

8-2016

The role of experience in processing foreign-accented speech

Fernando Llanos Lucas
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations



Part of the [Linguistics Commons](#)

Recommended Citation

Lucas, Fernando Llanos, "The role of experience in processing foreign-accented speech" (2016). *Open Access Dissertations*. 799.
https://docs.lib.purdue.edu/open_access_dissertations/799

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

**PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Fernando Llanos Lucas

Entitled

THE ROLE OF EXPERIENCE IN PROCESSING FOREIGN-ACCENTED SPEECH

For the degree of Doctor of Philosophy



Is approved by the final examining committee:

Alexander L Francis

Co-chair

Juana Gil Fernandez

Robert M Hammond

Co-chair

Amanda Seidl

Alejandro Cuza

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Rober M Hammond

Approved by: Madeleine M Henry

Head of the Departmental Graduate Program

7/1/2016

Date

THE ROLE OF EXPERIENCE IN PROCESSING FOREIGN-ACCENTED SPEECH

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Fernando Llanos Lucas

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2016

Purdue University

West Lafayette, Indiana

For all the Gods and Goddesses

To my cat

ACKNOWLEDGEMENTS

I would like to express my gratitude to my academic advisors, Dr. Alexander L. Francis and Dr. Robert R. Hammond, for their guidance, support and mentoring during my time at Purdue. I am especially indebted to Dr. Alexander Francis for bringing me the opportunity of learning from him and showing me the paths of science, perseverance and patience. Dr. Francis is one of the brightest and most tolerant minds that I had the pleasure to meet, and he taught me one of the most valuable pieces of knowledge that I learned in my career, which is that science is not about being right or wrong but about understanding why we are right or wrong.

I am also deeply indebted to Dr. Keith R. Kluender for his wise council, blind support, scientific insight, and excellent sense of humor. Dr. Kluender is one of those persons in which the classic Isaac Newton's quote "If I have seen further, it is by standing on the shoulders of giants" makes perfect sense, and I would like to express my gratitude to this 'giant' for treating me like a peer since the first moment that we started working together.

I would also like to express my gratitude to Dr. Amanda Seidl, who is one of the most precise and thorough scientific minds that I have ever meet, and Dr. Alejandro Cuza, for his scientific council, endless support and valuable friendship. I am also indebted to Dr. Juana Gil Fernandez for all her support and help with the experiments that I conducted

at her lab in Madrid, and to Steve Wehrle for helping me to keep my ‘mental sanity’ in good shape. None of my academic achievements would have been possible without their guidance and support.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS.....	x
ABSTRACT.....	xi
CHAPTER 1. INTRODUCTION	1
1.1 Preliminaries: Speech Variability and Phonetic Constancy	1
1.2 Deciphering a Foreign Accent: Statement of the Problem.....	3
1.2.1 The phonetic relaxation hypothesis	5
1.2.2 The phonetic calibration hypothesis	7
1.2.3 The phonetic switching hypothesis.....	7
1.3 Foreignness and Nativeness in the Communicative Framework	8
1.4 Outline of the Present Study.....	10
CHAPTER 2. BACKGROUND LITERATURE.....	12
2.1 Foreign Accents: Conceptualization, Development and Maintenance	12
2.1.1 Foreign-accented speech.....	13
2.1.2 The perceived quality of foreign accents.....	16
2.1.3 The development of foreign accents.....	18
2.1.4 The maintenance of foreign accents	24
2.2 Processing Cost and Adaptation to Unfamiliar Foreign Accents.....	27
2.2.1 Processing cost of unfamiliar foreign accents	27
2.2.2 Adaptation to foreign-accented speech.....	30
2.3 Perceptual Accommodation of Foreign-accented Speech Sounds	33
2.3.1 Accommodation of native speech variability	34

	Page
2.3.1.1 Speech invariance models	34
2.3.1.2 Speech variance models.....	35
2.3.2 Accommodation of speech variability from different languages	39
2.3.3 Accommodation of foreign-accented speech variability	41
2.3.1.1 The mechanism of phonetic relaxation.....	41
2.3.1.2 The mechanism of phonetic calibration.....	44
2.3.1.2 The mechanism of phonetic switching	46
2.4 Goals and Research Questions	47
CHAPTER 3. EXPERIMENT 1	50
3.1 Motivation and Overall Design	50
3.2 Methods.....	53
3.2.1 Participants	53
3.2.2 Speech Materials.....	55
3.2.3 Speech resynthesis	58
3.2.4 Procedure	60
3.3 Analysis	62
3.4 Results	63
3.5 Discussion	65
3.6 Summary and Further Research Questions	69
CHAPTER 4. EXPERIMENT 2	71
4.1 Motivation and Overall Design	71
4.2 Methods.....	71
4.3 Analysis, Results and Discussion.....	73
4.4 Summary and Further Research Questions	75
CHAPTER 5. EXPERIMENT 3	77
5.1 Motivation and Overall Design	77
5.2 Methods.....	77
5.3 Analysis, Results and Discussion.....	77
CHAPTER 6. EXPERIMENT 4	81

	Page
6.1 Motivation and Overall Design	81
6.2 Methods.....	82
6.2.1 Participants	82
6.2.2 Stimuli.....	82
6.2.3 Stimuli creation.....	84
6.2.4 Procedure	85
6.3 Analysis.....	88
6.4 Results and Discussion.....	88
CHAPTER 7. GENERAL DISCUSSION	92
7.1 Summary	92
7.2 Findings and Implications	94
7.2.1 Overarching question.....	95
7.2.2 General research questions	96
7.2.3 Specific research questions.....	97
7.2.3 Conclusions.....	98
7.2 Limitations and Future Directions.....	101
REFERENCES.	104
VITA.....	125
PUBLICATIONS.....	126

LIST OF TABLES

Table	Page
1 English-accent Experience in Experiment 1	53
2 Speech Items in Experiment 1	54
3 Predictions in Experiment 1	61
4 Boundaries in Experiment 1.....	64
5 Speech Items in Experiment 2	72
6 Boundaries in Experiments 1 and 2	73
7 Speech Items in Experiment 4	82
8 Target and Foil Trials in Experiment 4.....	86

LIST OF FIGURES

Figure	Page
1 Accommodation of Foreign-accented Sounds	6
2 Distorted Speech Samples.....	13
3 Statistical Learning of Tense and Lax Vowels.	21
4 Classification of Vowel Exemplars.	35
5 The Mechanism of Phonetic Relaxation.....	41
6 The Mechanism of Phonetic Calibration	43
7 VOT Boundaries	51
8 VOT Re-synthesized Samples	58
9 Cross-splicing Method	60
10 Voicing-identification Response Curves from Experiment 1	63
11 Voicing-identification Response Curves from Experiments 1 and 2.....	74
12 Voicing-identification Response Curves from Experiment 3.....	78
13 Voicing-identification Response Curves from Experiment 4.....	88
14 Performance by Block in Experiment 4.....	89
15 Research Questions and Experiments.....	93

LIST OF ABBREVIATIONS

AIC: Akaike information criterion
ANOVA: Analysis of the variance
EASC: English-accented Spanish and Spanish contact
ES: English followed by Spanish
ESC: English and Spanish contact
GQ: General research question
L1: First language
L2: Second language
LOR: Length of residence
M: Mean
ms: milliseconds
OQ: Overarching question
PAM: Perceptual assimilation model
R: Range
RMS: Root mean squared
SD: Standard deviation
SE: Spanish followed by English
SLM: Speech learning model
SNR: Signal to noise ratio
SQ: Specific research question
s: seconds
USA: The United States of America
VOT: Voice Onset Time

ABSTRACT

Llanos Lucas, Fernando. Ph.D., Purdue University, August 2016. The Role of Experience in Processing Foreign-accented Speech. Major Professors: Alexander L. Francis and Robert M. Hammond.

The present study examines the perceptual accommodation of the bilabial stop-consonant voicing contrast (i.e., /b/ vs. /p/), in several English- and Spanish-accented contexts, by native Spanish listeners with different degrees of experience with accented speech. In a series of four experiments, we confronted three potential mechanisms for the perceptual accommodation of foreign-accented sounds. According to the first mechanism (phonetic relaxation), listeners accommodate foreign-accented sounds by relaxing the phonetic boundary between native speech sound categories. According to the second mechanism (phonetic calibration), listeners accommodate foreign-accented sounds by adjusting the location of native perceptual boundaries according to the phonetic realization of native categories in the foreign-accented speech context. Finally, according to the third mechanism (phonetic switching), foreign-accented speech sounds are accommodated by switching to a non-native system of phonetic representations that was previously developed through long-term experience with the speech norm of the foreign accent. Experimental results indicate that Spanish listeners did not relax the phonetic boundary between /b/ and /p/ in an English-accented Spanish context (Experiments 1 and 3). However, they

accommodated English-accented Spanish voicing differently, depending on their degree of experience with the English-accented speech norm. When Spanish listeners had little or no experience with the English norm, they calibrated the location of the perceptual boundary between /b/ and /p/ according to the Spanish or English phonetic realization of these sounds in the speech context (Experiment 4). Alternatively, when they had a high degree of experience with English-accented speech, they accommodated English-accented Spanish /b/ and /p/ by using an English-like system of phonetic representations that was not predictable from the phonetic realization of /b/ and /p/ in the speech context (Experiments 1 and 2).

These experimental results contribute to a better understanding of the role played by non-native experience in the perceptual accommodation of foreign-accents. In particular, they indicate that native listeners may rely on previous long-term experience with the native language of the foreign-accented speaker to efficiently accommodate foreign-accented speech variability in a different way to which they accommodate speech variability from different native-accented speakers.

CHAPTER 1. INTRODUCTION

1.1 Preliminaries: Speech Variability and Phonetic Constancy

One of the most astonishing human faculties is the ability to structure a physical speech continuum into a sequence of discrete speech units. Speech sounds differ from one another in terms of multiple acoustic-phonetic properties; for instance, Lisker (1998) identified 16 acoustic correlates of the stop consonant voicing contrast (e.g., /p/ versus /b/) and McMurray and Jongman (2011) collected a total of 24 involved just in the articulation of different fricative consonants (e.g., /s/ vs. /f/). Despite this acoustic complexity we are able to categorize the speech stream very quickly, even in the presence of noise or a foreign accent.

This faculty is not easy to explain if we keep in mind all the contextual variables involved in the acoustic realization of one single correlate. The pronunciation of the same speech category may change according to different factors, such as the speaker, the phonetic context, or the speech rate, among others (Diehl, Lotto & Holt, 2004; Magnuson & Nusbaum, 2007; Nusbaum & Magnuson, (1997). For instance, the formant transition pattern that encodes place of articulation in stop consonants (e.g., /b/ vs. /d/) changes depending on the following vowel (Delattre, Liberman & Cooper, 1955), which in turn changes spectral quality depending on the talker (Peterson & Barney, 1952).

Similarly, the duration of the initial formant transition that encodes the difference between [b] (short transition) and [w] (long transition) is not independent of speech rate, such that a fast speaking rate [w] may present longer formant transitions than a normal speaking rate [b] (Miller & Baer, 1983).

