The results showed that repeated pseudowords received more correct responses than pseudo minimal pairs ($\beta = 0.848, SE = 0.146, z = 5.790, p < 0,001$). However, the interaction showed that accuracy fell when a repeated pseudoword was presented for the second time ($\beta = -0.638, SE = 0.209, z = -3.04, p < 0.01$). This means that incorrect responses were more likely to be given to a the second member of repeated pseudoword than to the second member of a pseudo minimal pair. This model also showed that pseudowords with open vowels were less likely to obtain a correct response ($\beta = -0.729, SE = 0.142, z = -6.89, p < 0.001$).

The same process was followed for the analysis of the results from the learner group. The responses to real words were analysed first. Stepwise backward selection was used to find the best models, using the same factors and interactions as in the native speaker models. An additive model was compared to a model with the *presentation* and *word pair* interaction using an ANOVA for the total deviance of each model. This test indicated that the interaction was not justified. During the selection process, it was found that the *voice repetition* and *presentation* did not have a significant effect. The optimal model obtained (see Appendix 4) revealed that as the frequency of the words increased, so did the odds ratio of the response being correct ($\beta = 0.279, SE = 0.112, z = 2.49, p = 0.013$). Furthermore, there was a higher likelihood that responses to repeated words would be correct ($\beta = 0.420, SE = 0.167, z = 2.51, p = 0.012$). A difference from the results for the native speakers was that vowel height had a significant effect. In other words, the learners were more accurate in their responses to words with close vowels ($\beta = -0.677, SE = 0.164, z = -4.125, p < 0.001$). This finding could be the result of pseudo-homophony in the learners. Since they have two lexical candidates in the same 'space' (e.g., ship and sheep), lexical items containing the merged FLEECE/KIT contrast would be highly activated, making a 'yes' response more likely. Likewise, since they were able to discriminate between the DRESS and TRAP vowels, they should have noticed that there was a larger number of items contrasting these categories, which may had caused them to be more hesitant to accept stimuli having these vowels as real words.

In the case of the learners' responses to pseudowords, an additive model was fitted and compared to a model that included the predictor variables *vowel height*, *voice repetition*, as well as the interaction between *word pair* and *presentation*. The latter model offered a better fit ($\chi^2(1) = 2.24, p < 0.05$). The results of this model (see, Appendix 5) showed no facilitatory effect in having the members of a pair presented in the same voice. Moreover, neither did *vowel height* contribute significantly to the model, which indicates that the higher accuracy previously discussed for real words with close vowels was related to their lexical representations sharing the same perceptual space. However, no such difference was observed for pseudowords as they lack lexical representations. However, the model returned a significant interaction between *presentation* and *word pair* ($\beta = -0.366, SE = 0.147, z = -2.494, p = 0.012$). As with the group of native speakers, there was an increased chance that the learners would give an incorrect response when they heard a pseudoword for the second time.

Summing up, in the case of real words, both groups performed better with the repeated words than with the minimal pairs. In the case of pseudowords, when they heard a stimulus a second time, they were more likely to report it as a real word. A significant interaction between presentation and word pair for both native speakers and learners suggests that both groups of participants were able to store stimuli they had heard only once in memory. Similarly, voice repetition had no effect on accuracy, neither in the case of real words nor in the case of pseudowords. The main difference between the speaker groups was the larger number of false positives given by the learners.

5.7.2 *Reaction time*

Two types of priming were examined in this experiment: form priming and speaker identity priming. The former refers to the facilitatory effect observed when a word is presented for a second time, while the latter refers to the additional facilitatory effect resulting from presenting a word the second time in the same voice. In the reaction time analysis, I investigate whether word repetition and voice repetition effects are present in the lexical decision task.

The statistical tool used for the analysis of reaction time was linear mixed-effects models. The software used was the package lme4 (D Bates & Bolker, 2013) that runs in the R language.

The reaction time was measured from the onset of the audio file until a button was pressed. Since the stimulus started 10msec into the audio file, the actual responses from the onset of the stimulus are 10msec shorter than those reported below. For the analysis of reaction time, only correct responses to real words were included. This is customary in lexical decision task experiments since priming is expected to happen in the case of real words, and incorrect responses may involve strategies unrelated to the lexical access process being studied (e.g., Barca & Pezzulo, 2012; Forbach et al., 1974; Harrington, 2006; Perea & Carreiras, 2003). Six responses with reaction times faster than 500 *ms* and more than two standard deviations from the mean (M = 1128.52, SD = 304.62) were excluded as these were likely to be given before lexical activation had taken place.

Before fitting models to the reaction time values, the data were explored to determine whether they satisfied the assumptions of linear models, namely, normal distribution of the dependent variable, independence of observations, homogeneity of the data, and no collinearity between the predictor variables. The dependent variable was not normally distributed (skewness = 1.39, kurtosis = 5.29). Consequently, the Box-Cox power transformations (Velleman & Hoaglin, 1981) were applied to determine the best transformation. The best transformation was the inverse transformation[1]. The transformed response time variable is referred to as *transRT* in the rest of this section.

In order to examine the possibility of collinearity between predictors, a variance inflation factor (VIF) was estimated for an ordinary least squares regression model fitted for the response variable *transRT* and the predictor variables *presentation*, *native language*, *duration*, *logfreq*, *word pair*, and *trial number*. None of the VIF values was greater 1.4, which indicates that there was no collinearity between the predictors. For comparison, Zuur, Ieno and Elphick (2010) propose a VIF threshold value of 10, and that a value of 3 can be used when a very stringent approach is needed.

It is a common practice to center predictor variables in linear regression. Centering helps in interpreting regression coefficients when the scale of the predictors does not contain zero, as is the case with word frequency or stimulus duration. Moreover, grand mean centering helps in reducing collinearity

1 The distribution of the reaction time data was positively skewed, so the inverse transform function was used to normalise the distribution of the data. However this function returns very small values with the opposite polarity of the original data. To make these values easier to understand, the results of the transform were multiplied by -1000000.

when models include interaction terms. However, centering requires careful consideration of the method used for centering. The methods for centering most frequently used are grand mean centering and group mean centering. In mixed-effects models, the predictor variables can be classified into levels. Level-1 variables are those that are related to the individual (e.g., age of a given participant), while level-2 variables are cluster or group variables (e.g, native language of the group of participants). Level-1 variables can be centered on the grand mean or on the group mean. Group centering may be used when "[...] the most appropriate form of centering in situations in which the primary substantive interest involves a Level 1 (i.e., person level) predictor" (Enders & Tofighi, 2007, p. 128). However, if the model includes interactions between level-1 variables and level-2 variables, only grand mean centering is used. Finally, level-2 variables can only be centered on the grand mean. Since the variables in the models fitted were level-2 or cluster-level covariates (i.e., predictor and control variables such as *voice repetition*, *word pair* and *duration*), and their interaction with level-1 factors (i.e., participant, word) were included in the preliminary models, centering on the grand mean was the only possibility. The variables centered were stimuli *duration* and *trial number*. In the following discussion, the analysis was carried out on subsets of the general data. According, centering was calculated based on the grand mean of each subset.

Form priming

The reason to include items presented in the same voice was to determine the effect of form priming. It was expected that when repeated words were presented in the same voice, reaction times to the second member of a pair would be faster, which would indicate an effect of form priming. On the contrary, in the case of minimal pairs no priming should be observed as the participants will activate two different lexical entries (provided of course that they hear the minimal pair members as distinct words).

A difference in performance was expected to be observed between the English native speakers and the Spanish native speakers given the differences between languages. As discussed in the Methodological considerations (Chapter 3), Spanish has only one vowel category in the close front vowel space. Hence, it was expected that the Spanish native speakers would have difficulties discriminating between the English minimal pair members that involved a contrast between the close front vowels. If this was the case, their responses to the second member of a minimal pair with close vowels should show a priming effect. However, no priming should be observed in the case of the minimal pairs containing open vowels. The Spanish language has vowel categories for front mid and open vowels, so it was hypothesised that Spanish learners of English would assimilate to these vowel categories the English vowels DRESS and TRAP.

Based on the reasoning presented above, the data was analysed in four subsets. The subsets were: repeated word pairs with close vowels, minimal pairs with close vowels, repeated word pairs with open vowels, and minimal pairs with open vowels. Table 11 shows the average reaction times to the items presented in the same voice.

Reaction time for open vowels

The analysis of the stimuli containing open vowels is presented first. It was expected that with the stimuli containing open vowels, both language groups would behave similarly. In other words, this set of stimuli functioned as a form of control to compare native speakers and learners because it was less likely that the learners would assimilate the English vowels DRESS and TRAP

Table 11 Mean reaction times in milliseconds for stimuli presented in the same voice.

Language	Height	Word pair	Presentation	N	Mean	Priming	SD
English	close	minimal	first	213	1061.28		247.65
English	close	minimal	second	206	1050.54	10	271.02
English	close	repeated	first	222	1004.38		252.43
English	close	repeated	second	225	970.12	34	208.90
English	open	minimal	first	208	1084.00		274.61
English	open	minimal	second	216	1103.04	-29	251.75
English	open	repeated	first	244	1032.27		229.62
English	open	repeated	second	249	1015.99	16	201.69
Spanish	close	minimal	first	171	1120.11		288.72
Spanish	close	minimal	second	173	1024.72	95	269.23
Spanish	close	repeated	first	180	1022.93		266.02
Spanish	close	repeated	second	183	975.48	48	235.13
Spanish	open	minimal	first	143	1132.01		291.71
Spanish	open	minimal	second	152	1103.76	28	265.35
Spanish	open	repeated	first	177	1118.59		255.64
Spanish	open	repeated	second	172	1061.00	57	232.72

into a single category. Consequently, no priming should be observed between minimal pairs for this stimuli set.

The first analysis aimed to find out whether the experiment produced a form priming effect. This effect was expected to happen when the participants were presented with the same word for a second time. If priming occurred, the reaction time to the second presentation of a repeated word should be faster than the reaction time to the first presentation.

To analyse the effect of listening to a word for a second time, a model was fitted for the response variable *transRT* which included as fixed factors the predictors *native language*, *presentation*, *logfreq* (logarithmic frequency of the words), *duration* of the stimulus in seconds, *trial number*, as well as an interaction between *native language* and *presentation*. The random structure included random intercepts for *participant* (the participants), *stimulus*, as well as random slopes of each of the predictors included in the fixed structure for participant. The output of this model is presented in Appendix 6. The optimal model reached for the repeated words with open vowels was an additive model which included the effects of *presentation*, *native language*, stimuli *duration* and *trial number*, as well as random slopes of *trial number* for participants.

The results of the optimal model presented in Appendix 7, do not include the predictor word *frequency* because it was found to be not significant during the model selection process. In fact, this was the case for all the other subsets of pairs. This was probably because the stimuli were selected from amongst the most common words in English. The *duration* of the stimuli, however, had a significant positive effect ($\beta = 517, SE = 132, t = 3.9$). The results also show that there was a significant negative effect for *presentation* ($\beta = -28.43, SE = 13.99, t = -2.03$), which indicates that the items presented for the second time were responded to more quickly. Also the learners in general had slower reaction times than the native speakers. However, the interaction between *native language* and *presentation* was not significant ($\beta = -34.80, SE = 21.79, t = -1.60$), meaning that despite the differences in how fast each language group was, both were primed by the repeated words. The complete output of the optimal model is included in Appendix 8.

The next subset analysed was that of minimal pairs containing open vowels. The same procedure described in the analysis of repeated words was followed. The optimal model included the predictors *presentation* and *native language*, an interaction between these predictors, as well as the control variables item *duration* and *trial number*. As in the previous model, random slopes for trial number were allowed for each participant. The results of the optimal model (Appendix 9) showed that the interaction was not significant. Only the factors *duration* of the stimuli had a significant effect ($\beta = 512.31, SE = 119.92, t = 4.27$). The absence of an effect for *presentation* indicates that no facilitatory effect occurred between the members of the minimal pairs. Contrary to the case of repeated words, the responses to the second member of a pair were not faster than responses to the first one. This means that both groups were capable of discriminating between the members of a minimal pair, since listening to one member does not activate the other. Furthermore, the results from this model confirmed that the presentation effect observed for repeated words was indeed due to form priming.

Reaction times for stimuli with close vowels

The next step was to analyse the reaction times for the stimulus set containing close vowels. In this case, we find significant differences between the language groups. Given the difficulty Spanish speakers have in discriminating between the English vowels FLEECE and KIT, a priming effect between minimal pairs was expected.

In the case of the subset of repeated words containing close vowels, the optimal model obtained was an additive model. This means that the interaction term between native language and presentation was not statistically justified. This model included the predictors *presentation*, *native language* and stimulus *duration*. The results of this model showed that there was a significant effect for *presentation* ($\beta = -46.243, SE = 12.574, t = -3.68$), which suggests again that the participants were primed by the first member of the pair of repeated words. Also, it showed that there was no difference in reaction time between the two language groups. The full output of this model is presented in Appendix 10.

Finally, the same protocol was followed in the analysis of the subset of the minimal pairs containing close vowels. The optimal model (Appendix 11) included the *presentation*, *native language*, stimulus *duration* and *trial number*, as well as the interaction between *presentation* and *native language*. The interaction between *presentation* and *native language* was significant ($\beta = -64.62, SE = 25.55, t = -2.53$). The direction of this interaction indicates that the Spanish speakers were more strongly primed by the first member of the minimal pairs. The difference in the effect of presentation can be observed in Figure 6.

5.7.3 *Voice priming*

The other question that this experiment looked at was the possibility that the participants were able to store fine phonetic detail of the stimuli they were presented with. In particular, I wanted to investigate the possibility that the participants would respond more quickly if the items were presented in the same voice, because this would indicate that the participants were capable of storing the phonetic detail associated with a particular speaker. To find out if this was the case, the effect of voice repetition on the reaction time

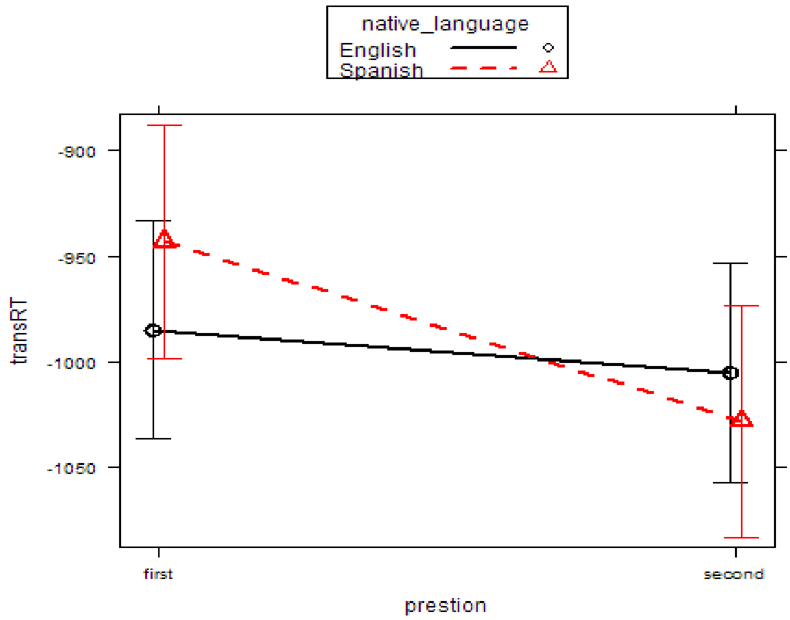

Figure 6 Effects of presentation by native language in minimal pairs containing close vowels

Table 12 Summary of reaction time values for second items of repeated word pairs.

Native language	Voice repetition	N	Mean	Δ RT	SD
English	dv	474	1018.11		265.91
English	sv	474	994.22	24	206.21
Spanish	dv	359	1066.06		292.29
Spanish	sv	355	1016.91	39	237.53

of the repeated word stimuli presented for the second time was analysed. Table 12 presents the average reaction time values for the second members of the repeated word pairs, collapsing across the two vowel contrasts.

As in the previous cases, the backward stepwise selection process was followed. The beyond optimal model included the predictors *voice repetition*, *native language*, an interaction between these two predictors, and the control variables *word frequency* (logarithmically transformed), stimulus *duration*, and *trial number*. The random structure comprised *participant* and *item* as well as random slopes for each of the fixed factors by *participant*. The results of this model are presented in Appendix 12.

After applying a backward stepwise selection process, an optimal model was fitted. This was an additive model with the following predictors: *voice repetition*, *native language*, *vowel height* and stimulus *duration*. The results of this model showed that the items presented in the same voice were responded to faster than those presented in different voices ($\beta = -18.157, SE = 8.512, t = -2.1$). The model also shows that items with open vowels showed longer reaction times than those with close vowels ($\beta = 64.085, SE = 20.228, t = 3.2$). This difference might at least in part be due to the difference in stimulus *duration* between the words with close vowels (M = 592.4 ms, SD = 89.9) and open vowels (*M* = 615.4 ms, *SD* = 88.44). The complete output of this model is included in Appendix 13.

Next, the analysis of reaction time was carried out separately for each language group. The optimal model fitted for the learners group returned significant effects for *vowel height* ($\beta = 89.783, SE = 23.83, t = 3.77$), for *voice repetition* ($\beta = -30.541, SE = 13.127, t = -2.33$), and for duration ($\beta = 513.704, SE = 130.576, t = 3.93$). The former result indicates that there was a similar voice facilitation effect. A model including an interaction between *vowel height* and *voice repetition* was fitted to find out if the voice priming was related to a particular vowel set. This model showed that this interaction was not significant ($\beta = -16.143, SE = 26.342, t = -0.61$).

Although the voice facilitation effect was observed for the learners, the model fitted for the L1 English group showed no significant effect for voice repetition ($\beta = -9.948, SE = 10.897, t = -0.91$). The predictor *vowel height* almost reached significance ($\beta = 45.493, SE = 24.121, t = 1.92$), which as mentioned before could reflect that duration of the stimuli having this contrast was longer. A model including an interaction between *voice repetition* and *vowel height* was fitted, but it returned a non-significant interaction ($\beta = -40.22, SE = 21.777, t = -1.84$). These results indicate that voice facilitation was present for the learners only. A possible cause of this difference is presented in the following section. The complete outputs of the models fitted for each language group are included in Appendices 13 and 14.

5.8 DISCUSSION

The results of this experiment show that the Spanish speakers were primed by minimal pairs that included a phonological contrast (FLEECE VS. KIT) that is not used in their native language. Another important finding was that the learners could recognise a repeated word more quickly if it was presented in the same voice. These results are similar to those obtained in the experiment presented in Chapter 4. However, one of the concerns in that experiment was that the instructions given to the participants, namely to indicate whether each stimulus was new or repeated, had conditioned them to be more sensitive to the similarities between the members of a pair. To rule out this possibility, a lexical decision task was used in the experiment discussed in this chapter.

Previous studies using the lexical decision paradigm (e.g., Brunellière, Dufour, & Nguyen, 2011; Pallier et al., 2001; Weber, Broersma, & Aoyagi, 2011) concluded that lexical representations were abstract, and so only linguistically relevant information was stored in the lexicon. Form priming by minimal pairs in second language speakers was seen as further evidence for abstract representations. Since second language speakers showed word priming effects between the members of minimal pairs, these studies concluded that second language speakers normalised L2 input according to their L1 phonological settings, which caused the inability to learn a new phonological contrast in the L2. However, the results of the present experiment showed that voice identity also had a facilitatory effect. This effect means that the participants were able to store and access more phonetic detail during lexical activation than would be predicted by an abstract model. This finding is particularly relevant in the case of the Spanish L1 learners of English.

Although the results of this experiment showed that L2 speakers can store fine phonetic detail in their phonological representations, it is important to highlight the difference in voice facilitation observed in this study between the language groups. L1 speakers did not show a same-voice facilitation effect. This result is similar to that obtained by McLennan and Luce (2005) who found that voice priming was related to the task the participants had to carry out. They found that voice repetition had an effect when the task was difficult or slow, for example when pseudowords were very word-like in a lexical decision task. They concluded that there were time course differences in the activation of episodic and abstract representations. They proposed that abstract representations are activated first. The differences between groups found in the experiment discussed in this chapter could also be explained by how difficult the task was for each group. It was certainly more difficult for the learners to identify pseudowords. They would therefore have depended more on their episodic representations to complete the task. On the other hand, native speakers found the task easier, as seen in the much higher accuracy rates and shorter reaction times. Consequently, they could have made the selection between real word and pseudowords by accessing their abstract representations only.

However, the capacity adult second language learners have in storing phonetic detail is at odds with the difficulty they experience in forming new phonological categories in a second language. From an exemplar perspective, one would predict that by itself, this capacity to store phonetic detail would eventually lead to the formation of new categories in the L2. This does not seem to be the case with the English learners who took part in this experiment. However, it is necessary to consider that their exposure to the English language had been limited as they were learning English in monolingual Spanish-speaking society.

Considering the results obtained in Experiment 1 and 2 which showed the ability L2 learners had to store fine phonetic detail from different speakers, it is necessary to investigate whether L2 learners can use this information when learning new categories in their L2. Likewise it is important to investigate whether this information is specific to a given speaker (which would indicate that L2 learners' phonological representations are episodic) or it can be generalised to other voices and words (which would indicate that L2 phonological representation are abstract). In the next chapter, I present the results of an experiment designed to find out whether voice variation can affect the formation of new phonological categories in a second language.

6

EFFECTS OF VOICE FAMILIARITY ON L2 PHONOLOGICAL LEARNING

6.1 INTRODUCTION

The results from Experiment 1 showed that Spanish L1 learners of English had difficulties distinguishing minimal pairs in English that contrast the close front vowels FLEECE and KIT. Moreover, it was found that this difficulty was exacerbated when the minimal pairs were presented in the same voice. This suggests that the participants might have attended more to the similarities in the voice of the speaker than to the differences in the formant frequencies of the vowels. The effect of voice might have been caused by the experimental task, which explicitly asked the participants to identify repeated words.

Experiment 2 was designed to study the effect of voice without directing the participant's attention to the voice in which a stimulus was presented. In that experiment, the lexical decision paradigm was used to explore whether indexical information (i.e., fine phonetic detail such as vowel formant, speech rate and voice quality that makes a particular speaker different from others) could be stored by second language learners. As in Experiment 1, the results of Experiment 2 revealed that distinguishing between minimal pairs contrasting FLEECE and KIT was difficult for the Spanish L1 learners of English. However, it was found that in the case of repeated words, the participants responded faster to a stimulus if it had been presented in the same voice previously. This finding is evidence that the participants were able to store phonetic detail of the stimuli they encountered.

A question arising from these results is whether adult L2 learners' ability to store indexical information is related to the acquisition of new phonological categories in the target language. It still remains unclear whether voice differences are helpful in learning new phonological categories. The capacity learners have to store phonetic detail in a second language would suggest that with sufficient exposure to the target language, new phonological categories would be acquired. However this does not seem to be the case for most adult learners.

From an exemplar point of view, the capacity learners have to store phonetic detail would suggest, as Port (2007) argued, that they store words as representations of concrete, richly detailed memory traces, in episodic memory. This richness in detail is at the base of Goldinger's (1998) model of speech perception, which argues that the activation of such memory traces is proportional to their acoustic similarity with the stimulus, and that the listener's percept is the aggregate of all memory traces that are activated. Hence, it can be expected that with sufficient exposure to the language, i.e., with a richer exemplar memory in the target language, new phonological categories would emerge. However, this is not necessary the case in L2 phonology acquisition. Several studies (e.g., Andersson, Ferreira, & Henderson, 2011; Nguyen, Dufour, & Brunellière, 2012; Pallier et al., 2001) have found that extended exposure to an L2 does not by itself result in phonological learning.

The experiment presented in this chapter investigates whether the capacity to store acoustically rich exemplars can facilitate the learning of phonological categories that are difficult to discriminate in a second language. The experiment was carried out with a group of 31 native Spanish speakers who learned English as adults. The mouse tracking methodology was used to collect their responses. This method was selected because it offers the possibility of exploring the simultaneous activation of the members of a minimal pair. See below for more discussion on mouse tracking.

The results from Experiment 1 and Experiment 2 indicate that one of the challenges L2 learners face is to identify which variation in the speech signal is linguistically relevant and which is due to speaker differences. In fact, the acoustic correlates used to identify vowels, such as duration and formant structure, vary from speaker to speaker. Consequently, an L2 listener has to be able to recognise vowel quantity despite the variation arising from changes in speech rate. The same can be said about the variation in the formants of vowels due to differences in the length of the oral tract in speakers. The differences in oral tract length are commonly related to the sex of the speaker (Barreda & Nearey, 2012; Coleman, 1971), but can also be due to age (Bennett, 1981; Xue & Hao, 2003). The difficulty in noticing vowel similarities among the variation created by different voices may be one of the reasons why L2 speakers fail to learn new phonological categories. This could explain the apparent contradiction presented by the capacity to store phonetic detail and the failure to learn new phonological categories.

