ABSTRACT

Title of dissertation: SIMILARITY IN L2 PHONOLOGY

Shannon L. Barrios, Doctor of Philosophy, 2013

Dissertation directed by: Professor William J. Idsardi
Department of Linguistics

Adult second language (L2) learners often experience difficulty producing and perceiving non-native phonological contrasts. Even highly proficient bilinguals, who have been exposed to an L2 for long periods of time, struggle with difficult contrasts, such as /r/-/l/ for Japanese learners of English. To account for the relative ease or difficulty with which L2 learners perceive and acquire non-native contrasts, theories of (L2) speech perception often appeal to notions of similarity. But how is similarity best determined?

In this dissertation I explored the predictions of two theoretical approaches to similarity comparison in the second language, and asked: [1] How should L2 sound similarity be measured? [2] What is the nature of the representations that guide sound similarity? [3] To what extent can the influence of the native language be overcome?

In Chapter 2, I tested a 'legos' (featural) approach to sound similarity. Given a distinctive feature analysis of Spanish and English vowels, I investigated the hypothesis that feature availability in the L1 grammar constrains which target language

segments will be accurately perceived and acquired by L2 learners (Brown [1998], Brown [2000]). Our results suggest that second language acquisition of phonology is not limited by the phonological features used by the native language grammar, nor is the presence/use of a particular phonological feature in the native language grammar sufficient to trigger redeployment. I take these findings to imply that feature availability is neither a necessary, nor a sufficient condition to predict learning outcomes.

In Chapter 3, I extended a computational model proposed by Feldman et al. [2009] to nonnative speech perception, in order to investigate whether a sophisticated 'rulers' (spatial) approach to sound similarity can better explain existing interlingual identification and discrimination data from Spanish monolinguals and advanced L1 Spanish late-learners of English, respectively. The model assumes that acoustic distributions of sounds control listeners' ability to discriminate a given contrast. I found that, while the model succeeded in emulating certain aspects of human behavior, the model at present is incomplete and would have to be extended in various ways to capture several aspects of nonnative and L2 speech perception.

In Chapter 4 I explored whether the phonological relatedness among sounds in the listeners native language impacts the perceived similarity of those sounds in the target language. Listeners were expected to be more sensitive to the contrast between sound pairs which are allophones of different phonemes than to sound pairs which are allophones of the same phoneme in their native language. Moreover, I hypothesized that L2 learners would experience difficulty perceiving and acquiring target language contrasts between sound pairs which are allophones of the same phoneme in their native language. Our results suggest that phonological relatedness may influence perceived similarity on some tasks, but does not seem to cause long-lasting perceptual difficulty in advanced L2 learners.

On the basis of those findings, I argue that existing models have not been adequately explicit about the nature of the representations and processes involved in similarity-based comparisons of L1 and L2 sounds. More generally, I describe what I see as a desirable target for an explanatorily adequate theory of cross-language influence in L2 phonology.

SIMILARITY IN L2 PHONOLOGY

by

Shannon L. Barrios

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor William J. Idsardi, Chair
Professor Naomi Feldman
Professor Ellen Lau
Professor Jeffrey Lidz
Professor Nan Jiang

Dedication

For my family and friends, old and new.

Acknowledgments

I owe my gratitude to all the people who have helped make my graduate school years a rich and memorable experience.

I would especially like to thank my advisors, Bill Idsardi and Naomi Feldman. I can't thank Bill enough for agreeing to take me on as his advisee when I decided to change my focus during my second year of the program, and for challenging, guiding, and encouraging me throughout my time at UMD. I am also indebted to Naomi Feldman who arrived just in time to get a bit of computational modeling into this dissertation and for providing me with endless guidance and support during the most challenging parts of my graduate career.

I would also like to thank my committee members, Prof. Nan Jiang, Ellen Lau and Jeffrey Lidz for reading and commenting on this thesis and for having been excellent teachers and mentors.

I would also like to thank several other professors and colleagues who also contributed to time as a student in several ways, either by encouraged me to pursue graduate studies in Linguistics, or by modeling excellence in teaching and scholarship, including: Howard Lasnik, Norbert Hornstein, Colin Phillips, Juan Uriagereka, David Poeppel, Peggy Antonisse, Tonia Bleam, Valentine Hacquard, Alexander Williams, Paul Pieotroski, Andrea Zukowski, Arsalan Kahnemuyipour, Anthony Lewis, Jaklin Kornfilt, William Ritchie, Tej Bhatia, Robert DeKeyser, Kira Gor, Rose McEwen, Joaquin Gomez, Beverly Evans.

Thank you to my classmates and collaborators for their endless support and quality discussion, including but not limited to Annie Gagliardi, Brian Dillon, Alex Drummond, Pedro Alcocer, Dave Kush, Wing Yee Chow, Sol Lago, Shevaun Lewis, Brad Larson, Alexis Wellwood, Dan Parker, Giovanna Morini, Sunyoung Lee Ellis, Anna Lukyanchanko, Candise Lin, Matt Winn, Akira Omaki, SoOne Huang, Ariane Rhone, Phillip Monahan, Eri Takahashi, Diogo Almeida, Julian Jenkins, Clare Stroud, Matias Scharinger, Kazuya Saito, Dayna Cueva de Alegria, KyoungSon Park, and Esra Keisici.

Thank you to Kathi Faulkingham, Kim Kwok, for their support, good spirits, and help navigating the university bureaucracy.

I am also grateful to Elizabeth Nyugen and Anna Bonnet for their help with the MEG study reported in this thesis.

I would also like to thank my non-linguist friends who have supported me, including Ariadna Pich, Jenny Fisher, Holly Walrath, Minda Wesley, Paul Prescott, Sara Houghtaling, and Kathleen Nugent.

I would like to thank my family (old and new) who have supported me throughout graduate school. I am especially grateful to Edison Barrios. This thesis would not have been possible without your encouragement and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Unlike typical child first language (L1) learners, adult second language (L2) learners rarely (if ever) achieve native-like production and perception of the target language (TL)[1] . One important goal of Second Language Acquisition (SLA) research is to identify the cause of these well-documented differences in outcome between first and second language learners, as well as to explore the upper limits on ultimate attainment in adult second language acquisition.

One fruitful approach to variable success among L2 learners has been to hypothesize that non-convergent outcomes in L2 phonology are due in large part to transfer (the recruitment of L1 representations and processes in the analysis of L2 items). From this perspective, a great deal is to be gained from empirical studies of cross-language speech perception, which systematically explore the way that the learner's native and non-native languages interact during L2 acquisition and/or processing. This approach has been prevalent in the L2 literature since the earliest systematic study of the acquisition of L2 phonology. Accounts of L2 segmental acquisition have been offered from a variety of theoretical perspectives, including

[1] A noteworthy exception to this common finding is reported in Bongaerts et al. [1997] who found that highly successful Dutch late learners of English perform within the native speaker range on a spontaneous production task.

[a] work within the Contrastive Analysis Framework (Weinreich [1953], Lado [1957], Lehn and Slager [1959], Stockwell and Bowen [1965]), [b] work within the generative framework (Ritchie [1968], Michaels [1973], Michaels [1974], Archibald [1993], Hancin-Bhatt [1994], Brown [1998], Brown [2000]), and [c] work outside the generative framework (Best [1994], Flege [1995], Flege [2003]). All of these perspectives share the common goal of establishing which acoustic, articulatory, phonetic or phonological properties are relevant for interlingual identification.

This dissertation joins this body of work in aiming to provide an account of cross-language influence in L2 phonology and speech perception. Like previous researchers, I assume that cross-language influence occurs because L2 learners establish interlingual mappings via a similarity-based comparison, in which new objects are categorized on the basis of their similarity to existing category representations. In this dissertation, I use the term "similarity" to refer to the psychological proximity of the mental representation of two perceptual/conceptual objects.

The goal of this dissertation is to pin down the appropriate measure of phonological similarity and to investigate whether this influence of the native language can be overcome. To this end, I present original evidence which bears on similarity as it relates to the learning of novel segmental contrasts in a second language. The findings of these experiments and model simulations suggest that existing approaches to similarity are insufficient to capture the observed behavior. In light of these findings, I sketch what I see as a desirable target and outline a framework for the systematic study of cross-language sound similarity comparisons.

1.2 Similarity

Similarity has played a central role in theories of behavior and knowledge in the cognitive sciences. It has been employed as an explanatory construct in accounts of [a] object recognition (Biederman et al. [1987]), [b] memory retrieval (Hintzman [1986]), and [c] reasoning (Rips [1975]). Similarity has also been thought to underly [d] transfer in learning (Gentner et al. [2003], Ross [1984]), [e] generalization (Shepard [1987]), and [f] categorization (Nosofsky [1986]), among other phenomena.

1.2.1 Approaches to Similarity

Several different theoretical approaches to similarity have been proposed in the literature. These models differ in their commitments about the structure of mental representations and the process involved in comparing pairs of those mental representations. Here I focus on two classic approaches to measuring similarity; [a] a spatial approach (Shepard [1962], Shepard [1987]), and [b] a feature-based approach (Tversky [1977]), since prominent models of (L2) speech perception are instantiations of these two types of approach.

1.2.2 Spatial models of similarity

In spatial models of similarity (Shepard [1962]), objects are represented as points in a multidimensional, continuous, coordinate space, and (dis-)similarity between two mental objects corresponds to metric distance between the corresponding points. The closer together two objects are in this representational space, the more similar they are, and the further apart two objects are in this representational space, the more dissimilar.

Spatial models have also been supported by multidimensional scaling techniques (MDS), which take proximity data as input and output a map that locates all the points in space (Shepard [1962], Torgerson [1965], among many others). Spatial representations paired with MDS techniques provide a powerful tool for mapping mental spaces.

1.2.3 Featural models of similarity

Feature-based models provide an alternative to the spatial models approach just described. In feature-based models of similarity comparison, objects are represented as sets of discrete features. Featural representations of this sort make it possible to compare pairs of objects using elementary set operations. For example, if object a is represented by the feature set A and object b is represented by the feature set B, as shown in Figure 1.1, it is relatively straightforward to determine which features of the represented object are common and which are distinctive. Common features are those which lie at the intersection of feature set A and B, labelled A∩B in Figure 1.1. Distinctive features are those features which are not at the intersections of the feature sets representing each object. These are labelled $A-B$ and $B-A$ respectively. In feature-based models of similarity comparison, such as Tversky's Contrast Model, the similarity of a pair of objects can be determined by comparing the respective sizes of the sets of common and distinctive features.

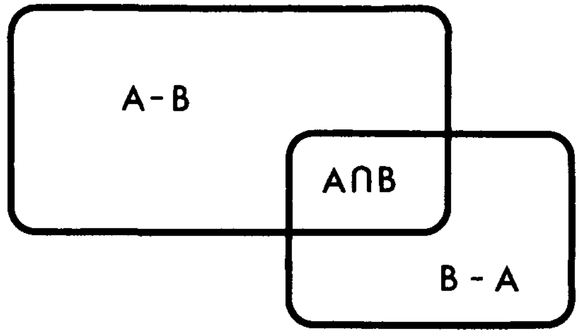

Figure 1.1: Feature overlap. Reprinted from Tversky [1977].

The model predicts that the similarity of a pair of objects will increase with the size of the set of common features. Dissimilarity is expected to increase as the size of the set of distinctive features grows.

1.2.4 Measuring sound similarity

Theories of phonology and speech perception have also made use of feature-based and spatial representations of phonological categories and either implicitly or explicitly assume that categorization is based on some feature-based or spatial model of similarity comparison.

The traditional view in linguistics is that words are encoded in long-term memory as discrete, abstract lexical representations (i.e. a series of segments consisting of a bundle of distinctive features, which indicate the articulatory configuration underlying the segment). These representations crucially do not contain the same type of acoustic variation which is present in the speech signal (Jakobson et al. [1952], Chomsky and Halle [1968]). The view that distinctive features play a privileged role as the primitives of lexical representation and phonological computation has also been adopted as the working assumption in speech perception (Studdert-Kennedy [1980]).

In contrast, recent models of spoken word recognition have proposed that speech perception involves a mapping of raw spectral properties directly onto stored long-term memory representations (Klatt [1989], Stevens [2002]). Moreover, in these exemplar-based/episodic models of word recognition a large amount of acoustic detail is stored in memory (Johnson [1997], Goldinger [1998], Bybee [2003], Pierrehumbert [2003]). These models are consistent with the type of representational theory assumed from a spaces perspective in which sounds/words are represented as points in a continuous, multidimensional space and are compared with existing representations.

1.3 Similarity and L2 Phonology

Not surprisingly, theories of nonnative and second language speech perception also appeal to notions of similarity to explain which contrasts will be easiest for non-native listeners to perceive and acquire. In what follows, I discuss the problem that the L2 learner faces in acquiring the phonology of a second language, as well as describe what a spaces and features approach each assume is involved in the process.

1.3.1 What must be learned?

In some respects the problem of acquiring an L2 phonology is much like the one a child learner faces as he/she acquires the phonology of his/her first language.

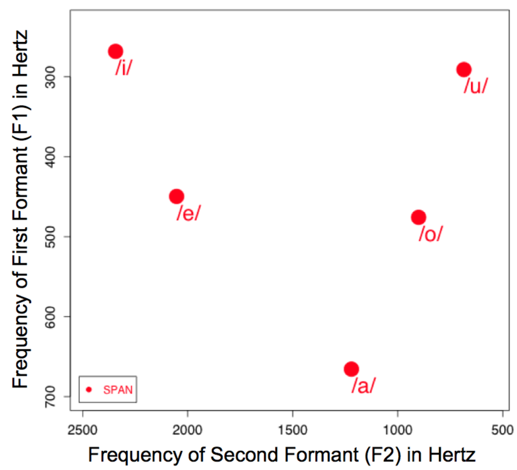

Figure 1.2: Spanish Vowel Means plotted in an F1 by F2 space. Values from Quilis and Esgueva [1983].

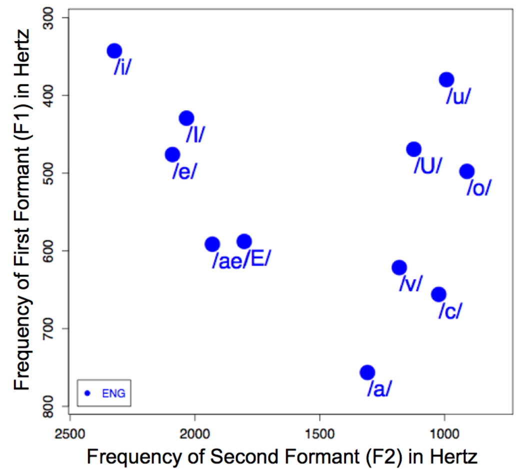

Figure 1.3: English Vowel means plotted in an F1 by F2 space. Values from Hillenbrand et al. [1995]. (The symbol I=/ɪ/, E=/ɛ/, U=/ʊ/, v=/ʌ/, c=/ɔ/, a=/ɑ/).

Using linguistic input as evidence of the variety of sounds that occur in the target language, the learner must determine the functional significance of each (i.e. whether a given sound is in contrastive or complementary distribution with the other sounds of that language). Moreover, the learner must develop phonemic representations for those segments which are in contrast in the target language, so as to distinguish them in his or her grammar.

In addition to the similarities mentioned above, there is an obvious difference between first and second language acquisition. Adult second language learners already have a fully developed linguistic system at their disposal. Along with this system comes a set of sounds and rules which act on them. A crucial question then is: how does having a fully developed L1 affect speech perception, and consequently, the development of L2 phonological knowledge? The answer to this question will bring us closer to a principled account of cross-language influence.

1.3.1.1 Acquiring new sound categories

To illustrate the need for acquiring new sound categories, take for example an adult L1 Spanish learner who is exposed to English. A adult Spanish speaker has acquired a grammar that allows him/her to make the five-vowel distinction required to distinguish word meanings in his or her native language. The mean of these five vowel categories are plotted in red in Figure 1.2 above. Figure 1.3 shows the mean of eleven English vowels taken from Hillenbrand et al. [1995]. From these two plots it is quite easy to see that the Spanish learner of English will encounter several sounds which he/she is not accustomed to. Some of which will be very similar or nearly identical to his/her Spanish language vowel categories, some less similar. In order to achieve native-like performance, the learner will need to establish new sound categories so as to distinguish them from existing ones.

Of course, the second language learner will need to acquire more than just sounds. The new phones which he/she is gaining experience with may be used in contrast with other segments in the inventory. That is, some new target language words will need to be distinguished by these new target language sounds. For example, the words [tʃip] 'cheap' and [tʃɪp] 'chip' are minimal pairs in English which rely on a contrast between English /i/ (a sound category which is very similar to a native language sound category) and /ɪ/ (a new sound category). Thus, the second language learner must learn which target language segments are in contrast (i.e. can change the meaning of words) and acquire distinct phonological representations for target language sounds, so as to distinguish minimally contrastive word pairs in the

L2 lexicon.

1.3.1.2 L2 phonology is hard

It is a widely accepted fact, supported both by anecdote and experimental evidence, that nonnative listeners and L2 learners experience difficulty perceiving and producing certain nonnative segmental contrasts. For example, the /r/-/l/ contrasts is notoriously difficult for Japanese learners of English (MacKain et al. [1981], Goto [1971], among many others). Likewise, Catalan learners of English reportedly experience difficulty with the English /i/-/ɪ/ contrast (Cebrian [2008]) and English-learners of French with front rounded vowels /y/ and /ø/ (Gottfried [1984]). Moreover, these perceptual difficulties have been argued to extend to the lexicon, resulting in homophonous representations of minimal pairs distinguished by difficult L2 contrasts (Pallier et al. [2001]).

1.3.1.3 But not uniformally so

A surprising finding, however, is that not all nonnative contrasts present the same degree of difficulty. That is, some nonnative contrasts are more difficult than others for adult second language learners to perceive and acquire. For example, while the /r/-/l/ contrast has been found to be particularly problematic for Japanese learners of English, these same learners have been shown to experience much less difficulty perceiving and acquiring other non-native sound contrasts such as /b/-/v/ (despite the fact that Japanese has neither contrast) (Brown [1998, 2000]). Thus, a descriptively and explanatorily adequate theory of second language phonology and speech perception will need to account for these more nuanced differences in the ease and difficulty with which second languages are learned. This raises the question, how are L2 learners measuring sound similarity so as to produce this asymmetry.

1.4 Two approaches to sound similarity

Similarity has been used, both explicitly and implicitly, as an explanatory construct in existing models of L2 phonology and speech perception. There have been several approaches which aim to account for the fact that L2 phonology is hard, but not uniformly so. Here I focus on two prominent approaches to this problem, which I will refer to as the 'rulers' approach and the 'legos' approach. The rulers approach is an instance of the classic spatial model perspective taken by (Shepard [1962], Shepard [1987]), whereas the legos story is an instance of a feature-based approach (Tversky [1977]).

1.4.1 A rulers approach

What I will be referring to as a 'rulers' approach pools together a number of spatial approaches to cross-language sound similarity. In this framework, L2 sound categories are mapped to the most similar L1 category, where similarity is measured as the distance between points in some space (articulatory, acoustic, perceptual) according to some metric. In these models, perceptual difficulty is thought to arise when two different nonnative sounds are mapped to a single L1 category (termed ‘single-category assimilation’ by Best [1994], Best [1995], Best and Tyler [2007]).

Under a rulers approach, acquiring new sounds involves learning where new target language sound categories lie in some space (which may be a collection of correlated subspaces, e.g. articulatory and auditory). For example, the Spanish learner will need to track the distribution of new English vowel categories relative to his/her existing Spanish vowel categories. Figure 1.4 shows the means of the vowel categories of Spanish and English plotted in (acoustic) F1 x F2 space. The intuition is that sounds will be grouped with the nearest L1 category, as measured by Euclidean distance in this space. Learning difficulty under a rulers approach is predicted when target language sounds are ‘equivalence classified’ (Flege [1995, 2003]). Equivalence classification is expected to occur when target language sounds are sufficiently similar to a native language phonetic category. In contrast, when new target language sounds are sufficiently different from existing phonetic categories, these models predict that the learner will experience little or no difficulty acquiring the new category. By distinguishing between ‘new’ and ‘similar’ phones (Flege [1995], Flege [2003]) or the degree of category goodness (Best [1994], Best [1995]), models of the rulers type make different predictions with respect to the perception and acquisition of various non-native sounds.

1.4.2 A legos approach

A second approach takes phonological features to be the representational building blocks. I will refer to this type of proposal as a ‘legos’ approach. From this

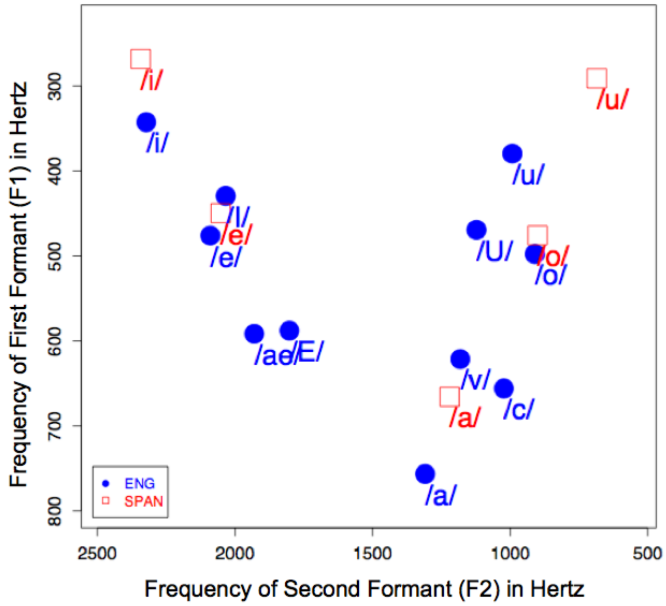

Figure 1.4: Spanish and English vowel category means plotted in F1 by F2 space. The English formant values come from Hillenbrand et al. [1995] and the Spanish data from Quilis and Esgueva [1983]. (The symbol I=/ɪ/, E=/ɛ/, U=/ʊ/, v=/ʌ/, c=/ɔ/, a=/ɑ/.)

perspective, learning a new target language sound means building a unique representation, so as to distinguish that sound from all segments in the inventory. In principle, this could be done either by acquiring a new feature or reusing existing features to uphold the novel contrast (see Brown [1998], Brown [2000]). The simple five vowel system from Spanish shown in Figure 1.5 can be used to illustrate this point.

The adult Spanish learner of English will have acquired this five vowel system, which we will assume has the following featural structure. It has a three-way height contrast (two high vowels /i /and /u/, two mid vowels /e/ and /o/, and a single low vowel /a/) and two-way contrast along the front/back dimension (Hualde [2005]). In order to distinguish the English vowel /æ/ from the vowel /a/ in his/her vowel

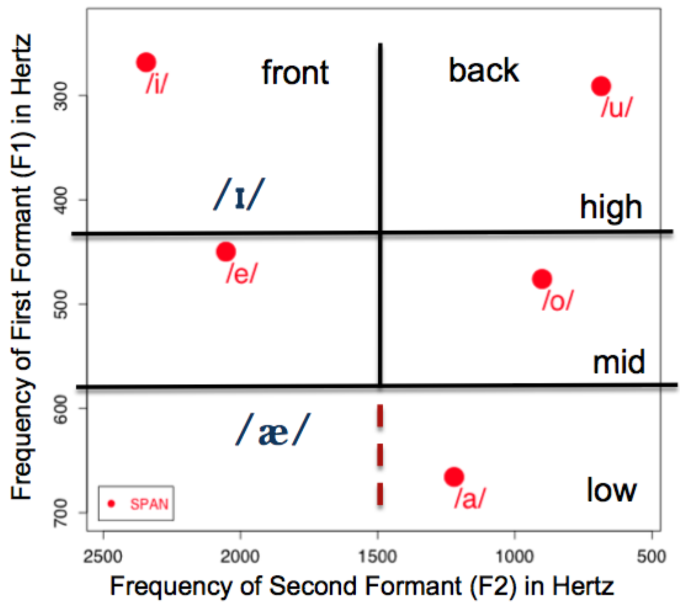

Figure 1.5: Spanish and English vowel categories plotted in F1 by F2 space. The English formant values come from Hillenbrand et al. [1995] and the Spanish data from Quilis and Esgueva [1983].

system, the learner will have to build a new representation. Under a particular instantiation of a feature-based approach, if a particular building block is used to distinguish word meanings in the native language then it is deemed 'available' and, thus, can be redeployed during second language learning. Thus, in this theory feature availability predicts which nonnative contrasts will be accurately perceived (and therefore acquired) (Brown [1998, 2000]). Notice that from this perspective the Spanish learner should be able to acquire the English /a/-/æ/ since he/she has a feature, namely [front], which can be reused for this purpose. In contrast, learning the distinction between /i/ and /ɪ/ would require acquiring a novel feature, since Spanish has no tense/lax distinction. The extent to which existing features are redeployed in L2 learning and whether or not new features can be acquired by L2 learners are both open questions in L2 phonology. I take up both of these issues in more detail in Chapter 2.

1.5 Tools for investigating similarity

There are several tools and techniques that have been used to investigate sound similarity.

This dissertation reports findings from a variety of methodologies, including behavioral, computational modeling and neuroimaging methods. The hope is that each of these tools will potentially shed light on different aspects of similarity in L2 phonology. As a general methodological point, I hope to demonstrate that there are a number of tools, some old and some new and exciting, which can be employed to help us characterize cross-language influence and better our understanding of how adults represent, and process various aspects of their second language.

1.6 Outline of this dissertation

Chapters 2, 3, and 4 report original evidence from experiments and simulations involving highly proficient Spanish late-learners of English, which can be brought to bear on the types of representation and processes involved in similarity comparison.

In Chapter 2, I use accuracy and reaction time measures to assess the abilities of a group of advanced L2 learners with new sounds and try to relate those abilities to distances in metric space (rulers) or featurally (legos). In particular, I investigate whether phonological features are redeployed to represent nonnative vowel contrasts in the second language (Brown [1998], Brown [2000]) and to what extent new phonological structure might be acquired. I explore this hypothesis by examining the phonetic perception and lexical representation of nonnative vowel contrasts by advanced L1 Spanish late-learners of English. I find converging evidence to suggest that second language acquisition of phonology is not constrained by the phonological features made available by the learner's native language grammar, nor is the presence or use of particular phonological features in the native language grammar sufficient to trigger redeployment. Based on these results, I conclude that feature availability is neither a necessary nor a sufficient condition to predict learning outcomes.

In Chapter 3, I investigate whether phonetic approaches can better explain the learners' performance on these vowel contrasts by extending a computational model proposed by Feldman et al. [2009] to nonnative speech perception. The model assumes that acoustic distributions of sounds control listeners' ability to discriminate a given contrast. By comparing the model's predictions with the behavior we observed in highly proficient bilinguals in the previous experiment, we can determine the extent to which these learners are relying on knowledge of categories in their native language, and how their knowledge has changed to incorporate knowledge of the target language. The results of this project will help us understand how people represent similarity among sounds and assess whether acoustic similarity is sufficient to account for their behavior.

In addition to encountering new sounds, second language learners must also acquire new phonological relationships. For example, [d] and [ð] are allophones of different phonemes in English, but not in Spanish. Moreover, the phonological relation between sounds in a language also impact listeners' perception, such that allophones of the same phoneme are perceived as more similar than allophones of separate phonemes. This finding has led some researchers to posit that "similarity is comprised of three components: [1] auditory similarity, [2] phonetic inventory, and [3] language-specific patterns of alternation" (Johnson and Babel [2010], p. 127). In Chapter 4, I use behavioral methods and magnetoencephalography (MEG), an electrophysiological technique that has the benefit of providing evidence for covert psychological processes that precede a listener's behavioral response, to investigate whether second language learners acquire new phonological relationships and how their ability to perceive and acquire these contrasts might be influenced by native language phonological relations. I hypothesize that if advanced Spanish learners of English have acquired knowledge of English, this knowledge should be reflected in their behavioral response, the MMN response (i.e. an automatic pre-attentive brain response to changes in stimulus), or both.