All these contextual variables conspire to increase the amount of acoustic variability of speech signals to the point that there are no two identical pronunciations of the same phoneme, and the same pronunciation can be perceived as two different phonemes depending on the context. For example, the same pronunciation [b] could be perceived as /b/, in a normal speech-rate context, or /w/, in a fast speech-rate context (Miller & Baer, 1983; for more examples of perceptual context effects see Lotto & Kluender, 1998; Lotto & Holt, 2006). With all that, listeners show phonetic constancy in their perceptual judgments, being able to accommodate speech variability from an early age (Jusczyk & Aslin, 1995; Schmale & Seidl, 2009). This suggests that our ability to develop reliable mental representations of speech categories is linked to our ability to cope with the acoustic variability characteristic of our native speech environment.

Interestingly, we can also handle uncommon types of speech variability in listening to speakers with hearing impairment (McGarr, 1983), synthetic speech (Schwab, Nusbaum & Pisoni, 1985), speech masked with noise (Mattys, Davis, Bradlow & Scott, 2012), simulated electrical-hearing (Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2011), sine-wave speech (Remez, Rubin, Pisoni & Carrel, 1981), time-compressed speech (Dupoux & Green, 1997), and foreign-accented speech (Bradlow & Bent, 2008).

These extreme cases of perceptual adaptation suggest that listeners' mental representations of speech categories are also highly flexible. Given the facility with which

our ears become tuned to the native speech environment, one may wonder how such a degree of cognitive flexibility is possible. The answer to this question is not straightforward. Perhaps, because variability is an essential component of any native speech environment, listeners develop perceptually flexible speech presentations to adapt to an environment in constant change. Perhaps, speech variability is systematic enough as to facilitate rapid perceptual adaptation or speech learning. The study of foreign-accented speech provides us with an optimal window to investigate these and other questions related to listeners' cognitive flexibility in speech processing.

1.2 Deciphering a Foreign Accent: Statement of the Problem

The trend toward globalization is causing an increase in the amount of language contact in the US, a country with an already strong immigrant tradition, and this increment is especially remarkable in some environments, such as academia or international companies, where the concentration of native speakers of different languages is typically higher than average. According to the US Census Bureau (2010), the percentage of foreign-born population in the US has increased from 9.7% to 40% since 1960. Considering that English is the lingua franca of this country, this increase can be also interpreted as an increment of the number of foreign-accented speakers of English.

Listening to a foreign accent can be demanding. For this reason, during the last several decades, speech scientists have tried to identify the speech features that hinder the efficient processing of foreign accents with the ultimate goal of improving the intelligibility of non-native speech and thus facilitating its processing by native listeners. With all that, and despite the contributions of this line of research in terms of language instruction, this

approach is sometimes dominated by a tacit perspective that puts the foreign-accented speaker into the center of the communicative problem, therefore introducing a realistic but incomplete version the communicative phenomenon.

Although the costs of processing unfamiliar accents are attested by experimental results (Cristia, Seidl, Vaughn, Schmale, Bradlow & Floccia, 2012; Munro & Derwing, 1995), other results suggest that listeners detect and adapt to foreign accents very efficiently (cf. Cristia et al., 2012; Bradlow & Bent, 2008; Clarke & Garret, 2004; Flege & Hammond, 1982). These other results underline the important role played by the listener in the communicative scene; a role that is examined in the present research, which does not focus on the identification of speech features that hinder the processing of foreign accents, but on the perceptual strategies used by native listeners to efficiently accommodate them.

While it is true that listeners adapt to foreign accents after receiving some experience with them, the way in which they actually do it is not yet well understood. This theoretical gap contrasts with the number of theories of perceptual accommodation proposed for native-accented speech (e.g., see Fowler & Magnuson 2012). Inspired by this previous work, the present research examines three different hypotheses on the perceptual accommodation of foreign-accented speech sounds by native listeners. These hypotheses are: 1) the phonetic relaxation hypothesis, 2) the phonetic calibration hypothesis, and 3) the phonetic switching hypothesis.

1.2.1 The phonetic relaxation hypothesis

This hypothesis is inspired by exemplar models of speech perception (Goldinger, 1998; Johnson, 2006; Nosofsky, 1988; Pierrehumbert, 2003; see sections §2.3.3.1 and

§2.3.1.2). According to this hypothesis, listeners relax their expectations on the pronunciation of native speech sound categories in an attempt to reconcile phonetic differences between the native- and foreign-accented speech norm, in a similar way to which they reconcile different native-accented pronunciations of the same speech category across different phonetic contexts and talkers. In particular, the phonetic relaxation hypothesis postulates that native listeners relax the perceptual boundaries between those native speech categories that are more frequently mispronounced in the foreign-accented norm that they are listening to.

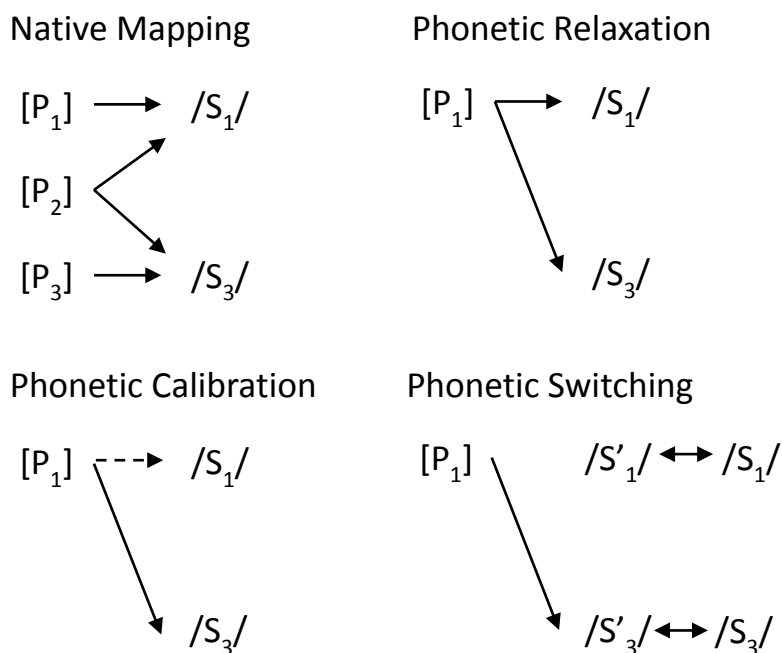


Figure 1. Accommodation of Foreign-accented Sounds. Top left: schematic representation of the prototypical perceptual mapping between native pronunciations (Ps) and speech sounds (Ss). Schematic representations of the phonetic relaxation hypothesis (top right), the phonetic calibration hypothesis (bottom left), and the phonetic switching hypothesis (bottom right), in listening to a hypothetical accent in which the native speech sound $/S_3/$ is pronounced as foreign-accented $[P_1]$ rather than native-accented $[P_3]$

This hypothesis is illustrated in Figure 1 (top-right panel). In this figure, the top-left panel shows a schematic representation of the typical perceptual mapping between pronunciations and speech sounds in listening to native speech. In this panel, the non-deterministic perceptual mapping from pronunciation [P₂] to sounds /S₁/ and /S₃/ was meant to illustrate the fact that the same pronunciation can be perceived differently depending on the context, just as was noted in section §1.1. The top-right panel of Figure 1 illustrates the phonetic relaxation hypothesis for a hypothetical native listener that has been also exposed to a foreign-accented pronunciation of /S₃/ as [P₁], rather than only as [P₃].

1.2.2 The phonetic calibration hypothesis

The phonetic calibration hypothesis is inspired in models of perceptual calibration (Holt, Lotto & Kluender, 2000; McMurray & Jongman, 2012; Sawusch & Pisoni, 1974; see sections §2.3.1.2 and §2.3.3.2). According to this hypothesis, listeners accommodate foreign-accented sounds by adjusting the short-term phonetic representation of native speech categories to their pronunciation in the speech context. Thus, instead of trying to reconcile different accented pronunciations of the same sound, they re-calibrate the location of native perceptual boundaries based on the information provided in the speech context, just as they do to accommodate idiolectal speech from different native-accented talkers. This hypothesis is illustrated in bottom-left panel of Figure 1, which represents the phonetic calibration of sound /S₃/ by a native speaker who is listening to a foreign-accented talker who is pronouncing /S₃/ as foreign-accented [P₁] rather than as native-accented [P₃].

1.2.3 The phonetic switching hypothesis

The phonetic switching hypothesis is inspired by models of bilingual phonetic switching (García-Sierra, Ramírez-Esparza, Silva-Pereyra, Siard and Champlin, 2012; Gonzales & Lotto, 2013; see sections §2.3.2 and §2.3.3.3). According to this hypothesis, foreign-accented sounds are not accommodated by temporary adjustments in the short-term representation of native speech categories, as suggested by the phonetic calibration hypothesis. Instead, native listeners rely on a whole new set of long-term speech representations acquired through long-term experience with the phonetic norm of the foreign accent that is being listened, in a similar way to which language learners and bilinguals accommodate sounds from their first (L1) and second (L2) language. This third hypothesis is schematically represented in the bottom-right panel of Figure 1, in which pronunciation [P₁] is mapped onto its corresponding speech sound /S'₃/ in the system of the foreign-accented talker that is being listened.

These three hypotheses (phonetic relaxation, phonetic calibration, and phonetic switching) were examined in a series of four experiments that investigated the processing of English-accented Spanish by monolingual speakers of Spanish and bilingual speakers of Spanish and English. Thus, the results of these experiments are also expected to contribute to a better understanding of bilingual experience in the processing of foreign-accented sounds.

1.3 Foreignness and Nativeness in the Communicative Framework

Before addressing the present study, we would like define two words that are frequently used in the next chapters: these words are the adjectives *foreign* and *native*. The

interpretation of these two adjectives is not always explained in literature specialized on foreign-accented speech because such interpretation is generally assumed to follow from a communicative framework in which a native speaker of L1 is listening to a L2 speaker of L1. Within this listener-oriented interpretation of the communicative framework, the native speaker of L1 and the L2 speaker of L1 are commonly referred as the native listener and the foreign speaker, respectively.

Although this is a very plausible interpretation of the communicative framework, it is not the only one. Alternatively, the L2 speaker of L1 could be viewed not as the foreign speaker of L1 but as the native speaker of L2. From this alternative, speaker-oriented interpretation, the native speaker of L1 could be rather viewed as a foreign listener of L2. This alternative interpretation of the same communicative framework described above reveals some of the referential ambiguities that may take place in a typical conversation between foreign- and native-accented speakers. In a certain sense, each of the two interpretations provided is just a reductionist version of a more realistic interpretation of the communicative framework, according to which communication flows in more than one direction, from speakers to listeners that in turn become speakers. With this exchange of communicative roles in mind, it could be argued that the foreignness of the foreign-accented speaker involves, and co-exists with, the foreignness of the native-accented one, in that both speakers address to each other with an accent that is foreign to their respective native speech norms.