In General American English, the close front vowel space is occupied by the phonemes of the lexical sets fleece and kit. In contrast, and as has been described above in connection with the preceding experiments, Spanish has only one phoneme in the close front vowel space (Obediente, 1998). So it is likely that the Spanish speakers are predisposed by this feature of their L1 to consider those phonetic details that fall inside the limits of phonetic space of their Spanish close front vowel as mere variants of the same phonological category. After all, the actual acoustic characteristics of the Spanish close front vowel vary according to the phonetic context in which it occurs. For instance, Spanish vowels in unstressed syllables tend to have a more centralised quality (Quilis & Esgueva, 1983). So a possible result of the exposure to English lexical items is not that new categories are formed but that both categories are assimilated into one L1 category.

Extending the Perception Assimilation Model to L2 perception (PAM-L2), Best and Tyler (2007) postulate that the type of assimilation that takes place during the learning of a second language depends on the similarities between the L2 and L1:

> Perceptual learning occurs for some L2 contrasts, but seems to depend on their phonological and phonetic relationship to the L1, specifically on perceived similarities vs. dissimilarities to L1 phonemes (Best & Tyler, 2007, p. 13).

Furthermore, PAM-L2 predicts that when two L2 phonological categories are initially perceived as equivalent but one of them is more deviant from a category in the L1, the less deviant category is more readily assimilated to the L1 category, while the more deviant would eventually be learnt as a new category. However, if both L2 categories are perceived to be equally similar to an L1 category, both L2 categories are assimilated to the L1 category and learning of a new category would be unlikely to happen. In this case, the L2 learner would not be able to discriminate between lexical entries contrasting in the assimilated categories. Another factor that is considered by PAM-L2 to contribute greatly to the learning of a new phonological category is the importance of detecting the difference between minimal pairs, as Best and Tyler (2007) explain:

> If they are high frequency words, or come from two dense phonological neighbourhoods which contain many minimally contrasting words, and if many of these words need to be discriminated for adequate interaction with the surrounding cultural environment, these factors would increase the communicatively relevant

> pressure to perceptually learn the distinction (Best & Tyler, 2007, p. 28).

This observation is relevant to the type of training used in the present experiment. Phoneme identification drills are common both in L2 teaching and in L2 perception studies. One disadvantage of this type of drill is that it is not necessary to activate a lexical entry to carry them out. For example, Dufour et al. (2010) carried out an experiment to explore whether a group of speakers of southern French could be trained to discriminate a phonological contrast not present in their dialect. In their experiment, a series of French pseudowords was presented simultaneously with a series of abstract shapes. The idea was that the participants would learn to associate a pseudoword with a shape. Although the results revealed that the participants learnt to make the association, they also demonstrated that no lexical activation was necessary. Similarly, in a study using several voices during training, Iverson and Evans (2009) trained a group of adult Spanish L1 learners of English to identify the vowel phonemes in a series of English words. Although they found that L2 learners improved in their ability to recognize the English vowels, they did not improve in the ability to discriminate minimal pairs. Although these studies showed that L2 speakers can learn to discriminate between vowels, they did not focus directly on the semantic component of the lexical entries during training. One could speculate that phoneme drilling increases learners' sensitivity to acoustic variation, but that lexical activation is not required to carry out this task. Hence, this type of phoneme drilling does not promote the association between the phonetic differences and lexical entries.

However, previous studies (e.g., Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Iverson et al., 2012; Norris et al., 2003) have found that the lexicon can provide a feedback mechanism that can aid phonological learning. For example, Norris et al. (2003) found that when listeners heard a stimulus containing an ambiguous sound, they were able determine which phonological category the sound belonged to by mapping this stimulus onto a lexical entry they already know. For this reason, I expected that minimal pair discrimination could be improved if the learner's attention were explicitly directed to meaning. The task used in this experiment consisted of matching the spoken form of an English word with its equivalent in Spanish. In this way, attention would be directed explicitly to meaning, and the learners would be more likely to notice the acoustic differences between the members of a minimal pair. Consequently, this training would create stronger links between the phonological and lexical components of the lexical entries.

6.2 MOUSE TRACKING

In the previous experiments, the main variable studied was latency, that is, the time participants took to respond to a stimulus. Although the method used in those experiments proved to have enough temporal resolution to pick up the effects of voice facilitation on reaction time, neither of them provides a way of investigating how a decision is made. In the case of the perception of minimal pairs, the methods fail to inform us about the levels of activation of the two competing lexical entries during processing. As a way to determine whether L2 learners activate the lexical entries of minimal pairs simultaneously, I decided to use the analysis of mouse trajectories. In mouse tracking, the participant sees a set of response boxes on a computer screen, and the movements of the mouse are tracked as a response is selected from that set, following the presentation of the stimulus. That way, the experimenter obtains information about the activation of a competing response option before a response is given. This method would help me to better understand the decision processes that take place before a response is given.

Recent studies on categorization tasks that have employed this paradigm (e.g., Freeman, Dale, & Farmer, 2011; Papesh & Goldinger, 2012; Spivey, Dale, Knoblich, & Grosjean, 2010) have found that the trajectory of the mouse in the course of a response can be affected by the simultaneous activation of a competitor. These studies have found that if a competitor is considered a possible response, the mouse trajectory deviates, at least temporarily, towards the response area for this competitor item (for more details on the methodology, see below).

This paradigm has been recently used in experiments in speech perception in the L1. In an experiment on lexical recall, Papesh and Goldinger (2012) found that a lower level of attraction to a competitor was related to a higher level confidence in the actual response given. Similarly, in a lexical decision task with L1 Italian speakers, Barca and Pezzulo (2012) found that with more ambiguous stimuli the mouse movements towards the correct response were less direct and there was stronger attraction towards the response area for the competitor.

Deviation in the mouse trajectories towards a competitor response is expected to occur when L2 learners assimilate two L2 phonological categories to one L1 category. In this case, minimal pairs are stored as if they were homophones, which means that the minimal pairs share the same phonological form but have different meanings. Consequently, for our L1 Spanish learners of English as an L2, both KIT and FLEECE should produce the same degree of attraction. For this reason, the use of the mouse tracking method provides a better insight into the kind of assimilation that takes place during L2 perception than what I have previously observed with the use of the gamepad.

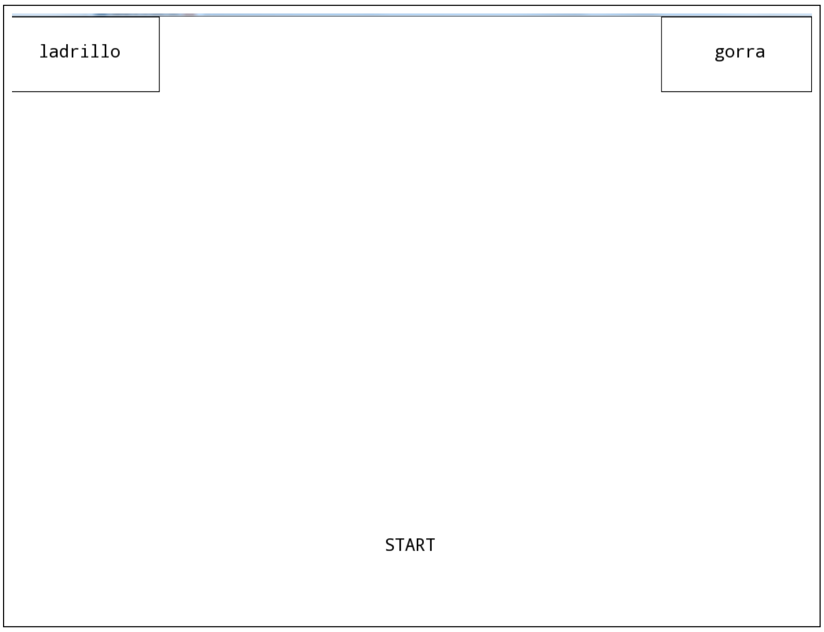

Figure 7 Example of a mouse tracking trial. In this practice block trial, the audio stimulus is /brɪk/ and the response options are "ladrillo" (brick) and "gorra" (cap).

6.3 METHOD

For each stimulus, the participants were asked to carry out a forced-choice decision between two alternatives. They heard an English word and then had to select a word in Spanish that was a translation of the word they heard (see Figure 7). The stimuli were presented over headphones and the participants used the computer mouse to give their responses. The experiment was presented using the application Open Sesame (Mathôt et al., 2012) running on a Dell Inspiron 1520 laptop. The input device was a Microsoft Express mouse connected to the USB port. This device offers an x-y resolution of 39.4 points per millimetre of mouse movement, a tracking speed of 1828 millimetres per second, and an imaging rate of 8000 frames per second. To collect the data from the mouse (i.e., x-y position and botton presses), I wrote a script in Python that run in Open Sesame (see Appendix 2).

6.4 participants

Thirty one native speakers of Spanish who spoke English as a second language took part in this experiment. The participants were recruited in Wellington but they started learning English in their home countries. All the participants were adults (mean age = 30 years, min=20, max= 51). They were all brought up in a Spanish speaking country. The average age at which they started learning English was 17.6 years (min = 5, max = 41, *SD* = 10.37). All the participants reported being familiar with American English either by being exposed to this dialect at school, in the media or having spent some time in the United States. This familiarity with American English was similar to that of the participants of the previous two experiments. The participants in those experiments were also exposed to this variety of English through the media and in the classroom, and no participants were excluded on the bases of having spent some time in an English speaking country.

6.5 experimental design

The experiment had a repeated measures design with two groups of participants. The experiment included a practice block, pre-training, training blocks and a post-training block. The practice block had eight trials with stimuli that were only used in this block but which had the same general characteristics as the rest of the stimuli. The post-training block had the same twenty minimal pairs (i.e., 40 words) as the pre-training block, but in a different random order. Comparison of performance in these two blocks can be used as an indicator of any learning that takes place during the training block. The training blocks included six of the minimal pairs that had been already presented in the pre-training block (i.e., 12 words), and which occurred again in the post-

Table 13 Number of stimuli used on the experiment blocks.

Block	N. stimuli	Voice	Feedback
Practice	8	male	Yes
Pre-training	40	male	No
Training	12	male/female	Yes
Post-training	40	male	No

training block. In order to minimise the differences in duration between the FLEECE and KIT vowel, so that participants were more likely to use differences in formant structure for discrimination, the words used in the training blocks ended in voiceless plosives.

The stimuli in each block were presented in a different random order for each participant. In the practice, pre-training and post-training blocks, the stimuli were presented in the male voice used in Experiment 2. However, in order to study the effect of voice familiarisation on perception, the voice condition was manipulated in the training blocks. In those blocks, half of the participants listened to the stimuli in the same male voice used in the pre-training and post-training blocks. The other half listened to the stimuli in the female voice from Experiment 2. Table 13 shows the number of stimuli used in each block.

6.6 STIMULI

The audio stimuli used in this experiment consisted of a subset of the stimuli used in Experiment 2. For Experiment 3, 20 English minimal pairs contrasting in the vowel sets FLEECE and KIT were used. The targets for the responses were the translations into Spanish of the minimal pair members, presented on the upper right and left corners of the screen. To ensure that the English words and their translations were likely to be known by the Spanish native speakers, a group of 20 Venezuelan students majoring in English at the University of Los Andes were asked to translate the English words used in the experiment into Spanish. There was high agreement in the translations. Nevertheless, only translations with more than 70% agreement were selected. For example, one minimal pair (*eel* vs. *ill*) was discarded due to the disagreement in the options of translations for *eel* ("anguila" = 55%, "morena" = 45%, "angula"= 5%). Appendix 3 presents a list of the stimuli used in this experiment.

6.7 procedure

The participants took part in the experiment individually in a quiet room. The instructions were presented on the computer screen and any questions were clarified by the researcher. The participants were encouraged to be as fast and accurate as possible. They were told they could take a short break between blocks if they wanted. The experiment took 20 to 30 minutes to complete. The participants were encouraged to start moving the cursor as soon as each stimulus starting playing, so that the movement did not start once a decision was made. If a participant took more than 1000 milliseconds to begin moving the mouse after the stimuli started playing, a message was displayed on the computer screen after the response was given. The message said "Please start moving earlier even if you are not certain about your response". There was a time-out of six seconds from the onset of the audio stimulus.

First a practice block was presented. This was followed by the pre-training block. Next the training blocks were presented. In the training blocks, half of the participants listened to the stimuli in the same male voice used in the pre-training block, while the other half heard a different, female voice.

Following Dufour et al. (2010), an 80% accuracy rate was considered to be sufficient indication that the participant could consistently discriminate the

Table 14 Number of blocks need to reach 80% accuracy threshold.

Number of blocks	Number of participants
6 (72 trials)	6
5 (60 trials)	5
4 (48 trials)	4
3 (36 trials)	2
2 (24 trials)	14

minimal pairs. Hence, after the first two training blocks, any participants with an accuracy rate of 80% or higher were given their score on the computer screen and continued to the post-training block. Any participants who scored less than 80% received one more training block. After the third training block, any participants who still scored less than 80% were asked whether they wanted to carry on with more training blocks or preferred to continue to the post-training block. The maximum number of blocks a participant could take was six. The reason for not having all the participants taking this number of blocks was to keep them motivated and focused on the task. Table 14 presents a summary of the number of blocks the participants needed to reach the 80% threshold.

After the training blocks, the same set of stimuli used in the pre-training block was presented. As in the pre-training block, the stimuli were presented in a different random order for each participant, and in a different random order from the pre-training block.

During the practice block and the training blocks, the participants received feedback after each trial. They did not receive feedback during the pre-training and post-training blocks. The sequence of the trials without feedback was as follows. At the beginning of each trial, the translations into Spanish of English minimal pairs were displayed in the top left and top right corners of the screen (the position was counterbalanced between participants). When the participants clicked the right mouse button, a cursor appeared at the bottom centre of the screen and the audio file corresponding to one of the minimal pairs was played. The responses were given by clicking on either of the Spanish words. There was a 6000 millisecond time out from the moment the audio file started to play.

For the blocks with feedback (i.e., the practice and training blocks), the sequence described above was followed. In the case of a correct response, the word CORRECT was displayed in green in the middle of the screen. If the response was incorrect, the word WRONG was displayed in red, then the correct translation was displayed and the audio file was played again.

6.8 RESULTS

The data collected during the experiment were:

- *Accuracy*: this indicated whether the Spanish word selected was the correct translation of the audio stimulus. Correct responses were coded 1 and incorrect responses 0.
- *Initiation time*: time in milliseconds corresponding to the first mouse movement registered from the beginning of the stimulus.
- *Response time*: the time in milliseconds from the beginning of the audio stimulus to the mouse click indicating the response selected.
- *Area under the curve (AUC)*: delimited by the trajectory of the mouse and a straight line between the cursor position at the beginning of the trial and the cursor position at the end of the trial. See Figure 8 for a graphic representation of calculations of the measures of the area under the curve and the maximal deviation.

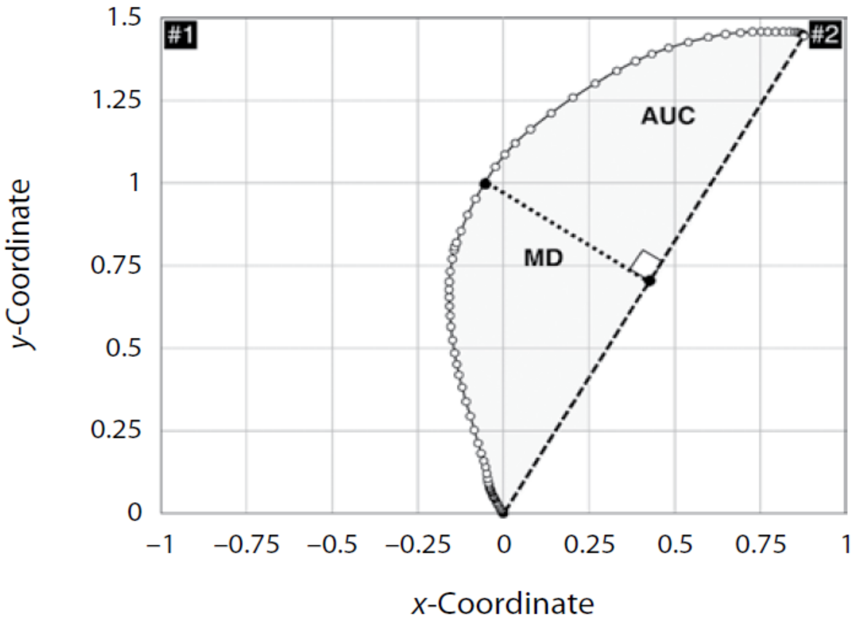

Figure 8 Diagram of mouse trajectory standard space and calculation of measures of spatial attraction to opposite response: maximal deviation (MD) and area under the curve (AUC). (Taken from Freeman & Ambady, 2010, p. 229)

- *Maximal deviation (MD)*: the longest perpendicular distance between the mouse trajectory and a straight line projected from the position of the mouse at the beginning of the trial and the final position of mouse at the end of the trial.

The predictor variables were:

- *Block*: The experiment included a pre-training block, up to six training blocks and a post-training block.

- *Voice*: The stimuli in the pre-training and post-training blocks were presented in a male voice for all the participants. However, for half of the participants the training blocks were presented in a female voice, and for the other half they were presented in the same male voice as in the pre- and post-training blocks.

- *Vowel*: one of the English close front vowels FLEECE or KIT
- *Training*: this variable indicates whether a stimulus used in the pre- and post-training blocks was also used during training.

The following control variables were included in the analysis:

- *Word frequency*: the logarithmic frequency of the stimulus.
- *Age*: the participant's age in years at the time of taking the test.
- *AgeLearning*: the participant's age in years at the time they started learning English.
- *Proficiency*: the participant's self-reported level of proficiency in the English language. The levels were: 1) beginner, 2) intermediate, 3) advanced, 4) near-native, and 5) native.
- *Training blocks*: the number of blocks a participant did during training.

The variables described above were analysed using mixed-effect models. The variables *age*, *ageLearning*, *proficiency*, *word frequency* entered in the models fitted for the analysis of the results were centred on the mean, as this would facilitate interpretation of the results and help reduce collinearity (see the discussion on centering procedures presented in Chapter 5). As in the previous experiments, the protocol proposed by Ieno et al. (2009) was followed (see Chapter 3 for a more detailed description of this protocol). For all the analyses, a beyond optimal model was first fitted. The beyond optimal model included all the predictor variables and the interactions of interest in the analysis, as well as the maximal random structure justified by the experiment. Following a backward stepwise selection, the optimal random structure was determined first and then the optimal fixed structure was obtained.

6.8.1 *Accuracy*

The research questions addressed in this study concern the effect of training on the ability of L2 speakers to discriminate English minimal pairs, and in particular whether any improvement in discrimination would a) extend to words not used during training, and b) be influenced by the use of a different voice during training from that used in the pre- and post-training blocks.

The basic expectation therefore was that if the training was effective, there should be an increase in the probability of a correct response after the training. If this training effect transfers to words that were not used in the training blocks, this would be an indication that the participants were capable of generalising across other lexical entries and therefore, arguably, of creating abstract representations.

If greater accuracy is observed in the post-training block only when the same voice is used in the training as in the pre- and post-training blocks, then this would indicate that the mental representations learned during training were specific to the exemplars that were encountered. Such a finding would indicate that the representations are based on exemplars and not abstract representations. On the other hand, if, as abstract models propose, all indexical information is discarded after pre-lexical processing, then there should be no difference between the training effects for stimuli presented in the same voice and those for stimuli presented in a different voice.

Table 15 presents the differences between the percentages of correct responses in the pre-training and post-training blocks in Experiment 3. The positive values for all conditions indicate that the probability of a correct response increased after training for all the conditions. However, the increase is clearly largest for responses to stimuli that were used in the training blocks ("yes" items) and which were presented in the same voice ("same" items).

A logistic regression model was fitted to the accuracy data in order to determine whether the differences observed in different blocks and across the training conditions were significant. The optimal model obtained (see Appendix 1) returned a significant interaction between the variable *block*, *voice* and *training* ($\beta = 0.631, SE = 0.296, z = 2.13, p = 0.033$). To facilitate the interpretation of the effects of training, the data were divided into pre-training and post-training blocks, and models were fitted to each subset.

The optimal model for the pre-training block (see Appendix 2) revealed effects of *proficiency* and *frequency*. The *proficiency* effect was such that the odds ratio for a correct response increased with proficiency ($\beta = 0.712, SE = 0.149, z = 4.77, p < 0.001$). More frequent words were more likely to have a correct response ($\beta = 4.401, SE = 0.064, z = 6.23, p < 0.01$). Furthermore, the stimuli containing the KIT vowel were more likely to get an incorrect response ($\beta = -0.357, SE = 0.089, z = -4.01, p < 0.01$).

However, the optimal model fitted to the data from the post-training block (see Appendix 3) showed no significant effect of vowel ($\beta = -0.193, SE = 0.214, z = -0.90, p = 0.37$). Considering the effect for vowel found in the pretraining block, this suggests that after training the participants might have improved their ability to identify lexical items containing the KIT vowel. As in the case of the pre-training block, there were also significant effects for *frequency* ($\beta = 0.316, SE = 0.155, z = 2.04, p < 0.05$) and for *proficiency* ($\beta = 0.748, SE = 0.134, z = 5.60, p < 0.001$). The results also showed an effect

Table 15 Percentage of correct responses in pre- and post-training blocks.

Training	Voice	Pre	Post	Δ accuracy
no	diff	66.45	69.13	2.67
no	same	62.92	64.49	1.57
yes	diff	72.61	74.70	2.08
yes	same	64.46	76.90	12.44

of the number *training blocks* needed to reach the accuracy threshold ($\beta = -0.009, SE = 0.004, z = -1.98, p < 0.05$), which means that those participants who required more training trials also made more errors in the post-training block. The model also showed that the words used in the training block had a higher probability of a correct response ($\beta = 0.549, SE = 0.237, z = 2.31, p < 0.05$). Despite the increase in accuracy observed when the same voice was used in training, the model showed that there was no significant effect of voice in the posttraining block ($\beta = -0.265, SE = 0.169, z = -1.56, p = 0.11$).

6.8.2 *Response time*

One of the aims of the present experiment was to determine the effect that familiarisation with a voice has on participant performance. In particular, I wanted to investigate whether the participants would exhibit a facilitation effect if they heard the words in the same voice during the training and the post-training blocks. If voice identity has a facilitatory effect on the recognition of the stimuli, a decrease in response times in the post-training block should be observed for the group of participants who had heard the training stimuli in the same voice as the pre- and post-training stimuli.

Two time variables were collected during the experiment. The first was *initiation time* which corresponds to the first mouse movement detected after the beginning of the stimulus. The other variable was *response time* which is the time from the beginning of the audio stimulus to the moment at which the participant clicked in one of the response boxes. The values of initiation time did not show any significant variation for the different conditions of the experiment, and so initiation time was not considered a meaningful variable for analysis. A probable explanation for the lack of variation is that the participants were told to start moving the mouse as soon as the audio

Table 16 Values of response times for pre- and post-training blocks. (The number of trials is presented in parentheses).

Accuracy	Vowel	*M* pre	*SD*	*M* post	*SD*	Δ RT
1	FLEECE	2182.9 (859)	873.7	1902.3 (879)	767.2	-280.6
1	KIT	2286.2 (768)	896.3	1946.1 (841)	752.7	-340.1
0	FLEECE	2581.9 (381)	1129.4	2322.2 (360)	1047.8	-259.7
0	KIT	2439.0 (472)	1133.5	2162.9 (399)	983.9	-276.1

started playing. The participants were encouraged to do so throughout the experiment. A message was displayed immediately after trials where the first mouse movement was detected later than 1000 milliseconds after the beginning of the audio file. The reason for including this message was to ensure that the mouse tracking would reflect the activation of an attractor during the selection process.

Before fitting a model, the data for the *response time* measure were plotted and visually inspected. Histogram plots of the distribution showed that this was positively skewed. The Box-Cox (1964) series of transformations was followed to normalise the distribution. The logarithmic transformation was found to be the best one. As can be observed in Table 16, there were larger differences in average response times between the pre-training and post-training blocks for the correct responses.

As in the previous experiments, only correct responses were used for the analysis of response time. The analysis of response time followed the procedure used for accuracy. That is, the responses for the pre-training block and the post-training block were carried out separately.