Finally, in Chapter 5, I discuss several issues pertaining to testing and modeling similarity more generally, and review general conclusions.

Chapter 2

Feature availability is neither necessary, nor sufficient

2.1 Introduction

L2 learners typically differ from native speakers of the target language in their production and perception of certain nonnative contrasts. Moreover, observed difficulties are not straightforwardly predicted from differences between L1 and L2 surface structures, as suggested by a Contrastive Analysis approach (Lado [1957], among others). Instead, adult second language learners have been found to perceive some nonnative segmental contrasts remarkably well (i.e. adult English listeners' discrimination of Zulu clicks, Best et al. [1988]), while other nonnative contrasts pose serious learning difficulties, such as /ɹ/-/l/ for Japanese learners of English or /i/-/ɪ/ for Spanish learners of English.

Despite a general consensus that at least some of these difficulties are attributable to the listener's native language, there is considerable debate about [1] the nature of the representations that guide the processing of nonnative input and [2] the extent to which the influence of the native language can be overcome.

2.1.1 Goal of this chapter

In this chapter we explore these issues by examining the phonetic perception and lexical representation of two nonnative vowel contrasts by advanced L1 Spanish late-learners of English. As mentioned in Chapter 1, models of nonnative speech perception and L2 sound learning have provided two different types of answers to these questions. Our goal is to provide additional empirical data to bear on the rulers and legos approaches to L2 sound learning. To this end, we investigate whether contrast difficulty is a simple matter of acoustic distance, whether phonological features are redeployed to represent nonnative contrasts in the second language (Brown [1998], Brown [2000]), and the extent to which new phonological structure can be acquired.

2.1.2 Outline of the chapter

In the section that follows, I first review a prominent rulers model of L2 speech perception (Flege's Speech Learning Model (SLM), Flege [1995], Flege [2003]), and highlight some important limitations of the existing model. I argue that in view of these limitations we have good reason to consider alternatives. In this chapter I pursue a particular instantiation of a legos approach to sound similarity.

2.1.3 A prominent rulers approach to L2 sound learning

Flege's Speech Learning Model (SLM) aims to account for the ease and difficulty with which nonnative sounds are perceived, and therefore, acquired by second language learners (Flege [1995, 2003]). Like other ruler approaches, such as Best's Perceptual Assimilation Model of nonnative speech perception (which will be discussed in more detail in Chapter 3), the SLM assumes that sounds are adequately represented as points in some multidimensional space and that sound learning involves learning where new target language sounds lie in that space.

The predictions of the SLM are based explicitly on the similarity/proximity of L2 sounds to existing phonetic categories in some representational space. According to the SLM, although adult L2 learners maintain the capacity to acquire new categories, their success can be predicted by whether or not a target language phone is 'new' or 'similar'. 'New' phones are those which are not equated to any native language phonetic category and, therefore, are perceived and acquired with little interference from the native language. 'Similar' phones, in contrast, are predicted to pose difficulties as a result of the process of '*equivalence classification*' whereby L2 sounds are equated to native language categories. From this perspective, acquiring new phonetic categories involves the learner discerning differences between target language sounds and existing phonetic categories. Thus, acquisition is expected to be slower and more difficult in the case of 'similar' phones.

2.1.4 Limitations of the SLM

While Flege's SLM is intuitively appealing, and a large body of experimental results have been interpreted in the framework, the proposal suffers from a significant limitation. The model relies heavily on the notion of similarity to make predictions about second language learners' ability to perceive and acquire various L2 sounds, but does not define or operationalize the construct, nor does it provide an *a priori* means of determining the similarity between L1 and L2 sounds independently of L2 perceptual differences which it aims to explain. Relatedly, the SLM lacks a means for determining when cross-language phonetic differences are large enough to support novel category formation[1]. However, in the absence of a characterization of cross-language similarity, the explanatory value of the construct is seriously diminished.

In sum, if a model of this sort is to be pursued, then one is faced with the burden of providing a metric by which similarity can be determined.

2.1.5 Acoustic similarity will not do

It is worth pointing out that existing evidence suggests that perceptual similarity cannot be predicted straightforwardly from simple acoustic similarity. For example, despite the fact that German /y/ and English /u/ are acoustically closer, Polka and Bohn [1996] report that L1 Canadian English speakers perceive German /u/ as a better exemplar of English /u/. Similarly, Strange et al. [2004] found that L1 American English listeners equate German /y/, /œ/ and French vowels /y/, /ʏ/, /œ/, /ø/ with English back vowels, despite them being acoustically closer to the front unrounded vowels of English. Moreover, English learners of French and French learners of English map French /ʁ/ to English /ɹ/, despite the fact that these sounds are acoustically quite different.

A similar point can be made if one attempts to predict which English vowel contrasts will be difficult for Spanish learners on the basis of acoustic similarity between L1 and L2 sound categories. Figure 2.1 shows the five Spanish vowels and

[1]We acknowledge that it may be possible to take a signal detection approach to this problem using the means and standard deviations to assess how much overlap there is in the tails of the distributions. To my knowledge this hasn't been explicitly proposed or adequately explored.

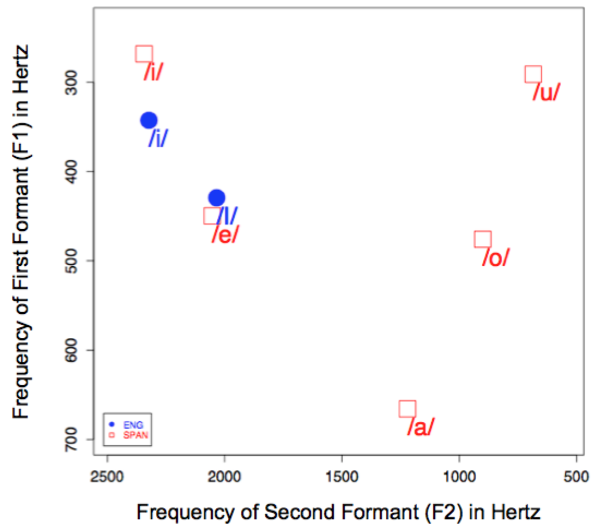

Figure 2.1: Five Spanish and two English vowel categories (shown in red and blue respectively) plotted in F1 by F2 space. The English formant values come from Hillenbrand et al. [1995] and the Spanish data from Quilis and Esgueva [1983].

two English vowels /i/ and /ɪ/ plotted in F1 x F2 space. The means of the Spanish categories are shown in red and the means of the two English vowel categories in blue. What can be readily observed from the plot is the fact that the mean of the English vowel /i/ lies between the Spanish vowels /i/ and /e/, although a bit closer to the Spanish vowel /i/, whereas the English vowel /ɪ/ lies remarkable close to the the Spanish vowel /e/. If acoustic similarity were a good predictor of perceptual similarity, we would expect that Spanish listeners would map Spanish/i/ onto English /i/ and English /ɪ/ onto Spanish /e/, as these are acoustically the nearest L1 categories. However, both anecdotally and empirically, the fact seems to be that Spanish speakers tend to map both English /ɪ/ and English/i/ onto Spanish/i/, not English /e/ (as the rulers approach would predict based on raw acoustic distance).

Thus, perceptual similarity does not seem to follow straightforwardly from acoustic similarity as the rulers approach would have it. Given the limitations just discussed, I argue that it is reasonable to consider whether or not an alternative approach fares any better. In what follows, I introduce Brown's Feature-based model of L1 interference.

2.2 A Legos Alternative

Brown's Feature-based Model (FBM) of L1 interference in L2 speech perception provides one of few phonological approaches to L2 sound perception and acquisition (Brown [1998, 2000]). Brown's proposal is formulated within a non-linear phonological framework in which phonemes are composed of a hierarchy of distinctive features known as a Feature Geometry (Clements [1985]). Following Rice and Avery [1991] Minimally Contrastive Underspecification, each phoneme of a particular language has a unique structural representation, which distinguishes it from other phonemes in the inventory[2].

The central premise of Brown's FBM of L1 interference is that "the learner's native grammar constrains which nonnative contrasts he or she will be able to accurately perceive and, therefore, limits which nonnative contrasts the learner will successfully acquire" (Brown [2000], p. 19). In particular, Brown hypothesizes that if a learner's grammar lacks the feature necessary to uphold a given phonological

[2]Additionally, given that languages differ with respect to their phoneme inventories, they may also differ with respect to the phonological features they use in order to uphold contrast. As a consequence, the representation for a given segment Z in language A may differ from its representation in language B.

contrast, he or she will be unable to accurately perceive and, therefore, acquire the nonnative distinction. In contrast, the mere presence of the necessary phonological feature in the native grammar will facilitate perception of the nonnative contrast, regardless of whether or not the target language segment can be found in the native inventory. Thus, Brown's FBM predicts that L1 phonological features may be "redeployed" by the learner in order to represent certain nonnative L2 contrasts (whether or not these features are the ones actually employed in the target language).

The predictions of Brown's FBM have been supported empirically by cross-language speech perception studies with L2 learners from various language backgrounds. For example, Brown [1998] used an auditory AX discrimination task and a forced choice picture identification task to compare the performance of Mandarin Chinese and Japanese learners on the English /ɹ/-/l/ contrast. She found that Chinese listeners were able to accurately perceive minimal pairs containing the English /ɹ/-/l/ contrast, whereas Japanese listeners were not. According to Brown, this is because the [coronal] feature that is needed to distinguish the contrast is available to be redeployed by Mandarin Chinese speakers, but not Japanese speakers.

In a second experiment, Brown further tested the redeployment hypothesis by comparing the acquisition of nonnative contrasts within a single language group. In this experiment, Brown used an auditory AX discrimination task and a forced choice picture identification task to investigate the perception and representation of a number of different English contrasts, including nonnative contrasts /ɹ/-/l/, /b/-/v/ and /f/-/v/, and native contrast /p/- /b/ by Japanese learners. Since the Japanese grammar contains both [continuant] and [voicing] contrasts, but does not employ the feature [coronal], it was predicted that Japanese learners would accurately perceive the English /b/-/v/, /f/-/v/ and /p/-/b/ contrasts, but not the /ɹ/-/l/ contrast. The results were consistent with Brown's prediction and were taken to support her feature-based model of L1 interference in which new features cannot be acquired, but new segmental representations can be built from existing L1 features.

Additional support for the view that features are redeployed during L2 learning comes from Matthews [1997], who similarly found that Japanese learners improved in their ability to perceive the nonnative contrasts /b/-/v/, /s/-/θ/, /θ/-/f/ for which their native language has the relevant feature, but not /ɹ/-/l/ for which the relevant feature is not used in the L1 grammar. Likewise, Atkey [2002] found that English speakers are able to perceive and acquire the Czech palatal stops (i.e. /c/-/ɟ/). According to the author this is due to the fact that English grammar already employs the [posterior] feature to distinguish the English /s/-/ʃ/ contrast, and so this feature is available for redeployment. These findings from Brown [1998], Brown [2000], Matthews [1997], and Atkey [2002] are all consistent with the idea that new target language representations are constrained by the available representational building blocks. In particular, these results suggest that new features cannot be acquired, but existing features can be redeployed to build new segmental representations for some contrasts.

2.2.1 Evidence for the acquisition of new structures

While the findings just mentioned appear to suggest that the acquisition of new structure in the L2 phonology is limited to cases where structures can be built from available L1 features via redeployment, several recent findings suggest that L2 learners are able to acquire new structures that require features which are inactive in the learner's native language.

For example, Curtin et al. [1998] conducted a perception training study to investigate the role of lexical and surface representations in phonological transfer by comparing native English and French speaking participants' ability to lexically represent the Thai three-way stop voicing distinction (e.g., /b/-/p/-/p^h/). They found that both English speakers and French speakers readily acquired the voicing contrast (presumably by redeployment of their native [voice] feature, which is used to distinguish lexical items in their L1). Additionally, the English-speaking group, but not the French-speaking group, showed evidence of acquiring the ability to lexically encode [aspiration] in Thai. This difference in outcomes is attributed to the presence of [aspiration] in surface representations in English, but not French. The authors concluded that L1 surface features can be lexicalized in L2 acquisition, even though they are not transferred initially[3]. These findings also suggest that L2 acquisition of phonology may not be limited by the distinctive features of the L1 phonology, but that experience with non-contrastive surface features may also play a facilitatory role in L2 sound learning.

LaCharité and Prévost [1999] also investigated whether L2 learners can ac-

[3]But see also Pater [2003] for contradictory findings.

quire phonological features which are inactive in their native language. The authors explored the acquisition of the the sounds /θ/ and /h/ by Canadian French speaking learners of English. They find that the French speaking participants showed better performance for /θ/ than they did for /h/. Since the representation of /θ/ relies on the learner's ability to acquire a new terminal feature [distributed] and the /h/ on adding a place feature [pharyngeal], the authors argue that new terminal features can be acquired, but that learners cannot acquire new place features.

In sum, a growing body of literature bears on the issues of feature redeployment and the acquisition of new phonological features by adult L2 learners. What is noteworthy however is that, as of yet, evidence regarding these issues has been limited to consonants. That is, no study (that I'm aware of) has brought data from nonnative vowel perception to bear on these issues of feature redeployment and the acquisition of new structure. We aim to fill this gap in the literature by conducting two experiments to test the predictions of Brown's feature-based model of L1 interference (Brown [1998, 2000]).

2.3 Motivation for two experiments

To test Brown's predictions regarding feature redeployment we chose three English vowel contrasts, namely /ɑ/-/æ/, /i/-/ɪ/, and /o/-/u/. Figure 1.5 is repeated here in 2.2 to illustrate the Spanish and English vowels of interest. The /o/-/u/ contrast is common to both Spanish and English, whereas the /ɑ/-/æ/ and /i/-/ɪ/, contrasts are English-specific. Although the Spanish vowel inventory has segments similar to English /ɑ/ and /i/, it lacks the phonemes /æ/ and /ɪ/. Given that the feature that distinguishes the /ɑ/-/æ/, but not the /i/-/ɪ/ contrast, is present in the L1 feature geometry of Spanish speakers, it is interesting to ask whether the highly proficient L1 Spanish late learners of English have acquired either or both of these non-native contrasts. Brown's model predicts that L1 features [± back, ± high] may be redeployed to represent the nonnative /ɑ/-/æ/ contrast, but that the /i/-/ɪ/ contrast will not be accurately perceived since the feature needed to make this tense/lax distinction is absent from the learner's L1 grammar (as are other potential contrasts which could be employed to this purpose, such as vowel length). Thus, Brown's model predicts that the /i/-/ɪ/ contrast will be unacquirable by Spanish learners, regardless of their experience. Two experiments were conducted using the same stimuli to explore the phonetic perception and lexical representation of the nonnative vowel contrasts by advanced L1 Spanish late-learners of English.

2.4 Experiment 1

In Experiment 1, we used an AX discrimination task to investigate the ability of advanced Spanish late-learners of English to discriminate minimally contrastive English words and nonwords. We manipulated whether the experimental pairs were comprised of a common vowel contrast (i.e. /o/-/u/) or an English-specific vowel contrast (i.e. /ɑ/-/æ/ or /i/-/ɪ/). It was hypothesized that if feature availability predicts the acquisition of nonnative contrasts as Brown's FBM suggests, we would find better performance for /ɑ/-/æ/ relative to /i/-/ɪ/. Table 2.1 summarizes these

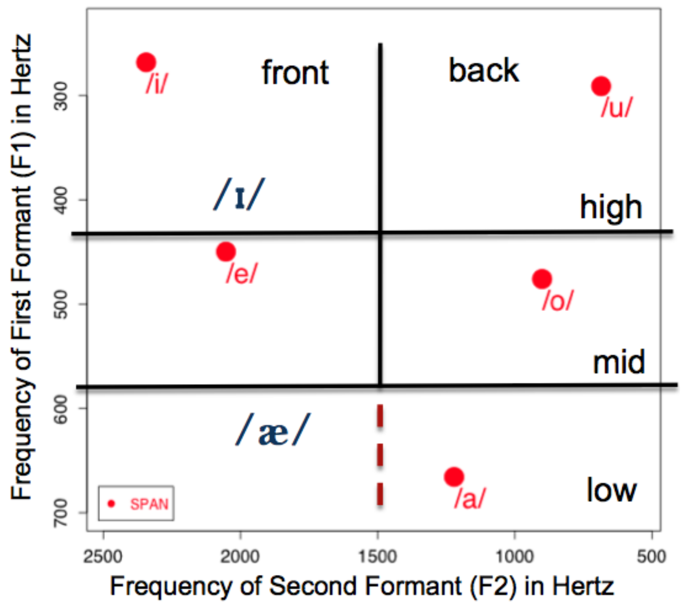

Figure 2.2: Spanish and English vowel categories plotted in F1 by F2 space. The English formant values come from Hillenbrand et al. [1995] and the Spanish data from Quilis and Esgueva [1983]. The lines which are overlaid represent the proposed featural partitions of this space.

Table 2.1: Summary of Predictions for Experiment 1

	Contrast		
Language Group	/o/-/u/	/ɑ/-/æ/	/i/-/ɪ/
Native Speakers (English)	good	good	good
Nonnative Speakers (Spanish)	good	good	poor

predictions.

2.4.1 Participants

Twenty-eight native English speakers (16 Male and 12 Female, Mean age= 21.25) and 28 advanced Spanish late-learners of English (11 Male and 17 Female, Mean age=27.36) from the University of Maryland, College Park campus participated in this experiment. Native English-speaking participants were recruited from

Table 2.2: Summary of Experimental Items(192 total)

Word: 48 Pairs			Nonwords: 48 pairs		
/u/-/o/	/ɑ/-/æ/	/i/-/ɪ/	/u/-/o/	/ɑ/-/æ/	/i/-/ɪ/
16 pairs	16 pairs	16 pairs	16 pairs	16 pairs	16 pairs

two introductory linguistics courses and were given course credit for their participation. The Spanish-English bilinguals were all students, post-docs or professors from various Spanish-speaking countries who were studying or working at the University of Maryland at the time of testing. All reported using English regularly throughout their day. On average the Spanish-English bilinguals who participated in this study began learning English in middle school at the age of 12 and they typically received 8 years of formal training. The mean length of residence in the United States was 6.5 years. All participants had very limited exposure to English in natural settings before arriving in the US. All Spanish-speaking participants were compensated $10 for taking part in this study.

2.4.2 Stimuli

The stimuli used in this experiment were monosyllabic English words and nonwords [4]. Each of the 192 experimental items (see *Appendix A* for full list of the experimental items) contained one of the following three vowel contrasts: /u/-/o/, /ɑ/-/æ/, and /i/-/ɪ/. Sixteen minimal pairs were constructed for each contrast for both words and nonwords (6 sets of 16 pairs), as summarized in Table 2.2.

In addition to the test stimuli described above, thirty-two additional words

[4]Some bisyllabic items were included where necessary.

and thirty-two nonwords served as filler items. Crucially, filler items did not include any of the vowel contrasts of interest. All stimuli were recorded by both one male and one female native speaker of American English. The condition in which each item appeared was counterbalanced across four lists such that a list contained either 'beat'-'beat', 'beat'-'bit', 'bit'-'beat', or 'bit'-'bit'. In each trial, the first member of a pair was followed either by the same word or nonword or by its minimal pair, spoken by a different speaker. All lists contained a block of English words, followed by a block of English nonwords.

2.4.3 Procedure

Participants were assigned to one of the four lists according to their order of arrival. Participants were tested individually in a sound-proof room seated in front of a Dell Inspiron 600m laptop computer. DMDX was used to present stimuli and record each participant's response (Forster and Forster [2003]). At the beginning of each trial a fixation cross + appeared in the center of the screen to warn the participant that a stimulus was about to be presented. Next, participants heard two auditory stimuli separated by a lag of 1500ms. For each pair, the participant was asked to decide, as quickly and accurately as possible, whether or not the stimuli they heard were the same or different. If the pairs were the same, the participant was asked to respond by pressing the F key, if the pairs were different, the participant was instructed to press the J key. Participants were allowed 4000ms to make their response. Response time was measured from the onset of the stimulus. Six practice items (with no feedback) preceded the test trials. The AX discrimination task was comprised of two 5-minute blocks and lasted approximately 10 minutes total.

2.4.4 Data Analysis

The data from one native English-speaking participant was excluded from analysis due to non-compliance with the experimental task (i.e. failing to respond during the entirety of second experimental block). Observations for which RTs exceeding the allotted time (4000ms) were also excluded. These excluded time-out responses made up 1.3% of the entire dataset. Response Accuracy and Errors were subsequently analyzed separately for our control contrast /o-u/ and English-specific contrasts /ɑ/-/æ/ and /i/-/ɪ/ using Generalized Linear Mixed Effects Models with subject as random effect[5]. Analyses of these response variables were carried out because the more traditional analysis based in Signal Detection Theory was not feasible, due to the fact that the A' scores computed were not normally distributed due to a ceiling effect[6]. This can be observed from the average A' scores for words and nonwords in Figure 2.3 and Figure 2.4, respectively. Errors counts were calculated from the raw data by summing the number of 'misses' and 'false alarms' for each experimental condition. 'Misses' were instances in which the stimuli differed, but the participant failed to detect the difference and 'false alarms' where observations in which the stimuli were identical, but the participant inaccurately responded that

[5]Analyses of RT were also conducted. However, no significant effect of RT was found for any condition, nor did we have predictions regarding the RT variable prior to the experiment. Thus, these analyses are not reported here.

[6]Essentially, the AX discrimination task we conducted was too easy overall to serve as an effective signal detection task.

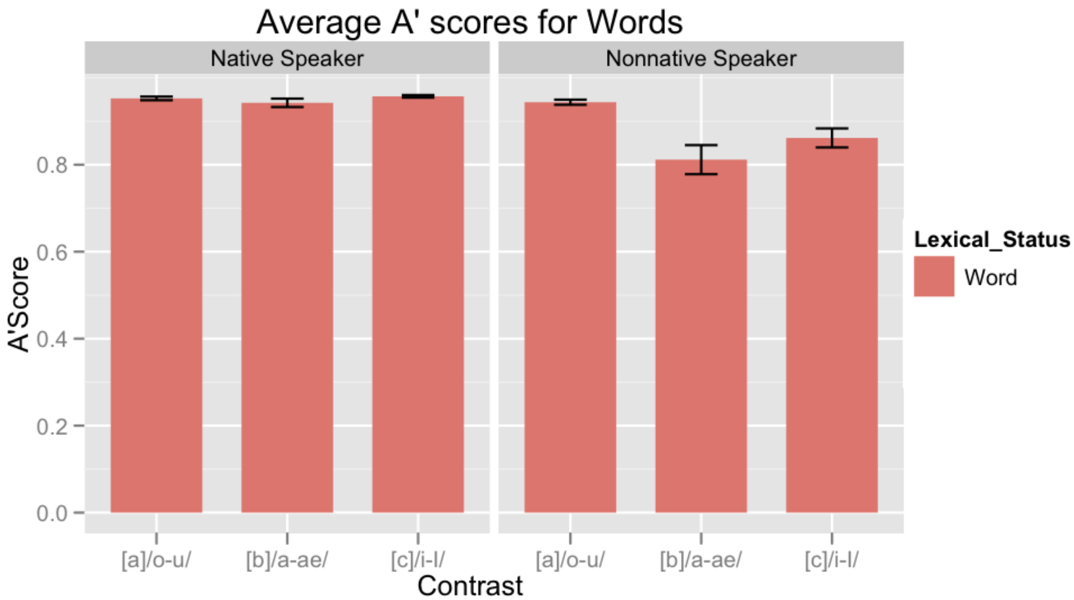

Figure 2.3: Average A' scores for words. Error bars represent standard error.

they were different.

2.4.5 Results

2.4.5.1 Response Accuracy

Statistical analyses were performed using GLMM binomial regression (using R package lme4) to assess the reliability of the effect of the experimental manipulation on response accuracy. Our analysis of response accuracy for the control contrast was comprised of fixed effects Language Group (native speaker vs. nonnative speaker), Lexical Status (word vs. nonword), Trial Type (same vs. different) and their interactions as fixed effects and subject as a random effect. In our analyses of the control condition we found a main effect Trial Type (z-value=-2.067, p=.0388), such that

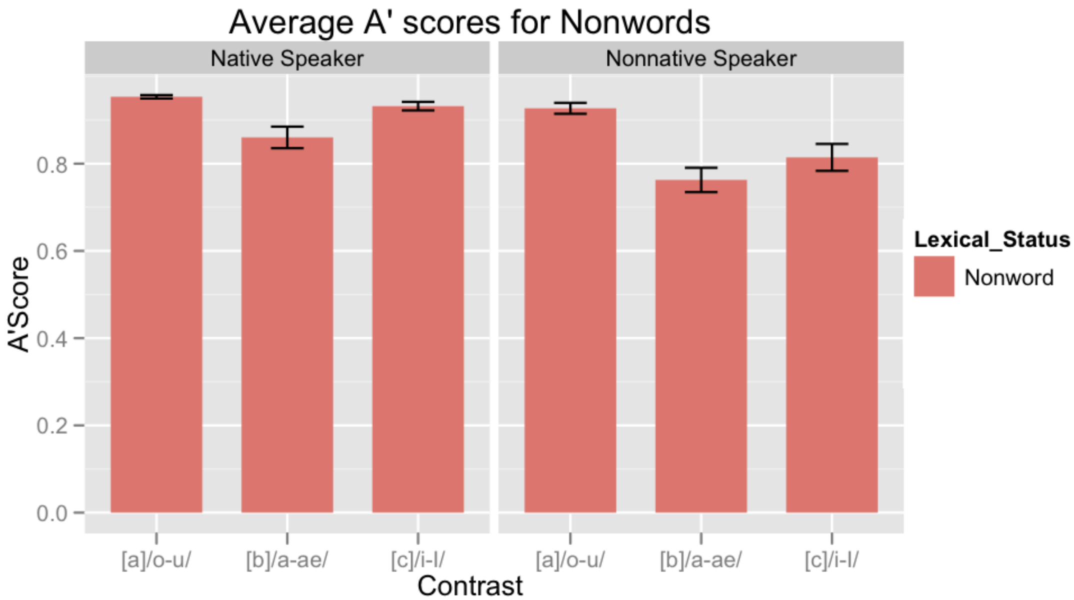

Figure 2.4: Average A' scores for nonwords. Error bars represent standard error.

participants were more accurate on same trials. Crucially, there was no main effect of Language Group (z-value=-1.302, p=.1931), nor did Lexical Status (z-value=0.368, p=.7129) or any of the interaction terms reach significance. This analysis confirms the effectiveness of our control condition for which no effect of Language Group was predicted. Thus, our subsequent analyses were restricted to the English-specific contrasts /i/-/ɪ/ and /ɑ/-/æ/.

Analyses of response accuracy for the English-specific contrasts were comprised of fixed effects Language group (native speaker vs. nonnative speaker), Contrast (/i/-/ɪ/ vs. /ɑ/-/æ/), Lexical Status (word vs. nonword), Trial Type (same vs. different), their interaction terms, and subject as a random effect. We again found a main effect of Trial Type (z-value=3.056, p=.0022), with same trials being more accurate than different trials. A significant effect of Language Group was also observed (z-value= -3.770, p=.0002). As expected, nonnative speakers were less accurate than native speakers on English-specific contrasts. The effect of Contrast was also significant (z-value=3.059, p=.0022). However, the effect was not in the predicted direction, as /ɑ/-/æ/ was found to be less accurate than /i/-/ɪ/. Also as expected, words were also more accurate than nonwords, (z-score=5.102, p <.0001). There was also a significant Language Group x Lexical Status interaction (z-value=-2.085, p=.0371), due to native speakers being less accurate on nonword stimuli. None of the other interactions reached significance.