This broader interpretation of the communicative framework may reveal some important variables that are not captured by the scope of a listener- or speaker-oriented interpretation. For example, foreign-accented speakers may alter the quality of their speech

based on their interaction with native-accented speakers. Similarly, native-accented speakers may accommodate the quality of their speech differently depending on the perceived quality of the foreign accent that they are listening to (Kim, Horton, & Bradlow, 2011; Smiljanić & Bradlow, 2011). The ability to perceptually accommodate accents will be argued to play an important role in the development and maintenance of foreign accents (sections §2.1.1, §2.1.3 and §2.1.4).

1.4 Outline of the Present Study

This study is structured in seven chapters. The goal of the second chapter (*Background Literature*) is to provide the reader with the theoretical and experimental background needed to address the reading of the rest of the chapters. This chapter includes a review of classic and contemporary studies on foreign-accented speech and speech perception, including the literature concerned with the acquisition of L2-speech sound categories. The end of that chapter provides a theoretically detailed description of the three hypotheses of perceptual accommodation of foreign-accented speech sounds that were first introduced in section §1.3, as well as the goals and research questions that motivate of the experiments included in the next chapters (Chapters 3 to 6). Except Experiment 3 (Chapter 5), the rest of the experiments herein have been already published in the journal *Language and Speech* (Llanos & Francis, 2016). Finally, Chapter 7 (*General Discussion*), presents a summary of the experimental results collected in the previous chapters, followed by a discussion of the main findings and implications of these results in the light of the research questions that motivated the experiments. This last chapter concludes with a section that discusses limitations and future directio

CHAPTER 2. BACKGROUND LITERATURE

This chapter is structured in four major sections. The first section (§2.1) reviews the characteristics of foreign-accented speech, and the main variables involved in the development and maintenance of foreign accents. The goal of this section is twofold: first, it tries to set background knowledge on foreign-accented speech; second, it tries to establish a theoretical connection between the maintenance of foreign accents and native listeners' ability to perceptually accommodate them (§2.1.4). The second section (§2.2) reviews the costs involved in the processing of unfamiliar accents, and the role played by accent experience in the reduction of these costs. The third section (§2.3) presents a theoretically detailed description of three hypotheses introduced in section §1.2. This description is, however, preceded by a brief review of the main theories and models of speech perception that inspired these hypotheses. Finally, the fourth section (§2.4) introduces the goals and research questions that motivate the experiments introduced in the next chapters.

2.1 Foreign Accents: Conceptualization, Development, and Maintenance

This section provides a characterization of foreign-accented speech, as compared to other types of unfamiliar or distorted speech (§2.1.1), and discuss the main variables

involved in the development and maintenance of foreign accents (§2.1.2 - §2.1.4). At the end of this section, it is argued that the lack of phonetic attainment in some foreign-accented speakers could be partially explained by native listeners' ability to perceptually accommodate them.

2.1.1 Foreign-accented speech

Foreign-accented speech can be broadly characterized as a systematic deviation from the native phonetic norm of a particular listener; such a deviation that is modulated by phonetic differences and similarities between the speech norm of the foreign-accented speaker and that of the native-accent listener. From this point of view, the main difference between foreign-accented speech and other types of unfamiliar or distorted speech, like speech in noise, or simulated electrical hearing (also known as vocoded speech; Gantz & Turner, 2003; Li & Loizou, 2008), is not that foreign-accented speech variability is systematic but that it results from the systematic interference of two phonetic norms.

This difference is illustrated in Figure 2, which shows spectrograms of the Spanish native-accented word *bata* 'robe' and four alternative versions of this word produced in English-accented Spanish, speech masked with noise, electrical simulated hearing, and English native-accented synthetic speech. As can be appreciated by comparing the spectrograms of these four versions of the same word, the main difference between the Spanish- and English-accented versions lies in the pronunciation of the stop consonants /b/ and /t/. In the Spanish native-accented version, /b/ starts with approximately 0.1 s of laryngeal pre-voicing that does not occur in English-accented Spanish version because, in English, the onset of voicing typically occurs after the consonant release. On the other

hand, the pronunciation of /t/ in the English-accented Spanish version results from the interference between the Spanish and English norm; while intervocalic /t/ is often flapped in English postonic syllables (*water*, [wɒ'rər]; Eddington & Elzinga, 2008; Hammond, 2011), in the English-accented Spanish spectrogram, /t/ is not flapped.

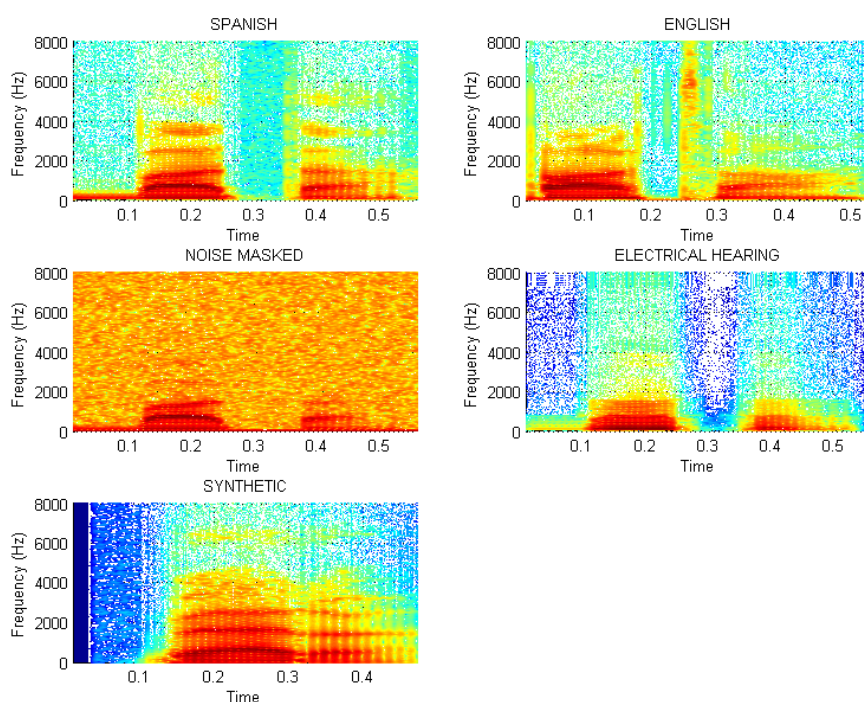


Figure 2. Distorted Speech Samples. Spectrograms of the Spanish word *bata* ‘robe’ in native-accented Spanish (*Spanish*), English-accented Spanish (*English*), masked with noise (*Noise*; 0.8 dB of signal-to-noise ratio), vocoded speech (*Electrical hearing*; 10 frequency-band vocoded channels), and English native-accented synthetic speech (*Synthetic*; based on Brookes, 2009)

In the noise-masked version, however, the spectral distortion seems to be clearly independent of the Spanish norm, although the noise degradation in the spectrogram looks quite systematic. Similarly, in the simulated electrical-hearing version, the spectral

degradation induced to simulate hearing in cochlear implants is systematically the same for all sounds across each of the 10 frequency channels defined by the vocoder filter (e.g., from 300 to 600 Hz). The case of synthetic speech is slightly different because some speech synthesizers actually incorporate phonetically-based distortion (Klatt, 1987; Story, 2003) but, even in this case, synthetic speech variability does not result from the interference of two speech norms.

The prevalence of phonetically-based speech variability in foreign-accented speech could explain why, in general, adapting to a novel accent is typically less challenging than adapting to other types of distorted speech, such as speech masked with noise, in which the distortion of the speech signal is mediated by a many-to-one mapping (many sounds, one type of distortion) that is perceptually more difficult to reverse than the typical acoustic mapping between native- and foreign-accented sounds, in which the distortion of the speech signal tends to affect differently to each sound.

Despite this, foreign-accented speech is sometimes dominated by the presence of other features, concerned with speech fluency (e.g., long pauses and/or speech interruptions), which can be actually more detrimental to speech intelligibility than the foreign accent by itself; although these other features tend to disappear at some point of the language learning period, whereas the accent may persist through the whole life (see sections §2.1.2 and §2.1.4). Furthermore, accent quality and speech intelligibility are not always correlated in the processing of foreign-accented speech. For example, while the foreign-accented talkers who are less intelligible are typically judged as the most heavily accented, some heavily accented speakers are sometimes judged as very intelligible (Derwing & Munro, 1997; Trofimovich & Isaacs, 2008). Thus, the degree of perceived

intelligibility cannot be always taken as a fundamental property of foreign-accented speech.

Another distinction that is commonly made in the study of foreign-accents is between dialects or regional accents, and foreign accents. This distinction is however inspired by geographical and sociolinguistic criteria that are not necessarily representative of the challenges faced by listeners (cf. Cristia et al., 2012). From the point of view of the listener, both regional and foreign accents bear a systematic departure from the native-accented norm, and the quality and quantity of such departure may vary across accents and speakers. However, foreign accents may sometimes exhibit a higher degree of variance across speakers from the same native background, who may speak with a milder or stronger accent depending on different variables. These variables are discussed in the next section.

2.1.2 The perceived quality of foreign accents

Basically, foreign accents exist because many foreign speakers do not fully converge to the accented norm of their listeners. In some cases, the lack of attainment is just provisional, meaning that some foreign-accented speakers will converge after a longer period of time. However, many non-native speakers do not converge to the native-accented speech norm of their listeners even after a long period of immersion (Flege, 2009; Flege, Munro, & McKay 1995), and this lack of attainment is intriguing when compared to L1-speech learners, who successfully converge to the native-accented norm of their language environment.

According to previous literature, the best predictors of phonetic convergence are the age of arrival in the foreign country (also known as onset of learning), the length of

residence in the foreign country, the frequency of L2-speech usage, and age (Flege, 2009; Piske, McKay, Flege, 2001). For instance, in two different studies taking into consideration these variables, one with 240 Italian immigrants living in Canada (Flege, Munro & McKay, 1995) and another with 240 Korean immigrants living in the USA (Flege, Yeni-Komshian & Liu, 1999), the highest correlate of accentness was age of arrival (Pearson's $r = -0.85$ in both Italians and Koreans), followed by L2 usage ($r = 0.60$ in Italians, $r = 0.61$ in Koreans), age ($r = -0.53$ in Italians, $r = -0.56$ in Koreans), and length of the residence ($r = 0.28$ in Italians, $r = 0.38$ in Koreans). In both studies, age of arrival seemed to cover a significantly higher portion of statistical variance, as measured by the partial eta-squared coefficient. Crucially, the percentage of speakers rated as free-of-accent was lower than 10% in the population of late bilinguals (learners with an onset of learning higher than 12 years old), in contrast to the 50% and 30% of Koreans and Italians rated as free-of-accent among some populations of bilinguals with an onset of learning lower than 12 years old (early bilinguals).