The optimal model fitted for the pre-training block (Appendix 5) showed that correct responses were slower for stimuli with the KIT vowel ($\beta = 0.047, SE = 0.016, t = 2.92$). This result, along with the lower accuracy for items with this vowel, indicates that KIT vowels were more difficult for learners.

In contrast, the model fitted to the post-training data (Appendix 6) showed no difference between vowels. The model also showed that the words used for training were responded to faster ($\beta = -0.073, SE = 0.022, t = -3.30$). Moreover, there was a significant interaction between *voice* and *training* ($\beta = 0.075, SE = 0.030, t = 2.50$). As seen in Figure 9, both same and different voice conditions showed similar response times for the words used during training. However, there were also significant differences for the words not used during training. These words had longer reaction times when presented in a different voice, which suggests that it was more difficult to generalise to other words when the voice was different. Also, the variable *AgeLearning* was significant ($\beta = 0.010, SE = 0.003, t = 2.97$), which indicates that the older the participants were at the time they started learning English, the slower they responded. This is probably an age effect since the current age of the participants is correlated with the age at which they started to learn English.

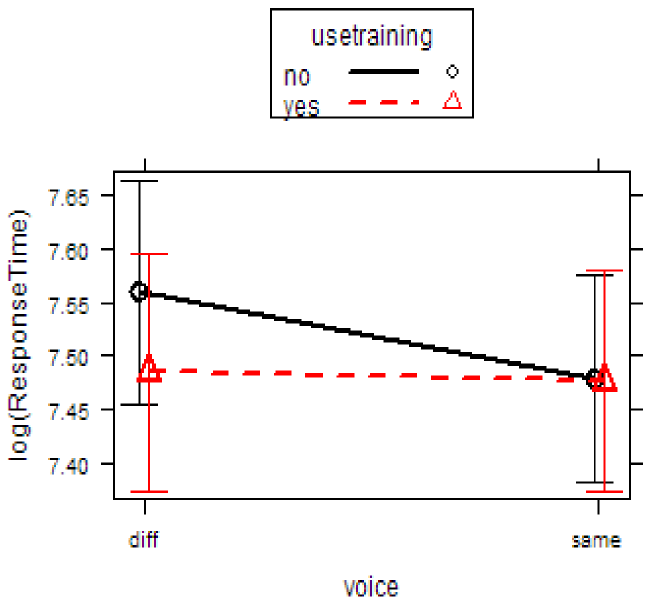

Figure 9 Response times to correct responses after training.

6.8.3 *Mouse track*

A variable collected during the experiment was the trajectory of the mouse from the bottom of the screen to the place on the screen where the response was selected. From these data, it was possible to calculate the area under the curve (AUC), which corresponds to the area circumscribed between the mouse trajectory and a straight line from the initial position of the mouse to the final position of the mouse. It was also possible to calculate the maximal deviation (MD), which is the maximal perpendicular distance from the straight line to the mouse trajectory. Both the area under the curve and the maximal deviation can be used to determine the degree of attraction of the alternative response during the selection of a response. The analysis of maximal deviation and area under the curve returned similar results. Note, however, that Freeman and Ambady (2010) have found that:

> [...] the AUC [Area Under the Curve] is a better index of the overall attraction toward the unselected alternative (incorporating all time steps), whereas MD [Maximal Deviation] is a better index of maximum attraction, but this attraction may be limited to fewer time steps. (p. 230)

The analysis of maximal deviation and area under the curve returned similar results. Given this observation, together with the fact that the purpose of this study was to determine the level of spatial attraction in general, the analysis of the mouse trajectories focused on the values of the area under the curve.

The analysis of the mouse trajectories had two purposes. Firstly, they were used to determine whether the English vowel categories FLEECE and KIT were equally assimilated into the Spanish vowel category /i/; and secondly, they help to investigate whether using different voices during training would have an effect on this assimilation process.

As mentioned before, Best and Tyler (2007) hypothesised that the ability to learn a new category would depend on how similar the L2 exemplars were to the L1 categories. They proposed that when exemplars of both L2 categories are perceived as equally similar to the L1 category, then new category learning is unlikely to take place. It was expected then that the analysis of the mouse trajectories would show how similar the FLEECE and KIT categories were perceived to be. More direct trajectories should be observed for stimuli perceived as better exemplars of the word selected in the response. On the other hand, a deviation towards a competitor would indicate this word was activated as well as the word that was finally selected.

Three aspects were investigated using the analyses of mouse trajectories. The first was whether the learners showed a bias in the spatial attraction between the FLEECE and KIT vowels. The second aspect was whether training would bring about changes in the spatial attraction patterns. The third was whether the use of a different voice during training would have an additional impact on the changes in spatial attraction.

The analyses of the pre-training and post-training blocks were carried out separately. The current analysis considers only the correct responses. Analysis of the incorrect responses is given further below. The results for the AUC (area under the curve) for the pre-training block showed that the more proficient the participants were, the more direct were the trajectories of their responses ($\beta = -0.080, SE = 0.033, t = -2.38$). Similarly there was a significant effect for *AgeLearning* ($\beta = -0.007, SE = 0.002, t = -3.17$). However there was no effect for *vowel* ($\beta = 0.005, SE = 0.019, t = 0.27$), which indicates that both options were equally attractive to the participants before training. The complete output of this model is included in Appendix 7.

The analysis of the data from the post-training block (Appendix 8) showed similarities to the pre-training block in *AgeLearning* ($\beta = -0.006, SE = 0.002, t = -2.75$) and proficiency ($\beta = -0.064, SE = 0.033, t = -1.96$). However, there was a significant interaction between *vowel* and *voice* ($\beta = -0.076, SE = 0.037, t = -2.03$). As seen in Figure 10, the main differences in AUC were observed for FLEECE stimuli when the training was carried out in a different voice. When the training was given in the same voice, there was more attraction to the Spanish response word that matched the meaning of the English word with the KIT vowel. However, when a different voice was used in the training block, the attraction to the response matching the KIT word was greatly reduced.

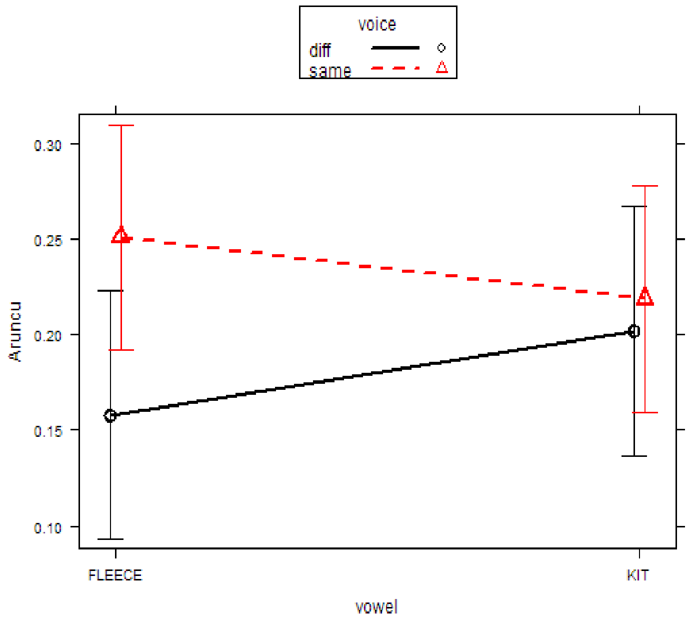

Figure 10 Effects of voice on the area under the curve for the stimuli presented in the post-training block (correct responses). Lower values of AUC indicate less attraction to the competing response.

A possible explanation for this pattern is that when training happened in the same voice, the participants felt that the two vowels were very similar and they therefore became more cautious when responding to FLEECE words. Another possibility is that the other voice used in training was a female speaker.

Since female voices show a larger acoustic vowel space than men (Hillenbrand & Clark, 2009), and particularly so with front vowels (Simpson & Ericsdotter, 2007), the participants receiving training with the female voice could have noticed the difference for FLEECE vowels more readily than those trained with the male voice. An acoustic analysis of the vowel formants of the stimuli used during the training block suggests that this could have been the case. As seen in Figure 11, there was no overlapping of the vowel formants between the male and the female voices. More importantly, the distance between F1 and F2 is greater for the female voice than for the male voice.

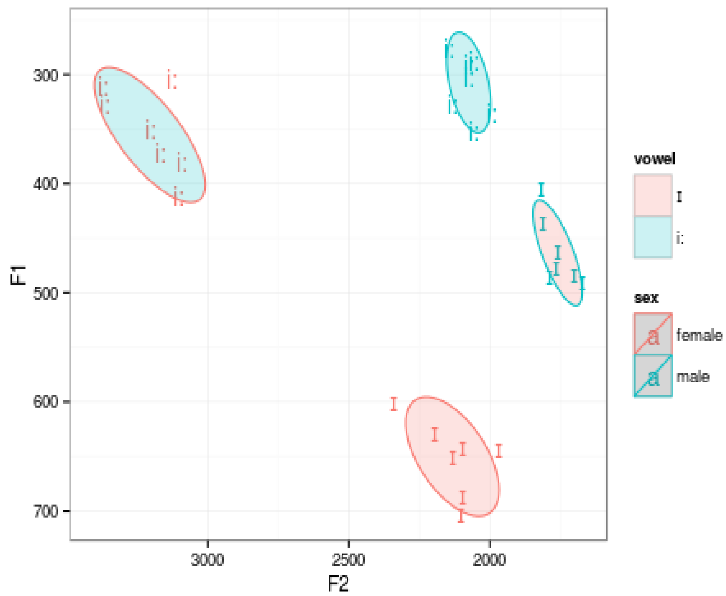

Figure 11 Vowel formant chart for the stimuli used in training.

Given the more direct trajectories to responses that matched the FLEECE vowel stimuli when training was given in a different voice, it is tempting to conclude that using a different voice facilitated the learning of English vowels. However, the main challenge for Spanish L1 learners of English is not to assimilate English FLEECE to their native category /i/, but to create a new category for KIT. Training in a different voice did not affect the AUC for responses to KIT vowels. Moreover, even though the analysis of accuracy

Table 17 Means of areas under the curve in the post-training block.

Accuracy	Vowel	Voice	*N*	Mean	*SD*
1	FLEECE	diff	409	0.173	0.368
1	FLEECE	same	470	0.238	0.413
1	KIT	diff	384	0.205	0.397
1	KIT	same	457	0.207	0.207
0	FLEECE	diff	151	0.309	0.437
0	FLEECE	same	209	0.241	0.459
0	KIT	diff	176	0.157	0.338
0	KIT	same	223	0.232	0.423

did return a significant effect of voice, Table 17 shows that there were fewer correct responses to KIT stimuli when training was done in a different voice.

An analysis of the incorrect responses showed that having training in a different voice produced more attraction to the response matching the FLEECE word. This is reflected in a significant interaction of vowel and voice ($\beta = 0.141, SE = 0.061, t = 2.31$). Figure 12 shows that incorrect responses to KIT stimuli (i.e., selecting the Spanish translation of the word "sheep" in response to the stimulus "ship") had the smallest AUC when training was given in a different voice. In other words, in some cases training in a different voice caused a stronger assimilation of English vowels to Spanish /i/. The full output if the model fitted to incorrect responses is presented in Appendix 9.

6.9 DISCUSSION

The main purpose of this experiment was to investigate the effect of voice on the learning of new phonological categories. In particular, I wanted to find out whether the variation in phonetic detail introduced by different voices would

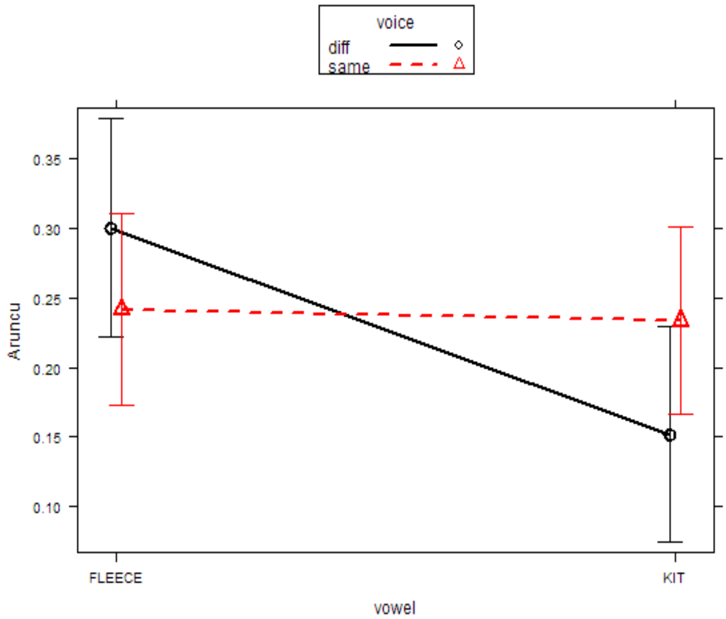

Figure 12 Effects of voice on the area under the curve for the stimuli presented in the post-training block (correct responses).

help the participants in learning the acoustic patterns needed to discriminate between the members of minimal pairs.

A theory of phonological representations is central to the discussion of the learning of new categories in a second language and the nature of phonological representations in general. As discussed in Chapter 2, these theories can be broadly classified into abstract models, those that claim that the information in store in the mental lexicon is poor in phonetic detail, and exemplar models, those that claim that phonological representations are rich in phonetic detail.

A strict abstract model of phonological representations would predict that voice variation has no effect in learning new categories, as all phonetic detail not relevant in the L1 would not be stored in memory. On the other hand, an episodic model of phonological representations would predict that phonetic detail is stored, and that new phonological categories eventually emerge from these phonetically rich representations.

It has been claimed, for example by Pallier et al., (2001), that phonological representations cannot be based on exemplars, given the difficulty L2 learners have in acquiring new phonological contrasts. The results of this experiment showed that after training the L2 learners were more accurate and faster at selecting the translation of the English words, which indicates that they were able to improve their ability to discriminate between minimal pairs. Moreover, their performance was better with the stimuli presented during training. This finding, along with the facilitation effect found in Experiment 2 for the stimuli presented in the same voice, provides evidence that L2 learners can store fine phonetic detail. Consequently, these findings reveal that the difficulty adult learners have in acquiring new phonological contrasts cannot be taken as solid evidence that phonological representations are abstract. The predictions of an extreme abstract model are not tenable, as the participants improved their ability to identify the members of the minimal pairs after training. This implies that with proper training, adult learners have the capacity to form new phonological categories in the target language even if these categories were previously assimilated to a single native language category.

The difficulty that learners have in acquiring L2 categories should not be attributed to an inability to store phonetic detail, but to the primacy that overlearned routines from the L1 have in speech perception. As discussed in Chapter 3, several studies have found that for Spanish native speakers, FLEECE is the more perceptually similar vowel to their native /i/. This was the also the case in the present experiment. However, the mouse tracking data revealed that the assimilation pattern was asymmetric. Although the attraction to FLEECE responses was very strong, the more indirect trajectory for KIT responses suggests that the learners could perceive differences in acoustic cues between both vowel categories. That is, they perceived that the KIT vowel was similar to Spanish /i/, but not as similar as the FLEECE vowel. This "goodness-of-fit analysis" would not be possible if their phonological representations did not contain phonetic detail. Furthermore, the finding that having the same voice during training mitigated the attraction to FLEECE suggests that fine phonetic detail is stored.

However, the ability to store fine phonetic detail does not rule out the possibility that L2 learners can use abstract representations. The ability to generalise to other words and other voices observed in the improvement in accuracy after training indicates that the learners were able to apply perceptual knowledge during training. However, their performance with words not used in training was poorer. This finding suggests that although they can generalise, the generalisation process does not spread fast across the lexicon. This could be due the nature of their lexical representations and not to perceptual deficiencies. Although learners can perceive fine phonetic detail, their limited experience and lack of awareness of the phonological contrast in the target language may indeed lead to merged lexical representations, which will require time to be re-categorised once the phonological contrast is acquired.

In the following chapter, I will review the findings of Experiments 1, 2 and 3 and discuss how these findings can answer the questions I set out in Chapter 2 about the nature of phonological representations in an L2.

7 GENERAL DISCUSSION AND CONCLUSION

7.1 OVERVIEW OF STUDY

The main objective of this thesis was to better understand the nature and development of phonological representations in an L2. Two approaches to phonological representations were considered, namely abstract and episodic. Abstract models claim that a normalisation process takes place during speech perception. This normalisation process is seen to take place in two stages. The first is a mapping of the acoustic signal onto pre-lexical representations, while in the second, pre-lexical representations are mapped onto abstract lexical representations. Pure abstract models propose that phonological representations bear no resemblance to the acoustic signal. In contrast, purely exemplar models propose that, instead of the normalisation mechanism included in abstract models, the acoustic signal is directly mapped onto the lexical representations. In exemplar models, this direct mapping is possible because phonological representations are seen as an aggregate of acoustically veridical exemplars of the words a person has encountered. As can be seen, the ability to store fine phonetic detail in the phonological representation lies at the centre of the debate between the abstract and the episodic approaches.

The ability to store fine phonetic detail is also crucial to the understanding of how phonology is learned in a L2. In order to investigate how fine grained L2 representations are, the experiments discussed in this study focused on the effects of speaker variation in the perception of minimal pairs in a second language. Since both linguistic and indexical information are transmitted as fine phonetic detail in the speech signal, a voice repetition facilitation effect would demonstrate not only that phonological representations contain fine acoustic detail, but also that they are based on episodic memories.

The research questions I set out at the beginning of this research project aimed to find out if this was the case. In the following sections I present the answers I found to these questions. Then I will present a general discussion of the findings and discuss the implications and the limitations of the study, as well as directions for future research.

7.2 ANSWERS TO THE RESEARCH QUESTIONS

Are L2 learners capable of storing acoustically rich representations in their L2?

The results from Experiment 1 and Experiment 2 showed that there was a facilitation effect when a stimulus was presented for the second time in the same voice. Following Goldinger's interpretation of the facilitation effect of voice repetition (Goldinger, 1998), the effect of voice repetition found in the second experiment is an indication that L2 learners are capable of storing acoustic detail in their phonological representations.

Are the L2 phonological representations of beginning learners predominantly episodic in nature?

Voice variation had a strong effect on the English learners who took part in the experiments carried out in this study. The results from Experiment 2 showed that, unlike native listeners, L2 learners experienced a facilitation effect when the stimuli were presented for a second time in the same voice. This difference between native speakers and learners suggests that, while the former rely more on abstract representation, the latter are more dependent on episodic memories.

Further evidence of the predominance of episodic representation in L2 comes from the results from Experiment 3. In the post-training block the learners performed better on the items they heard during training, which indicates that their representations are more reliant on episodic memories of the words they have experienced.

Can L2 learners generalise the perceptual learning acquired on a limited set of stimuli to the rest of their lexicon?

In Experiment 3, the learners showed an ability to generalise perceptual learning to other voices and words. After receiving training, the learners showed an increase in their ability to correctly identify minimal pair words not used during training. This generalisation shows that L2 learners can represent phonological contrasts in an abstract way. That is, they are able to extract the patterns that are necessary to discriminate vowel contrasts from the stimuli they experience and apply them to other words. However, these representations need not be abstract in the sense of not having any resemblance to the acoustic signal, but in terms of the lexical representations containing smaller discrete constituents that are shared across the lexicon.

Nevertheless, the learners' error rate on the words not used in training was significantly poorer than on the words used in training. Moreover, the increase in accuracy was greater when the voice used during training was the same, which shows that although the learners were able to generalise to other voices, their representations stored acoustic detail. This finding suggests that although L2 learners can generalise the learning they acquire from a limited set of words, they need more experience with their L2 before they can efficiently generalise across the lexicon.

7.3 DISCUSSION OF THE GENERAL FINDINGS

Previous studies of L2 perception which observed lexical priming between minimal pairs concluded that L2 representations did not store fine phonetic detail (e.g., Dufour & Nguyen, 2010; Pallier et al., 2001). However, the results from the present investigation revealed that the situation is more complex. The results presented in this thesis showed that although discriminating between minimal pairs was indeed difficult for L2 learners, there was a facilitatory effect when the stimuli were presented for a second time in the same voice. This voice repetition priming is evidence that L2 learners were able to store fine phonetic detail. Furthermore, the analysis of mouse trajectory carried out in Experiment 3 showed that there was an asymmetry in the attraction patterns between L2 categories. That is KIT stimuli were attracted to responses matching FLEECE words, but FLEECE stimuli were not as much attracted to responses matching KIT words. This finding shows that the L2 category that was acoustically closer to an L1 category (i.e., FLEECE) was dominant in the response pattern. This asymmetry indicates that fine phonetic detail was perceived. As Darcy, Daidone, and Kojima (2013) propose, this asymmetry is not necessarily the result of inability to perceive acoustic differences, but of lexical items being stored incorrectly under the same lexical entry.

How can we explain this contradiction between the ability to store phonetic detail and the difficulty to discriminate minimal pairs? A possible explanation is that during lexical access both indexical and abstract representations are involved. As McLennan and Luce (2005) proposed, both types of representations are supported in the lexical entries. Abstract units are prevalent early in the perception process. However, during the later stages of processing, episodic representations are more dominant. A consequence of this time course difference in the dominance of representations is that the type of representations used is related to the difficulty of the task at hand. In the case of fast, easy processing, for example when a lexical task includes very non-wordy pseudowords, the task can be carried out using abstract representations, which are the representations that are first activated. However, episodic representations are used when the task is difficult and requires more time to complete. For example, voice repetition effects have been found for low frequency words or when a lexical decision task included very word-like pseudowords (González & McLennan, 2007; McLennan & Luce, 2005). Such a view of phonological representations is also in line with the findings of Experiment 2. No voice priming effect was found for the L1 listeners, which is consistent with recent research which has found that voice facilitation effects are not strong for native speakers (e.g., Hanique, Aalders, & Ernestus, 2013; Orfanidou et al., 2011). For L1 listeners the high frequency of the stimuli would impose very little processing time. However, L2 learners must have found the lexical task much more difficult than the L1 listeners did. This difficulty is reflected in the learners' higher error rate and slower response times. Under these circumstances, L2 learners would have to activate their episodic representations.

Although voice priming is evidence that L2 learners store indexical, and hence fine phonetic detail in their phonological representations, it does not demonstrate that L2 learners can store veridical traces of the words they hear. Exemplar models need to take into account that the granularity of the representations is not simply determined by the capacities of the perceptual system (Pierrehumbert, 2001a). The result of the three experiments carried out in this thesis showed that the L1 had a strong impact on L2 listeners' ability to discriminate the English minimal pairs used. As the Native Language Magnet theory (Kuhl et al., 2008) predicts, sensitivity to acoustic detail is determined by how L1 categories warp the perceptual space. Hence how much phonetic detail is stored in the L2 depends on how perceptually close a stimulus is to a L1 phonological prototype. The granularity of phonological representations in an L2 seems to depend on the routines learned for the L1. These routines would determine what acoustic cues are to be attended, and which are ignored.

The difficulty in discriminating L2 phonological contrasts suggests that the models that propose phonological representations only contain veridical traces of experienced words are too powerful (e.g., Goldinger, 1998). Such models cannot explain why after prolonged exposure to the target language, learners fail to acquire L2 phonological contrasts. This difficulty in an L2 seems to arise from how the input was attended. In section 7.7, I will discuss how the Automatic Selective Perception (ASP) model (Strange, 2011) could be adapted to explain how episodic memories are stored in L2 phonology.

7.4 IMPLICATIONS FOR THE TEACHING OF A SECOND LANGUAGE

The effect of speaker variation was persistent in all three experiments. The results showed that it is more difficult for L2 listeners to identify the acoustic cues that signal different phonological categories when there was speaker variation. This finding is relevant for teachers of English as a second language. Although previous research has found that exposing learners to a variety of voices produced more robust learning (e.g., Barcroft & Sommers, 2005; Bradlow et al., 1997; Iverson et al., 2012; Lively et al., 1994; Trofimovich, 2005), the findings of this thesis indicate that voice variation can make the perception of phonological categories more difficult. Moreover, the results from Experiment 3 suggest that voice variation can affect the learning of new phonological categories.

The increase in difficulty that voice variation causes to L2 listeners should be taken into consideration in by L2 teachers, especially when teaching pronunciation and oral comprehension. Teachers should be aware that the more voices used in a given listening comprehension activity, the more difficult the task becomes. Given the difficulty L2 learners have in identifying linguistically relevant acoustic information from the random variation arising from different speakers, new voices in the teaching materials should be included progressively. In this way the learners have the opportunity to get accustomed to a particular speaker without being overwhelmed by differences between speakers. Field (2009) points out that a progression in the number of speakers and dialects is to be preferred in the L2 classroom. He proposes an emphasis first on using the teacher's voice and then gradually including other speakers of one standard variety and finally speakers of other dialects.