2.4.5.2 Errors

Statistical analysis was performed using GLMM Poisson regression with subject as a random effect (using R package lme4) to assess the reliability of the effect of the experimental manipulation on Error count (with the expectation that the error analysis should confirm the accuracy analysis of the previous section). Analyses of errors for the common contrast /o/-/u/ were comprised of fixed effects Language Group (native speaker vs. nonnative speaker), Lexical Status (word vs. nonword), as well as the Language Group x Lexical Status interaction with subject as a random effect. A main effect of Language Group (z-value=2.486, p=.0131), was observed which was not found in the Response Accuracy analysis above. Analysis of errors for the English-specific contrasts were conducted with fixed effects Language Group (native speaker vs. nonnative speaker), Lexical Status (word vs. nonword), and Contrast (/i/-/ɪ/ vs. /ɑ/-/æ/), as well as the Language Group x Lexical Status interaction with subject as a random effect. Consistent with the response accuracy analysis, a main effect of Language Group (z- value=4.363, $p<.0001$), Lexical Status (z-value=-5.208, $p<.0001$), Contrast (z-value=-3.864, $p=.0001$), as well as a Language Group x Lexical Status interaction (z-value=3.142, $p=.0017$) were observed.

2.4.6 Discussion

In Experiment 1, we tested the sensitivity of a group of advanced L1 Spanish late-learners of English with the expectation that the group's performance would differ on the English-specific contrasts as a function of feature availability. This prediction was based on the strong version of Brown's redeployment hypothesis that states that, when available, features manipulated by the native grammar will be redeployed in the L2. Thus, we expected superior performance (i.e. more accurate responses, fewer errors) for the /ɑ/-/æ/ contrast relative to the /i/-/ɪ/ contrast. We found no empirical support for Brown's predictions.

In general the results accord with previous findings. We found a reliable difference between native speakers and nonnative speakers for the English-specific contrasts /i/-/ɪ/ and /ɑ/-/æ/, but not the control contrast /o/-/u/. This result was expected given the large body of literature reporting the difficulty that nonnative speakers experience with certain nonnative contrasts.

Unexpectedly, however, we found that the Spanish-English bilinguals' performance on the /ɑ/-/æ/ contrast was less accurate than on the /i/-/ɪ/ contrast. While this was an unexpected result, there is a possibility that this effect may have been driven by factors that were external to the experiment, namely, the fact there are a number of English words, such as the word 'pasta', in which the initial vowel may be pronounced as either [ɑ] and [æ] as a function of dialect. Alternations of this sort have been shown to reduce the contrastiveness of an otherwise contrastive pair (Huang and Johnson [2011], Johnson and Babel [2010]). We speculate that this may have contributed to the decrement in accuracy observed for the /ɑ/-/æ/ for nonwords in the native speaker group. While the exact nature of the effect observed for /ɑ/-/æ/will have to be taken up in future investigation, it is clear that one would want to be cautious in interpreting the relative accuracy of the two English-specific contrasts for the nonnative speakers given that lower accuracy was also observed for /ɑ/-/æ/ for the native speaker group for whom the contrast was not expected to be problematic. This by itself suggests that the concept of "similarity" is not sufficiently covered by distance or featural metrics alone, as Johnson and Babel [2010] also suggest.

2.5 Experiment 2

Experiment 2 uses a medium-lag repetition priming paradigm (Pallier et al. [2001]) to investigate whether advanced Spanish learners of English show evidence of having acquired distinct phonological representations for minimal pairs in the L2 lexicon. In this task, participants are asked to perform an auditory lexical decision task on one of four counterbalanced lists of English words and nonwords. Test words and nonwords were followed 9 to 23 items later by either an identical

Table 2.3: Summary of Predictions for Experiment 2

		Contrast		
Language Group	Condition	/o/-/u/	/ɑ/-/æ/	/i/-/ɪ/
Native Speakers	Repetition	yes	yes	yes
	Minimal pair	no	no	no
Nonnative Speakers	Repetition	yes	yes	yes
	Minimal Pair	no	no	yes

stimulus (i.e. *Repetition*, such as 'sheep'-'sheep') or a minimal pair (i.e. *Minimal Pair*, such as 'sheep'-'ship'). Each stimulus pair was distinguished minimally by either a common vowel contrast (i.e. /o/-/u/) or an English-specific vowel contrast (i.e. /ɑ/-/æ/ or /i/- /ɪ/). It was predicted that if Spanish speaking participants have distinct lexical representations for test pairs containing minimal contrasts, then facilitation effects should be observed for the Repetition condition, but not for the Minimal Pair condition. However, if minimal pairs containing nonnative contrasts have homophonous lexical representations, facilitation effects of approximately the same magnitude should be observed for nonnative contrasts in both conditions. Crucially, if feature availability predicts the acquisition of nonnative contrasts as Brown's FBM suggests, we should find minimal pair priming for /i/-/ɪ/, but not /ɑ/-/æ/. These predictions are summarized in Table 2.3.

2.5.1 Methods

2.5.2 Participants

Participants in this experiment were the same 28 Spanish speakers and 28 native English speakers who participated in the AX discrimination task described above. All participants performed the auditory lexical decision task prior to performing the AX discrimination task.

2.5.3 Stimuli

The stimuli used in the medium-lag repetition priming experiment were the same 192 experimental items (96 words and 96 nonwords) and 64 filler items used in the auditory AX discrimination task detailed above. Four counterbalanced lists of 256 stimuli were created. Each list contained one member of each test pair followed, 9 to 23 items down the list, by either itself or its minimal pair, spoken by a different speaker.

2.5.4 Procedure

Participants were assigned to one of the four lists according to their order of arrival. Participants were tested individually in a sound-proof room seated in front of a Dell Inspiron 600m laptop computer. DMDX (Forster and Forster [2003]) was used to present stimuli and record participants' responses. Participants responded by pressing one of two buttons on the keyboard. At the beginning of each trial a fixation cross + appeared in the center of the screen to warn the participant that a stimulus was about to be presented. Next, participants heard an auditory stimulus. For each stimulus, the participant was asked to decide, as quickly and accurately as possible, whether or not the stimulus they heard was an English word. If the stimulus was a word, the participant was asked to respond by pressing the F key, if the sequence was not an English word, the participant was instructed to press the J key. Following Pallier et al. [2001], stimuli were presented with an interstimulus interval of 2.5 seconds. Response time was measured from the onset of the stimulus. The test trials were preceded by six practice items to ensure that the participant was clear on the experimental task. No feedback was provided on either the practice or test items. The auditory lexical decision task lasted approximately 20-25 minutes and was broken into two 10-12 minute blocks separated by a self-timed break.

2.5.5 Data Analysis

The data from two nonnative speakers who performed at chance were excluded from analysis. Data points which fell within the predetermined low and high cutoff of 300ms and 2500ms were retained for an analysis of *'Repetition Effect'* (86.2% of the original data)[7]. Following Pallier et al., we define a repetition effect as a reaction time decrease between the first and second occurrence of an item or between the occurrence of an item and its minimal pair. Repetition effects were computed by subtracting the RT of the second occurrence of an item from the initial occurrence, which served as its baseline. Thus, a positive value indicated a priming effect. Figure 2.5 shows repetition effects for words, whereas Figure 2.6 shows repetition effects for in response to nonword stimuli.

[7]This cutoff was determined based on previous report that normal latencies for lexical decisions based on spoken words typically range from 500-1000msec (Almeida [2009]).

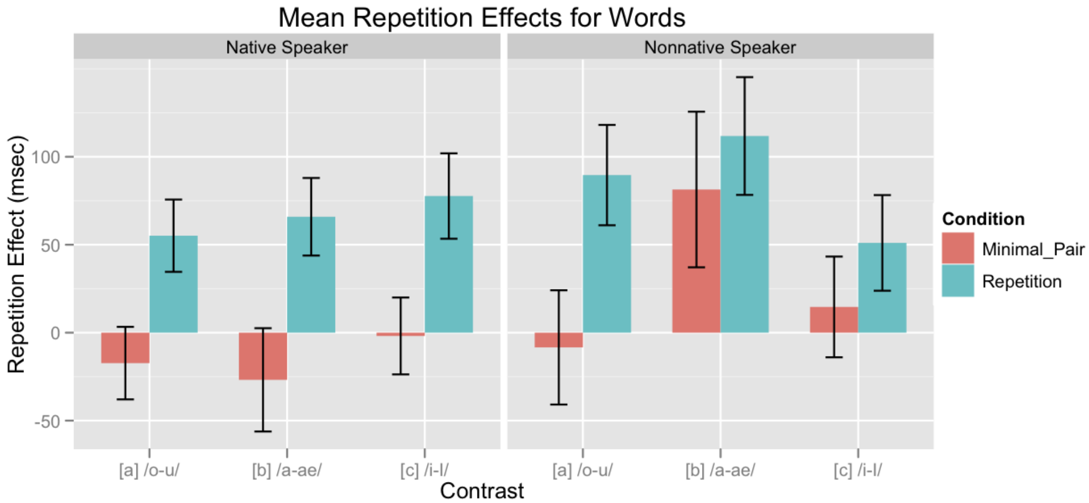

Figure 2.5: Mean repetition effects in response to word stimuli as a function of Language group (native vs. nonnative Speaker), Condition (repetition vs. minimal pair), and Contrast (/o/-/u/, /ɑ/- /æ/, and /i/-/ɪ/). Error bars represent standard error.

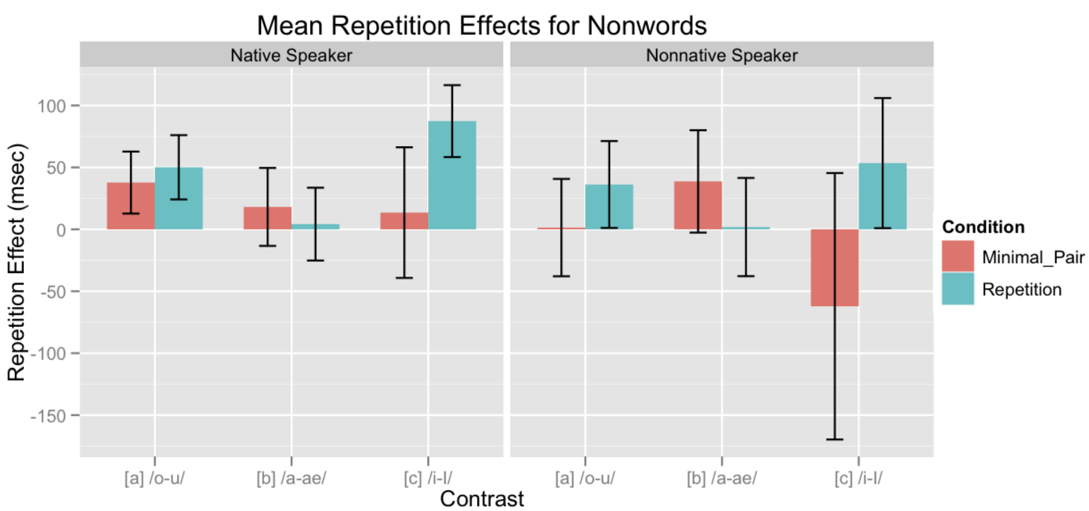

Figure 2.6: Mean repetition effects in response to nonword stimuli as a function of Language group (native vs. nonnative Speaker), Condition (repetition vs. minimal pair), and Contrast (/o/- /u/, /ɑ/-/æ/, and /i/-/ɪ/). Error bars represent standard error.

2.5.6 Results

2.5.6.1 Repetition Effects

Linear mixed effect modeling was employed to assess the reliability of the effects elicited by the experimental manipulations Baayen [2008]. Analyses of repetition effects for word stimuli for the control contrast /o/-/u/ comprised of fixed effects Language Group (native speaker vs. nonnative speaker), Condition (repetition vs. minimal pair), as well as the Language Group x Condition interaction and subject and item as random effects. A main effect of Condition (t=2.089, p=.0370) was observed. Repetition effects were larger for the repetition condition than for the minimal pair condition. Additionally, there was no effect of Language Group, nor did any of the interaction terms reach significance, suggesting that there was no difference between the native and nonnative group on their performance on common contrasts. This confirms the adequacy of the control condition for which no differences were expected.

Analyses of repetition effects for word stimuli for the English-specific contrasts were again comprised of fixed effects Language Group (native speaker vs. nonnative speaker), Condition (repetition vs. minimal pair), Contrast (/ɑ/-/æ/ vs. /i/-/ɪ/), as well as the Language Group x Condition x Contrast interaction and subject and item as random effects. A main effect of Condition (t=2.488, p=.0129) was again observed, with Repetition yielding more priming than Minimal Pairs, as expected. We also observed a main effect of Language Group (t=2.617, p=.0089), with nonnative speakers showing more priming overall. The Language Group x Contrast interaction did not reach significance (t=-1.676, p=.0939). No other significant effects were found.

Statistical analyses of nonword stimuli revealed no significant results for either the control or English-specific contrasts regardless of Language group, Contrast or Condition.

2.5.7 Discussion

Experiment 2 investigated whether advanced L1 Spanish late-learners of English have acquired distinct lexical representations for English minimal pairs which are distinguished by the English-specific vowel contrasts /ɑ/-/æ/ and /i/-/ɪ/ using a medium-lag repetition priming paradigm. We hypothesized that if learners are unable to acquire new features, but are able to redeploy existing features, the advanced Spanish late-learners of English should differ in their performance on these two non-native contrasts. Given that the Spanish learners' native grammar makes available the feature [$\pm$ back], but the language lacks a tense/lax distinction, the learners were expected to perform better (i.e. show less minimal pair priming) for the /ɑ/-/æ/ contrast relative to the /i/-/ɪ/ contrast. They did not. Although it is not statistically significant, the effect is in the wrong direction, so we find no support for Brown's hypothesis.

We replicate previous findings for nonwords and words that containing common contrasts. Like Pallier et al. [2001] we observe larger repetition effects for the repetition condition than the minimal pair condition for words containing contrasts which are common to both languages, but no reliable repetition effects for nonword repetitions or minimal pairs. Moreover, we observed no reliable difference between Spanish participants and native speaking controls for this control condition.

The results for words containing nonnative contrasts are less clear. Although we found a main effect of Condition and Language Group, we did not find a significant interaction between these factors. Therefore, we have little statistical support for claiming that nonnative speakers evidence minimal pair priming for nonnative contrasts. However, upon visual inspection of the data it appears that priming is found for the Spanish participants in the minimal pair condition for the /ɑ/-/æ/ contrast, whereas minimal pair priming is not found for the /i/-/ɪ/ contrast. Additional statistical tests will be required to confirm this possibility. However, it is important to point out that these results, if anything, would again run counter the predictions of Brown's model regarding the acquirability of nonnative contrasts. These results are also in contrast with previous findings from Pallier et al. [2001]. Both of these points will be taken up in greater detail in the section that follows.

2.6 General Discussion

Two experiments were conducted to investigate the phonetic perception and lexical representation of two English-specific vowel contrasts (i.e. /ɑ/-/æ/ and /i/-/ɪ/) by advanced Spanish late learners of English. These experiments were designed to explore the hypothesis that perception and acquisition of nonnative contrasts is contingent upon the availability of phonological features in the native grammar. In

Experiment 1, an AX discrimination task was employed to investigate the Spanish-participants' ability to discriminate minimally contrastive English words and non-words. In Experiment 2, we used medium-term repetition priming in an auditory lexical decision task to investigate word-recognition processes in the same participants and using the same stimuli. In both experiments it was expected that, if feature availability predicts the acquirability of nonnative contrasts by L2 learners, we should observe better performance (i.e. higher A'- prime scores in Experiment 1 and less priming in Experiment 2) for minimal pairs containing the /ɑ/-/æ/ contrast relative to the /i/-/ɪ/ contrast.

In both experiments we found that [1] the Spanish learners of English who participated in our study continued to differ from native speakers of English in their sensitivity to the two English-specific vowel contrasts which we tested. Yet, [2] these participants performed much better on both tasks than was expected based on previous findings Pallier et al. [2001] and the fact that the participants in the current study were late-learners [8]. On the AX discrimination task, the Spanish participants received A'-prime scores of .86 and .81 for the /i/-/ɪ/ and /ɑ/-/æ/ contrasts respectively. This is well above chance (.5) and much better than we would be expected from participants who were experiencing extreme difficulty with these contrasts. Moreover, contra Pallier et al. [2001], we did not observe reliable minimal pair priming by our Spanish learners of English for words differing in a single English-specific contrast.

[8]Note that better performance on the part of the Spanish-learners of English is surprising in light of the fact that early learners have been found to out-perform late learners in both perception and production tasks Flege and MacKay [2004]; Piske et al. [2002].

The first finding is consistent with previous results, since differences between native and nonnative speakers are ubiquitous in the literature. The second finding, however is much more interesting. The finding that our highly proficient Spanish late-learners of English performed better than the Spanish-dominant Spanish-Catalan bilingual participants in Pallier et al. [2001] might be taken to suggest better language separability on the part of the Spanish-English as opposed to Spanish-Catalan bilinguals. To test more directly whether this difference in performance is due to a language mode effect one could manipulate the language mode variable by including cognates in the list of English test words. If the presence of cognate words in the test stimuli caused the performance of the Spanish-Catalan bilinguals to be suboptimal in the study conducted by Pallier et al., then we should be able to coax Spanish-English bilingual participants into a more bilingual mode of processing and cause a similar performance decrement by introducing Spanish-English cognates among the English test words. Future research should address this possibility.

The finding that the Spanish participants in our study did not show reliable minimal pair priming for words containing difficult nonnative contrasts is also interesting in that it potentially tells us something about the nature of these bilinguals' lexical representations. In particular, it suggests that these learners cannot simply be assimilating two L2 phonemes to a single L1 phoneme as previously suggested to account for the Pallier et al. findings. That is, the Spanish participants cannot simply be representing /ʃip/ 'sheep' and /ʃɪp/ 'ship' as /ʃip/. This approach would be incompatible with our empirical findings. One possibility is that the Spanish learners in our study have begun to acquire one or more of the L2 categories; however, they do not always reliably categorize incoming sounds, such that minimal pairs containing the contrasts are perceived as homophones only some of the time. Yet another possibility is that whether or not Spanish learners have homophonous lexical representations for minimal pairs containing English-specific contrasts varies on a by-subject and a by-item basis, such that Subject 001 may have homophonous representations for ‘sheep’ and ‘ship’, but not ‘reach’ and ‘rich’. Our current data cannot distinguish between these possibilities.

With respect to the role of distinctive features in perception and acquisition, Experiment 1 and 2 were designed to test the predictions of Brown’s redeployment hypothesis, which states that when available phonological features employed by the native grammar will be redeployed in the L2. To investigate this, we manipulated feature availability, such that the native grammar of Spanish makes available the necessary features to represent /ɑ/-/æ/, but not /i/-/ɪ/. However, in both experiments, we found better performance for /i/-/ɪ/ than /ɑ/-/æ/(essentially the opposite pattern of results than were expected). The results of Experiment 1 and 2, therefore, provide converging evidence suggesting that the strong version of Brown’s redeployment hypothesis is likely to be incorrect. In particular, our results suggest that not only were available features not being redeployed as predicted, but our results are also compatible with new features having been acquired. In Chapter 3 we explore the extent to which a rulers approach can provide better coverage of the current data.

Finally, the finding that Spanish learners performed better on /i/-/ɪ/than /ɑ/-/æ/ is surprising given numerous accounts for the difficulty that Spanish learners typically experience with the /i/-/ɪ/contrast. This finding provides additional support for the view that new structures can be learned by late L2 learners (Curtin et al. [1998], LaCharité and Prévost [1999]). One possible explanation for the relatively good performance on this contrast is due to an effect of pedagogy, in the sense that Spanish-learners may have received considerable training on the /i/-/ɪ/ contrast. This, of course, would also set them apart from the Spanish-Catalan learners who presumably are exposed to their L2 in a more naturalistic environment. A second possibility is that we have the wrong analysis of Spanish vowels. Both of these possibilities should be taken up in future work. In particular, it will be important to collect data regarding the quantity and quality of training received for problematic nonnative contrasts and to consider the possible role of pedagogy in the acquisition of difficult nonnative contrasts.

2.7 Conclusion

In sum, this chapter we set out to use two different behavioral paradigms to investigate the nature of the representations that guide the processing of nonnative input, as well as the extent to which the influence of the native language can be overcome. In particular, we explored a particular instantiation of a Legos approach to similarity. We investigated whether phonological features are redeployed to represent non-native contrasts in the second language (Brown [1998], Brown [2000]). The results across both experiments converge to suggest that second language acquisition of phonology is not constrained by the phonological features made available by the learner's native language grammar, nor is the presence/use of particular phonological features in the native language grammar sufficient to trigger redeployment. These findings suggest that feature availability is neither a necessary, nor a sufficient condition to predict the observed learning outcomes.

Chapter 3

Ruler's Alone won't work

3.1 Introduction

As detailed in the previous chapters, we have seen that nonnative speakers and second language learners often experience difficulty producing and perceiving speech contrasts which are non-contrastive in their native language. Yet, not all nonnative contrasts prove to be equally challenging for second language learners (Brown [1998], *among others*). Instead, there is considerable evidence for gradient performance across various non-native contrasts (Best [1994, 1995]). Existing models of nonnative and second language phonology and speech perception provide various accounts of this phenomenon.

One approach to this asymmetry has been to take what I have been calling a 'Legos approach'. From a 'legos' perspective, phonological features are the representational building blocks which the learner may use to distinguish minimal contrasts. Moreover, nonnative contrasts differ in their perceptibility and acquirability depending on whether or not the learner's native language grammar employs the necessary phonological feature to distinguish words in his/her native language. For example, Japanese learners of English performed better on the English /b/-/v/ than /r/-/l/, despite having neither of these contrasts in the native language (Brown [1998]). In this view, nonnative segmental contrasts are predicted to cause difficulty when a representation for the input cannot be built from the active phonological features in the learner's native language (i.e. the existing building blocks). Alternatively, in the case that the necessary phonological features are active in the native language and may be redeployed to represent novel target language contrasts, the approach predicts relative perceptual ease and ultimate acquirability of a nonnative contrast.

An alternative approach, and one that has many instantiations in the cross-language speech perception literature is what I have been calling a 'Rulers Approach'. From a rulers perspective, acoustic-phonetic or articulatory-phonetic similarity can be used to predict the difficulty that nonnative speakers and L2 learners will encounter with nonnative contrasts. From a rulers perspective a nonnative or L2 sound is mapped to the nearest native language category. In this view, perceptual, and consequently, learning difficulty is predicted to arise when two different L2 sounds are "close enough" in some sense yet to be made precise, and are mapped to a single L1 category.

3.1.1 A prominent model of nonnative perception

A prominent approach to cross-linguistic speech perception has been to make predictions about the discriminability of diverse nonnative contrasts based on the way in which the input is perceptually assimilated to L1 sound categories. Best's Perceptual Assimilation Model (PAM, Best [1994, 1995]) aims to describe and account for patterns of cross-language perception. In particular, the model makes predictions regarding the *'naive listeners'*[1] performance in the perception of diverse nonnative contrasts.

The central premise of Best's PAM is that nonnative sounds are mapped onto a listener's native categories on the basis of articulatory similarities. According to the model, the discriminability of a given nonnative contrast is determined by the manner and degree to which the sounds can be assimilated to a listener's native categories. The PAM predicts that those nonnative sounds which are not assimilated to any L1 sounds (i.e. UNASSIMILATED) will be discriminated well, regardless of experience with the language[2]. Empirical support for this claim comes from studies which demonstrate adult English monolinguals' relatively good perception of Zulu clicks, which are presumably unlike any native sound category (Best et al. [1988]).

The PAM states that one of three patterns of assimilation is expected for nonnative sounds which are assimilated to native language sound categories. Nonnative contrasts can be assimilated as [A] SINGLE-CATEGORY (SC) ASSIMILATIONS, [B] CATEGORY-GOODNESS (CG) ASSIMILATIONS, AND [C] TWO-CATEGORY (TC) ASSIMILATIONS. Examples of each of these contrasts types will be provided in the next paragraphs. The discriminability of nonnative sounds that are assimilated to two distinct L1 categories (i.e. TWO-CATEGORY ASSIMILATIONS) is predicted to be quite good, more accurate than if the phones are assimilated to a single native language category. With respect to instances in which two nonnative sounds are per-

[1]Best and Tyler [2007] use the term *'naive listener'* to refer to an individual who is functionally monolingual. A naive listener can be contrasted with *'L2 learners'* who are in the process of acquiring a second language.

[2]Presumably because the perception of these sounds relies on systems which are not language-specific and, therefore, are not subject to the same kind of shaping as a result of language experience

ceived as instances of the same category. The model distinguishes between single category and category goodness assimilation and predicts superior discrimination of a pair of nonnative sounds which differ in degree of goodness of fit to an existing speech category (i.e. Category-Goodness Assimilation), relative to a pair of nonnative sounds which are judged to be equally good exemplars of a single L1 category (i.e. Single-Category Assimilation).

The Perceptual Assimilation Model's predictions about each of these assimilation types have been supported by numerous cross-language speech perception studies. For example, Best and Strange [1992] investigated the perceptual assimilation hypothesis by testing the identification and discrimination of three synthetic English approximant contrasts (i.e. /w-j/, /r-w/, and /r-l/) in syllable initial position by native English and native Japanese listeners. On the basis of the approximants' phonemic status and the articulatory/phonetic details of the nearest Japanese sound categories, the authors predict that the /w-j/ contrast should be perceptually assimilated as a two category assimilation, and therefore be discriminated very well by Japanese listeners. The /r-w/ contrast is predicted to pattern like a category-goodness assimilation with English /w/ being a better exemplar of a single native language category. In contrast, English /r/ and /l/ are predicted to be perceptually assimilated as a single category assimilation, since both are equally poor exemplars of a single Japanese category. As a result, the /r-l/ contrast is predicted to be discriminated poorly. Consistent with the predictions of the PAM, the authors found that Japanese listeners showed inconsistent categorization and poor discrimination of /r-l/. While Japanese listeners labeling and discrimination performance with the

/r-w/ contrast was still poorer than that of native listeners, performance on the contrast was still better than for the /r-l/ contrast[3].

Hallé et al. [1999] investigated the same English approximant contrasts by French listeners. On the basis of articulatory-phonetic considerations, the authors predict that /w-j/ will be a two category assimilation with corresponding good discrimination, whereas the /r-l/ and /r-w/ contrasts will be category goodness and single category assimilations, respectively. Consistent with the PAM predictions, the authors found that /r-w/ was the most difficult contrast in both identification and discrimination due to a predominantly /w/ identification. Poorer discrimination performance was also reported for /r-w/ than /r-l/ and /w-j/.

Additional support for the PAM predictions come from Best et al. [2001] who investigated the perception of the Zulu voiceless vs. voiced lateral fricative contrast (i.e. /ɬ/-/ɮ/), the voiceless aspirated vs. ejective velar stops (i.e. /kʰ/-/k'/), and plosive vs. implosive voiced bilabial stops (i.e. /b/-/ ɓ/) by English listeners, which were expected to be assimilated as Two-Category, Category-Goodness and Single-Category assimilations, respectively. As predicted, the authors observed the best discrimination performance for the two category assimilation, followed by the category goodness assimilation and single category assimilation (TC> CG >SC).

Several other cross-linguistic speech perception findings are also consistent with the predictions of the PAM. Polka [1991] reported better discrimination performance by native adult English listeners for four Hindi dental-retroflex stop contrasts

[3]Many others have reported poor discrimination for the /r-l/ contrast by Japanese listeners and so those findings are likewise consistent with the PAM predictions for *Single Category Assimilations*(Miyawaki et al. [1975]; Goto [1971]; MacKain et al. [1981], to name a few).

when these conformed to Two-Category assimilation types than Single-Category assimilations. Likewise, English listeners where better at discriminating Salish velar vs. uvular ejective contrast (i.e. /k'/-/q'/) than a Farsi velar vs. uvular stop contrast (i.e. /g/-/ G/), which correspond to a Two-Category and a Single-Category assimilation, respectively(Polka [1992]). Finally, studies conducted with bilinguals have also reported poor discrimination and limited plasticity for nonnative contrasts which are assimilated as Single-Category type assimilations in the native language (Pallier et al. [1997]; Bosch et al. [2000]; Weber and Cutler [2004]).