At a first glance, these results (see also Flege & Liu, 2001) seem to support the view that there is a critical period for learning (Johnson & Newport, 1989; Lenneberg, 1969), or a native phonological filter that suppresses any further acquisition of novel speech. However, a closer look at the data collected in the last two decades suggest that is not probably the case, and that there is a high degree of individual variability. First, although many late bilinguals do experience difficulty in losing their foreign accent, some are indeed able to do it. Similarly, although no all early bilinguals converge to the speech norm of their listeners (Flege, Birdsong, Bialystok, Mack, Sung, & Tsukada, 2005; Bongaerts, Van Summeren, Planken, & Schils, 1997). This high degree of individual variability suggests

that the lack of convergence is probably not the consequence of a biological landmark in cognitive or learning skills (see also Snedeker, Geren, & Shafto, 2012). Furthermore, in contrast to the hypothesis of the L1 phonological filter, language learners have been shown to learn speech contrasts that are suppressed in their native system (e.g., Flege & Port, 1981). What these results do not yet explain is how foreign accents are developed and/or maintained. This issue is discussed in the next section.

2.1.3 The development of foreign accents

In recent literature on L2-speech acquisition, there is a general consensus that the learning mechanisms that are used to acquire L1 speech categories remain intact over the life span, and are thus involved in the acquisition of novel sound systems (Best & Tyler, 2007). Most of the theoretical divergence between recent speech-learning models is not typically concerned with this point, but with the nature of the speech representations that are acquired; e.g., whether they are psychoacoustic representations (Aslin, Pisoni, Jusczyk, 1983; Flege, 1995), or articulatory gestures (Best, 1994; Fowler, 1986).

The most cited models of L2-speech learning in the last two decades are probably the perceptual assimilation model (PAM; Best, 1994; Best, Roberts, Goodell, 2001) and the speech learning model (SLM; Flege, 1995; Flege 2003). These are not the only models of L2-speech learning (e.g., the gradual learning model, Escudero & Boersma, 2003; the perceptual magnet effect model, Kuhl, 1991, Kuhl & Iverson, 1995). However, in this section, I will focus on the PAM and SLM because they have been extensively investigated in a wider range of phonetic contrasts.

The PAM and SLM both postulate that L2-speech acquisition of is highly influenced by the degree of phonetic similarity between L1 and L2 speech sounds, being this influence stronger at the onset of the speech learning period. According to these models, learners should have more difficulty learning a contrast between non-native sounds when these sounds are perceived as different phonetic realizations of the same native speech category. For instance, naïve Spanish learners of English who do not contrast lax and tense vowels in their native Spanish system, typically perceive English tense [i] (e.g., *beat*) and English lax [ɪ] (e.g., *bit*) as two different phonetic realizations of Spanish tense /i/ (Escudero & Boersma, 2009; Morrison, 2004).

The difficulty of perceiving the difference between two or more non-native sounds seems to play a determinant role in the development of a foreign accent. However, even when the perceptual mapping between L2 and L1 speech categories is one-to-one, learners may still have problems identifying the most prototypical pronunciations of L2-speech categories. For example, although both English and French have /u/ in their sound inventories, French /u/ is typically more tense, or peripheral, than English /u/ (Delattre, 1964; Hillenbrand, Getty, Clark & Wheeler, 1995). Thus, even when French /u/ is still the phonetically closest neighbor of English /u/, English learners of French typically struggle to produce a prototypical pronunciation of /u/ in French, especially during the onset of the learning period (Flege, 1987).

Ideally, after a significant amount of L2 experience, L2 learners should be able to dissimilate a whole new set of long-term representations that they can use to efficiently accommodate L2 speech variability. However, although both the PAM and SLM are very accurate at predicting how L2 speech categories are assimilated to L1 speech categories at

the onset of the learning period, the way in which L2 speech categories are finally dissimilated from L1 speech categories is not successfully explained by any of these models. Hypothetically, the dissimilation of L2 speech categories could be achieved through experience with the statistical distribution of acoustic-phonetic cues in L2 input. This idea, which is not incompatible with the SLM, has been explored by different models of statistical learning (Boersma & Hayes, 2001; Johnson, 2006; Maye, Werker & Gerken, 2002; Maye, Weiss & Aslin, 2008; McMurray, Aslin, Toscano, 2009; Pierrehumbert, 2003; Toscano & McMurray, 2010). The proposal that L2 speech categories can be dissimilated through sampling experience with the distribution of acoustic-phonetic cues in L2 speech is theoretically supported by simulated statistical-learning.

This idea is illustrated in Figure 3 (panel C), which shows that the statistical cost of assimilating English /i/ and /ɪ/ to Spanish /i/ increases in proportion to the amount of experience with the phonetic distribution of English /i/ and /ɪ/. Specifically, Figure 3 shows the results of two hypothesis-driven simulations on the statistical learning of the English tense/lax contrast /i/-/ɪ/ by naïve Spanish learners of English. The first hypothesis assumes that English tokens of /i/ and /ɪ/ are perceived as instances of Spanish /i/, as could be expected in naïve Spanish learners of English (perceptual assimilation hypothesis). The second hypothesis assumes that the same tokens are perceived as instances of two different vowel categories, as would be expected by native listeners of English (perceptual dissimilation hypothesis).

In each simulation, a series of 1000 vowel pairs were sampled from two realistic acoustic distributions of English /i/ and /ɪ/ in terms of their first two formants F1 and F2. At each sampling step, two independent Gaussian mixture models, of either one

distribution (perceptual assimilation hypothesis) or two distributions (perceptual dissimilation hypothesis), were fitted to all pairs previously sampled.

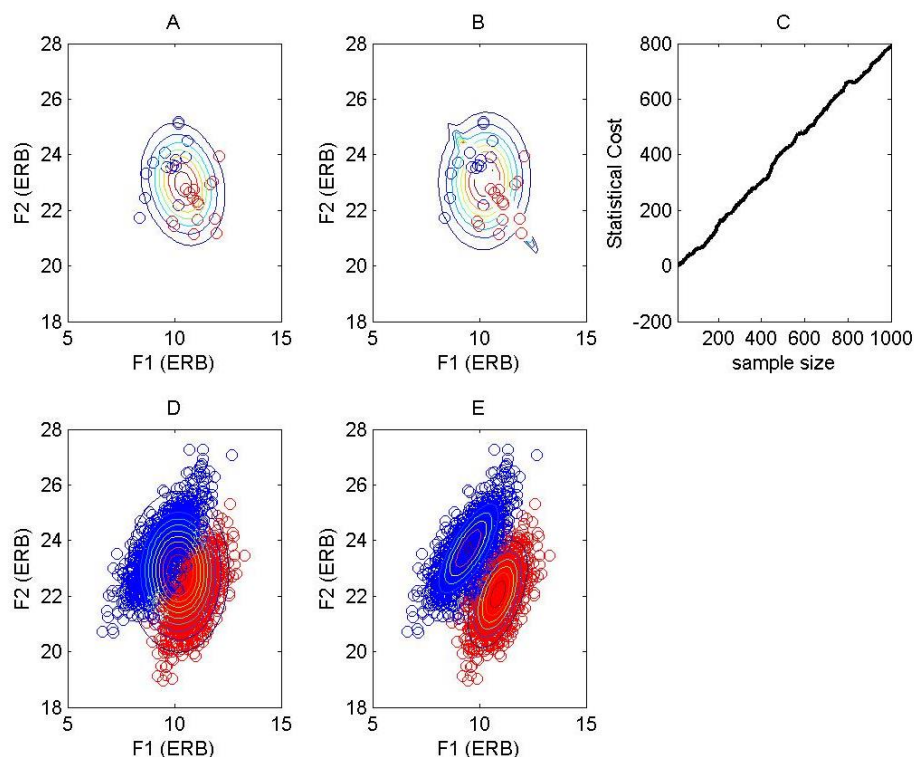


Figure 3. Statistical Learning of Tense and Lax Vowels. Gaussian mixture models fitted to a distribution of /i/ - /ɪ/ vowel pairs, sampled from prototypical equivalent rectangular bandwidth (ERB) formant values for American English (inspired by Hillenbrand et al., 2005). Panels A and B show two Gaussian mixture models of one (A) and two categories (B) fitted to a distribution of ten vowel pairs. Panels D and E shows the same models fitted to a distribution of 1000 vowel pairs. Panel C plots the statistical cost of the one-category model minus the cost of the two-category model across different sample sizes, from 1 to a total of 1000 vowel pairs. Statistical cost was measured by the AIC of the corresponding model. Models were fitted by means of the expectation maximization algorithm (Bilmes,1998)

In each simulation, the statistical cost of the corresponding hypothesis was measured as the Akaike information criterion (AIC) of the corresponding model's fit. The

AIC measures the cost of a statistical model fitted to a given set of data, such that the lower the AIC the lower the cost (Posada & Buckley, 2004). The results are shown in Figure 3 (panel C), which plots the statistical cost of the perceptual assimilation hypothesis minus the statistical cost of the perceptual dissimilation hypothesis. Thus, Figure 3 (panel C) can be interpreted as the cost of rejecting the dissimilation hypothesis as a function of the sample size. Such cost increases linearly with the sample size, meaning that the dissimilation of /ɪ/ from /i/ becomes statistically more reliable as the number of tokens sampled from /i/ and /ɪ/ increases. Interestingly, when the sample size is small (Figure 3, panels A and B), the dissimilation model fails to efficiently fit two clearly separated speech categories (Figure 3, panel B).

Although Spanish learners of the English tense/lax vowel contrast may rely on other cues than F1 and F2 (e.g., vowel duration; Escudero & Boersma, 2004; Morrison, 2009; Kondarouva & Francis, 2008, 2010), the results of the simulations introduced in Figure 3 suggest that the perceptual dissimilation of English /ɪ/ from Spanish /i/ is, at least, statistically reliable. Thus, the lack of convergence of L2 speech learners to the L2 native-accented speech norm may be related to other variables than the amount of experience with the distribution of cues or the degree of phonetic similarity between L1 and L2 speech categories. The role played by these other variables in the maintenance of foreign-accents is addressed in the next section.

2.1.4 The maintenance of foreign accents

One of the most intriguing aspects related to the study of L2 speech acquisition is the lack of convergence to the L2 speech norm exhibited by many L2 learners. This

phenomenon is not always successfully explained by models of L2 speech learning. As was noted in section §2.1.2, while some L2 speech learners lose their foreign accent after a period of time, many others maintain their accents even after 20 years of immersion in the non-native speech environment (cf. Flege et al., 1995). Ideally, one might expect a systematic correspondence between the amount of language experience and learning. However, this is not always the case in L2-speech acquisition.

A lack of convergence to the L2 speech norm might result from differences in the type of input and language interaction. Language users may alternate different ratios of L2/L1 depending on their age, socioeconomic, or immigrant status, among other variables that are not always taken into account, but that may play an important role in the understanding of individual learning differences (Gerken, Amengual, & Birdsong, 2014; cf. Flege et al, 2002; Perez-Leroux, Cuza, 2011). Also, as noted by Flege (2009), most of the studies on L2-speech acquisition base their measurements of language use on participants' self-reports who might not pay enough attention to the complexity of their language performance. Therefore, the lack of phonetic convergence may be caused by individual differences that are not appropriately captured by impressionistic variables, such as participants' self-reports of language use.