Another implication from this thesis is the kind of teaching materials to use for phonological training in an L2. Best and Tyler (2007) had previously noticed that an important factor in learning a phonological contrast in an L2 is that the words that exploit such a contrast need to be correctly discriminated for adequate social interaction. In other words, learners are more likely to learn a new contrast when they are aware of the semantic differences between minimal pairs. The task used in the training blocks in Experiment 3 showed that associating the words forming a minimal pair with their translation equivalents in the L1 is an effective training method. This task could easily be integrated into the classroom when students share the same L1.

7.5 Implications for research methodologies

Two behavioural methodologies were used in this experiment, i.e., measuring reaction time using a gamepad and recording mouse trajectories. Although the latter technique does not offer a time resolution as accurate as the gamepad, mouse tracking makes it possible for us to study the how a cognitive process develops over time. The use of the mouse tracking methodology proved to be a useful technique in speech perception. In Experiment 3, the analysis of mouse trajectories revealed that both members of minimal pair entered lexical competition before a response was selected. Moreover, an asymmetry was observed between the members of a minimal pair, which showed that one phonological category was more attractive (i.e., dominant) than the other. This valuable information concerning L2 perception would have not been obtained if only the more common reaction time methodology had been used.

As to the technologies used, I made a commitment not to use proprietary technologies. The cost involved in purchasing or developing these technologies can be a barrier to many researchers working with limited resources. Being aware of this limitation, the experiments presented in this thesis were carried out using technologies that are easily available and inexpensive. The software used was free (preferably open source and multiplatform), while the hardware was standard equipment available from most technology retailers. The satisfactory performance of these low-cost technologies in the experiments carried out in this thesis might encourage other researchers to use them more widely.

7.6 LIMITATIONS

The purpose of this thesis was to determine if L2 learners are able to store fine phonetic detail in their phonological representations. Although the results from Experiment 2 showed that L2 learners stored phonetic detail, the present study did not investigate how fine-grained the information contained in their episodic memories was. In other words, the experiments carried out in this study were not developed to find out the level of detail with which temporal and spectral information was stored in long term memory. This lack of information concerning the granularity of L2 representations limits the inferences that can be made about how episodic information is used during L2 phonology learning.

Another limitation of the present study was the availability of participants. There are relatively few native Spanish speakers in New Zealand. This lack of availability of participants had an impact on the design and implementation of the experiments conducted in this thesis. The small pool of possible participants made more difficult to control for age, language competence and learning environment (e.g., home country vs. English speaking country). This was particularly the case in Experiment 3, where due to unforeseen difficulties, the participants had to be recruited in Wellington only. Likewise, the limited availability of participants and time constraints made a longitudinal study impractical. A longitudinal study would have allowed me to investigate how a learner's performance changes as they become more experienced with the target language.

Finally, although several studies have found an effect of word frequency in speech perception and lexical access, the materials required for the tasks in the three experiments carried out in this study (i.e., discriminating minimal pairs) meant that it was not possible for me to study the effects of word frequency on lexical access. Given the limited vocabulary of English learners, only high frequency words were included amongst the stimuli used in this study.

7.7 TOWARDS A HYBRID MODEL OF L2 PHONOLOGY

The results obtained in this thesis showed an effect of voice facilitation for L2 listeners, but also showed that L2 listeners were not efficient in perceiving phonetic detail needed to discriminate between L2 phonological categories. The first finding offers support for an exemplar model, while the second finding is evidence for an abstract model of phonological representations in L2. How can these contradictory findings be reconciled? I propose that these results are compatible with a hybrid model. Hybrid models claim that phonological representations contain abstract representations as well as episodic memories of the words that the listeners experience. Moreover, these two types of representations are the result of different perception modes. I concur with the hypotheses proposed by the Automatic Selective Perception model (Strange, 2011) in that speech perception can operate in a phonetic or a phonological mode. The phonological mode is used to "detect sufficient phonologically relevant contrastive information for word-form identification" (Strange, 2011, p. 460). This mode is highly automatic, hence faster than the phonetic mode. It is the default mode used when listening to continuous speech in the L1. In contrast, the phonetic mode is used when the perception task requires attentional focus to contextual phonetic detail. This perceptual mode is slower and requires more cognitive resources. This is the mode that is used, for example, when listening to a different L1 dialect, or in the early stages of learning a new L1.

I propose that in a hybrid model of L2 phonology learning, the phonological mode is employed for fast processing of the speech signal in the L2, while the phonetic perception mode is used when attention is directed to fine phonetic detail, even if this is not relevant for L1 phoneme identification. This mode allows the listener to store a phonetically rich representations of the input. Therefore, phonological representations in a hybrid model, contain an abstract canonical form of a word linked to episodic traces of that word. These two types of representations (the abstract and the episodic) are available to the listener, and the purpose and/or the difficulty of the perceptual task determines which representations is more readily accessed.

A hybrid model predicts that when the stimuli are processed using the phonological mode, only abstract representations are stored. In this case no speaker identity priming should be observed. This explains one of the findings of this thesis. If the learners have not yet learned L2 phonological categories, the phonological mode would map the two vowel phonemes that differentiate the minimal pairs onto one abstract L1 phoneme, hence producing the effect of pseudo-homophony observed in this thesis. On the other hand, when the phonetic mode is used abstract representations and phonetically rich exemplars are stored. Accordingly, if the task presented to the learners requires that this mode is used, a voice facilitation effect would be observed.

The specific inclusion of episodic representations is a significant departure from ASP. The ASP model follows the premises of direct realism. Consequently, it is concerned with the ability to extract information from the acoustic signal, but remains agnostic about the format in which this information is stored in long term memory. However, in order to account for voice facilitation effects and generalisation across the lexicon, a model should include both and abstract and episodic representations.

7.8 SUGGESTED FURTHER RESEARCH

The findings of this study showed that L2 learners use abstract and episodic representations. More research is needed to better understand how indexical and linguistic information is integrated in the L2. There are a number of directions in which this research could be carried further. Firstly, more research is needed to determine how finely-grained is the phonetic detail that is stored in long term memory. The present study showed that phonetic detail was stored but it did not examine how fine grained this was. Further research is needed to determine how accurately temporal and spectral acoustic information is stored in long term memory. Secondly, more research is necessary to determine how abstract and episodic information are accessed and integrated during speech perception. In particular more research is needed in order for us to gain a better understanding of the time-course of the perceptual process in an L2. Thirdly, in order to confirm that hybrid representations are the result of different perception modes, future studies should focus on determining how voice facilitation effects and perceptual learning in an L2 are affected by the task given to the learners. Finally, future research should investigate how abstract and episodic information is used at different levels of language proficiency. This requires longitudinal studies on the effects of voice variation on L2 perception and spoken word recognition.

7.9 CONCLUSION

This thesis reviewed the nature of phonological representations in learners of a second language. The results of the experiments carried out showed that L2 learners store fine phonetic detail in their phonological representations. Nevertheless, the granularity of these representations depends heavily on the phonological categories established in the L1. These results revealed that L2 phonological representations are not exclusively abstract or purely episodic, but that both types are available to L2 listeners. The use of one or the other is determined by the task at hand.

The results obtained in this study are relevant for the debate on the general format of phonological representations. Researchers working on the L1 (e.g., Cutler & Weber, 2007; Cutler, 2012; Goldinger, 2007; McLennan & Luce, 2005; Pierrehumbert, 2006) propose that hybrid models are needed to better explain the complex interaction between abstract and episodic representations. The findings of this study indicate that is also the case in L2 phonology.

The findings presented in this study underscore that future models of how L2 phonology is learned need to include an exemplar level of representation in their models. The most influential current models of L2 phonology learning (Best & Tyler, 2007; Escudero, 2005; Flege, 1995; Strange, 2011) are purely abstract. Considering the impact that voice variation has on L2 perception, these models fail to accurately explain an important aspect of the learning of phonology in a second language. Hybrid models are better positioned to explained L2 phonology learning. The challenge for future models is to describe specific mechanism about how abstract and exemplar information are integrated in the learning of L2 phonology.

Appendices

MODELS FITTED IN EXPERIMENT 1

1 BEYOND OPTIMAL FITTED TO RESPONSE CHOICE

```
Family: binomial ( logit )
Formula: correct ~ voice_rep * contrast * word_type + trial_order
  + NBC +  distance + (1 + trial_order | participant) + (1 | word)
   Data: datasec
     AIC      BIC   logLik deviance df.resid
  3141.8   3267.0  -1549.9   3099.8     2855
Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.4176 -0.6756  0.3019  0.6012  3.3560

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 participant (Intercept) 0.038457 0.196106
             trial_order 0.000018 0.004243 -0.42
 word        (Intercept) 0.424320 0.651398
Number of obs: 2876, groups: participant, 57; word, 53

Fixed effects:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                               1.918624   0.777374   2.468  0.01358 *
voice_repsv                              -0.282298   0.987818  -0.286  0.77505
contrastbilabial                         -2.816190   0.904059  -3.115  0.00184 **
contrastvelar                            -2.443763   0.886637  -2.756  0.00585 **
contrastvowel                            -2.442661   0.610236  -4.003 6.26e-05 ***
word_typerw                              -0.183769   0.749574  -0.245  0.80633
trial_order                               0.001791   0.003470   0.516  0.60578
NBC                                       0.012775   0.190111   0.067  0.94642
distance                                 -0.043518   0.018778  -2.318  0.02048 *
voice_repsv:contrastbilabial              1.386649   1.222906   1.134  0.25684
voice_repsv:contrastvelar                 0.935420   1.193260   0.784  0.43309
voice_repsv:contrastvowel                -0.369795   0.788721  -0.469  0.63917
voice_repsv:word_typerw                   0.847769   0.916658   0.925  0.35505
contrastbilabial:word_typerw              1.671925   0.955585   1.750  0.08018 .
contrastvelar:word_typerw                 1.135838   0.978768   1.160  0.24585
voice_repsv:contrastbilabial:word_typerw -3.169948   1.300668  -2.437  0.01480 *
voice_repsv:contrastvelar:word_typerw    -0.576418   1.319628  -0.437  0.66225
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Correlation of Fixed Effects:
                  (Intr) vc_rps cntrstb cntrstvl cntrstvw wrd_ty trl_rd NBC    distnc vc_rpsv:cntrstb
voice_repsv       -0.665
contrstblbl       -0.776  0.594
contrastvlr       -0.765  0.610  0.656
contrastvwl       -0.790  0.644  0.688   0.701
word_typerw       -0.818  0.650  0.708   0.721    0.742
trial_order       -0.142 -0.245 -0.061  -0.111   -0.142   -0.120
NBC                0.238 -0.137 -0.187  -0.185   -0.250   -0.519  0.024
distance          -0.397  0.188  0.215   0.182    0.148    0.143 -0.085 -0.158
vc_rpsv:cntrstb    0.589 -0.782 -0.745  -0.489   -0.509   -0.516  0.045  0.118 -0.208
vc_rpsv:cntrstvl   0.556 -0.834 -0.497  -0.758   -0.537   -0.542  0.216  0.117 -0.174  0.652
vc_rpsv:cntrstvw   0.589 -0.818 -0.536  -0.554   -0.786   -0.572  0.253  0.157 -0.144  0.630
vc_rpsv:wr_        0.593 -0.935 -0.545  -0.564   -0.566   -0.698  0.257  0.155 -0.120  0.723
cntrstblb:_        0.526 -0.452 -0.795  -0.475   -0.452   -0.493  0.123 -0.174 -0.096  0.598
cntrstvlr:_        0.550 -0.435 -0.481  -0.794   -0.496   -0.643  0.012  0.185  0.029  0.349
vc_rpsv:cntrstb:_ -0.430  0.635  0.612   0.374    0.355    0.390 -0.112  0.089  0.174 -0.847
vc_rpsv:cntrstv:_ -0.398  0.615  0.366   0.603    0.393    0.526 -0.072 -0.226 -0.060 -0.482
                  vc_rpsv:cntrstvl vc_rpsv:cntrstvw vc_r:_ cntrstb:_ cntrstv:_ vc_rpsv:cntrstb:_
voice_repsv
contrstblbl
contrastvlr
contrastvwl
word_typerw
trial_order
NBC
distance
vc_rpsv:cntrstb
vc_rpsv:cntrstvl
vc_rpsv:cntrstvw   0.683
```

```
vc_rpsv:wr_         0.779             0.710
cntrstblb:_         0.378             0.376             0.476
cntrstvlr:_         0.585             0.376             0.476  0.394
vc_rpsv:cntrstb:_ -0.530             -0.466            -0.668 -0.743    -0.291
vc_rpsv:cntrstv:_ -0.784             -0.468            -0.673 -0.278    -0.762     0.413
```

2 OPTIMAL MODEL FITTED TO RESPONSE CHOICE

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) [glmerMod]
 Family: binomial ( logit )
Formula: correct ~ voice_rep * contrast + word_type + voice_rep + distance
+  (1 | participant)
   Data: datasec

     AIC      BIC   logLik deviance df.resid
  3288.4   3354.0  -1633.2   3266.4     2865

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.9561 -0.7685  0.3723  0.6453  2.2848

Random effects:
 Groups      Name        Variance Std.Dev.
 participant (Intercept) 0.04308  0.2075
Number of obs: 2876, groups: participant, 57

Fixed effects:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                   1.10237    0.18687   5.899 3.65e-09 ***
voice_repsv                   0.56989    0.16380   3.479 0.000503 ***
contrastbilabial             -1.27645    0.17569  -7.265 3.73e-13 ***
contrastvelar                -1.30520    0.19502  -6.693 2.19e-11 ***
contrastvowel                -1.83452    0.18184 -10.089  < 2e-16 ***
word_typerw                   0.45040    0.10334   4.359 1.31e-05 ***
distance                     -0.02674    0.00760  -3.518 0.000434 ***
voice_repsv:contrastbilabial -0.99042    0.23451  -4.223 2.41e-05 ***
voice_repsv:contrastvelar     0.01130    0.25457   0.044 0.964601
voice_repsv:contrastvowel    -0.88590    0.24230  -3.656 0.000256 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Correlation of Fixed Effects:
                                                                    vc_rpsv:  vc_rpsv:
                 (Intr) vc_rps cntrstb cntrstvl cntrstvw wrd_ty distnc cntrstb   cntrstvl
voice_repsv      -0.530
contrstblbl      -0.638  0.486
contrastvlr      -0.661  0.460  0.499
contrastvwl      -0.633  0.450  0.508   0.487
word_typerw      -0.533 -0.006  0.228   0.253    0.346
distance         -0.554  0.206  0.212   0.297    0.112   -0.033
vc_rpsv:cntrstb   0.369 -0.689 -0.703  -0.319   -0.324   -0.040 -0.099
vc_rpsv:cntrstvl  0.448 -0.683 -0.354  -0.741   -0.311    0.014 -0.327  0.463
vc_rpsv:cntrstvw  0.370 -0.666 -0.334  -0.314   -0.680   -0.067 -0.090  0.464    0.445
```

3 BEYOND OPTIMAL MODEL FITTED FOR REACTION TIME (ALL WORD PAIRS INCLUDED)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ presentation * voice_rep * contrast + NBC + duration +
    distance + ctrial + (1 + ctrial | participant)
   Data: dfcorrect

REML criterion at convergence: 21032.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7735 -0.6548 -0.0723  0.6116  3.2979

Random effects:
 Groups      Name        Variance  Std.Dev. Corr
 participant (Intercept) 2.7888053 1.66997
             ctrial      0.0002512 0.01585  -0.51
 Residual                7.8585948 2.80332
Number of obs: 4234, groups: participant, 57

Fixed effects:
                                               Estimate Std. Error t value
(Intercept)                                  -2.875e+01  3.701e-01  -77.69
presentationsecond                           -1.856e-01  3.449e-01   -0.54
voice_repsv                                  -3.543e-01  2.584e-01   -1.37
contrastcontrol                              -4.862e-01  2.401e-01   -2.03
contrastvelar                                -3.116e-01  2.784e-01   -1.12
contrastvowel                                 5.924e-02  2.587e-01    0.23
NBC                                           1.123e-01  5.145e-02    2.18
duration                                      3.099e-03  2.926e-04   10.59
distance                                     -1.235e-02  1.371e-02   -0.90
ctrial                                        1.600e-02  2.528e-03    6.33
presentationsecond:voice_repsv                2.245e+00  4.436e-01    5.06
presentationsecond:contrastcontrol            4.366e-01  3.798e-01    1.15
presentationsecond:contrastvelar              1.301e+00  4.897e-01    2.66
presentationsecond:contrastvowel              2.524e+00  4.700e-01    5.37
voice_repsv:contrastcontrol                   2.656e-01  3.130e-01    0.85
voice_repsv:contrastvelar                     1.503e+00  3.736e-01    4.02
voice_repsv:contrastvowel                     3.159e-01  3.516e-01    0.90
presentationsecond:voice_repsv:contrastcontrol -3.166e+00  5.164e-01   -6.13
presentationsecond:voice_repsv:contrastvelar   -2.737e+00  6.534e-01   -4.19
presentationsecond:voice_repsv:contrastvowel   -3.453e+00  6.652e-01   -5.19

Correlation of Fixed Effects:
                            (Intr) prsntt vc_rps cntrstc cntrstvl cntrstvw NBC
prsnttnscnd                 -0.229
voice_repsv                 -0.374  0.375
cntrstcntrl                 -0.287  0.414  0.596
```

```
contrastvlr                   -0.240  0.364  0.516  0.567
contrastvwl                   -0.228  0.391  0.554  0.626   0.542
NBC                           -0.276 -0.079  0.105  0.087  -0.047   -0.005
duration                      -0.570 -0.035 -0.051 -0.234  -0.174   -0.253    0.056
distance                       0.003 -0.420  0.009  0.003   0.010    0.007   -0.030
ctrial                        -0.168  0.019 -0.120 -0.053  -0.081   -0.063   -0.087
prsnttnsc:_                    0.172 -0.679 -0.523 -0.305  -0.274   -0.291    0.127
prsnttnscnd:cntrstc            0.158 -0.709 -0.335 -0.595  -0.338   -0.368    0.104
prsnttnscnd:cntrstvl           0.038 -0.666 -0.243 -0.303  -0.561   -0.302    0.220
prsnttnscnd:cntrstvw           0.085 -0.636 -0.261 -0.308  -0.282   -0.528    0.190
vc_rpsv:cntrstc                0.273 -0.314 -0.821 -0.752  -0.432   -0.467   -0.061
vc_rpsv:cntrstvl               0.213 -0.269 -0.700 -0.419  -0.756   -0.405    0.056
vc_rpsv:cntrstvw               0.202 -0.286 -0.742 -0.456  -0.410   -0.737    0.030
prsnttnscnd:vc_rpsv:cntrstc   -0.112  0.537  0.450  0.427   0.245    0.264   -0.123
prsnttnscnd:vc_rpsv:cntrstvl  -0.023  0.541  0.339  0.218   0.415    0.219   -0.205
prsnttnscnd:vc_rpsv:cntrstvw  -0.068  0.468  0.339  0.205   0.190    0.360   -0.161

                              duratn distnc ctrial prsn:_ prsnttnscnd:cntrstc
prsnttnscnd
voice_repsv
cntrstcntrl
contrastvlr
contrastvwl
NBC
duration
distance                      -0.001
ctrial                        -0.032 -0.028
prsnttnsc:_                    0.002  0.084 -0.043
prsnttnscnd:cntrstc            0.102 -0.101 -0.040  0.581
prsnttnscnd:cntrstvl           0.165  0.160 -0.060  0.490  0.545
prsnttnscnd:cntrstvw           0.112  0.044 -0.042  0.495  0.569
vc_rpsv:cntrstc                0.092 -0.007  0.082  0.437  0.445
vc_rpsv:cntrstvl               0.077 -0.015  0.125  0.373  0.243
vc_rpsv:cntrstvw               0.132 -0.013  0.112  0.394  0.262
prsnttnscnd:vc_rpsv:cntrstc   -0.060  0.045  0.043 -0.851 -0.736
prsnttnscnd:vc_rpsv:cntrstvl  -0.111 -0.213  0.066 -0.709 -0.404
prsnttnscnd:vc_rpsv:cntrstvw  -0.049 -0.072  0.049 -0.679 -0.399

                              prsnttnscnd:cntrstvl prsnttnscnd:cntrstvw
prsnttnscnd
voice_repsv
cntrstcntrl
contrastvlr
contrastvwl
NBC
duration
distance
ctrial
prsnttnsc:_
```

prsnttnscnd:cntrstc		
prsnttnscnd:cntrstvl		
prsnttnscnd:cntrstvw	0.489	
vc_rpsv:cntrstc	0.215	0.226
vc_rpsv:cntrstvl	0.406	0.204
vc_rpsv:cntrstvw	0.211	0.383
prsnttnscnd:vc_rpsv:cntrstc	-0.414	-0.428
prsnttnscnd:vc_rpsv:cntrstvl	-0.773	-0.379
prsnttnscnd:vc_rpsv:cntrstvw	-0.354	-0.712

	vc_rpsv:cntrstc	vc_rpsv:cntrstvl	vc_rpsv:cntrstvw
prsnttnscnd			
voice_repsv			
cntrstcntrl			
contrastvlr			
contrastvwl			
NBC			
duration			
distance			
ctrial			
prsnttnsc:_			
prsnttnscnd:cntrstc			
prsnttnscnd:cntrstvl			
prsnttnscnd:cntrstvw			
vc_rpsv:cntrstc			
vc_rpsv:cntrstvl	0.577		
vc_rpsv:cntrstvw	0.614	0.543	
prsnttnscnd:vc_rpsv:cntrstc	-0.575	-0.324	-0.345
prsnttnscnd:vc_rpsv:cntrstvl	-0.293	-0.533	-0.274
prsnttnscnd:vc_rpsv:cntrstvw	-0.287	-0.250	-0.490

	prsnttnscnd:vc_rpsv: :cntrstc	prsnttnscnd: vc_rpsv:cntrstvl
prsnttnscnd		
voice_repsv		
cntrstcntrl		
contrastvlr		
contrastvwl		
NBC		
duration		
distance		
ctrial		
prsnttnsc:_		
prsnttnscnd:cntrstc		
prsnttnscnd:cntrstvl		
prsnttnscnd:cntrstvw		
vc_rpsv:cntrstc		
vc_rpsv:cntrstvl		

```
vc_rpsv:cntrstvw
prsnttnscnd:vc_rpsv:cntrstc
prsnttnscnd:vc_rpsv:cntrstvl  0.593
prsnttnscnd:vc_rpsv:cntrstvw  0.580                   0.499

                             prsnttnscnd:vc_rpsv:cntrstvl
prsnttnscnd
voice_repsv
cntrstcntrl
contrastvlr
contrastvwl
NBC
duration
distance
ctrial
prsnttnsc:_
prsnttnscnd:cntrstc
prsnttnscnd:cntrstvl
prsnttnscnd:cntrstvw
vc_rpsv:cntrstc
vc_rpsv:cntrstvl
vc_rpsv:cntrstvw
prsnttnscnd:vc_rpsv:cntrstc
prsnttnscnd:vc_rpsv:cntrstvl
prsnttnscnd:vc_rpsv:cntrstvw  0.499
```

4 OPTIMAL MODEL FITTED FOR REACTION TIME (ALL WORD PAIRS INCLUDED)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ presentation * voice_rep * contrast + NBC + duration +
    ctrial + (1 + ctrial | participant)
   Data: dfcorrect

REML criterion at convergence: 21026.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7728 -0.6525 -0.0744  0.6086  3.2990

Random effects:
 Groups      Name        Variance  Std.Dev. Corr
 participant (Intercept) 2.7901507 1.67037
             ctrial      0.0002512 0.01585  -0.51
 Residual                7.8582226 2.80325
Number of obs: 4234, groups: participant, 57

Fixed effects:
                                                 Estimate Std. Error t value
```

```
(Intercept)                                     -2.875e+01  3.701e-01  -77.68
presentationsecond                              -3.161e-01  3.129e-01   -1.01
voice_repsv                                     -3.523e-01  2.584e-01   -1.36
contrastcontrol                                 -4.857e-01  2.401e-01   -2.02
contrastvelar                                   -3.091e-01  2.783e-01   -1.11
contrastvowel                                    6.090e-02  2.587e-01    0.24
NBC                                              1.110e-01  5.143e-02    2.16
duration                                         3.099e-03  2.926e-04   10.59
ctrial                                           1.593e-02  2.527e-03    6.30
presentationsecond:voice_repsv                   2.278e+00  4.421e-01    5.15
presentationsecond:contrastcontrol               4.021e-01  3.779e-01    1.06
presentationsecond:contrastvelar                 1.372e+00  4.833e-01    2.84
presentationsecond:contrastvowel                 2.542e+00  4.695e-01    5.41
voice_repsv:contrastcontrol                      2.637e-01  3.130e-01    0.84
voice_repsv:contrastvelar                        1.498e+00  3.736e-01    4.01
voice_repsv:contrastvowel                        3.117e-01  3.516e-01    0.89
presentationsecond:voice_repsv:contrastcontrol -3.145e+00  5.158e-01   -6.10
presentationsecond:voice_repsv:contrastvelar   -2.862e+00  6.384e-01   -4.48
presentationsecond:voice_repsv:contrastvowel   -3.496e+00  6.634e-01   -5.27