In so much as these assimilation patterns can and do make predictions about the discrimination behavior one would expect to observe with naive listeners upon encountering nonnative speech contrasts, it would be useful to try to formalize a model which can [a] capture these existing qualitative patterns and [b] could be use to make quantitative predictions about expected behavior of nonnative listeners from particular language backgrounds. These predictions could then be compared with actual behavior. Moreover, in as much as our model makes the right predictions about cross-linguistic speech perception by naive listeners, then we can begin to modify and extend the model in particular ways to capture aspects of L2 phonological development.

3.1.2 Goal of this chapter

In light of the fact that a strong version of the legos approach we pursued in the previous chapter appears to be inconsistent with our experimental results, the goal of this chapter is [a] to extend a computational model proposed by Feldman et al. [2009] to nonnative speech perception, and [b] to investigate whether a phonetic approach can better explain the performance of our advanced Spanish learners on the English-specific vowel contrasts. In particular, by comparing the predictions of the model with the behavior observed in our experiments we can determine the extent to which these learners are relying on knowledge of categories in their native language, and how their knowledge has changed to incorporate knowledge of the target language. The hope is that the results of this project will help us understand how people represent similarity among sounds and assess whether a more sophisticated story of a rulers flavor can provide better coverage of the data.

3.1.3 Outline of the chapter

The remainder of the chapter is structured as follows: In section 3.2.1, I introduce a computational model put forward by Feldman et al. [2009], followed by an overview of our model which extends Feldman et al.'s to multiple dimensions in section 3.2.2. In section 3.3.1, I provide a simple toy model which is intended to demonstrate that the model is capable of predicting the types of qualitative patterns described in Best's Perceptual Assimilation Model. In section 3.3.2, I demonstrate how the model can be used to predict the classification behavior one would expect to observe from native Spanish speakers when asked to identify English vowels in terms of their Spanish vowel categories. These predictions are then compared to the actual behavior of Spanish participants reported in a study conducted by Flege

[1991]. In section 3.3.3, we use the model to predict the discrimination behavior we should observe from nonnative Spanish listeners on the discrimination task involving English-specific vowel contrasts, with the goal of comparing these predictions to the behavior we observed by highly proficient bilinguals in the discrimination experiment described in the previous chapter.

3.2 A computational model of nonnative speech perception

3.2.1 Extending an existing model

Feldman et al. [2009] proposed a Bayesian model of speech perception. The basic idea behind the model is that during speech perception the listener infers not only category membership on the basis of his/her knowledge of native language phonetic categories, but also recovers acoustic detail regarding a speaker's target production from a noisy speech signal.

The listener, upon hearing a speech sound, S, must infer the target production, T, by making use of various sources of knowledge at his/her disposal. For the adult listener this includes knowledge of the phonetic categories of his/her native language, and knowledge that speakers tend to produce sounds near category centers/means. The model provides an account of the type of perceptual magnet effects that have been reported in the literature (Kuhl et al. [1992]; Kuhl [1991]; Grieser and Kuhl [1989]; Iverson and Kuhl [1995])[4], by formalizing the perceptual magnet effect as a bias toward the category mean. In the model the influence of each category is

[4] More recently, the model has been shown to accurately predicts perceptual data from stop consonants, and fricatives as well (Kronrod et al. [2012]).

weighted by the probability that it produced the speech sound. In inferring the acoustic target the listener must also take the category variance and speech signal noise into account. The model predicts that when the speech signal noise is high then the listener will rely less on the acoustic detail in the signal and more on prior knowledge of the phonetic categories in his/her language. On the other hand, when the meaningful variance is high then the listener will rely more on the acoustic detail in the signal.

The model that I overview below is an extension of the original Feldman et al. [2009] model. We make two main changes. First, given that the amount of the variance observed for vowel categories can vary quite dramatically, we drop this simplifying assumption to allow for unequal variances[5]. Second, we extend the original model, which was applied to vowel perception data that involved two speech categories (i.e. /e/ and /i/) along a single dimension (i.e. F1-F2 space), to multiple dimensions. This extension was necessary in order to take into account the acoustic dimensions which have been thought to be crucial for vowel perception (i.e. F1, F2, (F3, duration)). In the next section I overview the multidimensional model of which the original single dimensional model is a special case.

3.2.2 Overview of our extended model

In this model the listener's knowledge of phonetic categories is represented as Gaussian distributions centered around a category mean, μ_c, with some variance,

[5]This extension to the model was also made by Kronrod et al. [2012] to be able to handle voiced and voiceless stop consonants for which the assumption of equal variances could not be maintained.

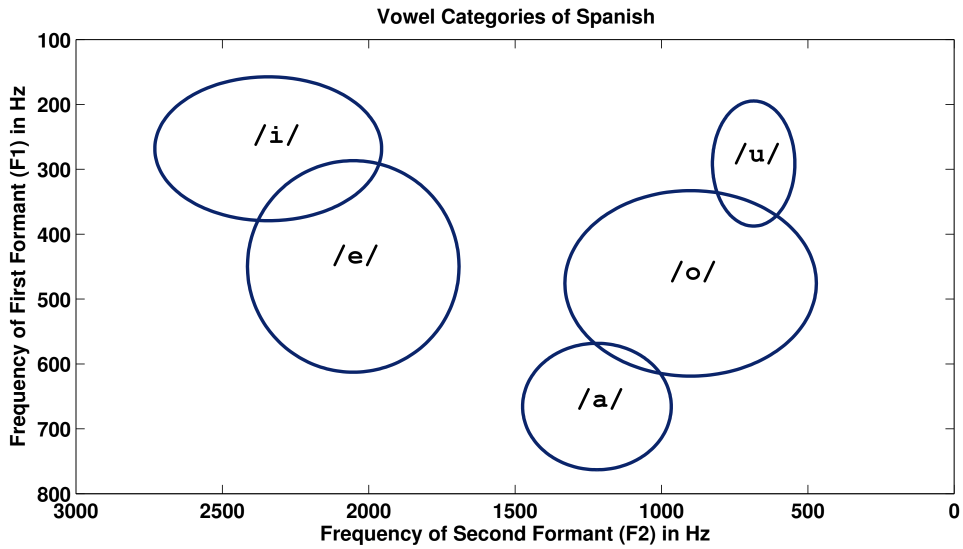

Figure 3.1: The five vowel system of Spanish plotted in F2 x F1 space. Ellipses represent the 95% confidence interval given the variance of each of the categories.

Σ_c. The category mean in the multidimensional case is a vector with values for each of the relevant dimensions. The category variance is a diagonal covariance matrix that provides information about the amount of variability there is around the mean in each of the dimensions. For example, a Spanish listeners' knowledge of Spanish /i/ produced by male speakers is represented as having a mean F1 of 268Hz and a mean F2 of 2342Hz (Quilis and Esgueva [1983]). The variance of this category is a diagonal covariance matrix which provides information about the amount of variability there is around the mean in both the F1 and F2 dimensions, 45^2 and 158^2, respectively(Quilis and Esgueva [1983]). It should be mentioned that it is unclear that our treatment of covariance as a diagonal covariance matrix is the appropriate one, but it is a simplifying assumption which we have made on the basis of the summary statistics for the materials we had access to, which lacked additional data about the way in which these dimensions covary. Figure 3.1 shows the full set of Spanish vowels with category means averaged over productions from 16 male native speakers of Peninsular Spanish(Quilis and Esgueva [1983]) plotted in F2 by F1 space. The ellipses represent the 95% confidence interval given the variance of each category (i.e. +/- 2 SDs of the mean). Thus, one can see that the category /u/ varies considerably in F1 and less so along the F2 dimension. This category variance can be thought of as meaningful variance introduced by co-articulation.

In this model, the speaker selects a target production, T, from a phonetic category, c, that has a mean, μ_c, and a variance, Σ_c. The target production, T, which the listener is trying to infer is also subject to an additional source of noise, speech signal noise, Σ_s. This statistical model can be written as:

$$T \mid c \sim \mathcal{N}(\mu_c, \Sigma_c) \tag{3.1}$$

$$S \mid T \sim \mathcal{N}(T, \Sigma_s) \tag{3.2}$$

Integrating over T will give you

$$S \mid c \sim \mathcal{N}(\mu_c, \Sigma_c + \Sigma_s) \tag{3.3}$$

The statistical model given in (3.3) captures the assumption that a stimulus observed by the listener is normally distributed around the category mean, μ_c, with a variance that is the sum of the category variance, Σ_c, and the noise variance, Σ_s.

Given this generative model, it is possible to model both identification and discrimination behavior. We assume that in an identification task listeners are sampling from a posterior distribution over category membership, $p(c \mid S)$. The $p(c \mid S)$ corresponds to the probability that a stimulus came from any given category and can be computed by applying Bayes' rule as follows:

$$p(c \mid S) = \frac{p(S \mid c)p(c)}{\sum_c p(S \mid c)p(c)} \tag{3.4}$$

Two values are needed to compute this posterior probability, $p(c \mid S)$, namely [i] $p(c)$ and [ii] $p(S \mid c)$. The $p(c)$ corresponds to the listener's prior belief in a given category before having heard a stimulus, S. This value can be estimated by the token or type frequencies of a category in a particular language. For our purposes, we make the simplifying assumption that each vowel category in a given language is equally likely. Thus, the prior in each of our simulations is set to 1/n, where n= the number of native language vowel categories.

The $p(S \mid c)$ is the likelihood of the stimulus under the assumption that it was generated by a given category. This probability distribution is normally distributed around the mean of a category, μ_c, with a variance which is the sum of the category and noise variance, $\Sigma_c + \Sigma_s$, as given in the statistical model in equation (3.3). This probability can be computed by summing over all possible target sounds, $p(S \mid c) = \int p(S \mid T)p(T \mid c)dT$.

The posterior probability of a sound belonging to a given category, $p(c \mid S)$ is computed by applying Bayes' Rule (equation (3.4)), by multiplying the likelihood, $p(S \mid c)$ and prior, $p(c)$, and then normalizing by taking the marginal probability of the data given all possible categories. This posterior distribution must be computed separately for each of the categories.

Following Feldman et al. [2009] we assume that the listener also uses Bayes' rule to infer the acoustic detail of the speaker's target production. In doing so, the listener should take each of the native language categories into account, but should weight their influence by the probability of each having generated the observed speech sound. This involves computing the posterior distribution on target productions, $p(T \mid S)$, by marginalizing over each of the categories as shown in (3.5).

$$p(T \mid S) = \sum_{c} p(T \mid S, c) p(c \mid S) \tag{3.5}$$

Note that the second term on the right-hand side of the equation is the posterior distribution over category membership, $p(c \mid S)$, already computed in (3.4). The first term on the right-hand side of the equation, $p(T \mid S,\ c)$ corresponds to the posterior distribution over T computed by assuming that it comes from category c, as in equation (3.6).

$$p(T \mid S,\ c) \propto p(S \mid T) p(T \mid c) \tag{3.6}$$

Recall that $p(S \mid T)$ and $p(T \mid c)$ are both Gaussians and are distributed as described in the statistical models given in (3.2) and (3.1), respectively. Substituting these values yields:

$$p(T \mid S,\ c) \sim \mathcal{N}((\Sigma_c^{-1} + \Sigma_s^{-1})^{-1}(\Sigma_c^{-1}\mu_c + \Sigma_s^{-1}S), (\Sigma_c^{-1} + \Sigma_s^{-1})^{-1}) \tag{3.7}$$

The posterior distribution over T computed assuming that it comes from a particular category, $p(T \mid S,\ c)$, results in a single Gaussian solution for that category.

The posterior distribution on target productions, $p(T \mid S)$, computed using the equation in (3.5), is a mixture of Gaussians where each Gaussian distribution represents a solution for one of the categories under consideration, $p(T \mid S,\ c)$, computed as in (3.6).

In inferring T, we assume that the listener computes the mean of the posterior distribution for the target production given the speech sound (i.e. $p(T \mid S)$), which corresponds to $\mathrm{E}[T|S]$ and can be computed using the equation in (3.8).

$$E[T \mid S] = \sum_c (\Sigma_c^{-1} + \Sigma_s^{-1})^{-1}(\Sigma_c^{-1}\mu_c + \Sigma_s^{-1}S)p(c \mid S) \tag{3.8}$$

The listener's best estimate of T then is a weighted average of the speech sound, S and the means μ_c of each of the categories that could have generated that speech sound. That is, each category makes a contribution to the mean of this posterior probability, but the contribution of a particular category mean is proportional to the $p(c \mid S)$. Each category mean biases perception toward the category center, where the strength of the bias is determined by the relationship between the meaningful category variance, Σ_c, and the noise variance, Σ_s. As a consequence of μ_c being weighted by Σ_s, the more noise in the signal, the more the listener will rely on his/her knowledge of phonetic categories. Similarly, S is weighted by Σ_c. This has the consequence that the greater the meaningful variability, the more weight the listener places on the acoustic detail.

The estimated value of T, $E[T \mid S]$, can be used to model discrimination behavior. According to the model, discriminating two stimuli comes down to inferring T for the pair and then comparing the values to see if the intended target productions were the 'same' or 'different', given some criteria for equivalence classification. The greater the distance between Ts, the more likely the stimuli will be 'different'.

3.3 Simulations

3.3.1 A toy model using a single dimension

To test whether our model is capable of making the types of qualitative predictions described by Best's PAM, we first created a toy model in which acoustic stimuli were drawn from one of five categories which varied along multiple dimensions (i.e. F1 and F2). For the purpose of this simulation, we set up a simple vowel system with five vowel categories (approximately /i, e, a, o, u/), with a mean F1 of 300, 500, 700, 500, and 300Hz and a mean F2 of 2500, 2000, 1500, 1000, and 500Hz, respectively. The category variance along the F1 and F2 dimensions was set to 75^2 and the off-diagonal values were set to 0. The noise variance is a symmetrical covariance matrix with a value of 100^2 along the diagonal. The stimulus, S, is a vector of values from 100 to 300Hz F1 and 2500 to 1500Hz in the F2 dimension.

Figure 3.2 shows the predicted relationship between acoustic and perceptual space in the case of multiple vowel-like categories plotted in an F2 x F1 space. As can be seen from Figure 3.2, although the stimuli are equidistant in acoustic space, their distance in perceptual space is warped. The figure shows that our multidimensional

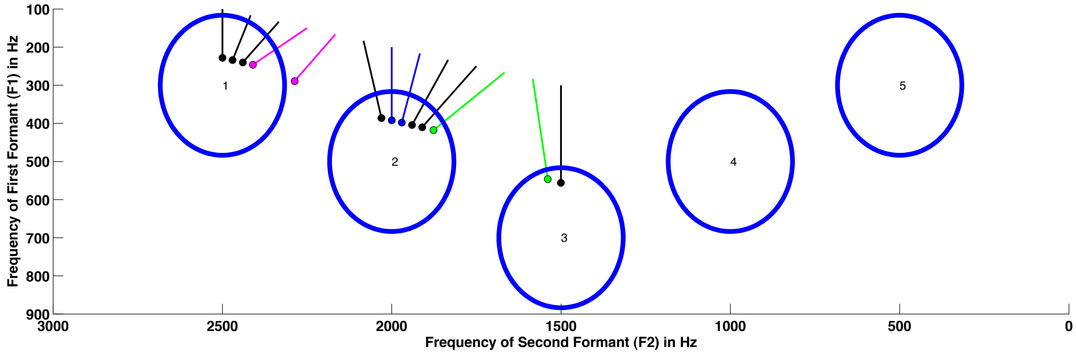

Figure 3.2: Predicted relationship between acoustic and perceptual space in the case of multiple vowel categories. Ellipses represent the 95% confidence interval given the variance of each of the categories. The tail of the line segments represents the "raw" production, and the circle represents the recovered value for the target production. The stimulus pair shown in blue correspond to Best's *Single-Category Assimilation*. The pair in magenta corresponds to a *Category-Goodness Assimilation*. A *Two-Category Assimilation* is shown in green.

model can produce the qualitative pattern of perceptual assimilation described by Best [1994, 1995]. The stimulus pair shown in blue are both equidistant from a single equally likely phonetic category (which centers around a mean F1 of 500Hz and mean F2 of 2000Hz) and so are pulled toward the mean of that category by roughly the same extent. The blue pair are thus perceived as equally good exemplars of the same phonetic category and can be said to correspond to Best's *Single-Category Assimilation*. The pair in magenta are both perceived as exemplars of the same phonetic category (which centers around a mean F1 of 300Hz and F2 of 2500Hz), yet the degree to which they are perceived as good exemplars of that category differs with one of the two members of the stimulus pair being a better exemplar of the category than the other. The pair corresponds to a *Category-Goodness Assimilation*. A *Two-Category Assimilation* is shown in green where a pair of acoustic stimuli are assimilated to one of two different native language categories.

It is worth mentioning that one limitation of our model is that it currently cannot account for Best's "non-assimilable" and "non-categorizable" patterns. This is because under the current model all sounds are drawn toward existing category centers. That is, there is no mechanism by which a stimulus will be considered too unlike other sounds to be assimilated or categorized as an existing sound. Ultimately the model may be extended to accommodate these patterns[6], but here it is our priority to explore the extent to which the current model can account for the Single-Category, Category-Goodness, and Two-Category patterns of perceptual assimilation described by Best.

3.3.2 Modeling existing identification data

As mentioned earlier, the model overviewed in 3.2.2 can be used to model identification data. We assume that in an identification task the listener is sampling from the posterior distribution over category membership $p(c \mid S)$. As a first approximation to real nonnative speech perception data we compared the model's predictions for identification with the classification patterns reported for English vowels by monolingual Spanish speaking listeners in a study conducted by Flege [1991].

Flege [1991] investigated interlingual identification by having Spanish speakers with varying degrees of English experience label tokens of English vowels as instances of their native language categories (i.e. /i, e, a, o, u/) or "none" if the vowel was not

[6]Learning of new categories could be modeled as a dirichlet process process (Navarro et al. [2006]; Ferguson [1973]).

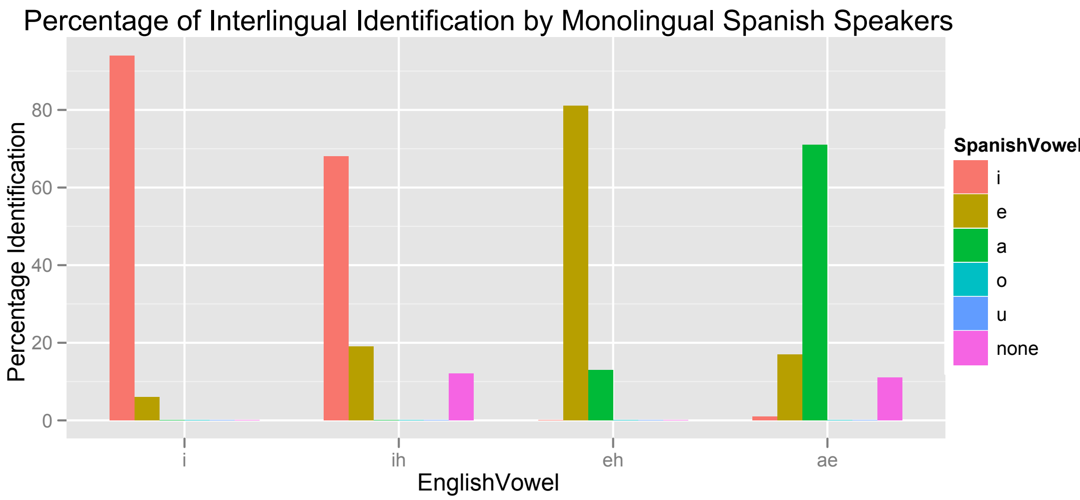

Figure 3.3: Percentage of Interlingual Identification by Monolingual Spanish Speakers. Reprinted from Flege [1991].

heard as a familiar Spanish one[7]. The author found that the majority of tokens of English /i/ were identified as Spanish /i/ (94% of the time). English/ɪ/ identification was more divided with English/ɪ/ being identified as Spanish/i/ (68% of the time), Spanish /e/ (19% of the time), and "none", (12% of the time). English /ɛ/ were divided between being labelled as Spanish /e/ (81% of the time) and Spanish /a/ (13% of the time). English /æ/ was more often identified with Spanish /a/ (71%), but also labelled as Spanish /e/ (17% of the time), "none" (11% of the time) and /i/ (1% of the time). Figure 3.3 shows the interlingual identification reported by Flege [1991] for monolingual Spanish listeners.

In order to determine whether the interlingual identification patterns observed by Flege [1991] could be explained by the model, we used the mean F1 and F2, μ_c, and category variances, Σ_c, reported by Quilis and Esgueva [1983], as in Figure

[7]It should be mentioned that our model has no means of predicting "none" responses, again since each category mean biases perception toward the category center. This is another way in which the model could be extended in future work.

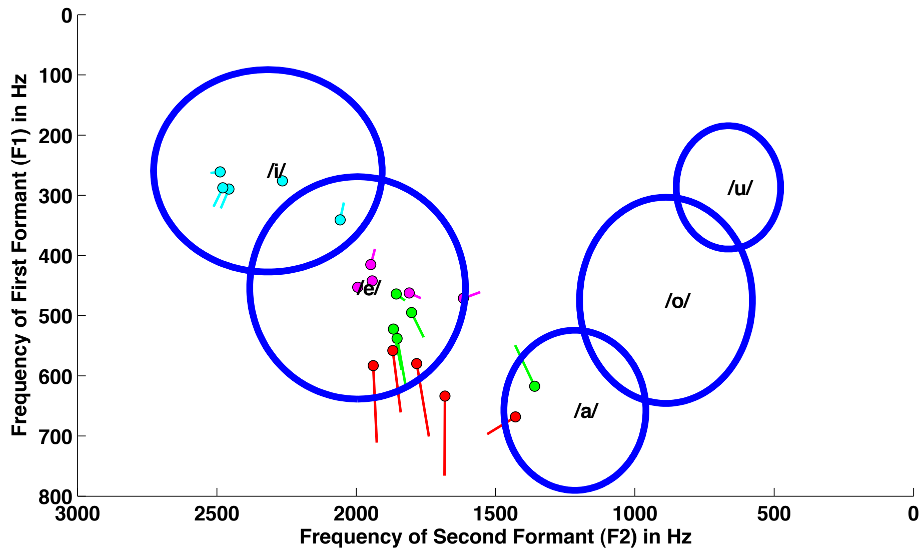

Figure 3.4: Predicted relationship between acoustic and perceptual space for subset of stimuli produced by male speakers. The tail of the line segments represents the "raw" production, and the circle represents the recovered value for the target production. The points shown in light blue correspond to tokens of English /i/, the points shown in magenta to tokens of English /ɪ/, those in green correspond to English /ɛ/, and those in red to English /æ/.

??. These values were measured from productions by 16 male and 6 female native speakers of Peninsular Spanish. Again, the prior, p(c) for this simulation is set to 1/n, where n= the number of native language vowel categories (i.e. 5). The noise variance was a diagonal covariance matrix with the variance set to 75^2. The speech sound, S, was a vector of values measured from the stimuli used by Flege [1991] split by gender.

Figures 3.4 and 3.5 show the predicted relationship between acoustic and perceptual space separately for the subset of the stimuli which were produced by male speakers and those produced by female speakers. These figures are useful because they provide a visual of the categorization predicted by the model, as the estimated value of T is expected to be pulled toward the most likely category to have produced

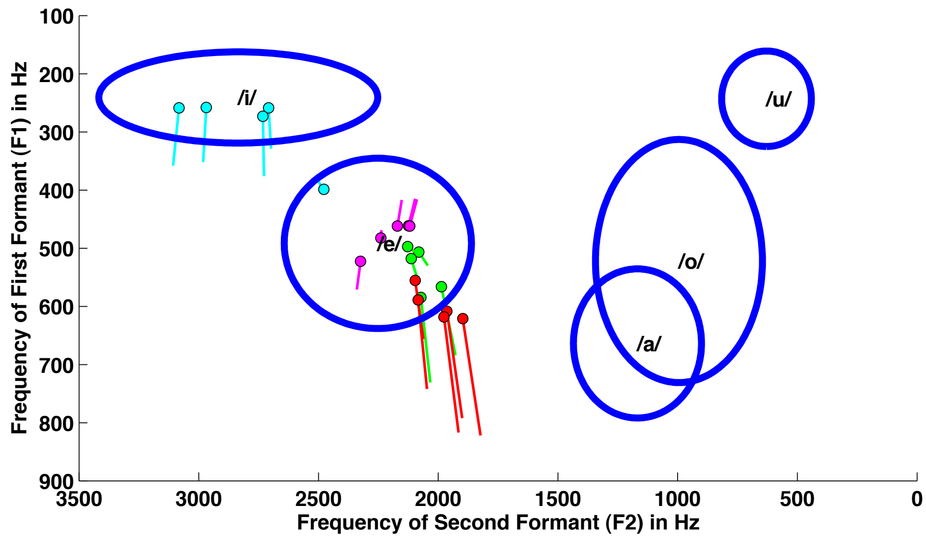

Figure 3.5: Predicted relationship between acoustic and perceptual space for subset of stimuli produced by female speakers. The tail of the line segments represents the "raw" production, and the circle represents the recovered value for the target production. The points shown in light blue correspond to tokens of English /i/, the points shown in magenta to tokens of English /ɪ/, those in green correspond to English /ɛ/, and those in red to English /æ/. Note that none of the /æ/ productions were identified as /a/ for female tokens.

the observed stimulus S, the category with the greatest p($c|S$). Thus, the direction of warping is indicative of the classification of each stimulus. The points shown in light blue correspond to tokens of English /i/, the points shown in magenta to tokens of English /ɪ/, those in green correspond to English /ɛ/, and those in red to English /æ/. As can be seen for Figure 3.4, productions of English /i/ were most often identified by the model as Spanish /i/, whereas the majority of tokens of English /ɪ/, /ɛ/, and /æ/ were identified with Spanish /e/ although with greater or less probability. The model produces a very similar pattern of results for female English productions as can be seen in Figure 3.5.

Figure 3.6 summarizes the models predictions for interlingual identification. The figure shows the average p($c|S$) for each of the five Spanish vowel categories

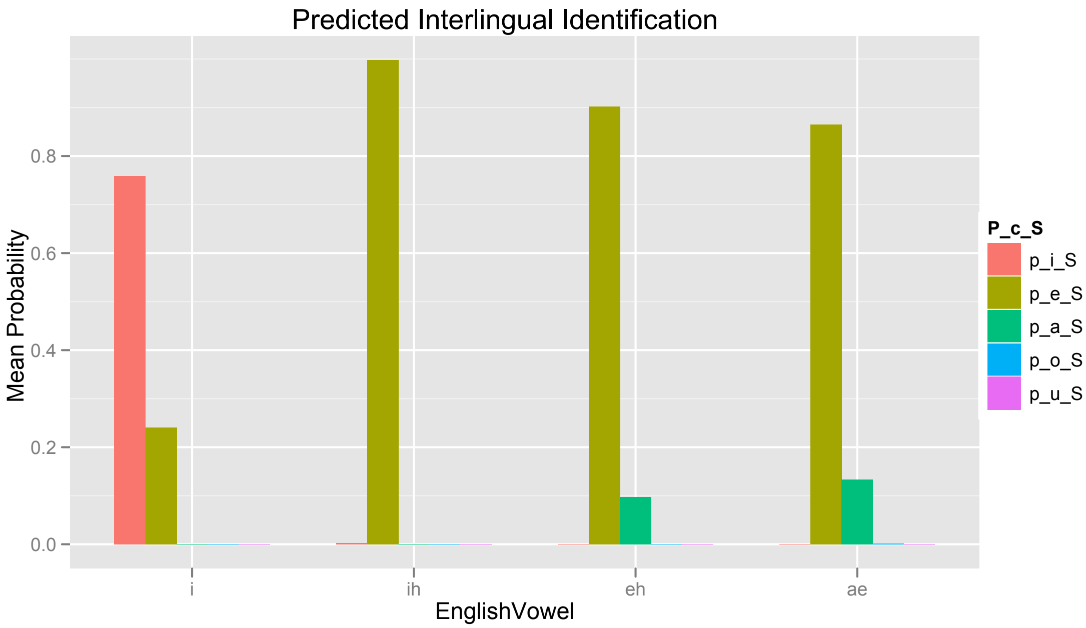

Figure 3.6: Predicted Interlingual Identification.

for each of the for English stimulus types /i, ɪ, ɛ, and æ/. While the model's predictions for tokens of English /i/ and English /ɛ/ are qualitatively similar, the model's predictions for English /ɪ/ and /æ/ are quite different from the observed behavior by monolingual Spanish speakers in Flege's experiment. In particular, while the monolingual participants who took part in Flege's study were more likely to identify tokens of English /ɪ/ with Spanish /i/ and English /æ/ with Spanish /a/, the model predicted that both sets of English tokens should be identified as tokens of Spanish /e/. The reason for this can be seen most clearly from Figure 3.4 and 3.5. In both cases tokens of English /ɪ/ tend to be near prototypical exemplars of the Spanish /e/ category (at least when the acoustic dimensions used are F1 and F2). The English /ɛ/ tokens are likewise generally quite good tokens of Spanish /e/. English /æ/ tokens which fall in an unused portion of the Spanish vowel space, are drawn toward the nearest category (in most cases Spanish /e/).