Alternatively, the lack of convergence could be also viewed as a consequence of listeners' ability to perceptually accommodate foreign-accented speakers. While the study of L2-speech acquisition is sometimes approached by experimental designs that evaluate the processing of L2 phonetic properties by an experimental group of language learners relative to a control group of native-accented speakers, speech signals are acoustically very redundant. In particular, sounds are cued by multiple phonetic properties that provide

different but functionally-equivalent ways of processing the same speech contrast. For example, in the production of the English tense/lax vowel contrast /i/-/ɪ/, differences in vowel spectral quality are typically correlated with differences in vowel duration, such that the English tense vowel /i/ is typically produced with a longer duration than its lax counterpart /ɪ/.

Speech redundancy can sometimes be a source of phonetic divergence, when native- and foreign-accented speakers rely on different cues to the same speech contrast. For example, while native speakers of English tend to prime spectral vowel-quality differences in the decoding of the English tense/lax vowel contrast, Spanish learners of English tend to prime vowel duration over vowel spectral quality (Kondaurova & Francis, 2010). However, as far as both duration and vowel quality are correlated in native-accented English, the choice of duration over quality by Spanish learners of English should not posit a major perceptual challenge for native English listeners, and vice versa. Therefore, so long as foreign-accented speakers are efficiently accommodated by native-accented listeners, they may not need to modify their accent to efficiently communicate, in a similar way to which regional-accented speakers (e.g., Irish-accented speakers of English) do not need to do it in order to be understood by native listeners from other regional accents (e.g., North American-accented speakers of English). The next section examines how native-accented listeners perceptually accommodate foreign-accented talkers.

2.2 Processing Cost and Adaptation to Unfamiliar Foreign Accents

This section focuses on the main variables involved in the process of adaptation to unfamiliar foreign accents (section §2.2.2) that usually takes place after a period of

familiarization with the accent during which native listeners experience several types of processing costs (section §2.2.1). Therefore, in this section, the term *adaptation* refers to the process by which native listeners overcome the initial costs involved in the processing of unfamiliar accents; the question of how they actually overcome these costs is addressed in section §2.3.

2.2.1 The processing cost of unfamiliar accented speech

Listening to an unfamiliar accent typically bears an extra processing cost in that it may slow processing speed and hinder speech comprehension (Cristia et al., 2012; Mattys, Davis & Bradlow, 2012). This is not surprising; speech processing slows almost every time that we are challenged with unexpected speech patterns (e.g., when listening to new voices; Goldinger, 1996; Kraljic & Samuel, 2005). However, foreign-accented speech may induce a higher processing cost than listening to novel native-accented voices because unfamiliar foreign-accents tend to present both indexical (talker) and phonetically related speech variability.

The processing of unfamiliar accented speech has been shown to increase response-time in different types of tasks, such as the sentence verification task, in which listeners evaluate the truth of sentences (Munro & Derwing, 1995), the visual probe matching task, in which they evaluate the matching between words and pictures (Clarke & Garrett, 2004), and the lexical decision task, in which listeners judge the lexical status, as word vs. non-word, of different speech items (Maye, Aslin & Tanenhaus, 2008; Floccia, Goslin, Girard & Konopczynski, 2006). In all these tasks, listeners tend to show slower performance when processing accented speech items.

Unfamiliar accents may also perturb speech comprehension, usually measured by the average percentage of words correctly transcribed per sentence (Well, 2001; Barlow & Bent, 2008). For instance, in a perceptual study that combined the use of foreign-accented speech with additive noise, Wjngaarden (2001) found that the overall difference in sentence intelligibility between native and foreign-accented speech was equivalent to a speech-to-noise ratio decay of 3 dB. In other words, the intelligibility cost caused by the accent was equivalent to masking the native-speech signal with a noise volume 3 dB higher than the volume of noise in the foreign-accented speech signal.

Besides processing speed and speech comprehension, foreign accents may also impact listeners' bias and attitudes toward non-native accented speakers (Chiba, Matsuura, & Yamamoto, 1995; McKenzie, 2008), who may be judged by their accent to be less truthful or worse teachers (Rubin and Smith, 1990). For example, Lev-Ari and Keysar (2010) found that trivia statements (e.g. *a giraffe can go without water longer than a camel can*) were judged as less truthful when they were produced with a foreign accent even when listeners were told that the statements were all written by specialized researchers.

Interestingly, the perceived quality of an accent seems to also be influenced by listeners' biases. For example, listeners tend to perceive speech as more accented when speech samples are paired with faces that are judged as foreign (Kang & Rubin, 2009; McGowan 2011; Rubin, 1992). In a recent study, Babel and Russell (2015) found that native-accented Canadian English speech masked with noise was judged as more heavily accented when speech samples were paired with Chinese faces, instead of white faces. Interestingly, speech comprehensibility was also lower when native-accented speech samples were paired with Chinese faces.

Although these experimental results are sometimes interpreted as willful misunderstandings of the speech signal caused by listeners' stereotype-driven biases (Lippi-Green, 1997; cf. Kang & Rubin, 2009), this interpretation is not incompatible with a perspective that takes into account potential interferences caused by the cost of processing unexpected relationships between speech and/or visual cues. For example, the credibility bias reported by Lev-Ari and Keysar (2010) was only significant for those foreign-accented speakers that were more difficult to understand; that is, the speakers that bore a higher processing cost. Similarly, the face stereotype bias reported by Babel and Russel (2015) could be a by-product of the noise -which might force listeners to rely more on their visual expectations- and the conflict of expectations provided by facial cues (foreign accent) and speech cues (native accent). The processing of speech properties is engaged by multiple phonetic expectations facilitated by the talker, the context or the topic of the conversation, and when the information provided in the speech signal deviates from these expectations, speech processing becomes more effortful thus altering the normal processing of speech (Adank, 2012; Yi, Phelps, Smiljanic, & Chandrasekaran, 2013).

2.2.2 Adaptation to foreign-accented speech

After an initial period of familiarization with the foreign accent, listeners typically go through a period of adaptation in which the initial processing cost is gradually reduced. Among the factors that contribute to this adaptation, the most relevant ones seem to be related to the degree of familiarization with the talker, the accent, and the topic of the conversation (cf. Bradlow & Bent, 2008; Gass & Varonis, 1984). However, listeners may also benefit from previous language background. For example, L1 speakers of L2 tend to

benefit from their common native background when listening to each other in L2 (Bent & Bradlow, 2003; Hayes-Harb, Smith, Bent & Bradlow, 2008; however, this interlanguage benefit was not replicated by Munro et al 2006).

Part of this success is due to listeners' ability to rely on prior accented experience to accommodate novel talkers and speech items from the same foreign-accented norm. Accent-specific experience has been shown to increase processing speed (Clarke & Garrett 2004; cf. Maye et al, 2008) and speech comprehension. For instance, in a study that investigated adaptation of English listeners to Chinese- and Slovakian-accented English, Bradlow and Bent (2008) found significant improvement in speech comprehension in those listeners that were tested on the same foreign-accented talkers that they were exposed to in the training phase. In a second experiment, English listeners were also able to generalize experience with multiple foreign-accented talkers to improve adaptation to a novel talker from the same accented norm. In a similar study, Sidaras, Alexander, and Nygaard (2015) found that English listeners were able to generalize prior experience with Spanish-accented English to improve speech comprehension in novel speech items.

Listeners' ability to generalize to new talkers and new utterances based on prior experience with other talkers with the same accent suggests that native listeners are doing something other than mere memorization of foreign-accented words. This generalization of foreign-accented speech patterns can sometimes be very rapid. For example, Clarke and Garrett (2004) found evidence of perceptual adaptation to novel Spanish-accented English items within the first minute of exposure. This rapid adaptation to foreign-accented speech contrasts with amount of time that is usually required to generalize non-native speech variability in an unfamiliar second-language context. Native listeners seem to experience

less difficulties to generalizing unfamiliar speech patterns in a foreign-accented speech context in which they count on lexical expectations that can be used to calibrate the difference between native- and foreign-accented pronunciations of the same words. In contrast, the generalization of foreign-accented speech variability is typically more difficult to achieve in an unfamiliar speech environment in which native listeners cannot count on lexical expectations. For example, Norris, McQueen and Cuttler (2003) found that listeners are able to calibrate unfamiliar pronunciations of native speech sounds only when those pronunciations are presented in native words (see also Reinisch, Weber & Mitterer, 2013).

The important role played by lexical expectations in the adaptation to foreign-accented speech is manifested in the characteristic increase of top-down lexical recruitment involved in the processing of unfamiliar accents, relative to such amount in the processing of native-accented speech (Lev-Ari, 2015; Mitterer & McQueen, 2009). In support of this, listeners seem to perceive fine-grained phonetic details more efficiently when those details are presented in a familiar lexical environment. For example, Pierrachione, Del Tufo and Gabrieli (2011) found that talker recognition skills decline in an unfamiliar foreign language environment.

With all that, the fact that listeners improve at the processing of novel foreign-accented speech items based on prior accent experience suggests that the extra amount of top-down recruitment may decline at certain point of the adaptation period; presumably, as a consequence of having developed more accurate phonetic representations of foreign-accented sounds. At this point, the mechanism(s) that mediates the generalization of a foreign-accented norm is not very well understood. For some authors (e.g. Nygaard, Sommers & Pisoni, 1994; Weil, 2001), the accommodation of foreign-accented speech

should not differ much from the accommodation of speech from unfamiliar but native-accented talkers. In this view, listeners should be able to accommodate talker differences independently of whether they are accented or not. However, even in this case, it is not clear which of the models proposed for accommodation of native-speech variability would better account for the perceptual accommodation of foreign-accented sounds. Furthermore, given the strong degree of phonetic dissimilarity between some native- and foreign-accented speakers, native listeners may alternatively opt for a different strategy and rely on a new system of phonetic representations previously developed through experience with the phonetic norm of the accent. From this perspective, the accommodation of foreign-accented speech variability would be more similar to the accommodation of speech variability in an L2-speech context. The next section reviews some of the classic models of perceptual accommodation of speech variability proposed to date.

2.3 Perceptual Accommodation of Foreign-accented Speech Sounds

This section provides a theoretically detailed description of the three hypotheses of perceptual accommodation of foreign-accented speech investigated in the present study. This description is however preceded by a brief review of the main models of speech perception which inspired them (sections §2.3.1 - §2.3.2). (For a more detailed discussion of these and other models, the reader can consult Diehl, Lotto & Holt (2004) and Fowler and Magnuson (2012).)

2.3.1 Accommodation of native-speech variability

Models of perceptual accommodation of native-speech variability can be broadly classified into two classes, depending on whether they focus on the processing of invariant or distributional-contextual speech properties. Models in the first class (speech invariance theories) propose that speech variability is accommodated by the identification of speech invariant features. Alternatively, models in the second class (speech variance theories) propose that speech variability is rather accommodated by the processing of distributional and/or contextual input properties.