Correlation of Fixed Effects:
                           (Intr) prsntt vc_rps cntrstc cntrstvl cntrstvw NBC
prsnttnscnd                -0.250
voice_repsv                -0.374  0.417
cntrstcntrl                -0.287  0.457  0.596
contrastvlr                -0.240  0.406  0.516  0.567
contrastvwl                -0.228  0.434  0.554  0.626   0.542
NBC                        -0.276 -0.101  0.105  0.088  -0.046   -0.005
duration                   -0.570 -0.039 -0.051 -0.234  -0.174   -0.253    0.056
ctrial                     -0.168  0.008 -0.120 -0.053  -0.081   -0.063   -0.087
prsnttnsc:_                 0.172 -0.712 -0.526 -0.306  -0.276   -0.293    0.130
prsnttnscnd:cntrstc         0.159 -0.832 -0.335 -0.597  -0.339   -0.370    0.101
prsnttnscnd:cntrstvl        0.038 -0.668 -0.248 -0.308  -0.570   -0.308    0.227
prsnttnscnd:cntrstvw        0.085 -0.681 -0.261 -0.309  -0.283   -0.529    0.191
vc_rpsv:cntrstc             0.273 -0.349 -0.821 -0.752  -0.431   -0.467   -0.061
vc_rpsv:cntrstvl            0.213 -0.304 -0.700 -0.419  -0.756   -0.405    0.056
vc_rpsv:cntrstvw            0.202 -0.321 -0.742 -0.456  -0.410   -0.737    0.030
prsnttnscnd:vc_rpsv:cntrstc -0.112  0.613  0.450  0.428   0.245    0.264   -0.122
prsnttnscnd:vc_rpsv:cntrstvl -0.023  0.510  0.349  0.223   0.427    0.226   -0.216
prsnttnscnd:vc_rpsv:cntrstvw -0.068  0.483  0.340  0.205   0.191    0.362   -0.163

                           duratn ctrial prsn:_ prsnttnscnd:cntrstc
prsnttnscnd
voice_repsv
cntrstcntrl
contrastvlr
contrastvwl
NBC
duration
```

```
ctrial                        -0.032
prsnttnsc:_                    0.003 -0.041
prsnttnscnd:cntrstc            0.102 -0.043  0.595
prsnttnscnd:cntrstvl           0.167 -0.056  0.485  0.572
prsnttnscnd:cntrstvw           0.112 -0.041  0.493  0.577
vc_rpsv:cntrstc                0.092  0.082  0.439  0.446
vc_rpsv:cntrstvl               0.077  0.125  0.375  0.243
vc_rpsv:cntrstvw               0.132  0.112  0.396  0.262
prsnttnscnd:vc_rpsv:cntrstc   -0.060  0.045 -0.859 -0.736
prsnttnscnd:vc_rpsv:cntrstvl -0.114  0.061 -0.710 -0.437
prsnttnscnd:vc_rpsv:cntrstvw -0.049  0.047 -0.677 -0.409

                              prsnttnscnd:   prsnttnscnd:
                              cntrstvl       cntrstvw
prsnttnscnd
voice_repsv
cntrstcntrl
contrastvlr
contrastvwl
NBC
duration
ctrial
prsnttnsc:_
prsnttnscnd:cntrstc
prsnttnscnd:cntrstvl
prsnttnscnd:cntrstvw           0.488
vc_rpsv:cntrstc                0.219          0.226
vc_rpsv:cntrstvl               0.414          0.205
vc_rpsv:cntrstvw               0.216          0.384
prsnttnscnd:vc_rpsv:cntrstc   -0.427         -0.431
prsnttnscnd:vc_rpsv:cntrstvl -0.766         -0.379
prsnttnscnd:vc_rpsv:cntrstvw -0.348         -0.711

                              vc_rpsv:  vc_rpsv:  vc_rpsv:
                              cntrstc   cntrstvl  cntrstvw
prsnttnscnd
voice_repsv
cntrstcntrl
contrastvlr
contrastvwl
NBC
duration
ctrial
prsnttnsc:_
prsnttnscnd:cntrstc
prsnttnscnd:cntrstvl
prsnttnscnd:cntrstvw
vc_rpsv:cntrstc
vc_rpsv:cntrstvl               0.577
```

```
vc_rpsv:cntrstvw              0.614     0.543
prsnttnscnd:vc_rpsv:cntrstc  -0.575    -0.323    -0.344
prsnttnscnd:vc_rpsv:cntrstvl -0.301    -0.548    -0.284
prsnttnscnd:vc_rpsv:cntrstvw -0.289    -0.252    -0.493

                             prsnttnscnd:vc_rpsv:  prsnttnscnd:vc_rpsv:
                             cntrstc               cntrstvl
prsnttnscnd
voice_repsv
cntrstcntrl
contrastvlr
contrastvwl
NBC
duration
ctrial
prsnttnsc:_
prsnttnscnd:cntrstc
prsnttnscnd:cntrstvl
prsnttnscnd:cntrstvw
vc_rpsv:cntrstc
vc_rpsv:cntrstvl
vc_rpsv:cntrstvw
prsnttnscnd:vc_rpsv:cntrstc
prsnttnscnd:vc_rpsv:cntrstvl  0.617
prsnttnscnd:vc_rpsv:cntrstvw  0.585                  0.496
```

5 optimal model fitted for reaction time (repeated words)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep * presentation + duration + ctrial +
(1 +  ctrial | participant)
   Data: dfcontrol

REML criterion at convergence: 9564.1

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7922 -0.6322 -0.0842  0.5450  3.5513

Random effects:
 Groups      Name        Variance  Std.Dev. Corr
 participant (Intercept) 2.9502683 1.71763
             ctrial      0.0001726 0.01314  -0.26
 Residual                7.1394867 2.67198
Number of obs: 1947, groups: participant, 57

Fixed effects:
                                 Estimate
                                 Std. Error t value
```

```
(Intercept)                      -3.112e+01  4.613e-01  -67.47
voice_repsv                       5.684e-02  1.715e-01    0.33
presentationsecond                4.288e-01  2.041e-01    2.10
duration                          5.305e-03  4.179e-04   12.69
ctrial                            7.790e-03  2.648e-03    2.94
voice_repsv:presentationsecond -1.123e+00  2.544e-01   -4.41

Correlation of Fixed Effects:
            (Intr) vc_rps prsntt duratn ctrial
voice_repsv -0.330
prsnttnscnd -0.356  0.531
duration    -0.821  0.133  0.201
ctrial       0.052 -0.065 -0.144 -0.130
vc_rpsv:prs  0.288 -0.684 -0.788 -0.169  0.050
```

6 OPTIMAL MODEL FITTED FOR REACTION TIME (MINIMAL PAIRS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ presentation * voice_rep * contrast + duration + ctrial
    + (1 + ctrial | participant)
   Data: dfminpair

REML criterion at convergence: 4273

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.8321 -0.6319 -0.0815  0.6249  3.0422

Random effects:
 Groups      Name        Variance Std.Dev. Corr
 participant (Intercept) 2.540537 1.59391
             ctrial      0.000204 0.01428  -0.39
 Residual                7.584349 2.75397
Number of obs: 852, groups: participant, 57

Fixed effects:
                                              Estimate Std. Error t value
(Intercept)                                 -2.536e+01  6.336e-01  -40.03
presentationsecond                          -1.109e+00  4.691e-01   -2.36
voice_repsv                                 -1.544e+00  5.248e-01   -2.94
contrastvelar                               -3.777e-01  4.992e-01   -0.76
contrastvowel                                4.018e-01  5.097e-01    0.79
duration                                    -2.218e-04  6.409e-04   -0.35
ctrial                                       2.370e-02  5.946e-03    3.99
presentationsecond:voice_repsv               3.540e+00  7.256e-01    4.88
presentationsecond:contrastvelar             1.130e+00  6.429e-01    1.76
presentationsecond:contrastvowel             1.710e+00  7.394e-01    2.31
voice_repsv:contrastvelar                    3.479e+00  8.769e-01    3.97
voice_repsv:contrastvowel                    4.815e-01  8.526e-01    0.56
```

```
presentationsecond:voice_repsv:contrastvelar -3.706e+00  9.592e-01   -3.86
presentationsecond:voice_repsv:contrastvowel -1.744e+00  1.105e+00   -1.58

Correlation of Fixed Effects:
                             (Intr) prsntt vc_rps cntrstvl cntrstvw duratn
prsnttnscnd                  -0.451
voice_repsv                  -0.485  0.381
contrastvlr                  -0.308  0.346  0.675
contrastvwl                  -0.173  0.304  0.623  0.669
duration                     -0.794  0.217  0.107 -0.115   -0.264
ctrial                        0.134 -0.041 -0.604 -0.577   -0.545    0.082
prsnttnsc:_                   0.333 -0.656 -0.412 -0.171   -0.135   -0.216
prsnttnscnd:cntrstvl          0.143 -0.680 -0.215 -0.483   -0.244    0.057
prsnttnscnd:cntrstvw          0.079 -0.579 -0.158 -0.191   -0.427    0.094
vc_rpsv:cntrstvl              0.336 -0.238 -0.832 -0.812   -0.593   -0.003
vc_rpsv:cntrstvw              0.319 -0.238 -0.835 -0.637   -0.795    0.021
prsnttnscnd:vc_rpsv:cntrstvl -0.128  0.462  0.309  0.326    0.158    0.000
prsnttnscnd:vc_rpsv:cntrstvw -0.128  0.408  0.263  0.131    0.274    0.025

                                            prsnttnscnd:  prsnttnscnd:
                            ctrial  prsn:_  cntrstvl      cntrstvw
prsnttnscnd
voice_repsv
contrastvlr
contrastvwl
duration
ctrial
prsnttnsc:_                  -0.048
prsnttnscnd:cntrstvl         -0.011  0.433
prsnttnscnd:cntrstvw         -0.041  0.374     0.446
vc_rpsv:cntrstvl              0.716  0.222     0.265       0.081
vc_rpsv:cntrstvw              0.712  0.223     0.130       0.220
prsnttnscnd:vc_rpsv:cntrstvl  0.001 -0.720    -0.664      -0.298
prsnttnscnd:vc_rpsv:cntrstvw  0.014 -0.636    -0.291      -0.660

                             vc_rpsv:   vc_rpsv:   prsnttnscnd:
                             cntrstvl   cntrstvw   vc_rpsv:cntrstvl
prsnttnscnd
voice_repsv
contrastvlr
contrastvwl
duration
ctrial
prsnttnsc:_
prsnttnscnd:cntrstvl
prsnttnscnd:cntrstvw
vc_rpsv:cntrstvl
vc_rpsv:cntrstvw               0.778
prsnttnscnd:vc_rpsv:cntrstvl  -0.389   -0.188
```

```
prsnttnscnd:vc_rpsv:cntrstvw  -0.151   -0.324     0.475
```

7 OPTIMAL MODEL FITTED FOR REACTION TIME (VOWEL MINIMAL PAIRS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep * presentation + ctrial
     + (1 + ctrial | participant)
   Data: dfminvowel

REML criterion at convergence: 1220.1

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.5920 -0.5362  0.0130  0.5899  2.3166

Random effects:
 Groups      Name        Variance  Std.Dev. Corr
 participant (Intercept) 2.404e+00 1.550488
             ctrial      6.169e-06 0.002484 1.00
 Residual                5.949e+00 2.438967
Number of obs: 251, groups: participant, 57

Fixed effects:
                                 Estimate Std. Error t value
(Intercept)                    -24.988996   0.395052  -63.25
voice_repsv                     -1.691230   0.593422   -2.85
presentationsecond               0.802632   0.535156    1.50
ctrial                           0.012121   0.008854    1.37
voice_repsv:presentationsecond   1.810971   0.759817    2.38

Correlation of Fixed Effects:
            (Intr) vc_rps prsntt ctrial
voice_repsv -0.724
prsnttnscnd -0.210  0.051
ctrial      -0.527  0.800 -0.201
vc_rpsv:prs  0.233 -0.289 -0.659 -0.013
```

8 OPTIMAL MODEL FITTED FOR REACTION TIME (VELAR MINIMAL PAIRS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep * presentation + duration + ctrial
+ (1 + ctrial | participant)
   Data: dfminvelar

REML criterion at convergence: 1580.8

Scaled residuals:
```

```
     Min       1Q   Median       3Q      Max
-1.91990 -0.71625 -0.08629  0.65677  2.27232

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 participant  (Intercept) 2.0671530 1.43776
          ctrial      0.0004218 0.02054  -0.03
 Residual             7.7674777 2.78702
Number of obs: 308, groups: participant, 57

Fixed effects:
                                 Estimate Std. Error t value
(Intercept)                    -20.870363   2.653474  -7.865
voice_repsv                      1.634008   0.941494   1.736
presentationsecond              -1.069769   0.944758  -1.132
duration                        -0.005909   0.003372  -1.752
ctrial                           0.021972   0.016217   1.355
voice_repsv:presentationsecond   1.066008   1.008228   1.057

Correlation of Fixed Effects:
            (Intr) vc_rps prsntt duratn ctrial
voice_repsv  0.423
prsnttnscnd -0.879 -0.438
duration    -0.986 -0.538  0.868
ctrial       0.591  0.876 -0.649 -0.676
vc_rpsv:prs  0.775  0.178 -0.871 -0.750  0.470
```

9 OPTIMAL MODEL FITTED FOR REACTION TIME (BILABIALS MINIMAL PAIRS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep * presentation + ctrial + (1 + ctrial | participant)
   Data: dfminbilabial

REML criterion at convergence: 1512.6

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.05679 -0.64750 -0.08387  0.64604  2.75905

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 participant  (Intercept) 3.705363 1.92493
          ctrial      0.000115 0.01072  -1.00
 Residual             7.975578 2.82411
Number of obs: 293, groups: participant, 57

Fixed effects:
                                 Estimate Std. Error t value
```

```
(Intercept)                     -24.830617   0.467636  -53.10
voice_repsv                      -2.881026   0.672360   -4.28
presentationsecond               -1.272119   0.478985   -2.66
ctrial                            0.045227   0.008915    5.07
voice_repsv:presentationsecond   3.354520   0.732674    4.58

Correlation of Fixed Effects:
            (Intr) vc_rps prsntt ctrial
voice_repsv -0.716
prsnttnscnd -0.424  0.332
ctrial       0.473 -0.786 -0.092
vc_rpsv:prs  0.211 -0.291 -0.638 -0.056
```

10 OPTIMAL MODEL FITTED FOR REACTION TIME (PSEUDO-MINIMAL PAIRS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ contrast * presentation * voice_rep
+ duration + ctrial + (1 + ctrial | subject)
   Data: dfpseupair

REML criterion at convergence: 7301.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.4947 -0.6888 -0.0228  0.6407  2.7750

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 subject  (Intercept) 2.7837150 1.66845
          ctrial      0.0003213 0.01792  -0.73
 Residual             8.4269229 2.90292
Number of obs: 1435, groups: subject, 57

Fixed effects:
                                              Estimate Std. Error t value
(Intercept)                                 -2.840e+01  4.947e-01  -57.41
contrastvelar                               -3.887e-01  4.114e-01   -0.94
contrastvowel                                2.034e-01  3.686e-01    0.55
presentationsecond                           1.468e-01  4.540e-01    0.32
voice_repsv                                 -2.696e-01  3.566e-01   -0.76
duration                                     2.482e-03  5.822e-04    4.26
ctrial                                       2.178e-02  3.252e-03    6.70
contrastvelar:presentationsecond             3.741e-01  7.529e-01    0.50
contrastvowel:presentationsecond             2.180e+00  6.327e-01    3.45
contrastvelar:voice_repsv                    1.253e+00  5.116e-01    2.45
contrastvowel:voice_repsv                    5.657e-01  4.821e-01    1.17
presentationsecond:voice_repsv               1.476e+00  5.986e-01    2.47
contrastvelar:presentationsecond:voice_repsv -1.295e+00  9.162e-01   -1.41
```

```
contrastvowel:presentationsecond:voice_repsv -3.539e+00  8.655e-01   -4.09

Correlation of Fixed Effects:
            (Intr) cntrstvl cntrstvw prsntt vc_rps duratn ctrial
contrastvlr -0.050
contrastvwl -0.029  0.601
prsnttnscnd -0.102  0.440    0.502
voice_repsv -0.157  0.566    0.639    0.475
duration    -0.729 -0.375   -0.453   -0.254 -0.302
ctrial      -0.182 -0.048   -0.034    0.011 -0.099 -0.016
cntrstvlr:p -0.055 -0.573   -0.365   -0.642 -0.316  0.304 -0.078
cntrstvwl:p  0.037 -0.328   -0.560   -0.729 -0.347  0.227 -0.043
cntrstvlr:_  0.060 -0.801   -0.478   -0.349 -0.722  0.280  0.084
cntrstvwl:_  0.001 -0.477   -0.783   -0.391 -0.787  0.380  0.067
prsnttnsc:_  0.087 -0.319   -0.365   -0.754 -0.563  0.171 -0.069
cntrstvl::_  0.041  0.463    0.293    0.526  0.395 -0.240  0.098
cntrstvw::_ -0.034  0.229    0.399    0.529  0.394 -0.151  0.076

             cntrstvl: cntrstvw:cntrstvl:_ cntrstvw:_ prsn:_ cntrstvl::_
contrastvlr
contrastvwl
prsnttnscnd
voice_repsv
duration
ctrial
cntrstvlr:p
cntrstvwl:p   0.480
cntrstvlr:_   0.447     0.255
cntrstvwl:_   0.282     0.432    0.577
prsnttnsc:_   0.490     0.554    0.402      0.444
cntrstvl::_  -0.823    -0.396   -0.554     -0.330     -0.683
cntrstvw::_  -0.353    -0.732   -0.282     -0.515     -0.700  0.485
```

11 OPTIMAL MODEL FITTED FOR REACTION TIME (VOWEL PSEUDO-MINIMAL PAIRS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep * presentation + ctrial + (1 + ctrial | participant)
   Data: dfpseuvowel

REML criterion at convergence: 2645.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.4270 -0.6086  0.0031  0.6409  2.6295

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 participant  (Intercept) 2.8239703 1.68047
```

```
          ctrial      0.0001104 0.01051  -0.97
 Residual             7.6352553 2.76320
Number of obs: 524, groups: participant, 57

Fixed effects:
                                 Estimate Std. Error t value
(Intercept)                    -26.057300   0.297376  -87.62
voice_repsv                      0.015879   0.275363    0.06
presentationsecond               2.372748   0.417149    5.69
ctrial                           0.010389   0.003609    2.88
voice_repsv:presentationsecond  -2.066971   0.596360   -3.47

Correlation of Fixed Effects:
            (Intr) vc_rps prsntt ctrial
voice_repsv -0.474
prsnttnscnd -0.316  0.338
ctrial      -0.261 -0.009 -0.107
vc_rpsv:prs  0.221 -0.461 -0.702  0.080
```

12 OPTIMAL MODEL FITTED FOR REACTION TIME (BILABIAL PSEUDO-MINIMAL PAIRS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep * presentation + duration
+ ctrial + (1 + ctrial | participant)
   Data: dfpseubilabial

REML criterion at convergence: 2410.5

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.58198 -0.68206 -0.02711  0.64784  2.48044

Random effects:
 Groups      Name        Variance  Std.Dev. Corr
 participant (Intercept) 3.2362414 1.79896
             ctrial      0.0002896 0.01702  -1.00
 Residual                9.0298236 3.00497
Number of obs: 459, groups: participant, 57

Fixed effects:
                                 Estimate Std. Error t value
(Intercept)                    -29.742578   0.712255  -41.76
voice_repsv                     -1.213259   0.437552   -2.77
presentationsecond              -0.412777   0.502228   -0.82
duration                         0.005254   0.001049    5.01
ctrial                           0.038598   0.005248    7.35
voice_repsv:presentationsecond   1.488616   0.635074    2.34
```

```
Correlation of Fixed Effects:
            (Intr) vc_rps prsntt duratn ctrial
voice_repsv  0.231
prsnttnscnd  0.153  0.537
duration    -0.859 -0.559 -0.402
ctrial      -0.398 -0.412 -0.145  0.397
vc_rpsv:prs -0.035 -0.501 -0.744  0.196 -0.073
```

13 optimal model fitted for reaction time (velar pseudo-minimal pairs)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep * presentation + ctrial
    + (1 + ctrial | participant)
   Data: dfpseuvelar

REML criterion at convergence: 2325

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.09427 -0.65219 -0.06612  0.60103  2.55846

Random effects:
 Groups      Name        Variance  Std.Dev. Corr
 participant (Intercept) 2.3862235 1.54474
             ctrial      0.0004964 0.02228  -0.59
 Residual                8.1409359 2.85323
Number of obs: 452, groups: participant, 57

Fixed effects:
                                 Estimate Std. Error t value
(Intercept)                    -26.600428   0.345436  -77.01
voice_repsv                      0.849045   0.348256    2.44
presentationsecond              -0.183386   0.579279   -0.32
ctrial                           0.031706   0.004987    6.36
voice_repsv:presentationsecond   0.909578   0.666101    1.37

Correlation of Fixed Effects:
            (Intr) vc_rps prsntt ctrial
voice_repsv -0.643
prsnttnscnd -0.391  0.371
ctrial      -0.205  0.054 -0.192
vc_rpsv:prs  0.341 -0.510 -0.860  0.153
```

MODELS FITTED IN EXPERIMENT 2

1 OPTIMAL MODEL FOR RESPONSE CHOICE

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: correct ~ native_language * word_type + (1|subject)
    + (1 | stimulus)
   Data: data

     AIC      BIC   logLik deviance df.resid
 11734.2  11779.7  -5861.1  11722.2    14438

Scaled residuals:
    Min      1Q  Median      3Q     Max
-7.4118  0.1857  0.2929  0.4396  4.1169

Random effects:
 Groups   Name        Variance Std.Dev.
 stimulus (Intercept) 0.6476   0.8047
 subject  (Intercept) 0.1473   0.3838
Number of obs: 14444, groups: stimulus, 234; subject, 46

Fixed effects:
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                          2.3233     0.1209  19.218   <2e-16 ***
native_languageSpanish              -1.8860     0.1323 -14.259   <2e-16 ***
word_typerw                          0.2772     0.1336   2.075    0.038 *
native_languageSpanish:word_typerw   0.9831     0.1006   9.771   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) ntv_lS wrd_ty
ntv_lnggSpn -0.564
word_typerw -0.528  0.167
ntv_lnggS:_  0.247 -0.335 -0.474
```

2 OPTIMAL MODEL FITTED FOR RESPONSE ACCURACY TO REAL WORDS (NATIVE SPEAKERS)

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
```

```
Formula: correct ~ +logfreq + word_pair + prestion + (1 | subject)
    + (1 | stimulus)
   Data: subset(data, word_type == "rw" & native_language == "English")

     AIC      BIC   logLik deviance df.resid
  2054.9   2092.5  -1021.4   2042.9     3894

Scaled residuals:
    Min      1Q  Median      3Q     Max
-8.9062  0.1480  0.2132  0.2997  1.7710

Random effects:
 Groups   Name        Variance Std.Dev.
 stimulus (Intercept) 0.9722   0.9860
 subject  (Intercept) 0.4166   0.6455
Number of obs: 3900, groups: stimulus, 116; subject, 25

Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.5398     0.5518   2.790  0.00526 **
logfreq          0.2688     0.1567   1.716  0.08624 .
word_pairsame    0.6063     0.2368   2.561  0.01044 *
prestionsecond   0.4104     0.1562   2.627  0.00861 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) logfrq wrd_pr
logfreq     -0.924
word_pairsm -0.121 -0.040
prestinscnd -0.089 -0.047  0.050
```

3 OPTIMAL MODEL FITTED FOR RESPONSE ACCURACY TO PSEUDOWORDS (NATIVE SPEAKERS)

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: correct ~ height + word_pair * prestion + (1 | subject)
    + (1 | stimulus)
   Data: subset(data, word_type == "pw" & native_language == "English")

     AIC      BIC   logLik deviance df.resid
  2318.1   2362.0  -1152.0   2304.1     3943