How can we explain this mismatch between model predictions and actual behavior? One possibility is that the values for the category means and variances, μ_c and Σ_c, are not very typical or are not a good match to the participants' representation of their native categories. However, when we compared these means and variances to other standard values reported in the literatures (i.e. Bradlow [1995], Godinez Jr [1978], Martínez Celdrán [1995]) we did not find dramatic differences. In particular, the values did not vary in the respects that would be necessary in order to capture the classification behavior observed in Flege [1991]. In order for the model to capture the classification behavior, it should be the case that the tokens of English /ɪ/ falls nearest to Spanish /i/ and English /æ/ falls nearest to Spanish /a/. But this was not the case, regardless of the source of the acoustic values.

Another possibility is that the dimensions we chose to use in our modeling (namely, F1 and F2) are not the ones that listeners are using in categorizing vowels tokens. It is worth mentioning, however, that simply adding information about additional dimensions (i.e. F3 and duration) does not appear to be enough to solve the problem. We also ran the model in which F3 was added, as well as F3 and duration and found a very similar pattern of results.

Another parameter that contributes to the posterior distribution that we assume the listener is sampling from in an identification task is the prior probability of the category, p(c). As mentioned earlier, we have made the simplifying assumption that the prior probability of the category is equal across categories. If instead the prior probability of the Spanish category /e/ were substantially less probable than the category /a/, for example, it is possible that the model would classify those tokens currently identified as /e/ as Spanish /a/. However, it is unlikely that even a small prior for Spanish /e/ will result in the desired classification predictions for exemplars of English /ɪ/, since these are (as I mentioned above) very good exemplars of the Spanish /e/ category. Furthermore, even if this were possible, it would have an undesired result when it comes to the model's predictions for English /ɛ/, since these tokens would no longer be identified as tokens of Spanish /e/.

In sum, it seems that the model at present does a poor job of predicting monolingual Spanish participants' behavior when asked to identify exemplars of English /i, ɪ, ɛ, æ/ as instances of their Spanish vowel categories.

3.3.3 Modeling existing discrimination data

The model overviewed in 3.2.2 can also be used to model discrimination behavior. As mentioned above, one goal of extending the computation model proposed by Feldman et al. [2009] to nonnative speech perception by L2 learners was to make explicit connections between category knowledge and observed patterns of perception. In this simulation, we use the model to make quantitative predictions about the pattern of discrimination behavior that one may expect to observe at the initial stage of L2 learning for the English /i/-/ɪ/ contrast[8], given learners' knowledge of their L1 sound categories. We then compare the predicted d' scores with the empirical d' scores computed from participants' same/different responses to experimental

[8] We restricted our modeling to a single English specific vowel contrast for which it is easiest to measure the steady state portion of the vowel and extract F1 and F2 values.

stimuli reported in Barrios et al.[In preparation]. By comparing these predictions to the behavior of highly proficient L2 learners in our experiment, we hope to determine the extent to which these learners are relying on knowledge of categories in their L1 and how their knowledge has changed to incorporate knowledge of their L2.

In the original Feldman et al. [2009] model discrimination is modeled as a comparison of inferred Ts. It is predicted that given a substantial perceptual distance the 'listener' will be able to discriminate the stimuli. If instead two stimuli are too close to one another in perceptual space, then the pair will be difficult to discriminate. This discrimination model is appropriate in cases where the listener's task involves discrimination and where it is reasonable to assume that the listener is trying to infer an acoustic target. However, recall that in the AX discrimination task used by Barrios et al. [In preparation], the experimental stimuli used in each trial were produced by a male and a female speaker in succession, and therefore the listener cannot infer a simple target acoustic production without taking into account gender factors. Thus, this discrimination model is not appropriate. Instead, we assume that listeners are comparing the posterior distribution over all known categories (i.e. the p($c|S$)) for both stimulus A and stimulus X. Listeners are more likely to respond 'same' when the $p(c \mid S)$ is greatest for the same category for both stimulus A and stimulus X.

In order to compute the model's predicted d' scores based on Spanish-like and English-like category knowledge, we first compute the posterior probability of category membership given a stimulus, p($c|S$) using equation(3.4), for both stimulus

A and stimulus X for each experimental trial given English and Spanish priors. The English category parameters were set to the adult male and female means and variances reported in Hillenbrand et al. [1995]. The Spanish values again come from Quilis and Esgueva [1983]. Again, we make the simplifying assumption that each of the known categories were equally likely. The prior, p(c) was set to 1/5 and 1/12 for Spanish and English respectively. The noise variance was fit to the English data so as to maximize the likelihood. The noise variance, Σ_s was a diagonal covariance matrix with the values 95^2 and 427^2 along the diagonal. We chose to fit to the English data since we expected the model to make the most accurate predictions in this case. It is worth mentioning that very similar values were obtained when the noise variance was fit based on the Spanish listener data (i.e. 92^2 and 566^2).

We then use this these probabilities (i.e. $p(c_A \mid S_A)$ and $p(c_X \mid S_X)$) to compute the expected proportion of same responses, where

$$proportionsame = \sum_i p(c_A = i \mid S_A) p(c_X = i \mid S_X) \tag{3.9}$$

The *proportionsame* corresponds to the probability that both stimulus tokens are assigned to the same category i, and is the sum over all possible values of i. The result corresponds to the proportion of same responses expected for an experimental trial (i.e. pairing of stimulus A and X) given knowledge of Spanish or English phonetic categories. The proportion same was computed for each of the experimental trials. We averaged the proportion of different responses (i.e. 1 – proportion same) separately, for all 'same' trials (i.e. trials in which stimulus A and X were

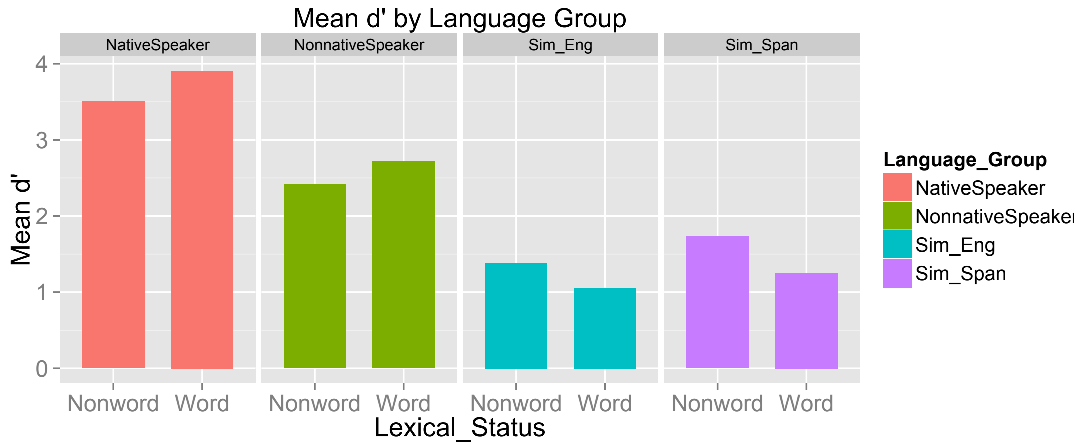

Figure 3.7: Model and empirical d' scores for Spanish and English Speakers

actually the same) and 'different' trials (i.e. trials in which stimulus A and X were actually different) in order to get the false alarm rate and hit rate, respectively[9]. These outputs were then used to compute d' scores according to Macmillan and Creelman [2005]'s Same-Different (Differencing model) and were compared to the data reported in Barrios et al. [In preparation].

The empirical d' scores from Barrios et al. [In preparation] and the predicted d' scores given knowledge of the phonetic categories of Spanish and English are shown in Figure 3.7 below.

3.4 Discussion

In this section, we set out to simulate the discrimination behavior of monolingual Spanish and English participants on an AX discrimination task involving the

[9] An alternative is to make the model deterministic by using a simple .5 rule to decide whether each trial is a hit, miss, false alarm or correct rejection. It is worth mentioning that conducting the analysis in this way leads to very similar results for this simulation.

English-specific /i/-/ɪ/ contrast, given native-like knowledge of Spanish or English vowel categories. We hoped that, by comparing the model's predictions to the actual behavior of the advanced Spanish learners of English who participated in our study, we would be able to access the degree to which these learners are relying on knowledge of their native language categories.

We expected that the model would make reasonably accurate predictions with respect to English participants' discrimination performance since the experimental stimuli were presumably tokens sampled from the same distributions used by the model for the purpose of classification. The classification based on Spanish categories only was expected to be less accurate since Spanish lacks an /ɪ/ category altogether and, while it has an /i/ category, it is unlikely that the distribution of the category overlaps completely with the English /i/ category.

Similarly, discrimination based on knowledge of English categories is expected to be quite good, whereas it is less clear how accurate discrimination behavior based on knowledge of Spanish categories would be. If tokens of English/ɪ/ are reliably categorized as Spanish /i/, discrimination is expected to be poor. If, on the other hand, exemplars of English /ɪ/ are classified as Spanish /e/, discrimination may in fact be quite good, although for the wrong reason.

As can be seen from Figure 3.7, the model predicts poor discrimination based on both Spanish and English-like category knowledge (shown in purple and blue, respectively). In contrast with our expectations, our simulations produced d' scores which are less than 1 for both Spanish and English. Moreover, the predicted d' scores for Spanish were slightly higher than d' scores based on English category

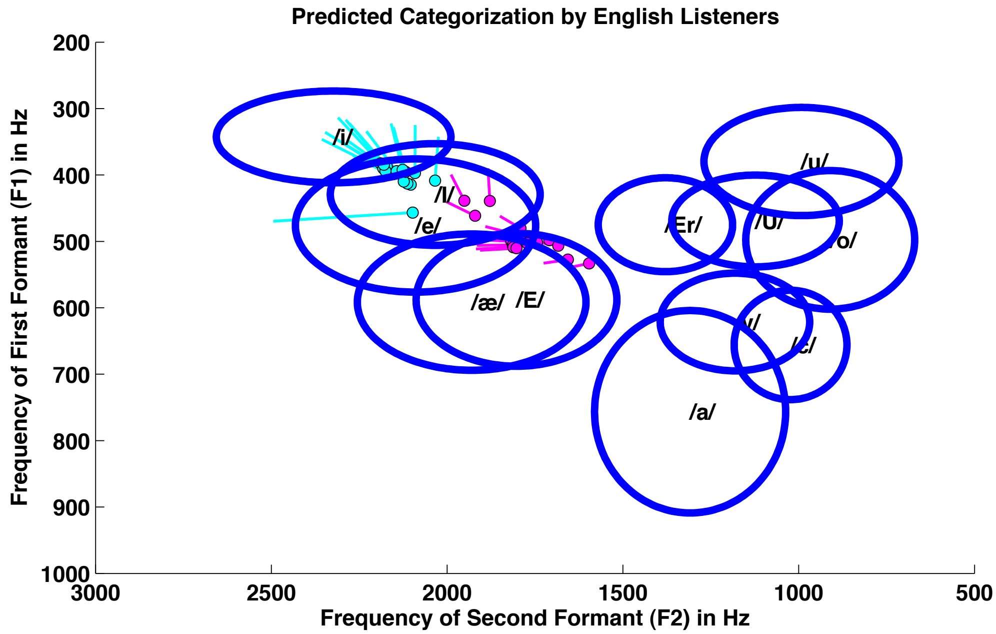

Figure 3.8: Predicted relationship between acoustic and perceptual space for subset of word stimuli produced by male speakers given knowledge of the English vowel categories. The tail of the line segments represents the raw production, and the circle represents the recovered value for the target production. The points shown in light blue correspond to tokens of English /i/, the points shown in magenta to tokens of English /ɪ/.

knowledge. The d' scores computed from the models output are also slightly higher for nonwords than for word stimuli. It is clear to see that in all three respects the model does a poor job of predicting/matching the real data. In Barrios et al. [In preparation], we observed higher d' scores for English than Spanish participants. We also found better discrimination for words than for nonword stimuli.

To what can we attribute the poor performance of the model? As with the simulations of Flege's interlingual identification data presented in the previous section, there are a number of possibilities. The first is that the actual English tokens produced by the male and female speaker in Barrios et al. [In preparation] may not

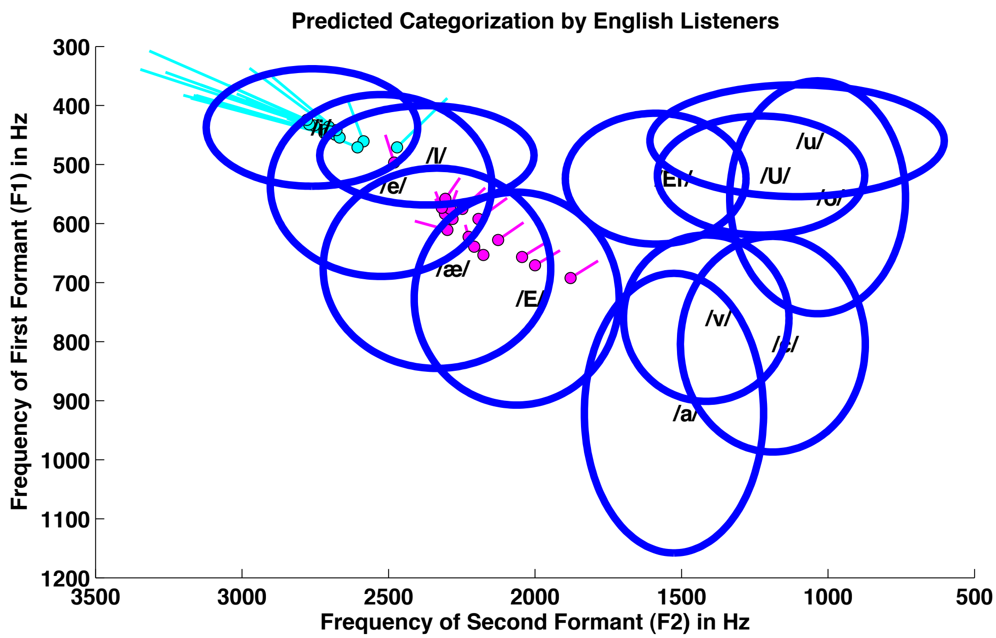

Figure 3.9: Predicted relationship between acoustic and perceptual space for subset of word stimuli produced by female speakers given knowledge of the English vowel categories. The tail of the line segments represents the raw production, and the circle represents the recovered value for the target production. The points shown in light blue correspond to tokens of English /i/, the points shown in magenta to tokens of English /ɪ/.

be a very good fit to the English categories from Hillenbrand et al. [1995]. Figure 3.8 and 3.9 show the categorization of the experimental stimuli with respect to English male and female categories, respectively. In the future it may be useful to measure the distributions from the speakers being used or, alternatively, create synthetic stimuli based on the "standard values" presented in Hillenbrand et al. [1995].

As mentioned above, it is possible that the dimensions we chose to use in our modeling (namely, F1 and F2) are not the ones that listeners are using in categorizing vowels tokens. Simply adding these dimensions to the model produced very similar results. There are other possible quantities, such as F1/F3 and F2/F3, which could also be tried and which have some plausible utility (Monahan and Idsardi [2010]).

Finally, it is likely that our model is doing a poor job of modeling English speakers discrimination behavior because real participants are doing something more with the input. For example, we know that real listeners are capable of speaker normalization (see Johnson [2005]). That is, they have no problem recognizing tokens of the same word produced by a male and a female speaker as instances of the same word. Yet, our model does not incorporate speaker normalization. Instead, we are assuming that it is reasonable to compare the acoustic measurements from tokens produced by a male speaker in our experiment to a male gold standard and the female productions against a female standard. This is likely to be an inappropriate assumption.

Additionally, real listeners also behave differently with respect to real words and nonword stimuli, showing an advantage for real words over nonwords (as in Barrios et al. [In preparation]. Of course, our model has no means of capturing this difference between the model and real participants' behavior, since category knowledge is represented as Gaussian distributions and there is no way in which words are different from nonwords in those dimensions. In the future it will be important to investigate how word-level information and phonetic category acquisition interact (Feldman et al. [2013]).

3.5 Conclusion

In this chapter we set out to use computational modeling as a tool for investigating the appropriate notion of similarity to capture L2 leaners' behavior. In particular, we explored a particular instantiation of a rulers approach to similarity in which similarity is conceived of distance between inferred T's in perceptual space.

We first created a toy model in order to demonstrate that is possible to simulate patterns of perception which are qualitatively similar to those described by Best and others in the speech perception literature.

We then used the model to simulate the identification behavior reported in Flege [1991]. We found that, while the model made reasonable predictions for exemplars of English /i/ and /ɛ/, it did a poor job of predicting Spanish listener's interlingual identification for tokens of English /ɪ/ and /æ/. The model has difficulty capturing the Spanish speakers' behavior since the experimental tokens of English /ɪ/ are essentially near perfect exemplars of Spanish /e/. This data presents a serious problem, not only for our model, but also for other metric/spatial models of similarity in which similarity corresponds to distance in some perceptual space.

Finally, it has become apparent that the current model would have to be extended in various ways to deal with real stimuli and to mimic real perceptual behavior.

Chapter 4

Acquiring new phonological relations

4.1 Introduction

In the previous chapters, I have tried to highlight the fact that acquiring the phonology of a second language entails learning much more than how the target language sounds are distributed in acoustic space. Like L1 learners, second language learners must also acquire a system of oppositions. This requires learning which of the sounds in the input are contrastive (can distinguish word meanings in the target language), and which are not. In addition to encountering new sounds, the L2 learner may also need to acquire new phonological relations among existing sounds. To illustrate this point, let's compare the phonological systems of Spanish and English.

Both Spanish and English have the following three phones [d], [ð], and [ɾ], but only /d/ and /ð/ are used to distinguish word meanings in English (with [ɾ] being one pronunciation of English /d/), and /d/ and /ɾ/ in Spanish (with [ð] being one pronunciation of Spanish /d/). Moreover, in American English a flapping rule causes /d/ (and /t/) to surface as [ɾ] in post-tonic intervocalic position, while a spirantization rule causes /d/ to be pronounced as [ð] in the same environment in Spanish. Figure 4.1 illustrates the relationship between allophones and phonemes in these languages. The Spanish learner of English must learn that some familiar

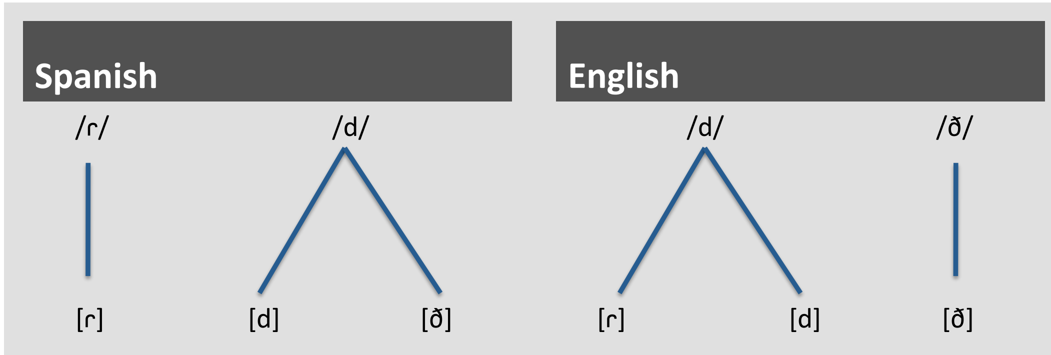

Figure 4.1: Relation between allophones and phonemes in Spanish and English.

sounds enter into new relationships in the target language. That is, the L2 learner must split native language allophones [d] and [ð] into two target language phonemes, which can be used to distinguish meaning in his/her second language.

Recent research at the intersection of speech perception and phonology has suggested that the phonological relatedness between sounds in a language also has an impact on their perceived similarity. For example, partial contrast due to phonological neutralization has been shown to reduce perceptual distinctiveness for native listeners (Huang [2001], Johnson and Babel [2010]), as have non-contrastive relations due to allophony (Boomershine et al. [2008]).

From the perspective of L2 speech perception and phonological development, this leads to the question of whether and to what extent perceptual similarity due to allophonic relations (as opposed to lack of experience) plays a role in L2 acquisition of phonology. We hypothesize that learned phonological relations between L1 allophones may influence the perceived similarity of a pair of contrastive target language sounds, and could consequently make them more difficult for L2 learners to perceive and acquire. While L2 researchers have investigated the difficulty that second language learners experience with L1 allophones in production (Lado [1957], Hammerly [1982], Hardy [1993], Eckman et al. [2001], Eckman et al. [2003]), little research has explored their role from the perspective of speech perception.

4.1.1 Goal of this chapter

This chapter explores the extent to which phonological relatedness among pairs of L1 positional allophones plays a role in the acquisition of nonnative contrasts. Specifically, this project investigates whether advanced L1 Spanish late-learners of English overcome automatic processing of L2 phones [d] and [ð] (which are both mapped to a single L1 phoneme /d/ in Spanish) and reanalyze these as separate phoneme categories (/d/ and /ð/) in the L2. We use previously established methods to assess the perception of the phones [d], [ð], and [ɾ] by three different participant groups, for whom the sounds of interest participate in different relationships. In English, [d] and [ɾ] are in complementary distribution, while [d] and [ð] are in contrastive distribution. In Spanish, [d] and [ɾ] are contrastive, while [ð] is a predictable variant of /d/. Thus, L1 Spanish learners of English must reanalyze an allophonic variant in the L1 as a phoneme in the L2 (i.e. /ð/).

The results of this project will bear on the long-held assumption that non-contrastiveness due to L1 allophonic relations results in learning difficulty when L1 phones must be reinterpreted as contrastive in the L2 phonology. In this way, such results will hopefully improve our understanding of the mapping between L1 and L2 phonology and highlight the role of contrast and allophony in L2 speech perception and acquisition.

4.1.2 Outline of the chapter

In the next section I review the literature on the role of contrast and allophony in shaping speech perception. Next, I present the findings of two experiments (one behavioral and one MEG) designed to investigate the role of L1 phonological relations in L2 speech perception and acquisition. In particular, they are aimed at determining whether Spanish learners of English succeed in acquiring new phonological relations.

4.2 Language experience & speech perception

Adult speakers command an impressive system of knowledge of the structure of their native language, which includes (among other things) [a] knowledge of the segmental inventory of the language, and [b] knowledge of the phonological processes which determine the mapping between predictable surface variants (i.e. allophones) and more abstract phonological representations (i.e. phonemes) by which morphemes and words are stored in long-term memory. Since languages may vary with respect to their phonemic and phonetic inventories, as well as the phonological rules which map between these levels of representation, infants must acquire this knowledge through exposure to the language of their environment.

Infant research has demonstrated that linguistic experience has an early and profound effect on the development of phonetic perception during the first year of life. Evidence for language-specific vowel categories has been reported at 6 months (Kuhl et al. [1992]; Polka and Werker [1994]), and for consonant categories at around 10-12 months (Werker and Tees [1984]).

4.2.1 Naive non-native listeners

Reports of language-specific behavior are robust and ubiquitous in the cross-language speech perception literature. Naive adult non-native listeners[1] experience difficulty identifying and discriminating many phonetic contrasts which are not used to distinguish lexical items in the learners' native language. The classic example of this phenomenon is the difficulty that Japanese listeners experience distinguishing the English /ɹ/-/l/ contrast (Goto [1971]; Miyawaki et al. [1975], etc.). The English contrast is non-contrastive in Japanese, which has a single liquid consonant (i.e. /ɾ/). Such results have been demonstrated using a range of speech contrasts, listener groups, and employing a variety of paradigms and experimental techniques (Abramson and Lisker [1970]; Werker et al. [1981]; Werker and Lalonde [1988]; Näätänen et al. [1997]; Winkler et al. [1999]; Phillips et al. [2000]; Kazanina et al. [2006], to name just a few).

In addition to experience with the speech contrasts, the relative ease and difficulty with which non-native contrasts are distinguished varies by the native language of the listener, as well as the gradient phonetic properties of the non-native stimuli (Best [1994]; Best et al. [1988, 2001, 2003]). For example, English speakers expe-

[1] Following Best and Tyler [2007], 'naive' is used here to distinguish these non-native listeners who are "not actively learning or using an L2" and are "linguisitically naive of the target language test stimuli", from 'L2 learners' who are "in the process of acquiring an L2 to achieve functional, communicative goals."

rience considerable difficulty perceiving Hindi stops /t̪/-/ t/ (Werker et al. [1981]), but perform extremely well with Zulu click contrasts (Best et al. [1988]), despite the fact that both contrasts are non-native for English listeners. English listeners also readily discriminate !Xóõ click contrasts which cause difficulty for speakers of other click languages, such as Isizulu and Sesotho (Best et al. [2003]). Best and colleagues argue that varied performance across stimulus contrast and listener language reflect properties of the non-native stimuli and those of the listener's native phonology (Best [1994]; Best and Tyler [2007])

Another important finding is that even difficult non-native contrasts can be discriminated under 'optimal' listening conditions (Werker and Logan [1985]. For example, English listeners experience difficulty discriminating the Hindi place contrast (i.e. /t̪/-/ t/) when the stimuli are separated by long interstimulus intervals (ISI), yet they discriminate this same contrast when the stimuli are separated by short ISIs (i.e. $<$ 500 msec). This finding suggests that language specific processing does not reflect the loss of perceptual ability, but rather a functional reorganization (Werker and Tees [1984]). Under this view, listeners have access to both universal and language-specific representations under certain listening conditions. More importantly from the perspective of L2 perception, if experience-related changes in perceptual processing do not reflect loss, then the difficulty that listeners encounter with non-native contrasts might be overcome with the appropriate kind of experience.

4.2.2 L2 learners

Much like cross-linguistic studies with monolingual listeners, L2 speech perception research has also produced considerable evidence that difficulties arise from differences in segmental inventory. For example, poor perception is often observed for vowel and consonant contrasts involving sounds which are absent from the learner's native language (i.e. the /ɹ/-/l/ contrast for Japanese learners of English (MacKain et al. [1981]; Brown [1998, 2000]), the /i/-/ɪ/ contrast for Catalan learners of English (Cebrian [2008]), and French front round vowels /y/-/ø/ for English learners (Gottfried [1984]), to name a few).