2.3.1.1 Speech invariance models

Some models of speech perception argue that distinctive features are decoded as speech-invariant articulatory gestures (direct realism; Best, 1995; Fowler, 1986; Studdert-Kennedy, 1991) or the neural-motor commands underlying the articulation of these gestures (motor theory; Liberman, 1985; Liberman & Mattingly, 1989). In the motor-speech theory of speech perception, speech variability is filtered out by a human-specific mechanism of speech processing that infers the articulatory gestures intended in the production of distinctive speech features. In the direct-realism model of speech perception, however, distal articulatory gestures are directly perceived in the lawfully-ordered acoustic structure of the speech signal.

In other speech invariance models, however, the accommodation of speech variability is directly facilitated by the identification of acoustic-invariant properties of the speech signal. The most cited work in this area is probably that of Blumstein and Stevens (1979; also Stevens & Blumstein, 1978), which revealed a systematic correspondence

between three spectrum configurations or templates (diffuse-falling, diffuse-rising, and compact) and three major places of articulation (labial, alveolar, and velar) in the production of English stop consonants (however, Blumstein 1998 found that this correspondence is not cross-linguistically universal).

2.3.1.2 Speech variance models

In speech variance models, contextual variability is considered to play a fundamental role in speech processing. The accommodation of native-speech variability in speech variance models is not based on the identification of speech invariant features but on the processing of contextual and distributional parameters that account for the variation and co-variation of speech properties in the input. Most of the work done in this area could be summarized in two general models or approaches, depending on whether they focus on the processing of distributional properties in the long or short term experience. These two approaches are well represented by exemplar and perceptual calibration models, respectively.

Exemplar models diverge from abstractionist models of speech processing in that they do not rely on the use of perceptual prototypes. In exemplar models (Goldinger, 1998; Johnson, 2006; Nosofsky, 1988; Pierrehumbert, 2003), every pronunciation leaves a fine-grained acoustic trace (or exemplar) that is stored in the long-term memory along with the lexical representation of the corresponding word. Therefore, in these models, speech categories are perceptually structured as dynamic distributions of exemplars collected by long-term experience with different pronunciations of the same sounds across different talkers and phonetic contexts. From this perspective, the accommodation of each new

pronunciation is determined by their distributional proximity to distributions of exemplars previously collected. This idea is illustrated in Figure 4, in which the proximity of four hypothetical vowels (black circles) to the exemplar distributions of English /i/ and /ɛ/ was measured as the average distance of each token to exemplars of the corresponding distribution (for a more sophisticated criteria of distributional proximity see Kruschke, 1992).

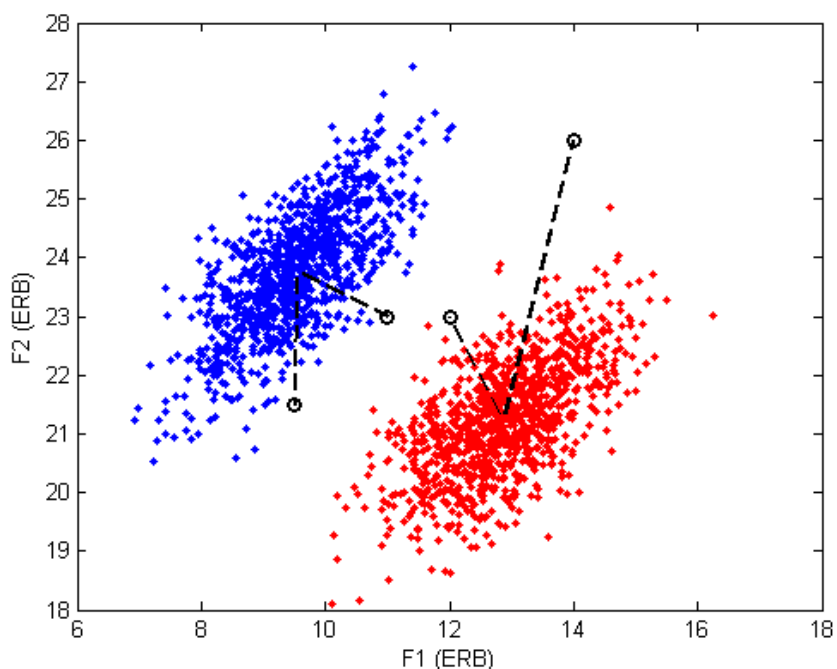


Figure 4. Classification of Vowel Exemplars. Classification of four vowel tokens (black circles) by their average proximity to all the exemplars included in the distributions of English /i/ (blue dots) and /ɛ/ (red dots) in an F1x F_2 formant space expressed in ERB units. Vowel distributional parameters are inspired by Hillenbrand et al. (1995)

The proposal that speech processing is determined by distributions of fine-grained acoustic information collected in the life span is supported by research. For example, Goldinger (1989) found that listeners rely on prior talker experience to speed up the

processing of words from familiar voices. Similarly, Mullenix and Pisoni (1990) showed that both phonetic cues and talker-specific features (e.g., talker voices) are highly integrated in speech processing, in that the perceptual evaluation of one of these dimensions (e.g., phonetic processing) is affected by the amount of variability provided along the other one (e.g., the number of talkers). Furthermore, recent research on bilingualism and second-language learning suggests that the processing of the native speech categories may change, across the life span, as a function of the amount and type of input from L1 and/or L2 (Chang, 2013; Flege, 1987; Mazzaro, Cuza & Collantoni, 2016; Schmid, 2013).

While exemplar models focus on the processing of distributional properties collected through long-term experience with the pronunciation of native speech categories, perceptual calibration models center on the variation and co-variation of speech cues in short-term experience with the speech context (Alexander & Kluender, 2010; Diehl & Kluender, 1988; Holt, Lotto & Kluender 2000; Lotto & Kluender, 1998; McMurray & Jongman, 2011; Sawusch & Pisoni, 1974). In perceptual calibration models, phonetic magnitudes are perceptually evaluated in contrast to their acoustic ranges in the previous speech context, such that the same phonetic magnitude (e.g., a specific formant frequency value) is perceived as lower or higher depending on whether its average range in the previous speech context is higher or lower, respectively. Therefore, in this type of model, speech variability is accommodated with respect to acoustic baselines, or expectations, that are calibrated on-line for each talker and phonetic context.

The perceptual calibration approach is also supported by evidence in the previous literature. For instance, the evaluation of temporal phonetic properties (e.g., voicing, vowel

or formant transition durations) is sensitive to the duration of segments in the immediate context (McMurray, Clayards, Tanenhaus, & Aslin, 2008; Miller & Dexter, 1988; Summerfield, 1981). Similarly, spectral speech properties are evaluated in contrast to their distribution in the previous speech context. For instance, in a seminal study on spectral calibration, Ladefoged and Broadbent (1957) found that acoustically ambiguous tokens between English [ɪ] (typically produced with a lower F1) and [ɛ] (typically produced with a higher F1) were more frequently perceived as /ɪ/ when F1 in the precursor sentence was shifted upward, making the test token F1 seem lower by comparison.

The spectral calibration of speech properties reported by Ladefoged and Broadbent (1957) has been replicated for different types of auditory contexts and targets, including non-adjacent speech precursors and non-speech targets (Holt 2005; Stephens & Holt, 2003). More recently, McMurray and Jongman (2011) found that the categorization of fricative consonants, which may be cued by up to approximately 24 acoustic correlates, can be replicated using a model employing a multiple regression analysis of speech cues (C-Cure; Jongman & McMurray, 2011) that calibrates the perceptual evaluation of acoustic cues based on their magnitude in a particular recent talker, sound, and phonetic context. Taken together, all these results support the view that native speech variability is accommodated by a rapid, context-dependent mechanism of perceptual calibration.

2.3.2 Accommodation of speech variability from different languages

In experienced second/foreign language learners and bilingual speakers, the perceptual accommodation of speech variability across different languages is typically achieved by the mechanisms of language switching and code switching (code-switching

typically refers to a locally-bounded switch; Pfaff, 1979; Lipski, 1982). The mechanism of language switching is not necessarily incompatible with any of mechanisms of perceptual accommodation reviewed in the previous section; listeners might still perceptually rely on the processing of long- or short-term distributional parameters, or the same set of speech invariant features. However, the mechanism of language switching establishes a fundamental difference with respect to these other mechanisms because it entails the ability to do it across two (or more) sets of language-specific phonetic representations. Another important difference between language switching and some of the models introduced above is that language switching is also modulated by factors that are not strictly related to the pronunciation of speech sounds in the previous speech context, such as the socio-pragmatic context (Auer, 2013; Blom & Gumperz, 2000; Bullock & Toribio, 2009; Toribio, 2004).

One of the earliest studies examining the mechanism of language switching in the processing of speech sounds across different languages was conducted by Elman, Diehl and Buchwald (1977). In their study, they investigated the effects of contextual language (English vs. Spanish) in the perceptual accommodation of stop-consonant voicing contrast (/b/ vs. /p/) by English-Spanish bilinguals. The selection of this contrast was motivated by the following cross-linguistic difference: while both Spanish and English include unaspirated stop-consonants (e.g., [p]) in their sound inventories, in English, unaspirated stops are treated as voiced (e.g. /b/) whereas in Spanish they are treated as voiceless (e.g., /p/) (Abramson & Lisker, 1972; Flege & Eefting, 1988). In their study, Elman et al. (1977) were able to induce listeners to employ either a Spanish- or English-specific voicing processing strategy to interpret the same stop consonant by changing the language of the precursor sentence to either Spanish or English.

Although the phonetic switching reported by Elman and colleagues was modulated by the manipulation of the speech context, other studies conducted in the last decade show that it may occur in the absence of auditory contextual cues. For example, García-Sierra, Ramírez-Esparza, Silva-Pereyra, Siard and Champlin (2012) recorded event-related potentials in Spanish-English bilinguals listening to syllables contrasting in velar stop consonant voicing (i.e., [ga], [ka], and [k^ha]) while silently reading Spanish or English passages. Results based on mismatch negativity revealed a language-specific (Spanish or English) mode of phonetic processing not predicted by speech properties alone.

2.3.3 Accommodation of foreign-accented speech variability

This section provides a theoretically detailed description of the three potential mechanisms of perceptual accommodation of foreign-accented speech sounds that were examined in the present study.

2.3.3.1 The mechanism of phonetic relaxation

This mechanism is inspired by exemplar models of speech processing (section §2.3.1). Like in exemplar models, it assumes that the lexical representations stored in the long-term memory contain fine-grained acoustic information about their pronunciation across different talkers and speech contexts in a particular language. According to this mechanism, foreign accents increase the amount of acoustic-phonetic variance of native-speech categories by mixing native- and foreign-accented exemplars in the lexical representations of the same words. As a consequence, the accommodation of foreign-accented speech sounds results in the relaxation of the native perceptual boundaries

between the speech categories that are phonetically mixed. In other words, as the acoustic-phonetic variance between speech categories increases, listeners adopt a less conservative perceptual criterion to minimize the chance of misclassifying foreign-accented tokens (for similar perceptual reactions to the increase of distributional noise in other modalities of perception, such as vision and touch, see Ernst and Banks, 2002).