Scaled residuals:
    Min      1Q  Median      3Q     Max
-7.7581  0.1489  0.2188  0.3144  2.7991
```

```
Random effects:
 Groups   Name        Variance Std.Dev.
 stimulus (Intercept) 1.540    1.2409
 subject  (Intercept) 0.331    0.5753
Number of obs: 3950, groups: stimulus, 118; subject, 25

Fixed effects:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                    2.0849     0.1422  14.658  < 2e-16 ***
heightopen                    -0.7294     0.1059  -6.886 5.74e-12 ***
word_pairsame                  0.8476     0.1464   5.790 7.02e-09 ***
prestionsecond                 0.4709     0.1354   3.478 0.000506 ***
word_pairsame:prestionsecond  -0.6387     0.2097  -3.046 0.002318 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) hghtpn wrd_pr prstns
heightopen  -0.452
word_pairsm -0.353 -0.032
prestinscnd -0.392 -0.016  0.391
wrd_prsm:pr  0.251  0.013 -0.698 -0.646
```

4 OPTIMAL MODEL FITTED FOR RESPONSE ACCURACY TO REAL WORDS (LEARNERS)

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: correct ~ +logfreq + word_pair + prestion + height
+ (1 | subject) + (1 | stimulus)
   Data: subset(data, word_type == "rw" & native_language == "Spanish")

     AIC      BIC   logLik deviance df.resid
  2821.6   2864.3  -1403.8   2807.6     3269

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.8308  0.2101  0.3418  0.4765  1.5710

Random effects:
 Groups   Name        Variance Std.Dev.
 stimulus (Intercept) 0.4561   0.6753
 subject  (Intercept) 0.4069   0.6379
Number of obs: 3276, groups: stimulus, 116; subject, 21

Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     1.02540    0.42304   2.424   0.0154 *
```

```
logfreq         0.27949    0.11205   2.494   0.0126 *
word_pairsame   0.42062    0.16720   2.516   0.0119 *
prestionsecond  0.04933    0.11737   0.420   0.6743
heightopen     -0.67666    0.16404  -4.125 3.71e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) logfrq wrd_pr prstns
logfreq     -0.874
word_pairsm -0.090 -0.062
prestinscnd -0.108 -0.036  0.003
heightopen  -0.285  0.082 -0.027 -0.001
```

5 OPTIMAL MODEL FITTED TO RESPONSE ACCURACY TO PSEUDOWORDS (LEARNERS)

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: correct ~ height + prestion * word_pair + (1|subject) + (1|stimulus)
   Data: subset(data, word_type == "pw" & native_language == "Spanish")

     AIC      BIC   logLik deviance df.resid
  4131.4   4174.2  -2058.7   4117.4     3311

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0089 -0.9267  0.4833  0.7223  3.5330

Random effects:
 Groups   Name        Variance Std.Dev.
 stimulus (Intercept) 0.5238   0.7237
 subject  (Intercept) 0.3631   0.6026
Number of obs: 3318, groups: stimulus, 118; subject, 21

Fixed effects:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.26232    0.14315   1.832  0.06688 .
heightopen                    0.12113    0.07345   1.649  0.09912 .
prestionsecond                0.15673    0.10397   1.507  0.13171
word_pairsame                 0.28623    0.10410   2.749  0.00597 **
prestionsecond:word_pairsame -0.36639    0.14689  -2.494  0.01262 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) hghtpn prstns wrd_pr
heightopen  -0.248
```

```
prestinscnd -0.359  0.002
word_pairsm -0.356 -0.006  0.493
prstnscnd:_  0.254 -0.003 -0.708 -0.709
```

6 beyond optimal model fitted for reaction time to repeated words (open vowels)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ native_language * prestion + cduration + logfreq +
    duration + trial_order + (1 + trial_order | participant) + (1 |stimulus)
   Data: sameopen

REML criterion at convergence: 10969.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.8954 -0.6326 -0.0350  0.5692  3.3387

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 participant  (Intercept) 12895.957 113.5604
          trial_order     0.162   0.4025 -0.74
 stimulus (Intercept)  3178.805  56.3809
 Residual             23928.683 154.6890
Number of obs: 842, groups: participant, 46; stimulus, 21

Fixed effects:
                                        Estimate Std. Error t value
(Intercept)                           -913.87961   88.54408 -10.321
native_languageSpanish                  83.12162   27.53504   3.019
prestionsecond                         -28.55444   15.03234  -1.900
cduration                              512.57495  135.83594   3.773
logfreq                                -27.91572   25.59602  -1.091
trial_order                              0.01256    0.16234   0.077
native_languageSpanish:prestionsecond  -34.80603   21.78495  -1.598

Correlation of Fixed Effects:
            (Intr) ntv_lS prstns cdurtn logfrq trl_rd
ntv_lnggSpn -0.133
prestinscnd -0.078  0.240
cduration    0.015 -0.005 -0.350
logfreq     -0.898 -0.003  0.016  0.054
trial_order -0.137 -0.006 -0.045 -0.154 -0.239
ntv_lnggSp:  0.051 -0.393 -0.599  0.005  0.000  0.001
```

7 OPTIMAL MODEL FITTED FOR REACTION TIME TO REPEATED WORDS (OPEN VOWELS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ native_language * prestion + cduration
+ (1 + trial_order |  participant) + (1 | stimulus)
   Data: sameopen

REML criterion at convergence: 10977.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.9192 -0.6409 -0.0396  0.5709  3.3261

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 participant  (Intercept) 1.286e+04 113.3909
          trial_order 1.609e-01   0.4011 -0.73
 stimulus (Intercept) 3.025e+03  55.0039
 Residual             2.393e+04 154.6949
Number of obs: 842, groups: participant, 46; stimulus, 21

Fixed effects:
                                      Estimate Std. Error t value
(Intercept)                           -1006.67      21.96  -45.84
native_languageSpanish                   82.96      27.56    3.01
prestionsecond                          -28.43      13.99   -2.03
cduration                               517.55     132.63    3.90
native_languageSpanish:prestionsecond   -34.80      21.79   -1.60

Correlation of Fixed Effects:
            (Intr) ntv_lS prstns cdurtn
ntv_lnggSpn -0.557
prestinscnd -0.316  0.240
cduration    0.043 -0.005 -0.358
ntv_lnggSp:  0.206 -0.392 -0.601  0.005
```

8 OPTIMAL MODEL FITTED TO REACTION TIME FOR MINIMAL PAIRS (OPEN VOWELS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ native_language + prestion + cduration
+ (1 + trial_order | participant) + (1 | stimulus)
   Data: diffopen

REML criterion at convergence: 9487.5

Scaled residuals:
     Min       1Q   Median       3Q      Max
```

```
-2.96178 -0.62128 -0.07782  0.59847  3.15575

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 participant  (Intercept) 8.786e+03  93.7338
          trial_order 1.216e-01   0.3487 -0.65
 stimulus (Intercept) 3.936e+03  62.7409
 Residual             2.785e+04 166.8800
Number of obs: 719, groups: participant, 46; stimulus, 38

Fixed effects:
                       Estimate Std. Error t value
(Intercept)             -945.11      23.09  -40.94
native_languageSpanish    20.63      25.03    0.82
prestionsecond           -20.37      24.53   -0.83
cduration                512.31     119.92    4.27

Correlation of Fixed Effects:
            (Intr) ntv_lS prstns
ntv_lnggSpn -0.477
prestinscnd -0.535  0.000
cduration    0.114 -0.006 -0.209
```

9 OPTIMAL MODEL FITTED TO REACTION TIME FOR REPEATED WORDS (CLOSE VOWELS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ native_language + prestion + cduration
+ (1 + trial_order |  participant) + (1 | stimulus)
   Data: sameclose

REML criterion at convergence: 10760.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1742 -0.6665 -0.0479  0.5828  3.3530

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 participant  (Intercept) 1.269e+04 112.6363
          trial_order 7.787e-02   0.2791 -0.51
 stimulus (Intercept) 3.654e+03  60.4510
 Residual             3.073e+04 175.2906
Number of obs: 810, groups: participant, 46; stimulus, 19

Fixed effects:
                        Estimate Std. Error t value
(Intercept)            -1024.487     26.248  -39.03
native_languageSpanish     7.252     31.374    0.23
```

```
prestionsecond           -46.243     12.574   -3.68
cduration                681.176    131.996    5.16

Correlation of Fixed Effects:
            (Intr) ntv_lS prstns
ntv_lnggSpn -0.544
prestinscnd -0.255  0.000
cduration    0.117  0.001 -0.187
```

10 OPTIMAL MODEL FITTED TO REACTION TIME FOR MINIMAL PAIRS (CLOSE VOWELS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ native_language * prestion + cduration
+ (1 + trial_order |  participant) + (1 | stimulus)
   Data: diffclose

REML criterion at convergence: 10137.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5834 -0.6389 -0.0481  0.6048  2.8101

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 participant  (Intercept) 1.522e+04 123.3495
          trial_order 2.197e-01   0.4687 -0.69
 stimulus (Intercept) 4.195e+03  64.7708
 Residual             3.024e+04 173.8888
Number of obs: 763, groups: participant, 46; stimulus, 38

Fixed effects:
                                       Estimate Std. Error t value
(Intercept)                             -982.63      26.23  -37.46
native_languageSpanish                    41.85      32.11    1.30
prestionsecond                           -20.25      27.15   -0.75
cduration                                317.77     159.09    2.00
native_languageSpanish:prestionsecond    -64.62      25.55   -2.53

Correlation of Fixed Effects:
            (Intr) ntv_lS prstns cdurtn
ntv_lnggSpn -0.552
prestinscnd -0.513  0.165
cduration    0.049 -0.010 -0.022
ntv_lnggSp:  0.216 -0.398 -0.424  0.015
```

11 BEYOND OPTIMAL MODEL FITTED TO REACTION TIME FOR SECOND ITEMS OF REPEATED WORDS

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ native_language + voice_rep + height + logfreq + distance +
    duration + ctrial_order + (1 + ctrial_order | subject) +  (1 | stimulus)
   Data: voicedf

REML criterion at convergence: 21969.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5218 -0.6470 -0.0768  0.5711  3.3527

Random effects:
 Groups   Name         Variance  Std.Dev. Corr
 subject  (Intercept)  8.623e+03  92.8615
          ctrial_order 2.328e-02   0.1526 -0.30
 stimulus (Intercept)  3.290e+03  57.3576
 Residual              2.982e+04 172.6960
Number of obs: 1662, groups: subject, 46; stimulus, 40

Fixed effects:
                         Estimate Std. Error t value
(Intercept)            -1.351e+03  1.337e+02 -10.100
native_languageSpanish  3.193e+01  2.856e+01   1.118
voice_repsv            -1.818e+01  8.512e+00  -2.135
heightopen              6.748e+01  2.097e+01   3.217
logfreq                -1.632e+01  1.803e+01  -0.905
distance               -2.521e+00  5.954e+00  -0.423
duration                6.012e+02  1.171e+02   5.133
ctrial_order           -4.465e-02  1.113e-01  -0.401

Correlation of Fixed Effects:
            (Intr) ntv_lS vc_rps hghtpn logfrq distnc duratn
ntv_lnggSpn -0.098
voice_repsv -0.035  0.001
heightopen   0.201  0.004  0.003
logfreq     -0.595  0.000  0.001 -0.134
distance    -0.697  0.001  0.005 -0.170  0.175
duration    -0.698  0.000  0.000 -0.257  0.121  0.256
ctrial_ordr  0.190  0.001  0.001  0.049 -0.220 -0.150 -0.048
```

12 OPTIMAL MODEL FITTED TO REACTION TIME FOR SECOND ITEMS OF REPEATED WORDS

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep + native_language + height + cduration +
    ctrial_order + (1 + ctrial_order | subject) + (1 | stimulus)
```

```
   Data: voicedf

REML criterion at convergence: 21983

Scaled residuals:
   Min     1Q Median     3Q    Max
-3.538 -0.643 -0.073  0.575  3.364

Random effects:
 Groups   Name         Variance   Std.Dev. Corr
 subject  (Intercept)   8618.0719  92.834
          ctrial_order     0.0231   0.152  -0.30
 stimulus (Intercept)   3165.0514  56.259
 Residual              29825.4457 172.700
Number of obs: 1662, groups: subject, 46; stimulus, 40

Fixed effects:
                         Estimate Std. Error t value
(Intercept)             -1071.6457    24.0651   -44.5
voice_repsv               -18.1569     8.5120    -2.1
native_languageSpanish     31.9402    28.5547     1.1
heightopen                 64.0851    20.2277     3.2
cduration                 621.5699   111.0880     5.6
ctrial_order               -0.0702     0.1062    -0.7

Correlation of Fixed Effects:
            (Intr) vc_rps ntv_lS hghtpn cdurtn
voice_repsv -0.178
ntv_lnggSpn -0.541  0.001
heightopen  -0.427  0.004  0.005
cduration    0.028 -0.001 -0.001 -0.218
ctrial_ordr -0.037  0.002  0.001  0.003  0.006
```

13 OPTIMAL MODEL FITTED TO REACTION TIME FOR SECOND ITEMS OF REPEATED WORDS (LEARNERS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ cvoice_rep + cheight + cduration + ctrial_order
+ (1 + ctrial_order | subject) + (1 | stimulus)
   Data: voiceSpdf

REML criterion at convergence: 9466.1

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.90165 -0.68229 -0.05881  0.62950  3.06815

Random effects:
 Groups   Name         Variance  Std.Dev. Corr
```

```
 stimulus (Intercept)  3.659e+03  60.4870
 subject  (Intercept)  1.157e+04 107.5790
          ctrial_order 6.067e-02   0.2463 -0.10
 Residual              3.032e+04 174.1400
Number of obs: 714, groups: stimulus, 40; subject, 21

Fixed effects:
               Estimate Std. Error t value
(Intercept)  -1013.1946   26.2644   -38.58
voice_rep      -30.5416   13.1266    -2.33
height          89.7835   23.8326     3.77
cduration      513.7042  130.5756     3.93
ctrial_order    -0.1116    0.1338    -0.83

Correlation of Fixed Effects:
            (Intr) vc_rps hghtpn duratn
voice_repsv -0.075
heightopen   0.067  0.009
duration    -0.834 -0.005 -0.222
ctrial_ordr -0.014  0.003 -0.001  0.004
```

14 OPTIMAL MODEL FITTED TO REACTION TIME FOR REPEATED WORDS (NATIVE SPEAKERS)

```
Linear mixed model fit by REML ['lmerMod']
Formula: transRT ~ voice_rep + height + cduration + ctrial_order
+ (1 +trial_order | subject) + (1 | stimulus)
   Data: voiceEndf

REML criterion at convergence: 12475

Scaled residuals:
   Min     1Q Median     3Q    Max
-3.706 -0.648 -0.056  0.502  3.395

Random effects:
 Groups   Name         Variance    Std.Dev. Corr
 stimulus (Intercept)  4347.59543  65.9363
 subject  (Intercept)  9023.21930  94.9906
          trial_order     0.00837   0.0915 -0.86
 Residual             27884.74043 166.9872
Number of obs: 948, groups: stimulus, 40; subject, 25

Fixed effects:
               Estimate Std. Error t value
(Intercept)  -1068.1762    23.9002   -44.73
voice_repsv     -9.9477    10.8972    -0.91
heightopen      45.4930    24.1210     1.92
cduration      709.1534   132.4656     5.40
```

```
ctrial_order    -0.0428     0.1251    -0.32

Correlation of Fixed Effects:
            (Intr) vc_rps hghtpn cdurtn
voice_repsv -0.228
heightopen  -0.513  0.002
cduration    0.032  0.001 -0.217
ctrial_ordr -0.080  0.002  0.003  0.007
```

1 MODEL FITTED TO RESPONSE ACCURACY (PRE- AND POST-TRAINING BLOCKS INCLUDED)

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: correct ~ +block * voice * usetraining + cproficiency
+ clogfreq + (1 | subject) + (1 | stimuli)
   Data: dferror

     AIC      BIC   logLik deviance df.resid
    5537     5615    -2757     5513     4947

Scaled residuals:
   Min     1Q Median     3Q    Max
-6.651 -0.845  0.403  0.660  2.352

Random effects:
 Groups  Name        Variance Std.Dev.
 stimuli (Intercept) 0.458    0.677
 subject (Intercept) 0.229    0.478
Number of obs: 4959, groups: stimuli, 40; subject, 31

Fixed effects:
                              Estimate Std. Error z value   Pr(>|z|)
(Intercept)                      0.929      0.201    4.62   0.00 ***
blockposttraining                0.150      0.119    1.26   0.208
voicesame                       -0.347      0.208   -1.67   0.095 .
usetrainingyes                   0.367      0.282    1.30   0.193
cproficiency                     0.855      0.136    6.29   0.00 ***
clogfreq                         0.401      0.163    2.47   0.014 *
blockposttraining:voicesame     -0.067      0.159   -0.42   0.673
blockposttraining:usetrainyes   -0.028      0.221   -0.13   0.899
voicesame:usetrainingyes        -0.248      0.204   -1.22   0.224
blockpost:vosame:usetrainyes     0.631      0.296    2.13   0.033 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
               (Intr) blckps voicsm ustrnn cprfcn clgfrq blckpsttrnng:v
blckpsttrnn    -0.290
voicesame      -0.573  0.280
usetranngys    -0.417  0.207  0.118
cproficincy     0.089  0.004 -0.128  0.005
```

```
clogfreq       -0.039  0.002 -0.002  0.110  0.007
blckpsttrnng:v  0.218 -0.750 -0.376 -0.156 -0.002 -0.001
blckpsttrnng:s  0.157 -0.539 -0.152 -0.383 -0.001 -0.001  0.405
vcsm:strnng     0.169 -0.287 -0.291 -0.417 -0.005  0.001  0.384
blckpsttr::    -0.117  0.403  0.203  0.288  0.004  0.001 -0.537

                 blckpsttrnng:s vcsm:s
blckpsttrnn
voicesame
usetranngys
cproficincy
clogfreq
blckpsttrnng:v
blckpsttrnng:s
vcsm:strnng       0.530
blckpsttr::      -0.748         -0.690
```

2 OPTIMAL MODEL FITTED TO RESPONSE ACCURACY (PRE-TRAINING BLOCK)

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) [glmerMod]
 Family: binomial ( logit )
Formula: correct ~ vowel + cproficiency + clogfreq + (1 | subject)
   Data: dferrorpre

     AIC      BIC   logLik deviance df.resid
  2987.1   3016.1  -1488.5   2977.1     2475

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7481 -1.0223  0.4693  0.7494  1.5065

Random effects:
 Groups  Name        Variance Std.Dev.
 subject (Intercept) 0.2599   0.5098
Number of obs: 2480, groups: subject, 31

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.9243     0.1132   8.167 3.15e-16 ***
vowelKIT      -0.3573     0.0892  -4.006 6.18e-05 ***
cproficiency   0.7116     0.1493   4.767 1.87e-06 ***
clogfreq       0.4006     0.0643   6.230 4.66e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
```

```
            (Intr) vwlKIT cprfcn
vowelKIT    -0.416
cproficincy  0.051 -0.016
clogfreq     0.038 -0.009  0.025
```

3 OPTIMAL MODEL FITTED TO RESPONSE ACCURACY (POST-TRAINING BLOCK)

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: correct ~ voice + usetraining + vowel + cproficiency + clogfreq +
    cproficiency + censayos + (1 | subject) + (1 | stimuli)
   Data: dferrorpost

     AIC      BIC   logLik deviance df.resid
    2766     2819    -1374     2748     2470

Scaled residuals:
   Min     1Q Median     3Q    Max
-4.814 -0.857  0.419  0.637  1.928

Random effects:
 Groups  Name        Variance Std.Dev.
 stimuli (Intercept) 0.367    0.606
 subject (Intercept) 0.136    0.369
Number of obs: 2479, groups: stimuli, 40; subject, 31

Fixed effects:
               Estimate Std. Error z value    Pr(>|z|)
(Intercept)     1.06284    0.20440    5.20       0.000 ***
voicesame      -0.26473    0.16968   -1.56       0.119
usetrainingyes  0.54943    0.23698    2.32       0.020 *
vowelKIT       -0.19320    0.21442   -0.90       0.368
cproficiency    0.74827    0.13366    5.60       0.000 ***
clogfreq        0.31606    0.15476    2.04       0.041 *
trainblock     -0.00912    0.00461   -1.98       0.048 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) voicsm ustrnn vwlKIT cprfcn clgfrq
voicesame   -0.464
usetranngys -0.341 -0.003
vowelKIT    -0.533  0.001  0.005
cproficincy  0.128 -0.216  0.011 -0.004
clogfreq    -0.025 -0.002  0.122 -0.014  0.010
ctrainblock  0.073 -0.172 -0.004  0.002  0.414 -0.004
```

4 OPTIMAL MODEL FITTED FOR RESPONSE TIME (PRE- AND POST-TRAINING BLOCKS INCLUDED)

```
Linear mixed model fit by REML ['lmerMod']
Formula: log(ResponseTime) ~ block * vowel + clogfreq + (1 | subject)
   Data: dfcorrect

REML criterion at convergence: 1886.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2950 -0.6894 -0.0919  0.6154  3.6748

Random effects:
 Groups   Name        Variance Std.Dev.
 subject  (Intercept) 0.03480  0.1865
 Residual             0.09863  0.3140
Number of obs: 3347, groups: subject, 31

Fixed effects:
                          Estimate Std. Error t value
(Intercept)               7.496776   0.035152  213.27
blockpretraining          0.132037   0.015084    8.75
vowelKIT                  0.022350   0.015183    1.47
clogfreq                  0.006701   0.007859    0.85
blockpretraining:vowelKIT 0.028054   0.021769    1.29

Correlation of Fixed Effects:
            (Intr) blckpr vwlKIT clgfrq
blckprtrnng -0.212
vowelKIT    -0.210  0.491
clogfreq    -0.005  0.000  0.030
blckprt:KIT  0.147 -0.692 -0.697 -0.020
```

5 OPTIMAL MODEL FITTED FOR RESPONSE TIME (PRE-TRAINING BLOCK)

```
Linear mixed model fit by REML ['lmerMod']
Formula: log(ResponseTime) ~ +vowel + cproficiency + clogfreq
+ cAgeLearning + (1 | subject)
   Data: dfcorrectpre

REML criterion at convergence: 1073.1

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.8532 -0.6541 -0.0544  0.6141  3.4807

Random effects:
 Groups   Name        Variance Std.Dev.
```

```
 subject  (Intercept) 0.02976  0.1725
 Residual             0.10568  0.3251
Number of obs: 1627, groups: subject, 31

Fixed effects:
              Estimate Std. Error t value
(Intercept)   7.618865   0.033239  229.21
vowelKIT      0.047220   0.016189    2.92
cproficiency -0.014597   0.051649   -0.28
clogfreq      0.013928   0.011649    1.20
cAgeLearning  0.006039   0.003493    1.73

Correlation of Fixed Effects:
            (Intr) vwlKIT cprfcn clgfrq
vowelKIT    -0.229
cproficincy  0.104 -0.003
clogfreq    -0.002  0.001  0.008
cAgeLearnng -0.029 -0.002  0.450 -0.005
```

6 OPTIMAL MODEL FITTED FOR RESPONSE TIME (POST-TRAINING BLOCK)

```
Linear mixed model fit by REML ['lmerMod']
Formula: log(ResponseTime) ~ voice * usetraining + vowel
+ cAgeLearning +  (1 | subject)
   Data: dfcorrectpost

REML criterion at convergence: 762.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.7911 -0.6933 -0.0968  0.5929  3.7030

Random effects:
 Groups   Name        Variance Std.Dev.
 subject  (Intercept) 0.03498  0.187
 Residual             0.08467  0.291
Number of obs: 1720, groups: subject, 31

Fixed effects:
                          Estimate Std. Error t value
(Intercept)               7.542867   0.052011  145.02
voicesame                -0.081985   0.070317   -1.17
usetrainingyes           -0.073411   0.022262   -3.30
vowelKIT                  0.021690   0.014095    1.54
cAgeLearning              0.010009   0.003367    2.97
voicesame:usetrainingyes  0.075318   0.030101    2.50

Correlation of Fixed Effects:
            (Intr) voicsm ustrnn vwlKIT cAgLrn
```

```
voicesame   -0.731
usetranngys -0.133  0.099
vowelKIT    -0.128 -0.001 -0.015
cAgeLearnng  0.028 -0.132  0.005  0.000
vcsm:strnng  0.101 -0.139 -0.739 -0.011 -0.006
```