Similar to naive listeners, L2 learners' categorization and discrimination performance varies with the L2 contrast in question and the L1 of the learner. For example, Brown [1998, 2000] demonstrated that Mandarin learners of English have less difficulty with the English /ɹ/-/l/ contrast than an otherwise matched group of Japanese learners, despite the fact that both languages lack the contrast. In a second experiment, Brown reported that Japanese learners performed better on the English /f/-/v/ and /b/-/v/ contrasts, than the /ɹ/-/l/ contrast, although both the /f/-/v/ and /ɹ/-/l/ contrast are non-native for the Japanese listener. According to Brown, this difference in performance is due to the availability of phonological features which can be redeployed for the purpose of processing and representing /f/-/v/, but not /ɹ/-/l/.

In sum, listeners often experience difficulty when they are asked to discriminate sounds which do not exist in their native language segmental inventory. Moreover, the relative ease or difficulty with which listeners perceive non-native stimuli is also influenced both by contrastive phonological and non-contrastive phonetic aspects of the L1 and the target language.

4.3 Perceptual correlates of allophony

Recent research at the intersection of phonology and speech perception has demonstrated that the presence and absence of particular sounds in the native language is not the only factor that influences the perception of speech. In addition, sounds which occur on the surface in a given language may also be non-contrastive, such as when they are context-dependent allophones of the same underlying phoneme.

Several studies have reported differences in the processing of phonemic and allophonic contrasts in the native language (Pegg and Werker [1997]; Whalen et al. [1997]; Harnsberger [2001]; Peperkamp et al. [2003]; Boomershine et al. [2008]). For example, Pegg and Werker [1997] reported finding worse performance by adult English listeners on the allophonic contrast between voiced [d] and the voiceless unaspirated [t] than on a phonemic contrast on an AX discrimination task. Similarly, Whalen et al. [1997] found that English listeners performed better when asked to discriminate phonemic [b-p] and [b-p^h] than allophonic contrasts [p-p^h] in an AXB discrimination task. Harnsberger [2001] found that Malayalam listeners perceive dental and alveolar nasals (i.e. [n̪] and [n]), which are allophonically related in the language, as highly similar to one another, despite the fact that they are no more similar acoustically than the palatal and velar nasals (i.e. [ɲ] and [ŋ]). Peperkamp et al. [2003] also reported a difference between French listeners' perception of phonemic [m]-[n] and allophonic contrasts [ʁ] - [χ], but only when the latter were presented in a phonological environment which conditions the application of the French voicing assimilation rule[2]. These results suggest that the perception of allophones is also sensitive to the phonological context.

In a recent study, Boomershine et al. [2008] investigated the impact of contrast and allophony by comparing the perception of the sounds [d], [ð], and [ɾ] by native English and Spanish listeners using a similarity rating and a speeded AX discrimination task. They predicted that language-specific exposure would result in greater perceptual similarity and longer RTs for allophonic than for phonemic contrasts. Their prediction was borne out. In particular, Spanish listeners produced higher similarity ratings and longer RTs than English speakers for the [d]-[ð] contrast, which is non-contrastive in Spanish, whereas English listeners rated [d]-[ɾ], which are related by an allophonic rule in their native language, as being perceptually less distinct.

Kazanina et al. [2006] investigated whether an early auditory brain response to an infrequent change in stimulus (i.e. the mismatch field, MMF response) is sensitive to the functional significance of native language sound categories. The authors studied the processing of the sounds [t] and [d] by Russian listeners, for whom the contrast is phonemic, and by Korean listeners, for whom the contrast

[2] [χ] is a predictable surface variant of the phoneme /ʁ/ which is found adjacent to a voiceless consonant.

is allophonic. Russian participants showed both behavioral evidence of categorical perception (i.e. a classic step-like identification function for the /ta/-/da/ VOT continuum and better between-category than within-category discrimination) and electrophysiological evidence of change detection in auditory cortex. In contrast, Korean participants showed neither behavioral, nor electrophysiological evidence of perceptual sensitivity. These results suggest that the auditory cortex groups sounds based on phonemic, but not allophonic categories, and that the functional significance of sounds factors into speech perception at a very early stage of processing (at least for this contrast).

In sum, all of these findings suggest that listener's perception of native and non-native sound contrasts is affected by the phonological status of the contrast. Moreover, auditory cortex is able to group the sounds on the basis of phonemic, but not allophonic contrasts. These findings suggest that listeners' perception of speech sounds is strongly and systematically constrained by the phonology of his or her native language. Moreover, these findings prompt the question of whether and to what extent these constraints can be overcome with experience. In particular, do L1 context-dependent allophones continue to play a role in L2 perception?

4.3.1 L1 allophones in L2 acquisition and processing

The bulk of L2 research on the acquisition of novel contrasts has focused on non-native distinctions which are non-contrastive due to differences in segmental inventory between the first and second language (i.e. the absence of one or more sound). Much less is known about the perception and acquisition of non-contrastive sound categories which are related by phonological rule in the learner's L1. The next two subsections are dedicated to a review of the existing literature on L1 allophones in L2 production and perception.

4.3.1.1 L1 allophones in L2 production

To date, nearly all research investigating the impact of L1 allophones on the acquisition of novel L2 contrasts has approached the problem from the perspective of production. Several researchers have found empirical support for Lado's early claim that when a learner must reinterpret two or more phones, which are positional variants of the same phoneme, as allophones of separate phonemes in the L2, he/she will encounter considerable difficulty, which is likely to persist in L2 production (Lado [1957]).

For example, in his research on L2 segmental production, Hammerly observed that "problems involving the use and non-use of allophones are more persistent than those involving the use and non-use of phonemes" ([Hammerly, 1982, p. 27]).

Similarly, Hardy [1993] investigated the production of three types of non-native contrasts by a single L1 Spanish-L2 English learner, and found that inventory constraints were easiest to overcome, followed by positional constraints, followed by allophonic rules.

The production research reported above clearly suggests that L2 learners experience difficulty when they must learn to associate allophones of a single L1 phoneme with two separate target language phonemes. However, they shed little light on the perception of these target language contrasts. Moreover, while some researchers have argued for a causal link between L2 perception and production (Flege [1995]), proposing that accented production is the result of inaccurate perception, other findings suggest that perception does not always precede production in L2 sound learning (Sheldon and Strange [1982]). It is also quite conceivable that a late learner's perception of non-native contrasts be quite good, but that he/she encounters difficulty producing the same contrast. Thus, in order to address questions regarding the role of L1 phonological relatedness in L2 perception, one must investigate the issue directly. I now turn to two training studies which can provide some insight into the perception of these novel target-language contrasts.

4.3.1.2 L1 allophones in L2 perception

In a perception training study, Curtin et al. [1998] investigated the role of lexical and surface representations in phonological transfer by comparing native English and French speaking participants' ability to lexically represent the Thai three way stop voicing distinction (e.g., /b/-/p/-/p^h/). They found that both English speakers and French speakers readily acquired the voicing contrast (presumably by transferring their native [voice] feature which is used to distinguish lexical items in their L1). Additionally, the English-speaking group, but not the French-speaking group, showed evidence of acquiring the ability to lexically encode [aspiration] in Thai. This difference in outcomes is attributed to the presence of [aspiration] in surface representations in English, but not French. The authors concluded that L1 surface features can be lexicalized in L2 acquisition, even though they are not transferred initially[3]. These findings also suggest that experience with non-contrastive surface features may actually play a facilitatory role in L2 sound learning.

In another training study, Herd [2011] evaluated the effectiveness of various training methods on L1 English-L2 Spanish learners' production and perception of Spanish /d/, /ɾ/, /r/. She conducted an MMN study both prior to and following perception training to establish the effects of the treatment on the automatic pre-attentive processing of auditory stimuli containing the Spanish /d/-/ɾ/contrast (i.e. [ede]-[eɾe]). As expected, Spanish participants showed a significant divergence between responses to standard and deviant categories. Unexpectedly, English participants also showed a significant MMN both at pre-test and post-test. These results are difficult to interpret, however, since the author does not report the performance of a monolingual English control group. Thus, there is no way to know how much learning has occurred. Additionally, assuming that learning has occurred, the participants in this training study were classroom L2 learners who are in their third semester of Spanish-language classes. Thus, it is unclear if Spanish-like performance is the result of learning in the classroom or during the training or some combination of the two.

Summing up, these training studies both suggest that perceptual learning can and does occur. Additionally these studies make valuable contributions to our understanding of how second languages are learned in a controlled classroom setting.

[3]But see also Pater [2003] for contradictory findings.

However, it is likely that they are less informative as to how and to what extent L2 perceptual learning takes place in a more naturalistic learning environment. In consequence, we are left with a number of unanswered questions regarding the perception and acquisition of novel L2 contrasts involving context-dependent allophones which are related by phonological rule in the learner's native language. For example, do novel target language contrasts involving L1 positional allophones pose any learning difficulty from the perspective of speech perception? Or, alternatively, does experience with the target language categories at the level of predictable surface variants actually facilitate acquisition? In so much as L1 positional allophones present learning difficulty, how persistent is this L1 interference? Can it be overcome with experience? In particular, will difficulty still be evident in the performance of advanced late-learners? And finally, is any learning on the part of these advanced learners manifest in their early automatic pre-attentive processing?

4.4 Motivation for two experiments

Two experiments were conducted to investigate questions regarding the representation and processing of the sounds [d], [ð], and [ɾ] by three participant groups: [a] English native speakers, [b] Spanish native speakers, and [c] advanced L1 Spanish late learners of English. These questions are:

1. Is perceptual sensitivity influenced by high-level factors that are relevant for the encoding of words in long-term memory?

2. If so, do we see the influence of this differential sensitivity in the representation and processing of advanced L1 Spanish late-learners of English?

We used a behavioral task and MEG recordings to investigate how perceptual space is impacted by higher-level factors that are relevant for the encoding of words in long term memory. To do so, we took advantage of the fact that three very similar phonetic categories exist in Spanish and English, but that the functional significance of these categories varies in these languages.

4.4.1 Cross-linguistic differences in the mapping between phones and phonemes: [d], [ð], and [ɾ] in English and Spanish

Although the phonetic categories [d], [ð], and [ɾ] occur in both Spanish and English, [d] and [ð] are used to distinguish word meanings in English only. For example, minimal pairs exist (i.e. [ðeɪ] 'they' and [deɪ] 'day') which demonstrate that the sounds [d] and [ð] are contrastive. On the other hand, the sound [ɾ] only occurs as a predictable surface variant of /d/ (and /t/) in English. That is, a productive phonological rule[4], the American English flapping rule, causes /d/ (and /t/) to surface as [ɾ] in post-tonic intervocalic position, and [d] elsewhere.

In Spanish, on the other hand, [d] and [ɾ] are contrastive, and [d] and [ð] are allophones of the same phoneme which occur in complementary distribution. A spirantization rule causes the voiced obstruents /b, d, g/ to surface as the approximants [β, ð, ɣ] intervocalically[5]. It is worth noting that while [d] and [ɾ] are contrastive

[4]Patterson and Connine [2001] report that flapped variant occurs 94% of the time in its conditioning environment.

[5]Like the flapping rule, the spirantization rule is a productive phonological process in Spanish.

in Spanish, these sounds never occur in the same context. [ɾ] occurs word medially and in free variation with [r] in word final position, but not word initially. [d] occurs word initially, as well as after nasals and [l].

4.4.2 Useful tools

We measured participants' perceptual sensitivity to sound pairs which are either contrastive or allophonic in the listeners' native language, as indexed by the d' scores on an AX discrimination task and MMN responses recorded using magnetoencephalography (MEG).

An AX discrimination task was employed in order to assess participants' sensitivity to the contrast between [idi], [iði], [iɾi] stimulus tokens. It was expected that listeners would discriminate phonemic contrasts more readily than allophonic contrasts, and that this difference in sensitivity should be reflected in participants' d' scores.

Unlike behavioral measures which may reflect late, conscious processes, electrophysiological responses such as EEG and MEG can be collected continuously from the onset of the stimulus without the necessity of an overt behavioral response. For this reason, in addition to the AX discrimination task, we also used magnetoencephalographic (MEG) recordings to measure the detailed time-course of brain activity in each of the three listener groups.

Of particular interest for our purposes are two auditory evoked components, namely the M100 and the MMN. The M100 is an evoked response which is produced

Waltmunson [2005] reports that the intervocalic spirantization of /d/ occurs 99% of the time.

whenever an auditory stimulus has a clear onset. The response can be observed regardless of task or attentional state (Näätänen and Picton [1987]). We use the M100 response to tones in order to select the auditory channels of interest for each of our participants for our MMN amplitude analyses.

The MMN component typically has a latency of 150–250 ms post-stimulus onset and its amplitude is a measure of 'percetual distance'. This early response component has also been shown to reflect early automatic, preattentive processes (Näätänen [1992]). The component is typically observed when there is an infrequent detectable change (i.e. deviant) in an otherwise frequent unchanging stimulus (i.e. standard). Following Phillips et al. [2000] we used a paradigm in which there is no acoustic standard. Instead, participants were presented multiple non-orthogonally varying tokens from each category. This is to avoid a purely acoustic interpretation of the elicited responses. The MMN has been shown to be sensitive to the functional significance of native language sound categories (Kazanina et al. [2006]). In the current study, we use the relative amplitude of the MMN as a measure of category identification.

4.5 Hypotheses & predictions

Our first hypothesis is that if early auditory brain responses are shaped by a more abstract analysis of the functional significance of those sounds (Kazanina et al. [2006]), then we should observe a different pattern of results depending on the native language phonology of the listener. Our predictions are summarized in the

	[d]-[ɾ] (Spanish Contrast)	[d]-[ð] (English Contrast)	[ð]-[ɾ] (Contrast for BOTH)
English	Poor discrimination/ small MMN	Good discrimination/ large MMN	Good discrimination/ large MMN
Spanish	Good discrimination/ large MMN	Poor discrimination/ small MMN	Good discrimination/ large MMN
Learners	?	?	Good discrimination/ large MMN

Figure 4.2: Summary of the predictions for the three participant groups.

Figure 4.2. English speakers are expected to more readily discriminate [idi], [iði], than [idi] and [iɾi]. In contrast, Spanish listeners should show better discrimination of [idi] and [iɾi], than [idi] and [iði]. [iði] and [iɾi] should be readily discriminated by all three groups since the contrast is phonemic in both languages.

Of interest is the performance of the advanced late learners of English. We expect that if the Spanish learners have made progress in acquiring the English [d]-[ð] contrast, they will outperform the monolingual Spanish speaker group on the AX discrimination task and possibly show evidence of perceptual sensitivity in their pre-attentive brain response. If they have not yet made progress in acquiring the target language contrast, we expect them to perform like the native Spanish speaker group. Additionally, (though unlikely) it is possible that the phonological relatedness of the sounds [d] and [ɾ] in the learners' target language may reduce the discriminability of the pair.

4.6 Experiments 3 & 4

4.6.1 Participants

Three groups of participants were recruited to participate in these experiments for monetary compensation; [a] 10 English native speakers (Male=3, mean age=23(SD= 3.8)), [b] 7 Spanish native speakers.[6] (Male=3, mean age=32 (SD=6.8)), and [c] 7 advanced/experienced L1 Spanish late learners of English.[7] (Male =3, mean age=32 (SD=5.4)). All participants tested strongly right-handed according to the Edinburgh Handedness Inventory (Oldfield [1971]) and reported no history of hearing or neurological disorder. All participants were recruited from in and around the University of Maryland. English speaking participants and the majority of the Spanish speaking learners were undergraduate and graduate students who studied or worked at the University of Maryland campus. The monolingual Spanish speakers were drawn from a neighboring community with a large Spanish speaking population. This group was largely comprised of immigrants from Central America who had recently arrived to the College Park area and continue to use Spanish as their primary mode of communication. These participants have had little exposure to English aside from what is heard on TV and the radio. We acknowledge that the participants in the monolingual Spanish speaking group and the other two language groups likely differ in SES, level of education, etc. While it may have been possible

[6] We have made every effort to find participants who have little or no experience with Spanish or English, respectively.

[7] Ultimately, it will be informative to include an inexperienced group that would allow us to further explore developmental predictions, but these initial three groups have been chosen as a starting point so that we might establish the general pattern and begin to explore the role of L1 allophones in L2 perception.

to find a better matched group of Spanish speakers elsewhere we were constrained by the fact that the MEG is not a portable device. Participants completed a language background questionnaire prior to participating in this experiment.

4.6.2 Stimuli

Materials for our experiments consisted of 10 natural tokens of the following VCV sequences: [idi], [iði], [iɾi] spoken by a single female speaker of American English with phonetic training. Multiple instances of each stimulus were recorded using a head-mounted microphone in a soundproof room. The vowel [i] was chosen for the vowel context because, Spanish [i] and English [i] have the greatest perceived similarity by listeners of both groups (Flege et al. [1994]). Moreover, the resulting stimulus set did not result in any known words.

Due to the fact that the phones [d] and [ð] and [d] and [ɾ] are in complementary distribution in Spanish and English, respectively, it is of course impossible to find contexts in which all three phones occur naturally. For this reason, the tokens (in particular the [idi] tokens) were produced with care by a native English speaker with phonetic training so as to avoid flapping. All tokens were later inspected by an additional trained phonetician to ensure that intervocalic [d]s were not produced as [ɾ]. To ensure that any observed differences in the measured auditory response could only be attributed to the consonant, as opposed to the preceding vowel or some other aspect of the stimulus tokens, the initial [i] from each token was removed and replaced with an identical [i] recorded in a neutral context (i.e. [isi]). To avoid introducing artifacts which might make the stimuli less natural, the initial vowel was cross-spliced such that the files were matched from positive going zero-crossing to positive going zero-crossing. Each was later inspected by two trained phoneticians for naturalness. The ten best stimulus tokens of each type were chosen on the basis of their perceived naturalness to native speakers of Spanish and English. The intent was that they would be equally good tokens of both languages. All stimuli were normalized for intensity using PRAAT (Boersma and Weenink [2009]) and were presented to participants at a comfortable listening level (~73dB).

4.6.3 Procedures

4.6.3.1 Informed consent & Language background questionnaire

Upon arriving to the lab, participants were provided with a overview of the procedures involved in the study. All participants provided informed written consent approved by the University of Maryland institutional review board as well as completed a Language Background Questionnaire (*Appendix B*) to ensure that he/she met the language background requirements for participants in the study.

4.6.3.2 MEG recordings

Magnetic fields were recorded in DC (no high-pass filter) using a whole-head MEG device with 157 axial gradiometers (Kanazawa Institute of Technology, Kanazawa, Japan) at a sampling rate of 1 kHz. An online Low Pass Filter of 200 Hz and a 60 Hz notch filter were applied during data acquisition. All stimuli were presented binaurally via Etymotic ER3A insert earphones at a comfortable listening level. MEG recording sessions included 4 runs: 1 screening run and 3 experimental blocks which are described in greater detail below. Participants passively viewed a silent movie during the experimental runs to avoid fatigue. In total each MEG recording session lasted approximately 90 minutes in total.

4.6.3.3 One-tone screening test

For the screening run, participants were presented approximately 100 repetitions of a 1 kHz sinusoidal tone. Each tone was separated by a randomly chosen ISI of 1000, 1400, or 1800ms. Data from the screening run were averaged and examined to verify a canonical M100 response. Twenty-one of twenty-four participants run across the three participant groups showed a reliable bilateral M100 response with a source/sink reversal between anterior and posterior channels in the left and right hemisphere. Three English participants were excluded based on the lack of a strong bilateral M100 response elicited by a 1-kHz pure tone at pretest.

4.6.3.4 MMN

In the experimental blocks stimuli were presented using a modified version of the passive oddball paradigm (Näätänen et al. [2004]). In each of the three experimental blocks one of the three stimulus types (i.e. [idi], [iði], or [iɾi]) was presented frequently (i.e. the standard) and was followed by infrequent stimuli of the other two types. For example, in Figure 4.3 the first block shows [idi] as the

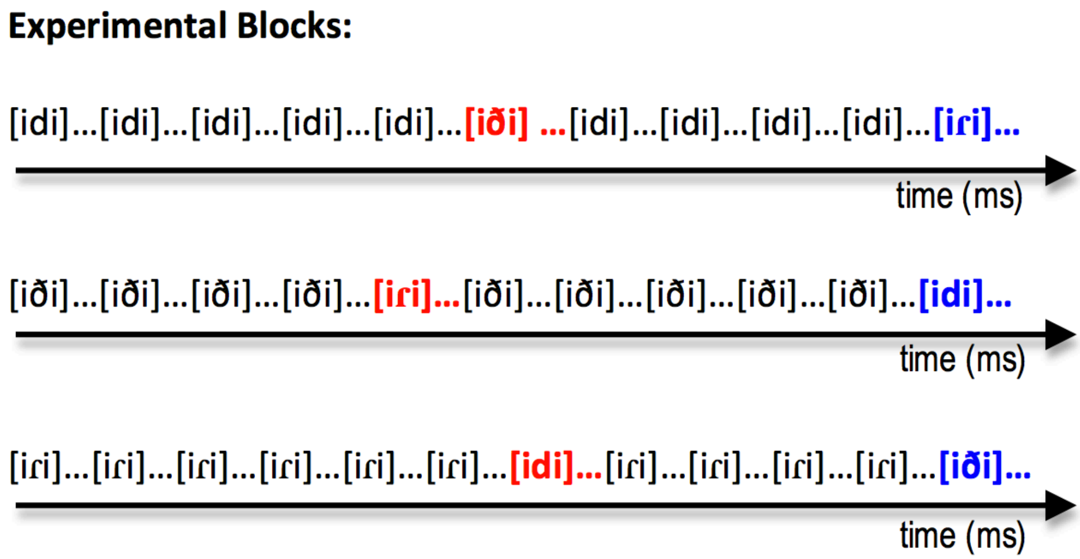

Figure 4.3: Illustration of the structure of each of the three experimental blocks in our modified passive oddball paradigm. Stimuli shown in black correspond to the standards for that experimental block. The two types of deviants for a particular block are shown in red and blue.

frequent standard and [iði] and [iɾi] as the less frequent intervening deviants. In any given block a deviant was presented after a minimum of 4 and a maximum of 6 standards with the probability of deviant (either deviant type A or B) = .167.

Each stimulus token was separated by an ISI that varied between 600 to 1000ms. The experiment consisted of three experimental blocks which lasted approximately 20 minutes. Participants were given a short break after each 10 minutes. Block order was counterbalanced across participants. Figure 4.3 shows the structure of each of the three blocks.

4.6.3.5 Behavioral tasks

During the behavioral portion of the experiment participants were seated in a quiet room in front of a computer with headphones placed and volume adjusted such that sounds played at a comfortable listening level. The behavioral portion of the experiment was controlled by DMDX (Forster and Forster [2003]) and consisted of two parts; [1] an AX discrimination task, and [2] an Identification and Rating task.

4.6.3.6 AX discrimination task

In the AX discrimination task participants were presented two of the experimental stimuli which were either different tokens of the same 'word' (i.e. [idi]-[idi], [iði]-[iði], [iɾi]-[iɾi]) or one of the six possible pairings of different 'words' (i.e. [idi]-[iɾi], [idi]-[iði], [iði]-[iɾi], [iði]-[idi], [iɾi]-[iði], [iɾi]-[idi]). Each stimulus was presented with an interstimulus interval (ISI) of 500ms. Participants were instructed to press the 'F' key on the keyboard with their left index finger if the two stimuli were two pronunciations of the same 'word' and to press the 'J' key with their right index finger if the paired stimuli consisted of two different 'words'. Participants were asked to respond as quickly and accurately as possible and had a maximum of 4 seconds to respond on each trial. The AX discrimination task lasted approximately 18 minutes and was divided into six 3 minute blocks with a participant controlled breaks in between each block.

4.6.3.7 Identification and rating task

After completing the AX discrimination task, participants were asked to perform an Identification and Rating task. During this portion of the experiment, they were presented with 40 stimuli (30 experimental items and 10 filler items). Each stimulus item was presented twice in a row, first for identification and later for rating. Participants were instructed to use the keys 1, 2, and 3 to identify the stimulus they heard as an instance of the 'word' 'eithee', 'eady' or something else in English and 'idi', 'iri', or something else in Spanish, respectively. After having made an identification response with respect to a particular stimulus, the participant would hear the stimulus again and were asked to rate that stimulus as a possible production of that 'word' in their native language using a scale from 1 to 5, where 1 = not very English-like/Spanish-like, and 5 = very English-like/Spanish-like. The identification and rating task took 7 minutes to complete.

4.6.4 Data analysis

4.6.4.1 AX discrimination data

Two Spanish and one English participants whose performance was below 60% accuracy over all experimental items were excluded from subsequent behavioral analyses. For the remaining participants, d' scores were computed according to the Same-Different Differencing Model (Macmillan and Creelman [2005]). 'Hits' were defined as instances when stimulus A and X were different and the participant responded that they were different, whereas a 'Miss' corresponded to a trial in which the participant responded same when stimulus A and X were different. Trials in which the participant responded different, but stimulus A and X were the same were coded as 'False alarms'. 'Correct rejections' were those trials in which the

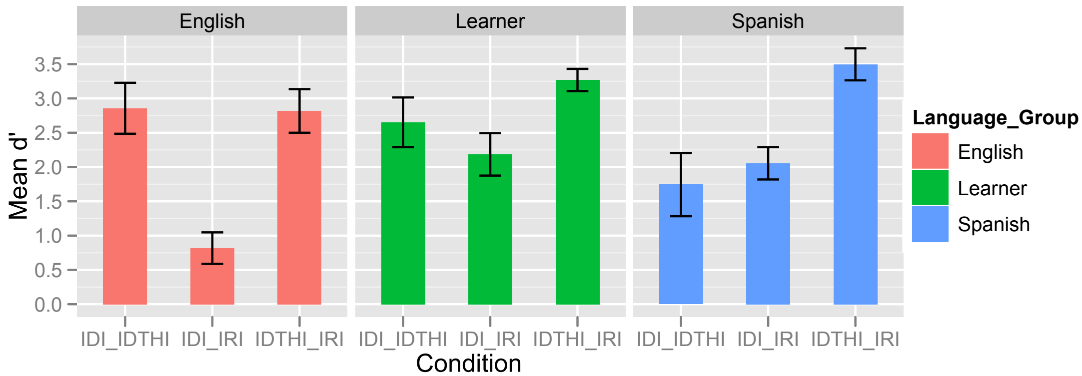

Figure 4.4: Mean d' scores by language group and stimulus pair. Error bars represent standard error.

participant responded same when the stimuli in a give trial were the same. The 'Hit rate' is the proportion of different trials to which participant responded different. The 'False alarm rate' is the proportion of same trials to which the participant responded different. We computed d' scores using the dprime.SD() function from the psyphy package in R. The result is a measure of sensitivity which factors out participants' response bias. These d' scores were subsequently analyzed using a Generalized Linear Mixed Effects Model (GLMM) with language group and condition as fixed effects and subject as a random effect. Figure 4.4 shows the mean d' score by language group and stimulus pair.

4.6.4.2 Identification data

We computed the percentage of accurate identifications for all three participant groups for each stimulus type. We excluded participants who misidentified filler items more than 70% of the time. Four participants were excluded overall (3 Spanish participants and 1 English participant). Across the remaining participants, stimulus

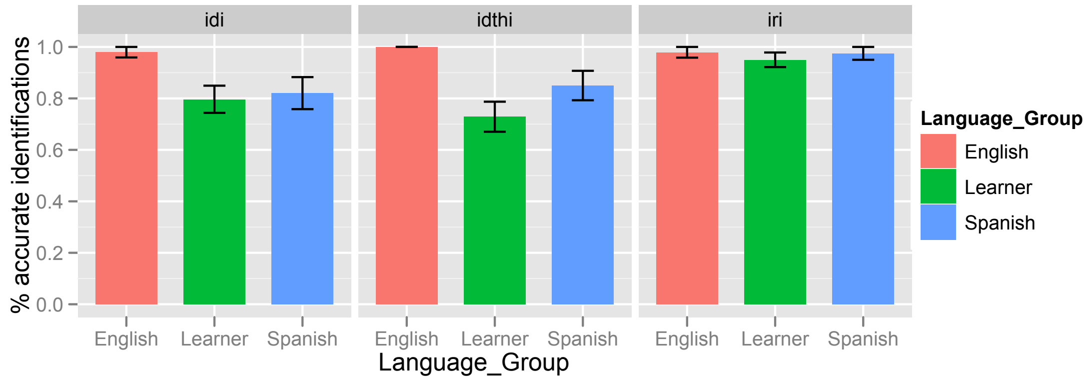

Figure 4.5: Percentage of accurate identifications by stimulus type and language group. Error bars represent standard error.

tokens were accurately identified as instances of their stimulus type 91% of the time. Figure 4.5 shows the mean percentage of accurately identified stimuli by group and stimulus type.