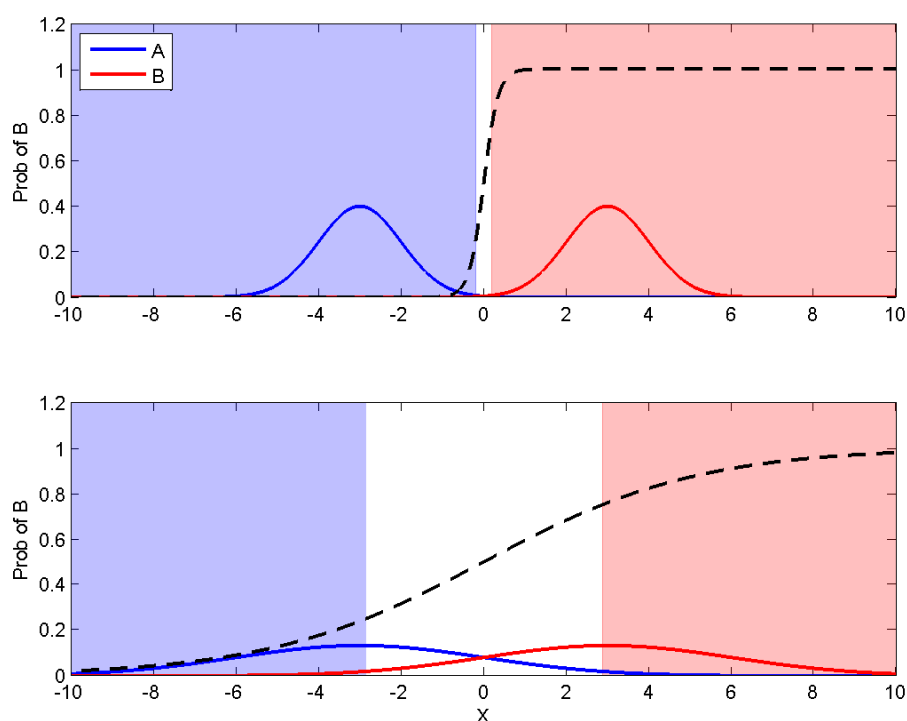


Figure 5. The Mechanism of Phonetic Relaxation. The top panel shows a schematic representation of the perceptual boundary (dashed line) between native distributions of speech categories A and B (solid lines), defined by two corresponding probability density functions along the same phonetic dimension X . The perceptual boundary can be interpreted as the probability of perceiving B along the same phonetic dimension X . White shaded areas bound the 25 – 75 percentile interval in the decision space. In the bottom panel, the perceptual boundary has been relaxed in proportion to the amount of distributional variance, or mixture, between the two categories, resulting in a less steep phonetic boundary

This idea is illustrated in Figure 5, in which the top panel shows a schematic representation of a typical perceptual boundary between two native distributions of speech categories, cued along one single phonetic dimension for the sake of simplicity. In the bottom panel, the perceptual boundary was smoothed, or relaxed, to account for the distributional mixture of conflicting pronunciations by native- and foreign-accented talkers.

The existence of a mechanism like phonetic relaxation is supported by research on foreign-accented speech. For instance, Baese-Berk, Bradlow, and Wright (2013) found that English listeners improved adaptation to a novel foreign accent (Slovakian-accented English) when they were previously exposed to five accents phonetically unrelated to the target (Mandarin, Thai, Korean, Romanian, Hindi) instead of just only one accent (Mandarin). Although it is difficult to estimate the individual contribution of each of the accents included in the study, it does not seem unreasonable to assume that the amount of distributional variance provided in the five-accent condition was potentially higher than in the one-accent condition. Thus, the results of Baese-Berk and colleagues could be explained by differences in amount of phonetic relaxation caused by exposure to phonetic tokens from multiple languages.

2.3.3.2 The mechanism of phonetic calibration

While the mechanism of phonetic relaxation is inspired by exemplar-based models of speech perception (section §2.3.1), the mechanism of phonetic calibration is inspired by perceptual calibration models. As in these models, the mechanism of phonetic calibration assumes that listeners calibrate the perceptual evaluation of phonetic magnitudes in

contrast to their acoustic ranges in the speech of a particular talker or in a particular speech context. As a consequence, the perceptual accommodation of foreign-accented speech sounds in this model is not facilitated by the relaxation of perceptual boundaries between native categories. Instead, the location of these boundaries is re-calibrated or adjusted to account for the acoustic-phonetic realization of these categories in the previous (foreign-accented) speech context.

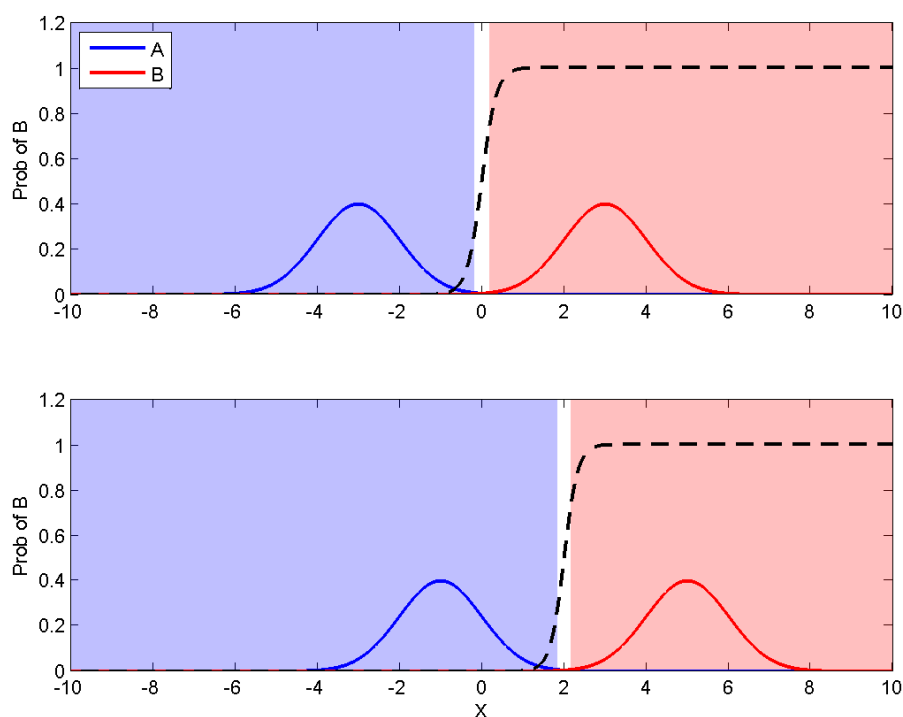


Figure 6. The Mechanism of Phonetic Calibration. The top panel shows a schematic representation of the perceptual boundary (dashed line) between native distributions of speech categories A and B (solid lines), defined by two corresponding probability density functions along the same phonetic dimension X . The perceptual boundary can be interpreted as the probability of perceiving B along the same phonetic dimension X . White-shaded areas bound the 25 – 75 percentile interval in the decision space. In the bottom panel, the perceptual boundary has been shifted from $X = 0$ to $X = 2$ or calibrated to account for one hypothetical non-native realization of the same categories

This idea is illustrated in Figure 6, which depicts a schematic representation of the calibration of the perceptual boundary between two hypothetical speech categories across two short-term distributions of these categories in a native-accented (top panel) and foreign-accented speech context (bottom panel).

The existence of a mechanism like phonetic calibration is also supported by research on the perception of foreign-accented speech. For instance, Maye et al. (2008) created an artificial foreign accent by lowering the spectral quality of front vowels (e.g. [ɪ] realized as [ɛ], and [ɛ] realized as [æ]). After being exposed to that accent, two groups of English listeners performed a lexical decision task in which the spectral quality of accented words was either lowered (group 1) or raised (e.g. [ɛ] realized as [ɪ]; group 2). Overall endorsement rate increased among those accented words that matched the pattern of phonetic deviation characteristic of the exposure, just as it would be expected from the phonetic (re)calibration of perceptual boundaries.

2.3.3.3 The mechanism of phonetic switching

Inspired by the mechanism of language switching (section §2.3.2), the mechanism of phonetic switching proposes that the accommodation of native- vs. foreign-accented speech is mediated by different language-specific modes of processing. However, an important difference between the mechanism of phonetic switching and language switching is that the mechanism of phonetic switching only involves switching between sound systems, but not necessarily syntactic or morphological properties.

According to this mechanism, listeners switch to the sound system that better accounts for the type of speech variability in the input. For example, native listeners may

switch to a non-native phonetic representation of, for example, /p/ and /b/ based on the non-native pronunciation of other sounds. Therefore, the mechanism of phonetic switching is less context-dependent than the mechanism of phonetic calibration because, by switching the sound system, listeners can resolve foreign-accented pronunciations that were not anticipated in the previous speech context, just as happens in code-switched speech production, which is not entirely predictable from the preceding speech (Pfaff, 1979; Lipski, 1982; Toribio, 2004; Bullok & Toribio, 2009).

In support of this last observation, Gonzales and Lotto (2013) found that the language-specific processing of speech categories can be modulated by the accent of contextual speech properties unrelated to the pronunciation of those categories. In their study, they examined the processing of bilabial stop-consonant voicing in non-words ([b]afri – [p]afri – [p^h]afri) in which the r-segment was edited to sound like a phonetic realization of either English or Spanish /r/. They found that English-Spanish bilinguals encoded voicing in an English- or Spanish-specific manner based on the accent quality of the /r/, which does not provide any direct clue about the pronunciation of /b/ in English or Spanish but does serve to indicate which language's phonetic system should be applied.

2.4 Goals and Research Questions

The goal of the present study is to determine how foreign-accented sounds are accommodated in listening to foreign-accented speech and to clarify the potential contributions of experience with the phonetic norm of the foreign accent and the speech context in such accommodation. As was noted in section §2.2, while foreign accents are initially harder to process, listeners adapt to them in a way that goes beyond the mere

memorization of foreign-accented words. This raises the question of how listeners actually accommodate foreign-accented speech sounds from previous experience. The present study attempts to fill this theoretical gap by addressing the following overarching question:

OQ: Does accent-specific experience changes the phonetic accommodation of sounds from that accent?

This overarching question is motivated by the proposal that experienced listeners are doing something other than inferring the intended pronunciation of foreign-accented words from the lexical, semantic or pragmatic context. In particular, if adaptation to foreign-accented speech is achieved by the generalization of the phonetic norm that is characteristic of the accent, then native listeners may give a different perceptual treatment to foreign-accented speech depending on the amount and type of experience with the phonetic norm of the foreign accent (Bradlow & Bent, 2008; Sidaras et al., 2015).

The overarching question is complemented with three general research questions:

GQ1: Do accent-experienced listeners relax the slope of perceptual boundaries in listening to foreign-accented speech?

GQ2: Do accent-experienced listeners calibrate the location of perceptual boundaries in listening to foreign-accented speech?

GQ3: Do accent-experienced listeners switch to a non-native set of phonetic boundaries in listening to foreign-accented speech?