7 OPTIMAL MODEL FITTED TO AUC (PRE-TRAINING BLOCK)

```
Linear mixed model fit by REML ['lmerMod']
Formula: Aruncu ~ vowel + cproficiency + cAgeLearning
  + (1 | subject)
   Data: dfcorrectpre

REML criterion at convergence: 1570

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2370 -0.6357 -0.2028  0.4784  4.8730

Random effects:
 Groups   Name        Variance Std.Dev.
 subject  (Intercept) 0.01055  0.1027
 Residual             0.14707  0.3835
Number of obs: 1627, groups: subject, 31

Fixed effects:
              Estimate Std. Error t value
(Intercept)   0.195767   0.022811   8.582
vowelKIT      0.005142   0.019090   0.269
cproficiency -0.079542   0.033495  -2.375
cAgeLearning -0.007259   0.002293  -3.165

Correlation of Fixed Effects:
            (Intr) vwlKIT cprfcn
vowelKIT    -0.393
cproficincy  0.086 -0.006
cAgeLearnng -0.021 -0.003  0.445
```

8 OPTIMAL MODEL FITTED TO AUC (POST-TRAINING BLOCK)

```
Linear mixed model fit by REML ['lmerMod']
Formula: Aruncu ~ vowel * voice + usetraining + cproficiency
+ cAgeLearning + (1 | subject)
   Data: dfcorrectpost

REML criterion at convergence: 1681.5

Scaled residuals:
```

```
    Min      1Q  Median      3Q     Max
-3.4399 -0.6443 -0.2280  0.4722  3.9223

Random effects:
 Groups   Name        Variance Std.Dev.
 subject  (Intercept) 0.009474 0.09733
 Residual             0.148327 0.38513
Number of obs: 1720, groups: subject, 31

Fixed effects:
                    Estimate Std. Error t value
(Intercept)         0.164232   0.033550   4.895
vowelKIT            0.043848   0.027512   1.594
voicesame           0.092938   0.044987   2.066
usetrainingyes     -0.019036   0.019835  -0.960
cproficiency       -0.064350   0.032798  -1.962
cAgeLearning       -0.006150   0.002236  -2.751
vowelKIT:voicesame -0.076058   0.037401  -2.034

Correlation of Fixed Effects:
            (Intr) vwlKIT voicsm ustrnn cprfcn cAgLrn
vowelKIT    -0.396
voicesame   -0.733  0.299
usetranngys -0.181 -0.014 -0.006
cproficincy  0.198 -0.031 -0.196  0.003
cAgeLearnng  0.122 -0.020 -0.191  0.003  0.468
vwlKIT:vcsm  0.296 -0.735 -0.408 -0.012  0.025  0.018
```

9 OPTIMAL MODEL FITTED TO AUC (INCORRECT RESPONSES IN POST-TRAINING BLOCK)

```
Linear mixed model fit by REML ['lmerMod']
Formula: Aruncu ~ vowel * voice + cproficiency
+ cAgeLearning + (1 | subject)
   Data: subset(dferrorpost, correct == 0)

REML criterion at convergence: 846.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.7404 -0.6595 -0.2344  0.5290  4.1234

Random effects:
 Groups   Name        Variance Std.Dev.
 subject  (Intercept) 0.005309 0.07286
 Residual             0.168050 0.40994
Number of obs: 759, groups: subject, 31

Fixed effects:
```

```
                   Estimate Std. Error t value
(Intercept)         0.315794   0.040357   7.825
vowelKIT           -0.148836   0.046202  -3.221
voicesame          -0.058749   0.053958  -1.089
cproficiency        0.027592   0.036031   0.766
cAgeLearning       -0.003477   0.002196  -1.583
vowelKIT:voicesame  0.140690   0.060870   2.311

Correlation of Fixed Effects:
            (Intr) vwlKIT voicsm cprfcn cAgLrn
vowelKIT    -0.571
voicesame   -0.765  0.420
cproficincy  0.246  0.098 -0.234
cAgeLearnng  0.164  0.069 -0.254  0.521
vwlKIT:vcsm  0.431 -0.760 -0.562 -0.084 -0.056
```

LIST OF STIMULI

1 LIST OF STIMULI USED IN EXPERIMENT 1

Table 18 Stimuli used in Experiment 1.

Item	Word	Word type	Word pair	Contrast	Frequency (COCA, raw)	BNC List
17	back	rw	minpair	bilabial	145956	1
17	pack	rw	minpair	bilabial	1619	1
19	bill	rw	minpair	bilabial	49420	1
19	pill	rw	minpair	bilabial	1091	5
21	bowl	rw	minpair	bilabial	1859	2
21	pole	rw	minpair	bilabial	718	3
22	big	rw	minpair	bilabial	53245	1
22	pig	rw	minpair	bilabial	721	2
27	card	rw	minpair	velar	4959	1
27	guard	rw	minpair	velar	5019	2
28	curl	rw	minpair	velar	126	3
28	girl	rw	minpair	velar	12382	1
31	cap	rw	minpair	velar	1785	3
31	gap	rw	minpair	velar	2047	2
33	cold	rw	minpair	velar	6929	1
33	gold	rw	minpair	velar	3713	2
6	seek	rw	minpair	vowel	2345	2
6	sick	rw	minpair	vowel	5434	1
7	feel	rw	minpair	vowel	42489	1
7	fill	rw	minpair	vowel	2604	1
9	seat	rw	minpair	vowel	4618	1

Continued on Next Page...

Table 18 – Continued

Item	Word	Word type	Word pair	Contrast	Frequency (COCA, raw)	BNC List
9	sit	rw	minpair	vowel	10132	1
11	sheep	rw	minpair	vowel	651	2
11	ship	rw	minpair	vowel	3466	2
13	bork	pw	pseudopair	bilabial	NA	NA
13	pork	rw	pseudopair	bilabial	880	6
14	beace	pw	pseudopair	bilabial	NA	NA
14	piece	rw	pseudopair	bilabial	10416	1
15	book	rw	pseudopair	bilabial	31380	1
15	pook	pw	pseudopair	bilabial	NA	NA
16	bake	rw	pseudopair	bilabial	388	2
16	pake	pw	pseudopair	bilabial	NA	NA
18	bast	pw	pseudopair	bilabial	NA	NA
18	past	rw	pseudopair	bilabial	25113	1
20	boat	rw	pseudopair	bilabial	2997	1
20	pote	pw	pseudopair	bilabial	NA	NA
23	beast	rw	pseudopair	bilabial	408	4
23	peast	pw	pseudopair	bilabial	NA	0
25	court	rw	pseudopair	velar	33738	1
25	gort	pw	pseudopair	velar	NA	NA
26	cost	rw	pseudopair	velar	11281	1
26	gost	pw	pseudopair	velar	NA	NA
29	ginde	pw	pseudopair	velar	NA	NA
29	kind	rw	pseudopair	velar	79709	1
30	gift	rw	pseudopair	velar	2952	2
30	kift	pw	pseudopair	velar	NA	NA
32	get	rw	pseudopair	velar	198039	1
32	ket	pw	pseudopair	velar	NA	NA

Continued on Next Page...

Table 18 – Continued

Item	Word	Word type	Word pair	Contrast	Frequency (COCA, raw)	BNC List
34	gas	rw	pseudopair	velar	6268	1
34	kass	pw	pseudopair	velar	NA	NA
1	mees	pw	pseudopair	vowel	NA	NA
1	miss	rw	pseudopair	vowel	6711	1
2	skeel	pw	pseudopair	vowel	NA	NA
2	skill	rw	pseudopair	vowel	777	2
3	feesh	pw	pseudopair	vowel	NA	NA
3	fish	rw	pseudopair	vowel	3477	1
4	brick	rw	pseudopair	vowel	478	2
4	breek	pw	pseudopair	vowel	NA	NA
5	bleed	rw	pseudopair	vowel	223	2
5	blid	pw	pseudopair	vowel	NA	NA
8	feed	pw	pseudopair	vowel	2019	1
8	fid	pw	pseudopair	vowel	NA	NA
10	greek	rw	pseudopair	vowel	816	3
10	grick	pw	pseudopair	vowel	NA	NA
12	please	rw	pseudopair	vowel	11769	1
12	plizz	rw	pseudopair	vowel	NA	NA
24	part	rw	reppair	control	51743	1
24	part	rw	reppair	control	51743	1
35	act	rw	reppair	control	12399	1
35	act	rw	reppair	control	12399	1
37	church	rw	reppair	control	10565	1
37	church	rw	reppair	control	10565	1
38	firm	rw	reppair	control	3528	2
38	firm	rw	reppair	control	3528	2
39	force	rw	reppair	control	17228	1

Continued on Next Page…

Table 18 – Continued

Item	Word	Word type	Word pair	Contrast	Frequency (COCA, raw)	BNC List
39	force	rw	reppair	control	17228	1
40	hair	rw	reppair	control	6509	1
40	hair	rw	reppair	control	6509	1
41	hate	rw	reppair	control	5674	1
41	hate	rw	reppair	control	5674	1
42	help	rw	reppair	control	41674	1
42	help	rw	reppair	control	41674	1
43	home	rw	reppair	control	48612	1
43	home	rw	reppair	control	48612	1
44	job	rw	reppair	control	32558	1
44	job	rw	reppair	control	32558	1
45	land	rw	reppair	control	7890	1
45	land	rw	reppair	control	7890	1
46	life	rw	reppair	control	64386	1
46	life	rw	reppair	control	64386	1
47	light	rw	reppair	control	8449	1
47	light	rw	reppair	control	8449	1
48	look	rw	reppair	control	91595	1
48	look	rw	reppair	control	91595	1
49	move	rw	reppair	control	20077	1
49	move	rw	reppair	control	20077	1
50	north	rw	reppair	control	15958	1
50	north	rw	reppair	control	15958	1
51	race	rw	reppair	control	14124	2
51	race	rw	reppair	control	14124	2
52	rock	rw	reppair	control	6233	2
52	rock	rw	reppair	control	6233	2

Continued on Next Page. . .

Table 18 – Continued

Item	Word	Word type	Word pair	Contrast	Frequency (COCA, raw)	BNC List
53	sad	rw	reppair	control	4190	2
53	sad	rw	reppair	control	4190	2
54	save	rw	reppair	control	7961	1
54	save	rw	reppair	control	7961	1

2 LIST OF STIMULI USED IN EXPERIMENT 2

Table 19 Stimuli used in Experiment 2.

Word type	Height	Stimuli	Frequency COCA (raw)
pw	close	beeth	NA
pw	close	bith	NA
pw	close	breeb	NA
pw	close	breesh	NA
pw	close	brib	NA
pw	close	brish	NA
pw	close	cleel	NA
pw	close	creent	NA
pw	close	crint	NA
pw	close	dreek	NA
pw	close	drick	NA
pw	close	feek	NA
pw	close	feesp	NA
pw	close	fick	NA
pw	close	fisp	NA
pw	close	flib	NA
pw	close	geech	NA
pw	close	gid	NA

Continued on Next Page. . .

Table 19 – Continued

Word type	Height	Stimuli	Frequency COCA (raw)
pw	close	glick	NA
pw	close	keeg	NA
pw	close	kig	NA
pw	close	kleesh	NA
pw	close	kwig	NA
pw	close	leesk	NA
pw	close	lisk	NA
pw	close	meeve	NA
pw	close	mip	NA
pw	close	miv	NA
pw	close	neef	NA
pw	close	neent	NA
pw	close	niff	NA
pw	close	nint	NA
pw	close	preef	NA
pw	close	preezz	NA
pw	close	priff	NA
pw	close	prizz	NA
pw	close	seef	NA
pw	close	shing	NA
pw	close	shreen	NA
pw	close	shrin	NA
pw	close	siff	NA
pw	close	skilch	NA
pw	close	smeesh	NA
pw	close	sneeg	NA
pw	close	sneeth	NA
pw	close	snith	NA
pw	close	spim	NA

Continued on Next Page…

Table 19 – Continued

Word type	Height	Stimuli	Frequency COCA (raw)
pw	close	spreel	NA
pw	close	spreem	NA
pw	close	spril	NA
pw	close	stit	NA
pw	close	thrick	NA
pw	close	treence	NA
pw	close	tweep	NA
pw	close	weeb	NA
pw	close	weent	NA
pw	close	zeech	NA
pw	close	zees	NA
pw	close	zich	NA
pw	close	ziss	NA
pw	open	balf	NA
pw	open	basp	NA
pw	open	belf	NA
pw	open	blam	NA
pw	open	blan	NA
pw	open	blem	NA
pw	open	bramp	NA
pw	open	cheb	NA
pw	open	dwen	NA
pw	open	fask	NA
pw	open	fep	NA
pw	open	fesk	NA
pw	open	gak	NA
pw	open	geth	NA
pw	open	gretch	NA
pw	open	jelth	NA

Continued on Next Page…

Table 19 – Continued

Word type	Height	Stimuli	Frequency COCA (raw)
pw	open	maff	NA
pw	open	masp	NA
pw	open	mesp	NA
pw	open	nank	NA
pw	open	nedge	NA
pw	open	nelk	NA
pw	open	nenk	NA
pw	open	paff	NA
pw	open	pamp	NA
pw	open	peff	NA
pw	open	pemp	NA
pw	open	plaff	NA
pw	open	plang	NA
pw	open	prab	NA
pw	open	prash	NA
pw	open	preb	NA
pw	open	presh	NA
pw	open	rance	NA
pw	open	rass	NA
pw	open	rence	NA
pw	open	ress	NA
pw	open	sast	NA
pw	open	scass	NA
pw	open	shalp	NA
pw	open	shang	NA
pw	open	shelp	NA
pw	open	sheng	NA
pw	open	smad	NA
pw	open	smed	NA

Continued on Next Page…

Table 19 – Continued

Word type	Height	Stimuli	Frequency COCA (raw)
pw	open	snaff	NA
pw	open	snam	NA
pw	open	sneff	NA
pw	open	snem	NA
pw	open	snet	NA
pw	open	spatch	NA
pw	open	spren	NA
pw	open	strab	NA
pw	open	streb	NA
pw	open	tamb	NA
pw	open	temb	NA
pw	open	thrap	NA
pw	open	threp	NA
rw	close	beast	453
rw	close	beat	6181
rw	close	big	57948
rw	close	bit	34804
rw	close	breeze	258
rw	close	brick	512
rw	close	cheek	370
rw	close	chick	254
rw	close	chief	14443
rw	close	clean	5590
rw	close	cream	1875
rw	close	deep	6276
rw	close	dip	455
rw	close	feel	45063
rw	close	feet	7870
rw	close	fill	2788

Continued on Next Page…

Table 19 – Continued

Word type	Height	Stimuli	Frequency COCA (raw)
rw	close	fish	3642
rw	close	fit	4225
rw	close	gym	927
rw	close	kiss	1644
rw	close	lead	11296
rw	close	leak	1267
rw	close	leap	776
rw	close	leave	18967
rw	close	lick	199
rw	close	lid	314
rw	close	lip	557
rw	close	live	35406
rw	close	meal	1462
rw	close	mill	494
rw	close	niece	355
rw	close	peach	181
rw	close	peek	345
rw	close	pick	9840
rw	close	pig	751
rw	close	pitch	1245
rw	close	please	12568
rw	close	queen	2916
rw	close	reach	6366
rw	close	read	20731
rw	close	rich	6936
rw	close	rid	3991
rw	close	ridge	1040
rw	close	ring	2981
rw	close	screen	4741

Continued on Next Page…

Table 19 – Continued

Word type	Height	Stimuli	Frequency COCA (raw)
rw	close	seat	4948
rw	close	seek	2461
rw	close	sheep	661
rw	close	ship	3619
rw	close	sick	5730
rw	close	sit	10723
rw	close	sleep	5560
rw	close	slip	994
rw	close	steal	1339
rw	close	stick	3991
rw	close	still	78234
rw	close	thief	526
rw	open	bad	28072
rw	open	band	4155
rw	open	bed	5409
rw	open	bell	2261
rw	open	bend	666
rw	open	black	22542
rw	open	breath	1824
rw	open	cap	1873
rw	open	cat	2018
rw	open	cattle	865
rw	open	check	10157
rw	open	chess	448
rw	open	dad	8441
rw	open	dead	12672
rw	open	deaf	615
rw	open	expand	1592
rw	open	expend	71

Continued on Next Page…

Table 19 – Continued

Word type	Height	Stimuli	Frequency COCA (raw)
rw	open	fetch	90
rw	open	flag	2513
rw	open	flash	957
rw	open	flat	2218
rw	open	flesh	573
rw	open	gas	6982
rw	open	gem	89
rw	open	grass	1024
rw	open	guess	26630
rw	open	jam	470
rw	open	jazz	2303
rw	open	kettle	68
rw	open	lack	4321
rw	open	land	8228
rw	open	lend	535
rw	open	mass	5337
rw	open	mat	233
rw	open	match	2772
rw	open	mention	4061
rw	open	mess	2588
rw	open	met	13012
rw	open	pack	1710
rw	open	paddle	109
rw	open	pan	1443
rw	open	pat	7952
rw	open	peck	332
rw	open	pedal	158
rw	open	pen	771
rw	open	pet	1310

Continued on Next Page...

Table 19 – Continued

Word type	Height	Stimuli	Frequency COCA (raw)
rw	open	press	17373
rw	open	rack	250
rw	open	rat	775
rw	open	sad	4504
rw	open	said	170736
rw	open	sand	1364
rw	open	sell	7485
rw	open	send	11240
rw	open	tally	271
rw	open	telly	20
rw	open	web	7625
rw	open	west	13040
rw	open	wreck	609

3 LIST OF STIMULI USED IN EXPERIMENT 3

Table 20 Stimuli used in Experiment 3.

Pair	Word	Vowel	Translation	Training	Frequency (logarithmic)
10	leave	FLEECE	irse	no	4.28
10	live	KIT	vivir	no	4.55
11	meal	FLEECE	comida	no	3.16
11	mill	KIT	molino	no	2.69
12	peach	FLEECE	durazno	no	2.26
12	pitch	KIT	lanzar	no	3.10
13	peek	FLEECE	asomarse	no	2.54
13	pick	KIT	escoger	no	3.99
14	reach	FLEECE	alcanzar	no	3.80
14	rich	KIT	rico	no	3.84

Continued on Next Page…

Table 20 – Continued

Pair	Word	Vowel	Translation	Training	Frequency (logarithmic)
15	read	FLEECE	leer	no	4.32
15	rid	KIT	eliminar	no	3.60
16	seat	FLEECE	asiento	no	3.69
16	sit	KIT	sentarse	no	4.03
20	steal	FLEECE	robar	no	3.13
20	still	KIT	todavia	no	4.89
4	feel	FLEECE	sentir	no	4.65
4	fill	KIT	llenar	no	3.45
5	feet	FLEECE	pies	no	3.90
5	fit	KIT	caber	no	3.63
6	heat	FLEECE	calor	no	3.56
6	hit	KIT	golpear	no	4.26
7	lead	FLEECE	liderar	no	4.05
7	lid	KIT	tapa	no	2.50
8	leak	FLEECE	gotear	no	3.10
8	lick	KIT	lamer	no	2.30
9	leap	FLEECE	saltar	no	2.89
9	lip	KIT	labio	no	2.75
1	beat	FLEECE	golpe	yes	3.79
1	bit	KIT	pedazo	yes	4.54
17	seek	FLEECE	buscar	yes	3.39
17	sick	KIT	enfermo	yes	3.76
18	sheep	FLEECE	oveja	yes	2.82
18	ship	KIT	barco	yes	3.56
19	sleep	FLEECE	dormir	yes	3.75
19	slip	KIT	resbalar	yes	3.00
2	cheek	FLEECE	mejilla	yes	2.57
2	chick	KIT	pollito	yes	2.40
3	deep	FLEECE	profundo	yes	3.80

Continued on Next Page. . .

Table 20 – Continued

Pair	Word	Vowel	Translation	Training	Frequency (logarithmic)
3	dip	KIT	salsa	yes	2.66

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

MEMORANDUM

Phone 0-4-463 5676
Fax 0-4-463 5209
Email Allison.kirkman@vuw.ac.nz

TO	Erwin La Cruz
COPY TO	Paul warren
FROM	Dr Allison Kirkman, Convener, Human Ethics Committee
DATE	20 October 2010
PAGES	1
SUBJECT	**Ethics Approval: No 17979 Listening in a second language**

Thank you for your applications for ethical approval, which have now been considered by the Standing Committee of the Human Ethics Committee.

Your applications have been approved from the above date and this approval continues until 02 August 2012. If your data collection is not completed by this date you should apply to the Human Ethics Committee for an extension to this approval.

Best wishes with the research.

Allison Kirkman
Convener

PYTHON SCRIPTS USED IN OPEN SESAME

1 SCRIPT FOR THE GAMEPAD PLUGIN USED IN EXPERIMENT 2

```
'''
Created on 10/12/2011

This file is part of OpenSesame.

OpenSesame is free software: you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation, either version 3 of the License, or
(at your option) any later version.

OpenSesame is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
GNU General Public License for more details.

You should have received a copy of the GNU General Public License
along with OpenSesame.  If not, see <http://www.gnu.org/licenses/>.

'''

import libqtopensesame
import pygame
from libopensesame import item, generic_response, exceptions
from libqtopensesame import qtplugin
import openexp.keyboard
import imp
import os
import os.path
from PyQt4 import QtGui, QtCore


class joystick_sesame (item.item, generic_response.generic_response):
    def __init__(self, name, experiment, string=None):
        '''
        Joystick class constructor
        Args: name, experiment, string
        '''
        self.flush = "yes"
        self.item_type = "joystick_sesame"
```

```
        self.description = "Collects joystick responses"
        self.timeout = "infinite"
        self.dummy = "no"
        self.response = None
        self.end_response_interval = None
        self.allowed_responses = "1,2"

        item.item.__init__(self, name, experiment, string)

    def prepare(self):
        '''
        Preparing the joystick
        '''

        # Pass the word on to the parent
        item.item.prepare(self)
        generic_response.generic_response.var_info(self)
        self._flush = self.get("flush") == "yes"

        # Prepares the timeout
        if self.has("timeout"):
            self.prepare_timeout()


        # Prepares the allowed responses
        if self.has("allowed_responses"):
            self._allowed_responses = [ ]
            for r in str(self.get("allowed_responses")).split(";"):
                if r.strip() != "":
                    try:
                        r = int(r)
                    except:
                        raise exceptions.runtime_error("'%s' is
                        not a valid response in joystick'%s'.
                        Expecting a number in the range 0 .. 5."
                        % (r, self.name))
                    if r < 0 or r > 255:
                        raise exceptions.runtime_error("'%s' is
                        not a valid response in joystick '%s'.
                        Expecting a number in the range 0 .. 5." % (r, self.name))
                    self._allowed_responses.append(r)
            if len(self._allowed_responses) == 0:
                self._allowed_responses = None
        else:
            self._allowed_responses = None

        #Prepares the joystick. Raises an error if no
        #joystick is detected.
        pygame.joystick.init()
```

```
    if(pygame.joystick.get_count() == 0):
        raise AttributeError
    self.__joystick = pygame.joystick.Joystick(0)
    self.__joystick.init()

    # Report success
    return True



def run(self):
    '''
    Handles joystick
    '''

    # Set the onset time
    # Mark the onset of the item (as 'time_[my_item]')
    # self.set_item_onset()

    # Flush responses, to make sure that earlier responses
    # are not carried over
    # if self._flush:
        # self._joystick_sesame.flush()

    # self.set_sri()


    time = pygame.time.get_ticks()
    go = True
    RT = None
    boton = None
    # c = self.experiment.time()
    # t = c
    # while timeout == None or t - c < timeout:
    while go:
        if pygame.time.get_ticks() - time >= self.timeout:
            go = False
        for event in pygame.event.get():
            if event.type == pygame.JOYBUTTONDOWN:
            #or event.type == pygame.KEYDOWN:
                RT = pygame.time.get_ticks () - time
                boton = event.button+1
                print boton
                print RT
                go = False


            elif event.type == pygame.KEYDOWN:
                if event.key == pygame.K_ESCAPE:
```

```
                        raise exceptions.runtime_error("The escape
                        key was pressed.")
                        go = False

        self.experiment.response_time = RT
        self.experiment.response = boton

        generic_response.generic_response.response_bookkeeping(self)
         self.experiment.set("response_time",
         self.experiment.end_response_interval
         - self.experiment.start_response_interval)
         return True
    # def var_info(self):
         return generic_response.generic_response.var_info(self)


class qtjoystick_sesame(joystick_sesame, libqtopensesame.qtplugin.qtplugin):
    def __init__(self, name, experiment, string=None):
        '''
        Constructor
        '''
        joystick_sesame.__init__(self, name, experiment, string)
        libqtopensesame.qtplugin.qtplugin.__init__(self, __file__)

    def init_edit_widget(self):
        '''
        Creates widget
        '''
        self.lock = True
        qtplugin.qtplugin.init_edit_widget(self, False)



        self.add_line_edit_control("timeout",
        "Timeout", tooltip = "Expecting a value in milliseconds
        or 'infinite'", default = "infinite")
        self.add_line_edit_control("correct_response",
        "Correct response", tooltip = "Expecting a
        button number (1 .. 8)")
        self.add_line_edit_control("allowed_responses",
        "Allowed responses", tooltip = "Expecting a
        semicolon-separated list of button numbers, e.g., 1;3;4")
        self.add_combobox_control("dummy",
        "Dummy mode (use keyboard instead)",
        ["yes", "no"], tooltip = "Enable dummy mode
        for testing purposes")

        self.edit_vbox.addStretch()
        self.lock = False
```

```
    def apply_edit_changes(self):
        """
        Set the variables based on the controls
        """

        # Abort if the parent reports failure of if the
        # controls are locked
        if not qtplugin.qtplugin.apply_edit_changes(self, False)
         or self.lock:
            return False

        # Refresh the main window, so that changes
        # become visible everywhere
        self.experiment.main_window.refresh(self.name)
        # Report success
        return True

    def edit_widget(self):
        self.lock = True
        qtplugin.qtplugin.edit_widget(self)
        self.lock = False
        return self._edit_widget
```