4.6.4.3 Rating data

Only those stimulus tokens which were accurately identified by our participants were entered into our rating analyses. The mean rating for each stimulus type by each participant group is shown in Figure 4.6. Both the identification and rating data were subsequently analyzed using a Generalized Linear Mixed Effects Model (GLMM) with language group and condition as fixed effects and subject as a random effect.

4.6.4.4 MEG data

MEG data were imported into Matlab and de-noised using a multi-shift PCA noise reduction algorithm (De Cheveigné and Simon [2007], De Cheveigné and Si-

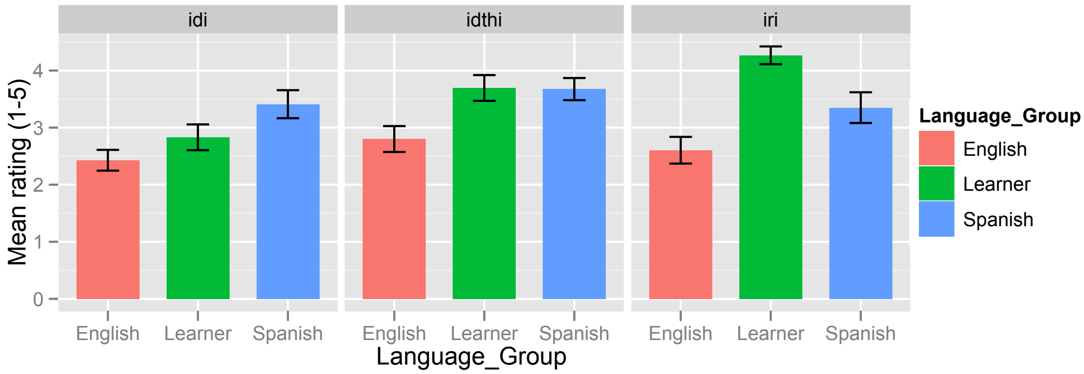

Figure 4.6: Mean stimulus ratings (1=not English-like/Spanish-like, 5=very English-like/Spanish-like) by stimulus type and language group. Error bars represent standard error.

mon [2008]). Epochs included 100ms prior to stimulus onset to 600ms post onset. Artifact rejection was conducted to exclude trials containing responses exceeding a pre-determined threshold value of 3 pT. All epochs were averaged, baseline corrected over the 100ms pre-stimulus interval, and filtered using a .03 to 30-Hz band-pass filter.

MMN amplitude analysis

We selected the 10 strongest left hemisphere channels (identified visually in MEG160) from the peak of the average M100 response to 1 kHz tones elicited during the pre-screening test for each participant. Three English speaking participants did not show a typical M100 response to the tone pretest and were excluded from all subsequent analyses.

For each participant in each condition we calculated the root mean square (RMS) amplitude of the MEG temporal waveforms from the left hemisphere channels selected on the basis of the auditory localizer pre-screening test. We then computed the mean power over a single 100ms time window from 310-410ms for each of the participants for each of the experimental conditions. This time window was chosen because the vowel offset and consonant onset occurred at 160ms and the MMN is expected to occur about 150-250ms post-stimulus onset. The grand average of the mean power over the 310-410ms time window was computed by averaging across participants for each condition (n = 7). These grand averages are shown in Figure 4.7. The mean RMS amplitudes over the 310-410ms time window were subsequently analyzed using a Mixed Effects Model with subject as a random effect (Baayen [2008]).

4.6.5 Results

4.6.5.1 d' scores

Statistical analyses were performed using a Generalized Linear Mixed Effects Model (GLMM) (using R package lme4) with factors Language Group (English, Learner, Spanish) and Condition ([idi]-[iɾi], [idi]-[iði], [iði]-[iɾi]), and the Language Group x Condition interaction as fixed effects and subject as a random effect. We observed a main effect of Language Group (t=3.3, p=.001), which is likely driven by the fact that the learner group differs from the Spanish group. We observed a main effect of Condition (t=5.47, p<.001), likely due to the scores being greater for the [iði]-[iɾi] contrast, suggesting that this contrast is an adequate control against other contrasts. We observed a Language Group by Condition interaction (t=4.2, p<.05), which is probably due to the fact that the Spanish and English groups differed.

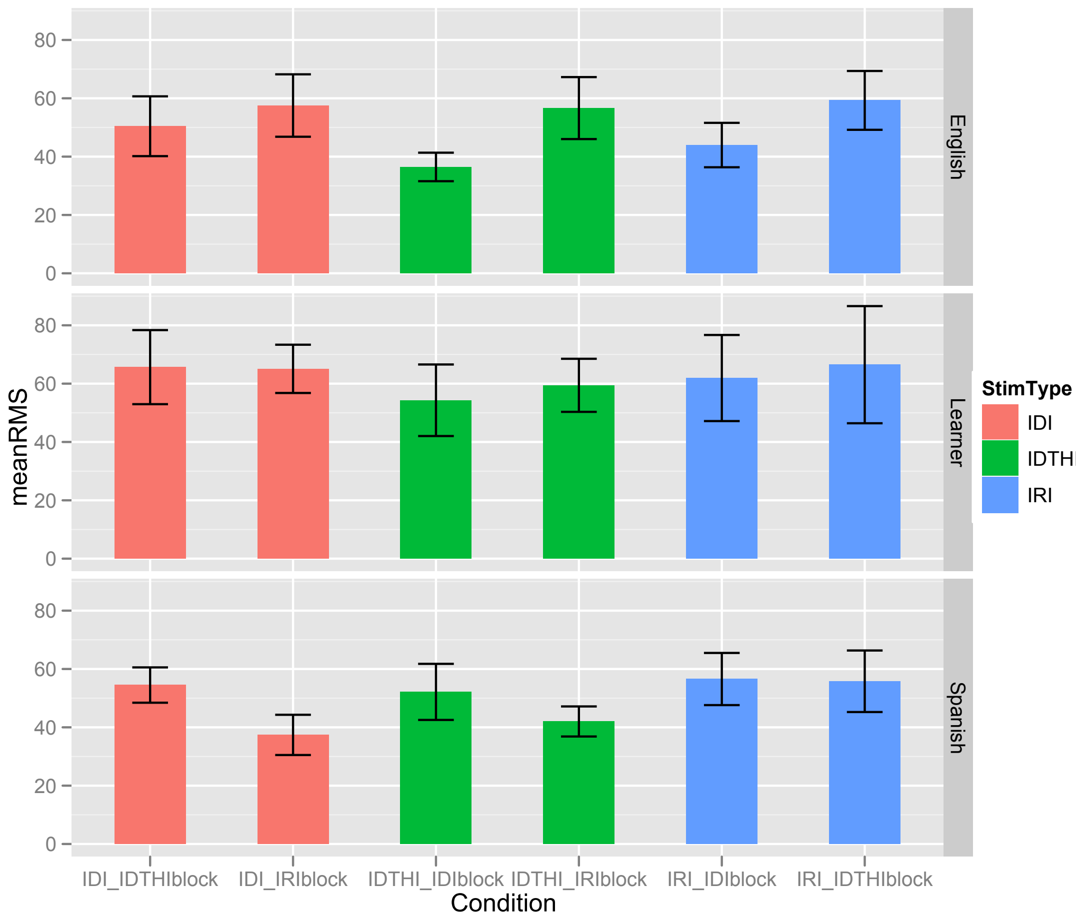

Figure 4.7: Grand averaged RMS amplitudes to deviants for each language group by condition for the 310-410ms time-window. Responses to [idi] tokens are shown in coral, [iði] in green, and [iɾi] in blue. Error bars represent standard error.

We conducted six planned tests of our experimental hypotheses regarding listeners' sensitivity to allophonic vs. phonemic contrasts using simultaneous tests for general linear hypotheses (multcomp package in R). We hypothesized that performance on allophonic contrasts would differ from the phonemic control contrasts for the Spanish and English groups, whereas the other phonemic contrast would not. For the learner group, we hypothesized that if the learners had acquired the contrast they would show no difference in d' scores for either pair. For the English group we found a significant difference in d' scores for the allophonic [idi]-[iɾi] contrast relative to the control contrast [iði]-[iɾi] (z=5.679, p<.001). There was no difference between the phonemic [idi]-[iði] and the control [iði]-[iɾi] contrast for the English speaker group (z=.141, p=1.00). Taken together these findings suggest that English listeners are less sensitive to allophonic than to phonemic contrasts. As expected, we find a similar pattern of results for the Spanish group. Again, we find a significant difference between the allophonic [idi]-[iði] and the [iði]-[iɾi] control contrast (z=3.969, p<0.001). We also found an unexpected difference in d' scores for the phonemic [idi]-[iɾi] contrast relative to the control [iði]-[iɾi] contrast (z=3.057, p=.02), which we will discuss in the following section. Interestingly, we find no difference between the [idi]-[iði] vs. [iði]-[iɾi] (z=1.697, p=0.53) for the learner group, suggesting that the learners have acquired the allophonic contrast. We also found a small but significant difference between [idi]-[iɾi] vs. [iði]-[iɾi] (z=2.839, p=0.04).

4.6.5.2 Percentage of accurate identifications

Statistical analyses of percentage of accurate identifications were performed using a Generalized Linear Mixed Effects Model (GLMM) (using R package lme4) with factors Language Group (English, Learner, Spanish) and Stimulus Type ([idi], [iði], and [iɾi]), and the Language Group x Stimulus Type interaction as fixed effects and subject as a random effect. While no main effect of Stimulus Type was found, the main effect of Language group was found to be reliable ($t=2.472$, $p=.01$), suggesting that the language groups differed from each other. The Language group x Stimulus Type interaction was also significant ($t=1.984$, $p<.05$), suggesting that Language group performance varies by Stimulus type.

We conducted nine planned comparisons (simultaneous tests for general linear hypotheses using multcomp() package in R) to test whether each participant group differed for each stimulus type. For the [idi] stimuli we found that English speakers performed differently from both the Spanish ($z=9.04$, $p<.001$) and Learner groups ($z=9.04$, $p<.001$). The Spanish group did not differ from the Learner group ($z=.33$, $p=.99$). For the [iði] stimuli, we find that the learners differ from the English group ($z=3.66$, $p<.01$). The English group did not differ from the Spanish group ($z=1.8$, $p<.39$), nor did the Spanish group differ from the Learner group ($z=1.5$, $p=.60$) for this comparison. Finally, for the [iɾi] contrast we find no difference for English vs. Spanish ($z=.04$, $p=1$), English vs. Learner ($z=.38$, $p=.99$), or Spanish vs. Learner ($z=.318$, $p=.99$).

4.6.5.3 Mean ratings

Statistical analyses of mean ratings were performed using a Generalized Linear Mixed Effects Model (GLMM) (using R package lme4) with factors Language Group (English, Learner, Spanish) and Stimulus Type ([idi], [iði], and [iɾi]), and the Language Group x Stimulus Type interaction as fixed effects and subject as a random effect. We found no main effect of Language group or stimulus type. The Language group by stimulus type interaction did reach significance ($t=4.379$, $p<.001$).

We conducted nine planned comparisons (simultaneous tests for general linear hypotheses using multcomp() package in R) to test whether each participant group differed for each stimulus type. For the [idi] stimuli we found that English speakers performed differently from both the Spanish ($z=4.23$, $p<.001$) and Learner groups ($z=4.23$, $p<.001$). The Spanish group did not differ from the English group ($z=1.09$, $p=.82$). For the [iði] stimuli, we found that the English group did not differ from the Spanish group ($z=1.08$, $p<.82$) or the Learner group ($z=.33$, $p=.99$), nor did the Spanish group differ from the Learner group ($z=2.39$, $p=.10$) for this comparison. Finally, for the [iɾi] contrast we observed no difference for English vs. Spanish ($z=1.32$, $p=.67$), English vs. Learner ($z=.57$, $p=.98$), or Spanish vs. Learner ($z=1.88$, $p=.30$).

4.6.5.4 MMN amplitudes

MMN amplitudes were determined as the average RMS amplitudes in our window of interest (i.e. 310-410ms). Again, this time window was chosen because the vowel offset and consonant onset occurred at 160ms and the MMN is expected to occur about 150-250ms post-stimulus onset. Linear mixed effects modeling was employed to assess the reliability of the effects elicited by our experimental manipulation.

Analyses of MMN amplitude for the standards consisted of fixed effects Language Group (English, Learner, Spanish) and Condition ([idi] standard, [iði] standard, [iɾi] standard), as well as Language Group x Condition interaction and subject as random effect. These statistical analyses revealed no significant results, suggesting that the mean power elicited by standard stimuli did not differ for any of the language groups or conditions, nor did these factors interact.

Analyses of the MMN amplitude for the deviants consisted of fixed effects Language Group (English, Learner, Spanish) and Condition ([idi] Deviant-[iði] Block, [idi] Deviant-[iɾi] Block, [iði] Deviant-[idi] Block, [iði] Deviant-[iɾi] Block, [iɾi] Deviant-[idi] Block, and [iɾi] Deviant-[iɾi] Block), as well as Language Group x Condition interaction and subject as random effect. These statistical analyses also revealed no significant results.

4.6.6 Discussion

In this experiment we explored the hypothesis that listeners form equivalence classes on the basis of phonemes and that this grouping would be reflected in their behavioral and MEG responses to the stimuli [idi],[iði], and [iɾi]. Given this hypothesis we predicted that English speakers would show greater sensitivity (higher d' scores and greater MMN amplitudes) for the [idi]-[iði] and [iɾi]-[iði] contrasts, which are phonemic in English relative to the allophonic [idi]-[iɾi] contrast. Spanish speakers should perform equally well on the [idi]-[iɾi] and [iɾi]-[iði] contrasts, but show a decrement in performance for the contrast which is allophonic in Spanish [idi]-[iði].

Our behavioral findings partially support this hypothesis. We found that English speaking participants show comparable sensitivity to the [idi]-[iði] and [iði]-[iɾi] contrasts, whereas they are less sensitive to the contrast between [idi] and [iɾi]. This finding is consistent with our hypothesis, since [d] and [ɾ] are allophones of the same phoneme in English, whereas the contrasts between [d]-[ð] and [ɾ]-[ð] are phonemic. Spanish speakers were also found to perform better on the [iði]-[iɾi] relative to the [idi]-[iði] contrast. These behavioral findings replicate previous results from Boomershine et al. [2008], and can be taken as evidence that phonological relatedness among sounds reduces their perceptual similarity.

However, we found only partial support for our hypothesis for the Spanish speaker group. Based on our hypothesis we expected to observe comparably large d' scores for the two phonemic contrasts [idi]-[iɾi] and [iði]-[iɾi] and a smaller d' value for the allophonic contrast [idi]-[iði]. Visual inspection of the data shows that the observed trend is as expected based on our hypothesis. The Spanish group exhibited the least sensitivity for the sound pair which is allophonic in Spanish (i.e. [idi]-[iði]). One possible explanation for this decrement in d' for the [d]-[ɾ] contrast is the fact that, although a phonemic contrast in Spanish, these sounds never contrast on the surface in the language. This is due to the fact that /ɾ/ does not occur in word-initial position, and /d/ is realized as [ð] word-medially, as pointed out by Boomershine et al. [2008]. A second possibility is that, as realized by an English speaker for whom the pair are allophones of the same phoneme, the contrast has been neutralized to some degree.

With respect to our second question regarding L2 learners' ability to overcome the perceived similarity of L1 allophones and acquire new target language phonological relations, we found no difference between learners' ability to discriminate between L1 allophones and L1 phonemes. This finding suggests that the advanced L2 learners have acquired adequate knowledge of the L2 phonological system to distinguish between the sounds [d] and [ð] which are phonemic in English, but not Spanish.

The results of our identification and rating task showed that on average our experimental stimuli were identified as instances of the intended category. Not surprisingly, we found some differences, however. For example, English speakers were more likely to identify our tokens of [idi] and [iði] as instances of [idi] and [iði] than Spanish speakers and Spanish learners of English. This was expected given that these stimuli were made from recordings from an English speaker. There were no differences between the three listener groups for the [iɾɪ] stimuli, suggesting that English [iɾɪ] are readily identified with Spanish [iɾɪ]. Interestingly, when we look at ratings for those stimuli which were accurately identified, we find that the Spanish and Learner groups gave higher ratings than the English listeners across the board.

Let us now turn to the MEG results. Again, based on our hypothesis that listeners are establishing equivalence classes on the basis of phonemes, we expected to observe a larger MMN when a deviant is in contrast with the standard in the listener's native language, relative to when the standard and deviant are phonologically related. Thus, the MMN elicited by an [idi] stimulus was expected to be larger when presented in a [iði] block than when presented in a [iɾi] block for the English speaking group. The opposite was expected for the Spanish speaking group. For the [iði] deviants in an [idi] and [iɾi] block, we expected no difference in MMN amplitude for the English group, since [ð] and [d] are both phonemic in English. The MMN to [iði] in an [idi] block was expected to be smaller than the MMN elicited for [iði] deviants among [iɾi] standards. Finally, for [iɾi] stimuli, we expected a larger MMN for English speakers when [iɾi] was a deviant in an [iði] block, than in an [idi] block. For Spanish speakers the MMN elicited by [iɾi] deviants were expected to be comparable in both blocks.

We found no statistical support for our hypothesis. Visual inspection of our MEG data shows that trends in the predicted direction are only observed in the case of [iɾi] stimuli. Unexpectedly, the MMN amplitudes elicited by [idi] stimuli in an [iɾi] block do not appear to differ much from [idi] in an [iði] block for English speakers. For Spanish speakers, we observe a difference for these two conditions where no differences was expected. For the [iði] stimuli, we observed a trend toward an unexpected difference in MMN amplitude for English speakers and no difference or a difference in the opposite direction for Spanish listeners.

How can we explain these unexpected findings? I can think of three possibilities, each of which can explain different portions of the observed native speaker data, but none that provide an explanation in full. I will discuss each of these in turn.

One possibility is that some deviants are more salient to listeners than others due to their phonotactic illegality. If phonotactic illegality were driving the effect, we would expect larger MMN amplitudes on conditions which involve [idi] stimuli, since [d] is not the allophone that is to occur intervocalically in either language. This explanation could account for the large MMN observed for Spanish speakers when [idi] is presented in an [iði] block and when [iði] is presented in an [idi] block, as well as the larger MMN for [idi] in an [iɾi] block for English speakers. Yet, not all conditions involving [idi] stimuli elicit a large MMN. For example, when [idi] is a deviant among a block of [iɾi] stimuli, Spanish speakers do not show a large MMN as would be predicted. Likewise, [idi] deviants in a block with [iði] as standard, elicit a relatively small MMN for English speakers.

Another possibility is that the observed effects are driven by low level acoustic differences across the stimulus types. That is, participants could be performing the task by tracking the difference in consonant duration. These differences were unavoidable as /ɾ/ is a shorter consonant than both /d/ and /ð/. If the MMN were tracking duration differences we would expect the largest MMNs for conditions which involve [iɾi] stimuli. This story, like the others, provides only partial coverage of the data. For example, we could claim that the large MMN for [idi] deviants among [iði] standards for English speakers is driven by the large differences between the duration of the intervocalic consonants. However, we should then be surprised to find differences across the language groups. For example, a relatively small MMN is elicited for the same condition in Spanish speakers.

A final possibility is that deviations from the expected pattern are due to the

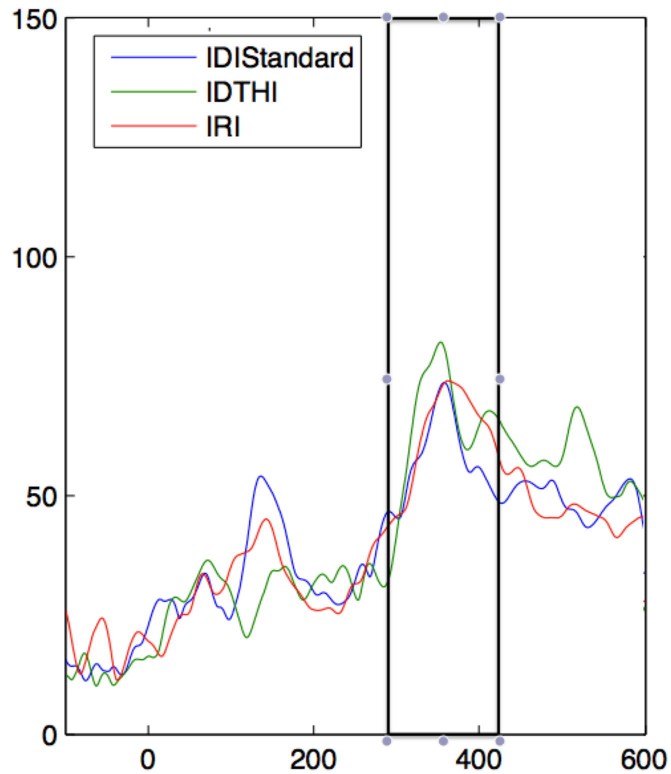

Figure 4.8: Grand averaged RMS amplitudes for [idi] stimuli for the learner group. Responses to *standards* are shown in blue, *deviant* responses are shown in red and green.

lack of surface contrast between [d] and [ɾ] in Spanish. This explanation provides an account for why [idi] in an [iɾi] block elicits a relatively small MMN. However, if we are to follow this line of reasoning, it is not clear why [iɾi] deviants in an [idi] block should illicit such a large MMN. This explanation is also specific to Spanish and so it will not help explain the puzzling English data (i.e. [idi] in an [iɾi] block, and [iði] in an [idi] block).

Despite the fact that the MEG responses observed for native speakers of Spanish and English did not conform to our hypotheses, the MEG data for the advanced Spanish Learners of English looks more promising. With respect to the learner data we asked whether native speakers of Spanish could acquire new phonological

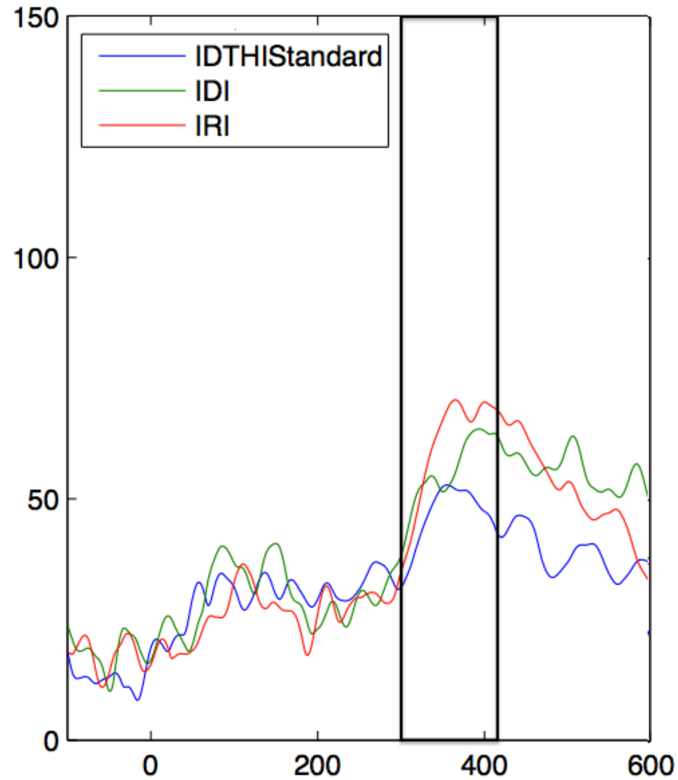

Figure 4.9: Grand averaged RMS amplitudes for [iði] stimuli for the learner group. Responses to *standards* are shown in blue, *deviant* responses are shown in red and green.

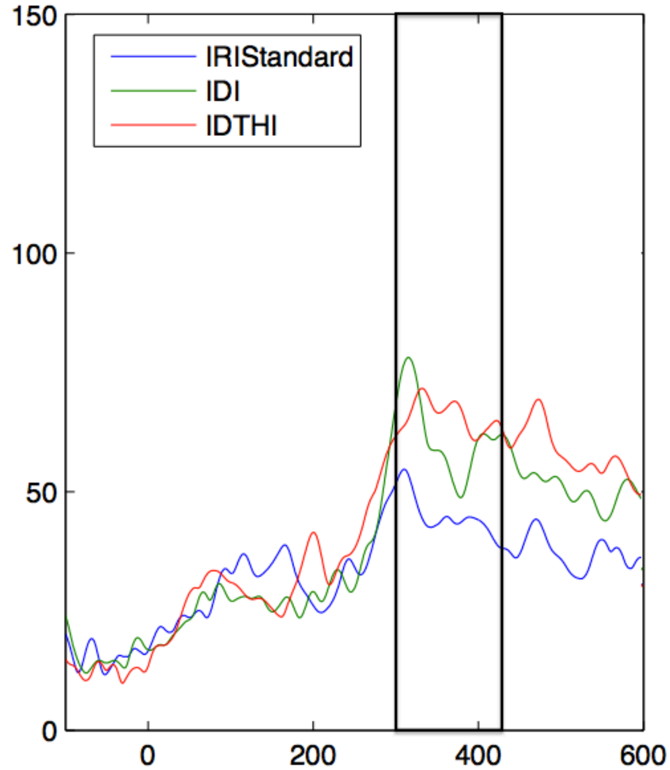

Figure 4.10: Grand averaged RMS amplitudes for [iri] stimuli for the learner group. Responses to *standards* are shown in blue, *deviant* responses are shown in red and green.

relationships. Unlike either of the native speaker groups, we observed large MMNs of comparable size for each deviant type, which can be observed from the grand averaged wave forms shown in Figures 4.8, 4.9, and 4.10. This could be interpreted as evidence that these L2 learners have been able to learn that [d] and [ð] are allophones of different phonemes in the target language. Of course, this pattern should be taken with caution since the native speakers data only partially confirmed our initial hypothesis, and also because the large MMNs where likewise observed for the native speaker group for both the [idi] deviants in an [iði] block and the [iði] deviants presented among [idi] standards.

4.6.7 Conclusion

In this chapter we used both behavioral and electrophysiological methods to investigate whether phonological relatedness between sounds within a language has an effect on their perceived similarity. We also aimed to provide evidence as to whether L2 learners can establish new phonological relationships. Our behavioral and MEG results (at best) provide only partial support for the role of phonological relatedness in perceived similarity. With respect to the learning of new phonological relationships, our data looks more promising. We show that L2 learners are able to more readily discriminate contrasts which are allophonic in their native language but phonemic in the target language than are Spanish listeners. Our MEG findings may also reflect this ability.

Chapter 5

Conclusion

5.1 Overview

This dissertation set out to explore the role of perceptual similarity in second language speech perception and phonological development. Second language learners have been shown time and again to differ from native speakers of the target language with respect to their perception and production of target language contrasts. A surprising finding in the L2 literature, and one that put the Contrastive Analysis approach into disfavor (Lado [1957]), is that the difficulties that second language learners experience cannot be solely due to differences in segmental inventory. L2 learners have been shown to perform differently with respect to various nonnative contrasts which are absent from their native language inventory. To account for these differences in perception across various novel target language contrasts, researchers have turned to the notion of similarity.

In this dissertation I explored the predictions of two theoretical approaches to similarity comparison in the second language, and asked:

1. How should L2 sound similarity be characterized/measured?

2. What is the nature of the representations that guide sound similarity?

3. To what extent can the influence of the native language be overcome?

5.2 Synthesis of empirical findings

In Chapters 2 to 4, I presented original empirical evidence from a series of behavioral experiments, and an MEG study in advanced Spanish late-learners of English, as well as a series of model simulations. The specific findings were summarized in the respective chapters. Here I synthesize the main empirical findings to address the research questions above.