2 SCRIPT USED TO CAPTURE MOUSE TRACKING DATA DURING EXPERIMENT 3

```
import pygame
from pygame.locals import *
pygame.init()
import pygame

## config
pygame.mouse.set_visible(True) # Make the mouse cursor visible
pygame.mouse.set_pos([510, 670]) # Mouse position at the beginning of each trial:
pos1=(190, 95) # Screen upper-left corner. Determines response 1
pos2=(830, 95) # Screen upper-right corner. Determines response 2
timeout = 6000
launch_warning = 1000

# Tracking mouse coordinates over time:
track_x=[] # x axis
track_y=[] # y axis
track_time=[] # time_stamp

#Time cotrols:
start_time = self.time()
trial_duration = start_time + timeout
RT = 0
```

```
response = 0
t=0

from openexp.sampler import sampler
audio = self.get("audio")
my_sampler = sampler(self.experiment, self.experiment.get_file(audio))
#my_sampler.pitch(0.94)
my_sampler.play()

while t < trial_duration and response == 0:
    t = self.time()
    time_stamp = t - start_time
    for event in pygame.event.get():
        x, y = list(event.pos)
        track_x.append(x)
        track_y.append(y)
        track_time.append(time_stamp)
        if event.type == MOUSEBUTTONDOWN:
            if  x < 190 and y < 95:
                response = 1
            elif x > 830 and y <95:
                response = 2
        if y < 660 and RT == 0:
            RT = time_stamp

# Prepare feedback controls:
tiempo_resp =  max(track_time)

# Set warning for late response startings
warning = "no"
if RT > launch_warning and RT !=0:
    warning = "yes"


# Determine if the response was correct
self.experiment.response = response
if not self.has('correct_response'):
    self.experiment.correct = 'undefined'
elif self.experiment.response == self.get('correct_response'):
    self.experiment.correct = 1
else:
    self.experiment.correct = 0

# Do response bookkeeping
# Set response time:
self.experiment.response_time =  max(track_time)

self.experiment.total_responses += 1
if self.experiment.correct == 1:
```

```
    self.experiment.total_correct += 1

self.experiment.acc = 100. * self.experiment.total_correct
    / self.experiment.total_responses
self.experiment.total_response_time
    += self.experiment.response_time
self.experiment.avg_rt = self.experiment.total_response_time
    / self.experiment.total_responses
self.experiment.accuracy = self.experiment.acc
self.experiment.average_response_time = self.experiment.avg_rt

# Transform list to string so that Open Sesame
# can read the variables:
str_time = str(track_time).strip('[]')
str_x = str(track_x).strip('[]')
str_y = str(track_y).strip('[]')

# Record the script variables in logger:
self.experiment.set("track_time", str_time)
self.experiment.set("track_x", str_x)
self.experiment.set("track_y", str_y)
self.experiment.set("RT", RT)
self.experiment.set("warning", warning)
self.experiment.set("tiempo_resp", tiempo_resp)
```

REFERENCES

Abramson, A. S., & Lisker, L. (1972). Voice-timing perception in Spanish word-initial stops. *Haskins Laboratories Status Report on Speech Research*, *29*(30), 15–25.

Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, *137*(2), 208–216. doi:10.1016/j.actpsy.2011.01.007

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, *39*(4), 558–570. http://doi.org/10.1016/j.wocn.2011.03.001

André, C., Ghio, A., Cavé, C., & Teston, B. (2003). PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. *Proceedings of XVTH ICPHS, Barcelone, Spain*, 1421–1424.

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barca, L., & Pezzulo, G. (2012). Unfolding Visual Lexical Decision in Time. *PLoS ONE*, *7*(4), e35932. doi:10.1371/journal.pone.0035932

Barcroft, J., & Sommers, M. S. (2005). Effects of Acoustic Variability on Second Language Vocabulary Learning. *Studies in Second Language Acquisition*, *27*(03), 387–414. doi:10.1017/S0272263105050175

Barreda, S., & Nearey, T. M. (2012). The direct and indirect roles of fundamental frequency in vowel perception. *The Journal of the Acoustical Society of America*, *131*(1), 466–477. doi:10.1121/1.3662068

Bates, D., & Bolker, M. M. and B. (2013). lme4: Linear mixed-effects models using S4 classes (Version 0.999999-2). Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Bates, R. A., Ostendorf, M., & Wright, R. A. (2007). Symbolic phonetic features for modeling of pronunciation variation. *Speech Communication*, *49*(2), 83–97.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*, *69*(1), 231–238. doi:10.1121/1.385343

Best, C. T. (1984). Discovering messages in the medium: Speech perception and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Theory and research in behavioral pediatrics.* New York, NY: Plenum Press.

Best, C. T., & Hallé, P. A. (2010). Perception of initial obstruent voicing is influenced by gestural organization. *Journal of Phonetics*, *38*(1), 109–126.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. doi:10.1121/1.1332378

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In M. J. Munro & O. Bohn (Eds.), *Second language speech learning: the role of language experience in speech perception and producction* (pp. 13–34). Amsterdam: John Benjamins Publishing Company.

Blevins, J. P. (2003) Stems and paradigms. *Language*, *79*, 737–767.

Boersma, P. (1998). *Functional phonology*. Holland Academic Graphics.

Boersma, P., & Weenink, D. (2009). Praat: dong phonetics by computer (Version 5.3.39). Retrieved from http://www.praat.org/

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

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

Browman, C., & Goldstein, L. (1995) Dynamics and articulatory phonology. In: R. Port & T. van Gelder (Eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*, 175–93. Cambridge, MA: MIT Pres

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. doi:10.1111/j.2044-8295.1986.tb02199.x

Brunellière, A., Dufour, S., & Nguyen, N. (2011). Regional differences in the listener's phonemic inventory affect semantic processing: A mismatch negativity (MMN) study. *Brain and Language*, *117*(1), 45–51. doi:10.1016/j.bandl.2010.12.004

Bybee, J. L. (2001). *Phonology and language use*. Cambridge, UK: Cambridge University Press

Centeno-Cortés, B. (2003). *Private speech in the second language classroom: Its role in internalization and its link to social production*. Pennsylvania State University, University Park.

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, *70*(4), 604–618.

Coleman, R. O. (1971). Male and Female Voice Quality and Its Relationship to Vowel Formant Frequencies. *Journal of Speech and Hearing Research*, *14*(3), 565–577.

Colomé, À., & Miozzo, M. (2010). Which Words Are Activated During Bilingual Word Production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 96–109. doi:10.1037/a0017677

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1057–1074. doi:10.1037/0278-7393.32.5.1057

Creel, S. C., & Dahan, D. (2010). The Effect of the Temporal Structure of Spoken Words on Paired-Associate Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 110–122. doi:10.1037/a0017527

Cruttenden, A. (2001). *Gimson's pronunciation of English* (6th ed., Arnold International student ed.). London: Arnold.

Cutler, A. (2008). The abstract representations in speech processing. *The Quarterly Journal of Experimental Psychology*, *61*(11), 1601–1619. doi:10.1080/13803390802218542

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Cutler, A., & Weber, A. (2007). Listening experience and phonetic-to-lexical mapping in L2. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 43–48).

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory Phonology*, *10*, 91–111.

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*(3), 372–420. doi:10.1075/ml.8.3.06dar

Davies, M. (2008). The Corpus of Contemporary American English: 450 million words, 1990-present. Retrieved from http://corpus.byu.edu/coca/

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241.

Decoene, S. (1993). Testing the speech unit hypothesis with the primed matching task: Phoneme categories are perceptually basic. *Perception & Psychophysics*, *53*, 601–616

Dell, G. S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, *93* (3), 283–321. doi:10.1037/0033-295X.93.3.283

Dufour, S., & Nguyen, N. (2008). *L'influence de la langue maternelle sur les capacités de l'auditeur dans la perception de la parole.* Retrieved August 16, 2011, from http://hal.archives-ouvertes.fr/index.php?halsid=itk2oli 4c0hf8cc9su8dnkerd3 &view_this_doc=hal-00454447&version=1

Dufour, S., & Nguyen, N. (2010, May). *Un changement de voix affecte t-il le processus de reconnaissance des mots parlés?* Retrieved February 27, 2011, from http://hal.archives-ouvertes.fr/hal-00541977

Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2007). The perception of phonemic contrasts in a non-native dialect. *The Journal of the Acoustical Society of America*, *121*(4), EL131–EL136. doi:10.1121/1.2710742

Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2010). Does training on a phonemic contrast absent in the listener's dialect influence word recognition? *The Journal of the Acoustical Society of America*, *128*(1), EL43. doi:10.1121/1.3431102

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008).Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, *106*(2), 682–706. doi:10.1016/j.cognition.2007.04.001

Dupoux, E., Pallier, C., Kakehi, K., & Jacques Mehler, J. (2001) New Evidence for Prelexical Phonological Processing in Word Recognition. *Language and Cognitive Processes*, *5*(16):491–505, 2001

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America,119*, 1950–1953.

Ellis, R., & Humphreys, G. W. (1999). Connectionist Psychology: A Text with Readings. *Psychology Press.*

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121–138. doi:10.1037/1082-989X.12.2.121

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition. Explaining the attainment of optimal phonological categorization*. Utrecht University.

Escudero, P., & Boersma, P. (2004). Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition*, *26*(04), 551–585. doi:10.1017/S0272263104040021

Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America*, *128*(5), EL254–EL260. doi:10.1121/1.3488794

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, *40*(2), 280–288. doi:10.1016/j.wocn.2011.11.004

Eulitz, C. & Lahiri, A. (2004). Neurobiological Evidence for Abstract Phonological Representations in the Mental Lexicon during Speech Recognition. *Journal of Cognitive Neuroscience*, *16*(4), 577–583

Field, J. (2009). *Listening in the Language Classroom*. Cambridge University Press.

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: orthographic evidence. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *43*(3), 701–31. doi:10.1080/14640749108400993

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research.* (pp. 233–277). Timonium, MD: York Press.

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

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of english. *Speech Communication*, *6*(3), 185–202. doi:10.1016/0167-6393(87)90025-2

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134. doi:10.1121/1.413041

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, *2*(2), 337–339. doi:10.3758/BF03209005

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*(1), 3–28.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. doi:10.3758/BRM.42.1.226

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in Motion Reveals Mind in Motion. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00059

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448–458.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183. doi:10.1037/0278-7393.22.5.1166

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. doi:10.1037/0033-295X.105.2.251

Goldinger, S. D. (2007). A complimentary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences*, 49–54.

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics*, *31*(3-4), 305–320. doi:10.1016/S0095-4470(03)00030-5

Goldstein, L., & Fowler, C. (2003). Articulatory phonology: a phonology for public language use. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 159–207). Berlin: Mouton de Gruyter.

González, J., & McLennan, C. T. (2007). Hemispheric differences in indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 410–24.

Griffen, T. D. (1976). Toward a nonsegmental phonology. *Lingua*, *40*(1), 1–20. doi:10.1016/0024-3841(76)90029-2

Hanique, I., Aalders, E., & Ernestus, M. (2013). How robust are exemplar effects in word comprehension? *The Mental Lexicon*, *8*(3), 269–294. doi:10.1075/ml.8.3.01han

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. *Frontiers in Psychology*, *4*, 563. doi:10.3389/fpsyg.2013.00563

Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. *EUROSLA Yearbook*, *6*, 147–168.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*(3-4), 373–405. doi:10.1016/j.wocn.2003.09.006

Hawkins, S. (2010). Phonetic variation as communicative system: Per ception of the particular and the abstract. In C. Fougeron, B. Kuehnert, M. Imperio, & N. Vallee (Eds.), *Laboratory Phonology 10*. Walter de Gruyter.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. doi:10.1016/j.wocn.2005.10.001

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

Hazan, V., & Simpson, A. (2000). The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. *Language and Speech*, *43*(Pt 3), 273–294.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and wome. *Attention, Perception, & Psychophysics*, *71*(5), 1150–1166.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America*, *109*(2), 748–763. doi:10.1121/1.1337959

Ieno, E. N., Walker, N., Zuur, A. F., Savelieve, A., & Smith, G. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York, NY: Springer.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877.
doi:10.1121/1.3148196

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278.
doi:10.1121/1.2062307

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(01), 145–160.
doi:10.1017/S0142716411000300

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.
doi:10.1016/j.wocn.2005.08.004

Johnstone, D. E. (1999). *Cool Edit Pro V. 1.2.* Retrieved from
http://www.syntrillium.com

Jong, K. J. de, Silbert, N. H., & Park, H. (2009). Generalization across segments in cecond language consonant identification. *Language Learning*, *59*(1), 1–31. doi:10.1111/j.1467-9922.2009.00499.x

Ju, M., & Luce, P. A. (2006). Representational Specificity of Within-Category Phonetic Variation in the Long-Term Mental Lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 120–138. doi:10.1037/0096-1523.32.1.120

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*, 1-–15.

King, W. (2014, February 2). COUNTS AND PROPORTIONS. Retrieved from http://ww2.coastal.edu/kingw/statistics/R-tutorials/proport.html

Kirsner, K., & Smith, M. C. (1974). Modality effects in word identification. *Memory & Cognition*, *2*(4), 637–640. doi:10.3758/BF03198132

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221. doi:10.1121/1.380986

Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics*, *7*, 279–312

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: comparison of three training methods. *Journal of Phonetics*, *38*(4), 569–587. doi:10.1016/j.wocn.2010.08.003

Konopka, K. (2011, July). *The Vowels of Mexican Heritage English in a Chicago Community*. Northwestern University. Retrieved from http://gradworks.umi.com/34/56/3456676.html

Kuhl, P. K., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Pholosophical Transactions of the Royal Society*, (363), 979–1000. doi:10.1098/rstb.2007.2154

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 9096–9101. doi:10.1073/pnas.1532872100

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A Theory of Lexical Access in Speech Production. *Behavioral and Brain Sciences*, *22*(01), 1–38. doi:10.1017/ S0140525X99001776

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *The Journal of the Acoustical Society of America*, *130*(6), 4053–4062. doi:10.1121/1.3651816

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*,*21*(1), 1–36. doi:10.1016/0010-0277(85)90021-6

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076–2087. doi:10.1121/1.410149

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. doi:10.1121/1.1894649

López, V. G. (2012). Spanish and English word-initial voiceless stop production in code-switched vs. monolingual structures. *Second Language Research*, *28*(2), 243–263. http://doi.org/10.1177/0267658312439821

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, *79*, 124–145.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. doi:10.3758/s13428-011-0168-7

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words,phonemes, and features. *Psychological Review*, *101*, 653–675.

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

McLennan, C. T., & Luce, P. A. (2005). Examining the Time Course of Indexical Specificity Effects in Spoken Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306–321. doi:10.1037/0278-7393.31.2.306

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*, 1113–112

Mehler, J., Segui, J., & Frauenfelder, U. (1981). The role of the syllable in language acquisition and perception. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 295–305). Amsterdam: North-Holland.

Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*(4), 527–545. doi:10.1016/j.jml.2013.07.002

Morais, J. (2010). Flows of information in and between systems of lexical access. In L. Scliar-Cabral (Ed.), *Psycholinguistics: scientific and technological challenges*. Porto Alegre, Brasil: EDIPUCRS.

Morrison, G. S. (2008). L1-Spanish Speakers' Acquisition of the English /i/-/I/ Contrast: Duration-based Perception is Not the Initial Developmental Stage. *Language and Speech*, *51*, 285.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, *60*(1), 65–91. http://doi.org/10.1016/j.jml.2008.07.002

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378.

Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved February 6, 2014, from http://www.victoria.ac.nz/lals/about/ staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf

Nearey, T. M. (2001). Phoneme-like units and speech perception. *Language and Cognitive Processes*, *16*(5-6), 673–681. http://doi.org/10.1080/01690960143000173

Nguyen, N., Dufour, S., & Brunellière, A. (2012). Does imitation facilitate word recognition in a non-native regional accent? *Frontiers in Cognition*, *3*, 480.
doi:10.3389/fpsyg.2012.00480

Nguyen, N., Wauquier, S., & Tuller, B. (2009). The dynamical approach to speech perception: from fine phonetic detail to abstract phonological categories. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to Phonological Complexity* (pp. 193–217). Berlin: Mouton de Gruyter.

Norris, D., & Cutler, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, *43*, 541–550.

Norris, D. (1999). The merge model: Speech perception is bottom-up. *The Journal of the Acoustical Society of America*, *106*(4), 2295–2295.
doi:10.1121/1.427854

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.
doi:10.1016/S0010-0285(03)00006-9

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech Perception as a Talker-Contingent Process. *Psychological Science*, 5(1), 42–46. doi:10.1111/j.1467-9280.1994.tb00612.x

Obediente, E. (1998). *Fonética y fonología*. Mérida, Venezuela: Universidad de los Andes, Consejo de Publicaciones.

Ohala, J. (1996). Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1718–1725.

Orfanidou, E., Davis, M. H., Ford, M. A., & Marslen-Wilson, W. D. (2011). Perceptual and response components in repetition priming of spoken words and pseudowords. *The Quarterly Journal of Experimental Psychology*, *64*(1), 96–121. doi:10.1080/17470211003743794

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*(2), 282–302. http://doi.org/10.1016/j.cognition.2010.01.001

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based Versus Abstract Lexical Entries. *Psychological Science*, *12*(6), 445–449. doi:10.1111/1467-9280.00383

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328. doi:10.1037/0278-7393.19.2.309

Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, *19*(5), 906–913. doi:10.3758/s13423-012-0281-3

Pattamadilok, C., Morais, J., De Vylde, O., Ventura, P., & Kolinsky, R. (2009). The Orthographic Consistency Effect in the Recognition of French Spoken

Words: An Early Developmental Shift from Sublexical to Lexical Orthographic Activation. *Applied Psycholinguistics*, *30*(03), 441–462. doi:10.1017/S0142716409090225

Perea, M., & Carreiras, M. (2003). Sequential effects in the lexical decision task: The role of the item frequency of the previous trial. *The Quarterly Journal of Experimental Psychology*, *56A*(3), 385–401. doi:10.1080/02724980244000387

Peperkamp, S., Le Calvez, R., Nadal, J.-P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition*, *101*(3), B31–B41. http://doi.org/10.1016/j.cognition.2005.10.006

Pierrehumbert, J. (2001a). Exemplar dynamics: Word frequency, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins Publishing Company.

Pierrehumbert, J. (2001b). Stochastic phonology. *GLOT*, *5*(6), 1–13.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.),*Laboratory phonology 7* (pp. 101–139). Berlin: de Gruyter.

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, *34*(4), 516–530. doi:10.1016/j.wocn.2006.06.003

Pierrehumbert, J., Tamati, & Clopper. (2006). The statistical basis of an unnatural alternation. In *Laboratory phonology VIII*.

Port, R. F. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, *25*(2), 143–170. doi:10.1016/j.newideapsych.2007.02.001

Port, R. F., & Leary, A. P. (2005). Against formal phonology. *LANGUAGE*, *81*, 927–964.

Quilis, A., & Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal. In M. Canterro & M. Esgueva (Eds.), *Estudios de fonética I* (pp. 137–252). Madrid: Consejo Superior de Investigaciones Científicas.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 651–666. doi:10.1037/0096-1523.23.3.651

Ryalls, J., Simon, M., & Thomason, J. (2004). Voice Onset Time production in older Caucasian- and African-Americans. *Journal of Multilingual Communiction Disorders*, *2*(1), 61–67.

Samuel, A. (1989). Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different. *Perception & Psychophysics*, *45*, 485–493

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1–17. doi:10.1037/0096-1523.3.1.1

Schwarz, C. J. (2011). Sampling, Regression, Experimental Design and Analysis for Environmental Scientists, Biologists, and Resource Managers.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306–3316. doi:10.1121/1.3101452

Simpson, A. P., & Ericsdotter, C. (2007). Sex-specific differences in F0 and vowel space. In *Proceedings of the XVIth ICPhS* (pp. 933–936). Saarbrücken.

Spivey, M. J., Dale, R., Knoblich, G., & Grosjean, M. (2010). Do curved reaching movements emerge from competing perceptions? A reply to van der Wel et al. (2009). *Journal of Experimental Psychology. Human Perception and Performance*, *36*(1), 251–254. doi:10.1037/a0017170

Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Brain Research. Cognitive Brain Research*, *18*(2), 162–171.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, *111*(4), 1872–1891. doi:10.1121/1.1458026

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2013). Effects of Age of Learning on Voice Onset Time: Categorical Perception of Swedish Stops by Near-native L2 Speakers. *Language and Speech*, 57, 425-450. http://doi.org/10.1177/0023830913508760

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. doi:10.1016/j.wocn.2010.09.001

Siu, M., Gish, H., Chan, A., Belfield, W., & Lowe, S. (2014). Unsupervised training of an HMM-based self-organizing unit recognizer with applications to topic classification and keyword discovery. *Computer Speech & Language*, *28*(1), 210–223. http://doi.org/10.1016/j.csl.2013.05.002

Sullivan, K. P. H., & Schlichting, F. (2007). Speaker discrimination in a foreign language: first language environment, second language learners. *International Journal of Speech Language and the Law*, 7(1), 95–112.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, *2*(3), 339–363.
doi:10.3758/BF03210972

Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, *1*(2), 121–131. doi:10.1002/acp.2350010205

Toscano, J. C., & McMurray, B. (2010). Cue Integration With Categories: Weighting Acoustic Cues in Speech Using Unsupervised Learning and Distributional Statistics. *Cognitive Science*, *34*(3), 434–464. http://doi.org/10.1111/j.1551-6709.2009.01077.x

Trofimovich, P. (2005). Spoken-Word Processing in Native and Second Languages: An Investigation of Auditory Word Priming. *Applied Psycholinguistics*, *26*(04), 479–504. doi:10.1017/S0142716405050265

Välimaa-Blum, R. (2009). The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? *CogniTextes*, *2*. http://doi.org/10.4000/cognitextes.211

Varadarajan, B., Khudanpur, S., & Dupoux, E. (2008). Unsupervised learning of acoustic sub-word units. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers* (pp. 165–168). Association for Computational Linguistics.

Velleman, P. F., & Hoaglin, D. C. (1981). *Applications, basics, and computing of exploratory data analysis*. Boston, MA: Duxbury Press.

Wadnerkar, M. B., Cowell, P. E., & Whiteside, S. P. (2006). Speech across the menstrual cycle: a replication and extension study. *Neuroscience Letters*, *408*(1), 21–24. doi:10.1016/j.neulet.2006.07.032

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*(4), 479–491. doi:10.1016/j.wocn.2010.12.004

Wedel, A. B. (2007). Feedback and Regularity in the Lexicon. *Phonology*, *24*(01), 147–185. doi:10.1017/S0952675707001145

Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

Whiteside, S. P., & Irving, C. J. (1998). Speakers' sex differences in voice onset time: A study of isolated word production. *Perceptual and Motor Skills*, *86*(2), 651–654. doi:10.2466/pms.1998.86.2.651

Whiteside, S. P., & Marshall, J. (2001). Developmental trends in voice onset time: some evidence for sex differences. *Phonetica*, *58*(3), 196–210. doi:56199

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America*, *123*(6), 4524–4538. doi:10.1121/1.2913046

Xue, S. A., & Hao, G. J. (2003). Changes in the Human Vocal Tract Due to Aging and the Acoustic Correlates of Speech Production: A Pilot Study. *Journal of Speech, Language, and Hearing Research*, *46*(3), 689–701. doi:10.1044/1092-4388(2003/054)

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. doi:10.1111/j.2041-210X.2009.00001.x