5.2.1 Feature availability is neither necessary, nor sufficient

In Chapter 2, we tested a legos (featural) approach to sound similarity. Given a distinctive feature analysis of Spanish and English vowels, we investigated the hypothesis that feature availability in the L1 grammar constrains which target language segments will be accurately perceived and acquired by L2 learners (Brown [1998], Brown [2000]). While we can't rule out the possibility that existing distinctive feature analyses of English or Spanish are incorrect, our evidence suggests that second language acquisition of phonology is not limited by the phonological features used by the native language grammar, nor is the presence/use of a particular phonological feature in the native language grammar sufficient to trigger redeployment. We take these findings to imply that feature availability is neither a necessary, nor a sufficient condition to predict learning outcomes.

5.2.2 Rulers alone won't work

In Chapter 3, we extended a computational model proposed by Feldman et al. [2009] to nonnative speech perception in order to investigate whether a sophisticated rulers approach to sound similarity can better explain existing interlingual identification and discrimination data from Spanish monolinguals, and advanced L1 Spanish late-learners of English, respectively. The model assumes that acoustic distributions of sounds control listeners' ability to discriminate a given contrast. We found that, while the model succeeded in imitating certain aspects of human behavior, the model at present is incomplete and would have to be extended in various ways to capture several aspects of nonnative and L2 speech perception.

5.2.3 The acquisition of new phonological relations

In Chapter 4 we explored whether the phonological relatedness among sounds in the listener's native language impacts the perceived similarity of those sounds in the target language. Listeners were expected to be more sensitive to the contrast between sound pairs which are allophones of different phonemes than sound pairs which are allophones of the same phoneme in their native language. Moreover, we hypothesized that L2 learners would experience difficulty perceiving and acquiring target language contrasts between sound pairs which are allophones of the same phoneme in their native language. Our results suggest that phonological relatedness may influence perceived similarity on some tasks (Boomershine et al. [2008]), but does not seem to cause long-lasting perceptual difficulty in advanced L2 learners.

5.3 Theoretical Implications

This dissertation reviews existing evidence and presents novel empirical findings that pose new demands of two prominent views of L2 phonology and speech perception, and has three main theoretical implications.

5.3.1 Representations and processes in L2 phonology

Elaborating an explanatory adequate theory of L2 speech perception and phonological development at an algorithmic level of description (Marr [1982]) requires that we characterize both the representational format the computational system uses, as well as the computations that act on those representations to generate the observed behavior. Existing models of L2 phonology and speech perception differ in their degree of explicitness regarding both aspects of the system. On the one hand, the rulers approach is vague about the nature of the representations involved in similarity comparison, but the approach's assumption that spatial representations are appropriate descriptions has the advantage of being easily amenable to many sophisticated mathematical techniques which can provide plausible stories about the process involved in comparing spatial representations of the sort posited.

The legos approach, on the other hand, provides a detailed characterization of the nature of the representations involved (which have their basis in phonological theory). What remains to be detailed from a legos perspective is precisely how the posited featural representation might be used in the process of similarity comparison. One of the main points of this dissertation is that neither of the existing approaches have made sufficiently explicit what is involved in similarity-based comparison and how similarity should be measured. Theoretical approaches to similarity comparison in second language phonology, therefore, need to be revisited in order to further understand the nature of the representations which guide sound similarity and computations that act on them.

5.3.2 L1 constraints on L2 phonology

Our feature availability and phonological relatedness findings can be brought to bear on the debate around the redeployment of L1 features and the acquisition of new target language structures. In contrast with previous reports (Brown [1998], Brown [2000], Matthews [1997], Atkey [2002]), our findings that advanced second language learners appear to acquire a phonological features (i.e. [+/-tense/lax]), which are not used in the native language grammar suggests that the acquisition of the second language sound system is not limited by the availability of phonological features. Similarly, we found that the phonological relatedness between target language sounds in the listeners L1 (i.e. allophonic pair [d]-[ð]) does not appear to cause long-lasting interference in speech perception, as has been reported for production (Lado [1957], Hammerly [1982], Hardy [1993]).

In sum, our data suggests that although advanced L2 learners do not perform at native levels on all non-native contrasts, given the appropriate type of evidence, L2 learners can acquire new phonological contrasts. This general conclusion is quite surprising in light of numerous reports that the phonological system of the native language strongly constrains L2 learning,(Pallier et al. [1997], Bosch et al. [2000], Pallier et al. [2001], Sebastián-Gallés and Soto-Faraco [1999], Navarra et al. [2005]) even when the learner has had extensive experience and ample opportunity for development. Future work will have to investigate the cause of differences of this sort.

5.4 Recommendations for future research

5.4.1 Developing explicit models of similarity comparison

As discussed in the previous section, this dissertation has highlighted the need for developing explicit models of similarity comparison, which will involve characterizing both the representational format the computational system uses, as well as the computations that act on those representations to generate the observed behavior.

As a general methodological point, I believe that using computational modeling as a tool to study similarity is one step in this direction. The computational modeling portion of this dissertation was a first attempt at making more precise how similarity could be determined under certain representational assumptions. Although our model does not capture many aspects of the data at present, the general technique shows promise for elucidating how it is that people represent the similarity between the sounds of their first and second language. Future work should also try to construct an explicit model of sound similarity comparison on the basis of phonological features.

5.4.2 Describing changes with L2 proficiency

The research which I have described in this dissertation has focused on relatively advanced late-learners of a second language. This raises the question of whether the representations that guide sound similarity vary with learners' proficiency. One possibility is that learners rely increasingly on abstract phonological representations as they gain experience with the target language and develop a substantial lexicon. The answer to this question will begin to provide a more complete picture of second language phonological development across various proficiency levels.

5.4.3 Articulated theory of learning mechanisms

In a similar vein, another future direction lies in providing a detailed account of how the learner uses input and prior knowledge to make inferences about the nature of the target language. One hypothesis is that this learning is error-driven. In the coming years I hope to provide an articulated theory of the mechanisms by which plasticity is exhibited.

5.4.4 Taking a closer look at L2 input

Finally, while we have made no direct comparison between Spanish-Catalan bilinguals and Spanish late-learners of English, the picture seems to be that the Spanish-Catalan bilinguals actually fare worse with respect to the types of nonnative contrasts which have been reported in the literature. It is, thus, interesting to consider what can be causing some L2 contrasts to be so difficult.

One possibility is that the Spanish-Catalan bilinguals are exposed to more variability with respect to the difficult contrasts as a result of living in a bilingual community. This variability may slow/prevent the acquisition of certain non-native contrasts. We have speculated that the decrement in performance we observed for the /ɑ/-/æ/ contrast may have been due to exposure to similar types of variability. In sum, more emphasis needs to be placed on characterizing the properties of the actual input to L2 learning and exploring the way that variability in the input may be influencing outcomes.

5.5 Conclusion

In this dissertation I have highlighted the importance of sound similarity comparisons in L2 phonology as a means of accounting for the ease and difficulty with which L2 learners perceive and acquire various nonnative contrasts. I have provided original evidence that bears on assumptions about the representations and processes involved in assessing the similarity of L1 and L2 sounds. On the basis of those findings, I have argued that existing models have not been adequately explicit about the nature of the representations and processes involved in similarity-based comparisons of L1 and L2 sounds. More generally, I have attempted to articulate what I see as a desirable target for an explanatorily adequate theory of cross-language influence in L2 phonology.

Appendix A

Stimuli used in Experiments 1 & 2

Word stimuli by Contrast

	/o/	/u/		/a/	/æ/		/i/	/ɪ/
1	blow	blue	1	odd	add	1	bean	bin
2	boat	boot	2	bond	band	2	beat	bit
3	coal	cool	3	bottle	battle	3	cheap	chip
4	crow	crew	4	block	black	4	deep	dip
5	float	flute	5	blond	bland	5	feel	fill
6	glow	glue	6	con	can	6	feet	fit
7	hope	hoop	7	cop	cap	7	jean	gin
8	known	noon	8	cot	cat	8	lead	lid
9	mode	mood	9	hot	hat	9	leak	lick
10	pole	pool	10	mop	map	10	leap	lip
11	road	rude	11	pot	pat	11	leave	live
12	role	rule	12	rock	rack	12	reach	rich
13	show	shoe	13	rot	rat	13	scene	sin
14	soap	soup	14	sock	sack	14	seek	sick
15	toll	tool	15	shock	shack	15	sheep	ship
16	wrote	root	16	top	tap	16	sleep	slip

Nonword stimuli by contrast

	/o/	/u/		/a/	/æ/		/i/	/ɪ/
1	botʃ	butʃ	1	baf	bæf	1	tʃim	tʃIm
2	bof	buf	2	bav	bæv	2	div	dIv
3	bok	buk	3	fam	fæm	3	gid	gId
4	fop	fup	4	fap	fæp	4	gip	gIp
5	gob	gub	5	gam	gæm	5	kib	kIb
6	kog	kug	6	gan	gæn	6	kig	kIg
7	log	lug	7	kag	kæg	7	mip	mIp
8	lov	luv	8	maf	mæf	8	nib	nIb
9	mof	muf	9	maz	mæz	9	pib	pIb
10	moz	muz	10	nas	næs	10	riv	rIv
11	nob	nub	11	naz	næz	11	sig	sIg
12	nog	nug	12	paf	pæf	12	ʃig	ʃIg
13	pon	pun	13	san	sæn	13	tid	tId
14	pov	puv	14	ʃaθ	ʃæθ	14	tig	tIg
15	rok	ruk	15	taf	tæf	15	θig	θIg
16	tof	tuf	16	zat	zæt	16	θip	θIp

Appendix B

Language Background Questionnaire

TO BE COMPLETED BY EXPERIMENTER		
DATE	TIME	PART ID #
EXPT NAME	GROUP	SCRIPT#/LIST
CONSENT FORM:	COMPLETE	

TO BE COMPLETED BY THE PARTICIPANT		
AGE	SEX MALE FEMALE	OCCUPATION

IF YOU ARE A STUDENT, PLEASE INDICATE YOUR YEAR	FRESHMAN SENIOR	SOPHOMORE GRADUATE	JUNIOR N/A

WHAT WAS THE FIRST LANGUAGE YOU LEARNED AS A CHILD?	

PLEASE ANSWER THE FOLLOWING QUESTIONS REGARDING YOUR EXPERIENCE WITH BOTH **SPANISH** AND **ENGLISH**.

ENGLISH	SPANISH
I STARTED LEARNING **ENGLISH** WHEN I WAS	I STARTED LEARNING **SPANISH** WHEN I WAS
I STARTED LEARNING **ENGLISH**: AT HOME IN SCHOOL IN COLLEGE/UNIVERSITY IN THE COMMUNITY OTHER (*PLEASE SPECIFY*)	I STARTED LEARNING **SPANISH**: AT HOME IN SCHOOL IN COLLEGE/UNIVERSITY IN THE COMMUNITY OTHER (*PLEASE SPECIFY*)
I STARTED LEARNING **ENGLISH**: IN AN ENGLISH-SPEAKING COUNTRY IN A SPANISH-SPEAKING COUNTRY	I STARTED LEARNING **SPANISH**: IN AN ENGLISH-SPEAKING COUNTRY IN A SPANISH-SPEAKING COUNTRY

I HAD FORMAL INSTRUCTION IN **ENGLISH** IN GRADE SCHOOL FOR IN COLLEGE/UNIVERSITY FOR OTHER FOR	I HAD FORMAL INSTRUCTION IN **SPANISH** IN GRADE SCHOOL FOR IN COLLEGE/UNIVERSITY FOR OTHER FOR
I HAVE LIVED IN A **ENGLISH-** SPEAKING COUNTRY FOR YEARS MONTHS	I HAVE LIVED IN A **SPANISH-** SPEAKING COUNTRY FOR YEARS MONTHS

PLEASE USE THE FOLLOWING SCALES TO RATE YOUR PROFICIENCY IN BOTH ENGLISH AND SPANISH.

ENGLISH	**Minimal**	**---**	**---**	**---**	**---**	**---**	**---**	**---**	**---**	**Native-like**
Speaking	1	2	3	4	5	6	7	8	9	10
Pronunciation	1	2	3	4	5	6	7	8	9	10
Listening	1	2	3	4	5	6	7	8	9	10
Reading	1	2	3	4	5	6	7	8	9	10
Writing	1	2	3	4	5	6	7	8	9	10
Grammar	1	2	3	4	5	6	7	8	9	10

SPANISH	**Minimal -- Native-like**									
Speaking	1	2	3	4	5	6	7	8	9	10
Pronunciation	1	2	3	4	5	6	7	8	9	10
Listening	1	2	3	4	5	6	7	8	9	10
Reading	1	2	3	4	5	6	7	8	9	10
Writing	1	2	3	4	5	6	7	8	9	10
Grammar	1	2	3	4	5	6	7	8	9	10

IF YOU SPEAK ANY ADDITIONAL LANGUAGES (OTHER THAN ENGLISH AND SPANISH), PLEASE LIST THEM BELOW ALONG WITH YOUR PROFIENCY IN EACH.

OTHER LANGUAGES	PROFICIENCY

PARA USO DE EL/LA INVESTIGADOR(A)		
FECHA	HORA	No DE IDENTIFICACIÓN
EXPERIMENTO	GRUPO	No DE SCRIPT/LISTA
FORMULARIO DE CONSENTIMIENTO:		

PARA USO DE EL/LA PARTICIPANTE		
EDAD	SEXO MASC FEM	OCCUPACIÓN
SI ES ESTUDIANTE, MARQUE SU AÑO.	FRESHMAN SOPHOMORE JUNIOR SENIOR POSTGRADO N/A	
¿CUÁL ES SU IDIOMA MATERNO?		

POR FAVOR RESPONDA LAS SIGUIENTES PREGUNTAS CON RESPECTO A SU EXPERIENCIA CON EL ESPAÑOL Y EL INGLÉS.

INGLÉS	ESPAÑOL
EMPECÉ A APRENDER **INGLÉS** CUANDO TENÍA AÑOS	EMPECÉ A APRENDER **ESPAÑOL** CUANDO TENÍA AÑOS
EMPECÉ A APRENDER **INGLÉS**: EN CASA EN LA ESCUELA EN LA UNIVERSIDAD EN LA COMUNIDAD OTRA (*ESPECIFIQUE*)	EMPECÉ A APRENDER **ESPAÑOL**: EN CASA EN LA ESCUELA EN LA UNIVERSIDAD EN LA COMUNIDAD OTRA (*ESPECIFIQUE*)
EMPECÉ A APRENDER **INGLÉS**: EN UN PAÍS ANGLOPARLANTE EN UN PAÍS HISPANOPARLANTE	EMPECÉ A APRENDER **ESPAÑOL**: EN UN PAÍS ANGLOPARLANTE EN UN PAÍS HISPANOPARLANTE

RECIBÍ INSTRUCCIÓN FORMAL EN **INGLÉS** EN	RECIBÍ INSTRUCCIÓN FORMAL EN **ESPAÑOL** EN
LA ESCUELA POR AÑOS LA UNIVERSIDAD POR AÑOS OTRA POR AÑOS	LA ESCUELA POR AÑOS LA UNIVERSIDAD POR AÑOS OTRA POR AÑOS
EN TOTAL, HE VIVIDO AÑOS MESES EN UN LOS EEUU U OTROS PAÍSES DE HABLA INGLESA.	EN TOTAL, HE VIVIDO AÑOS MESES EN PAÍSES DE HABLA HISPANA.

POR FAVOR USE LAS SIGUIENTES ESCALAS PARA INDICAR SU NIVEL DE DOMINO EN INGLÉS Y ESPAÑOL.

INGLÉS	**Mínimo -- Casi nativo**									
Habla	1	2	3	4	5	6	7	8	9	10
Pronunciación	1	2	3	4	5	6	7	8	9	10
Comprensión del habla	1	2	3	4	5	6	7	8	9	10
Lectura	1	2	3	4	5	6	7	8	9	10
Escritura	1	2	3	4	5	6	7	8	9	10
Gramática	1	2	3	4	5	6	7	8	9	10

ESPAÑOL **Mínimo --- Casi nativo**

Habla	1	2	3	4	5	6	7	8	9	10
Pronunciación	1	2	3	4	5	6	7	8	9	10
Comprensión del habla	1	2	3	4	5	6	7	8	9	10
Lectura	1	2	3	4	5	6	7	8	9	10
Escritura	1	2	3	4	5	6	7	8	9	10
Gramática	1	2	3	4	5	6	7	8	9	10

SI HABLA OTRO IDIOMA (ADEMÁS DEL INGLÉS O EL ESPAÑOL), POR FAVOR INDÍQUELO ABAJO JUNTO CON SU NIVEL DE DOMINO.

OTRO IDIOMA	NIVEL DE DOMINO

Bibliography

Abramson, A. and Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the sixth international congress of phonetic sciences*, pages 569–573. Academia Prague, Czechoslovakia.

Almeida, D. (2009). *Form, meaning and context in lexical access: MEG and behavioral evidence*. PhD thesis, University of Maryland.

Archibald, J. (1993). *Language learnability and L2 phonology: The acquisition of metrical parameters*, volume 19. Springer.

Atkey, S. (2002). The acquisition of non-native segmental contrasts: A look at English speakers' acquisition of Czech palatal stops. In *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech, Klagenfurt, Austria: University of Klagenfurt*, pages 22–31.

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

Barrios, S. L., Jiang, N., and Idsardi, W. J. (In preparation). Interlanguage mappings.

Best, C., McRoberts, G., and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners native phonological system. *The Journal of the Acoustical Society of America*, 109:775.

Best, C., McRoberts, G., and Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14:345–360.

Best, C., Traill, A., Carter, A., Harrison, K., and Faber, A. (2003). Xóõ click perception by English, Isizulu, and Sesotho listeners. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 853–856.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Goodman, J. and Nusbaum, H., editors, *The development of speech perception: The transition from speech sounds to spoken words*, pages 167–224. MIT Press.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In Strange, W., editor, *Speech perception and linguistic experience: Issues in cross-language research*, pages 171–204. MIT Press.

Best, C. T. and Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of phonetics*, 20(3):305–330.

Best, C. T. and Tyler, M. D. (2007). Nonnative and second-language speech perceptoin: Commonalities and complementarities. In Bohn, O.-S. and Munro, M. J., editors, *Language experience and second language speech learning: In honor of James Emil Flege*, pages 13–34. John Benjamins.

Biederman, I. et al. (1987). Recognition-by-components: A theory of human image understanding. *Psychological review*, 94(2):115–147.

Boersma, P. and Weenink, D. (2009). Praat: doing phonetics by computer (version 5.1. 05)[computer program]. retrieved may 1, 2009.

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

Boomershine, A., Currie Hall, K., Hume, E., and Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In *Contrast in Phonology: Theory, Perception, Acquisition*, volume 13 of *Phonology & Phonetics*, pages 145–171. Mouton de Gruyter.

Bosch, L., Costa, A., and Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*.

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

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. *Second language acquisition and linguistic theory*, pages 4–63.

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, 14(2):136–193.

Bybee, J. (2003). *Phonology and language use*, volume 94. Cambridge University Press.

Cebrian, J. (2008). The effect of perceptual factors in the acquisition of an L2 vowel contrast. In *Contrast in Phonology: Theory, Perception, Acquisition*, volume 13 of *Phonology & Phonetics*, pages 303–321. Mouton de Gruyter.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. Harper and Row.

Clements, G. N. (1985). The geometry of phonological features. *Phonology yearbook*, 2:225–252.

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

De Cheveigné, A. and Simon, J. Z. (2007). Denoising based on time-shift pca. *Journal of neuroscience methods*, 165(2):297–305.

De Cheveigné, A. and Simon, J. Z. (2008). Sensor noise suppression. *Journal of Neuroscience Methods*, 168(1):195–202.

Eckman, F. R., Elreyes, A., and Iverson, G. K. (2001). Allophonic splits in L2 phonology: The question of learnability. *International Journal of English Studies*, 1:21–51.

Eckman, F. R., Elreyes, A., and Iverson, G. K. (2003). Some principles of second language phonology. *Second Language Research*, 19(3):169–208.

Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological review*, 116(4):752.

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., and Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, 127(3):427–438.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230.

Flege, J. and MacKay, I. (2004). Perceiving vowels in a second language. *Studies in second language acquisition*, 26(1):1–34.

Flege, J., Munro, M., and Fox, R. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, 95(6):3623–3641.

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental Psychology*, 43(3):701–731.

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

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In Schiller, N. O. and Meyer, A. S., editors, *Phonetics and Phonology in Language Comprehension and Production*, pages 319–355. Mouton de Gruyter.

Forster, K. and Forster, J. (2003). Dmdx: A windows display program with millisecond accuracy. *Behavior Research Methods*, 35(1):116–124.

Gentner, D., Loewenstein, J., and Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, 95(2):393.

Godinez Jr, M. (1978). A comparative study of some romance vowels. *UCLA Working Papers in Phonetics*, 41:3–19.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological review*, 105(2):251.

Goto, H. (1971). Auditory perception by normal Japanese adults with sounds "l" and "r". *Neuropsychologia*, 9:317–323.

Gottfried, T. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*.

Grieser, D. and Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, 25(4):577.

Hallé, P. A., Best, C. T., and Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, 27(3):281–306.

Hammerly, H. (1982). Contrastive phonology and error analysis. *International Review of Applied Linguistics in Language Teaching*, 20(1):17–34.

Hancin-Bhatt, B. (1994). Segment transfer: a consequence of a dynamic system. *Second Language Research*, 10(3):241–269.

Hardy, J. (1993). Phonological learning and retention in second language acquisition. *Confluence: Linguistics, L*, 2:235–247.

Harnsberger, J. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetics*, 29:303–327.

Herd, W. (2011). *The perception and production training of /d/, /ɾ/, and /r/ in L2 Spanish: Behavioral, psycholinguistic, and neurolinguisitic evidence*. PhD thesis, University of Kansas.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, 97:3099.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological review*, 93(4):411–428.

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.

Huang, T. (2001). The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. *Working papers in Linguistics - Ohio State University Department of Linguistics*, pages 23–42.

Huang, T. and Johnson, K. (2011). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, 67(4):243–267.

Iverson, P. and Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, 97(1):553–562.

Jakobson, R., Fant, G., and Halle, M. (1952). *Preliminaries to speech analysis. The distinctive features and their correlates*. MIT Press, Cambridge MA.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In Johnson, K. and Mullennix, J. W., editors, *Talker variability in speech processing*, pages 145–165. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Johnson, K. (2005). Speaker normalization in speech perception. In *The handbook of speech perception*, pages 363–389. Blackwell.

Johnson, K. and Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, 38(1):127–136.

Kazanina, N., Phillips, C., and Idsardi, W. J. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103(30):11381–11386.

Klatt, D. H. (1989). Review of selected models of speech perception. In *Lexical Representation and Process*, pages 169–226. The MIT Press.

Kronrod, Y., Coppess, E., and Feldman, N. H. (2012). A unified model of categorical effects in consonant and vowel perception. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044):606–608.

Kuhl, P. K. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, 50(2):93–107.

LaCharité, D. and Prévost, P. (1999). The role of L1 and teaching in the acquisition of English sounds by francophones. In *Proceedings of BUCLD*, volume 23, pages 373–385.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.

Lehn, W. and Slager, W. R. (1959). A contrastive study of Egyptian Arabic and American English: The segmental phonemes. *Language Learning*, 9:25–33.

MacKain, K. S., Best, C. T., and Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2(4):369–390.

Macmillan, N. and Creelman, C. (2005). *Detection Theory: A Users Guide*. Lawrence Erlbaum Associates.

Marr, D. (1982). *Vision*. Freeman, San Francisco.

Martínez Celdrán, E. (1995). En torno a las vocales del español: análisis y reconocimiento. *Estudios de Fonética experimental*, 7:195–218.

Matthews, J. (1997). The influence of pronunciation training on the perception of second language contrasts. *International review of applied linguistics*, 35(2):223–229.

Michaels, D. (1973). Sinhalese sound replacement and feature hierarchies. *Linguistics*, pages 14–22.

Michaels, D. (1974). Sound replacements and phonological systems. *Linguistics*, pages 69–81.

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

Monahan, P. J. and Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalisation. *Language and cognitive processes*, 25(6):808–839.

Näätänen, R. (1992). *Attention and brain function*. Psychology Press.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., and Kimmo, A. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615):432–434.

Näätänen, R., Pakarinen, S., Rinne, T., and Takegata, R. (2004). The mismatch negativity (mmn): towards the optimal paradigm. *Clinical Neurophysiology*.

Näätänen, R. and Picton, T. (1987). The n1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24(4):375–425.

Navarra, J., Sebastián-Gallés, N., and Soto-Faraco, S. (2005). The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5):912.

Navarro, D. J., Griffiths, T. L., Steyvers, M., and Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of mathematical Psychology*, 50(2):101–122.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology*, 115(1):39–57.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1):97–113.

Pallier, C., Bosch, L., and Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64(3):B9–B17.

Pallier, C., Colomé, A., and 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.

Pater, J. (2003). The perceptual acquisition of Thai phonology by English speakers: task and stimulus effects. *Second Language Research*, 19(3):209–223.

Patterson, D. and Connine, C. (2001). Variant frequency in flap production. *Phonetica*, 58(4):254–275.

Pegg, J. and Werker, J. F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America*, 102:3742–3753.

Peperkamp, S., Pettinato, M., and Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In Beachley, B., Brown, A., and Conlin, F., editors, *Proceedings of the 27th Annual Boston University Conference on Language Development*, volume 2, pages 650–661. Cascadilla Press.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience*, 12(6):1038–1055.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, 46(2-3):115–154.

Piske, T., Flege, J., MacKay, I., and Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, 59(1):49–71.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, 89:2961.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & psychophysics*, 52(1):37–52.

Polka, L. and Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, 100:577.

Polka, L. and Werker, J. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2):421.

Quilis, A. and Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal. *Estudios de fonética*, 1:159–252.

Rice, K. and Avery, P. (1991). On the relationship between laterality and coronality. *Phonetics and phonology*, 2:101–124.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of verbal learning and verbal behavior*, 14(6):665–681.

Ritchie, W. (1968). On the explanation of phonic interference. *Language Learning*, 18(3-4):183–197.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive psychology*, 16(3):371–416.

Sebastián-Gallés, N. and Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72(2):111–123.

Sheldon, A. and Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(03):243–261.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27(2):125–140.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820):1317–1323.

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*, 111:1872.

Stockwell, R. and Bowen, J. D. (1965). *The sounds of English and Spanish*. The University of Chicago Press.

Strange, W., Levy, E., and Lehnhoff, J. (2004). Perceptual assimilation of French and German vowels by American English monolinguals: Acoustic similarity does not predict perceptual similarity. *The Journal of the Acoustical Society of America*, 115:2606.

Studdert-Kennedy, M. (1980). Speech perception. *Language and Speech*, 23(1):45–66.

Torgerson, W. S. (1965). Multidimensional scaling of similarity. *Psychometrika*, 30(4):379–393.

Tversky, A. (1977). Features of similarity. *Psychological review*, 84(4):327–352.

Waltmunson, J. (2005). *The relative difficulty of Spanish/t, d/, trill and tap by L1 English speakers: Auditory and acoustic methods of defining pronunciation accuracy*. PhD thesis, Doctoral dissertation, University of Massachusetts, Amherst.

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

Weinreich, U. (1953). *Languages in contact: Findings and problems*. Mouton.

Werker, J., Gilbert, J., Humphrey, K., and Tees, R. (1981). Developmental aspects of cross-language speech perception. *Child development*, pages 349–355.

Werker, J. and Lalonde, C. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental psychology*, 24(5):672.

Werker, J. and Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, 37:35–44.

Werker, J. and Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63.

Whalen, D., Best, C., and Irwin, J. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, 25:501–528.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, 7(3):357–369.