Each of these research questions is related to each of the hypotheses of perceptual accommodation of foreign-accented speech sounds that were introduced in section §2.3.3 (phonetic relaxation, phonetic calibration, and phonetic switching). For instance, if accent-experienced listeners accommodate foreign-accented speech sounds by means of phonetic relaxation, then they are expected to relax native perceptual boundaries in a foreign-accented speech context. Alternatively, if foreign-accented speech sounds are accommodated by means of phonetic calibration, experienced listeners are expected to calibrate the location of target perceptual boundaries to match their native- or foreign-accented distribution in the previous speech context. Finally, if experienced listeners accommodate foreign-accented speech sounds by phonetic switching, then they are expected to switch between two language-specific perceptual boundaries across accents, even in the absence of clear contextual cues to the phonetic realization of target speech in the native- or foreign-accented speech norm. These three research questions were initially addressed in Experiment 1 (Chapter 3).

CHAPTER 3. EXPERIMENT 1 - THE EFFECTS OF EXPERIENCE WITH ENGLISH-ACCENTED SPEECH ON THE PROCESSING OF ENGLISH-ACCENTED SPANISH VOICING BY NATIVE LISTENERS OF SPANISH

3.1 Motivation and Overall Design

Experiment 1 examined the effects of having experience with English-accented speech on the processing of stop-consonant voicing in a native- and English-accented Spanish context. Two groups of native Spanish listeners with a low and high degree of experience with English-accented speech were tested in the accommodation of the bilabial stop-consonant voicing contrast (i.e., /b/ vs. /p/; like in the Spanish minimal pair *bata* ‘robe’ vs. *pata* ‘paw’).

Stop consonant voicing contrasts provide an optimal way of investigating the accommodation of English-accented Spanish sounds by native Spanish listeners because, in word-initial position, English-accented Spanish /b/ is typically pronounced as Spanish /p/ (i.e., as an unaspirated English [p]). For English listeners, /b/ and /p/ are perceptually distinguished by a perceptual VOT boundary located at around 20 ms ($/b/ < 20$ ms, $/p/ \geq 20$ ms), whereas the native Spanish boundary is closer to 0 ms VOT ($/b/ < 0$ ms, $/p/ \geq 0$ ms) (Abramson & Lisker, 1972; Flege & Eefting, 1988; García-Sierra et al., 2012; Llanos, Dmitrieva, Shultz & Francis, (2013). The top and middle panels of Figure 7 show the typical VOT distribution of /b/ and /p/ in Spanish (top panel) and English (middle panel

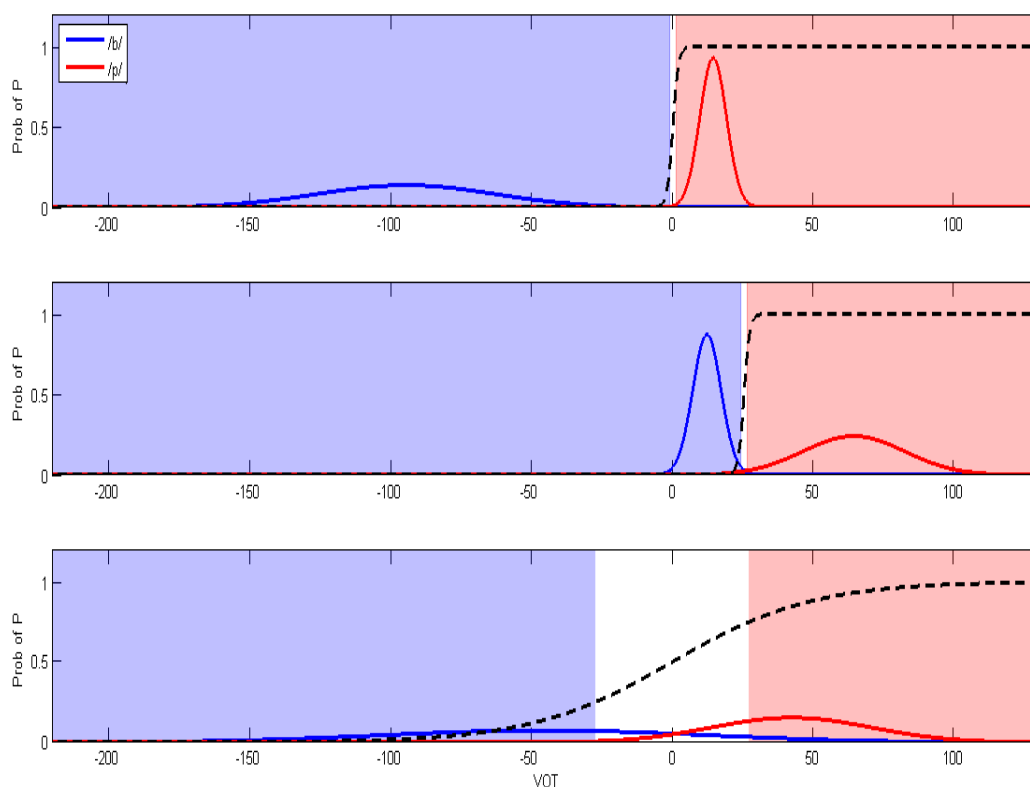


Figure 7. VOT Boundaries. Top: Schematic VOT boundary (dashed line) between two typical VOT distributions of /b/ and /p/ in Spanish. Middle: Schematic VOT boundary (dashed line) between two typical VOT distributions of /b/ and /p/ in English. Bottom: Schematic relaxed VOT boundary between two distributions fitted to production of /b/ and /p/ in both English and Spanish. VOT normal distributions were fitted to real production data collected by Dmitrieva et al. (2015). For the sake of visualization, distributions were re-scaled by a factor of 10. With shaded areas bound the 25 - 75 percentile interval corresponding to the probability of identifying /p/, as a function of the VOT boundary

Hypothetically, the exposure to the English-like pronunciation of a target /b/ as a Spanish /p/ might encourage Spanish listeners to relax the difference between /b/ and /p/ in an English-accented Spanish context, thus relying on a less steep VOT perceptual boundary. This hypothesis (the phonetic relaxation hypothesis) is schematically

represented in the bottom panel of Figure 7. On the other hand, experienced listeners might adjust the location of the Spanish VOT boundary to account for the VOT realization of /b/ and /p/ in the previous speech context. If this hypothesis is true (the phonetic calibration hypothesis), then Spanish listeners are expected to rely on either a Spanish- or English-like VOT boundary depending on whether contextual /b/ and /p/ are phonetically realized as either in Spanish (Figure 7, top panel) or English (Figure 7, middle panel). Finally, based on the phonetic realization of other sounds than /b/ and /p/, experienced listeners might switch between a Spanish or an English speech representation of /b/ and /p/. If this hypothesis is true (the phonetic switching hypothesis), then listeners are expected to switch between a Spanish- and English-like VOT boundary across accents, even in the absence of clear contextual cues to the perceptual calibration of /b/ and /p/ in Spanish or English.

In Experiment 1, participants classified syllables with variant VOT ranging from /b/ to /p/ presented in an English-accented Spanish context, and also in a baseline context of native-accented Spanish. To test the phonetic switching hypothesis, none of the speech contexts included exemplars of /b/ or /p/ that could be used to calibrate the location of the VOT boundary between /b/ and /p/. Speech contexts consisted of simple, natural Spanish sentences to maximize the realism of the context so that listeners had the greatest possible basis for recognizing the accent as either Spanish or English.

3.2 Methods

3.2.1 Participants

Two groups of participants were recruited and tested in West Lafayette, Indiana. The first group (the experienced group) consisted of 16 Spanish-English bilinguals (9 women,

7 men) that were also instructors of Spanish at Purdue University ($M = 32.9$ years; $R = 24 - 43$ years, $SD = 4.3$ years). This selection of participants was made to maximize the potential amount of contact with the English-accented norm via natural experience with native-accented English and English-accented Spanish. All of them were native speakers of Spanish from different Hispanic countries: five from Colombia, five from Spain, three from Chile, and three from Argentina, the Dominican Republic, and Venezuela. The second group (the inexperienced group) consisted of 16 native listeners of Castilian Spanish (11 women, 5 men), recruited and tested in Madrid, Spain ($M = 34.7$ years, $R = 22 - 40$ years, $SD = 3.2$ years).

Although Castilian Spanish and some Latin-American Hispanic dialects may differ in terms of prototypical VOT in production, this difference is very small (on the order of 30 ms of prevoicing for /b/ and 4 ms for /p/, based on Rosner et al., 1998, and Williams, 1977) as compared to the difference between Castilian and English VOT (of the order of 100 ms for /b/ and 45 ms for /p/; cf. Dmitrieva et al., 2015).

Table 1. Language Experience in Experiment 1

	Experienced Group	Inexperienced Group	Two-sample T-Test
LOR	$M = 7.2$ years	$M = 0.2$ years	$t(30) = -6.01$ $p < 0.001$, $\eta^2 = 0.54$
ESC	$M = 3.1$ Likert 1-5	$M = 1.4$ Likert 1-5	$t(30) = -6.06$ $p < 0.001$, $\eta^2 = 0.55$
EASC	$M = 4.8$ Likert 1-5	$M = 1.6$ Likert 1-5	$t(30) = 11.57$ $p < 0.001$, $\eta^2 = 0.81$

Note. First and second columns: average scores for the experienced and the inexperienced group with respect to length of residence (LOR), relative English/Spanish contact (ESC), and English-accented Spanish contact (EASC). Third column: t-statistics, p-values and size effects (η^2) for the between-group comparison included in the corresponding row

Given that all of the participants of the experienced group arrived in the USA at a similar age, as second language adult learners ($M = 23.2$ years) the amount of experience with English-accented speech was quantified in terms of the following three variables (cf. Flege, 2009; Piske, McKay, Flege, 2001): length of residence (LOR), measured by the number of years spent in the USA; the amount of relative English/Spanish contact (ESC), measured in a Likert scale from 1 *all contact in Spanish* to 5 *all contact in English*; and the amount of contact with English-accented Spanish (EASC), measured in a similar Likert scale from 1 *almost never* to 5 *almost always*. Table 1 shows the mean scores obtained by each group on LOR, ESC and EASC. A battery of three two-sample t-tests revealed that the experienced group scored significantly higher than the inexperienced group across the three variables (Table 1, 3rd column).

3.2.2 Speech materials

Speech items consisted of one Spanish minimal pair (*bata* ‘robe’, *pata* ‘paw’) and four Spanish precursor sentences (Table 2) that were provided to each speaker in a written format. Speech materials were recorded from a male native speaker of American English and a male native speaker of Castilian Spanish. To assess the strength of the accent, the native speaker of American English had little background experience on Spanish. To ensure the fluency of English-accented Spanish materials, the English talker was provided with a Spanish native-accented model of each item prior to each reading.

Although hearing a native Spanish model may have encouraged the English speaker to produce stimuli with an accent somewhat more native-like than otherwise, this was deemed an acceptable risk in order to ensure that the sentences were fluent, as fluency may

drastically affect the processing of foreign-accented speech (Ginther, Slobodanka, & Jang, 2010; Pinget, Bosker, Quené, & de Jong, 2014; Riazantseva, 2001). However, to confirm that listeners perceived recorded stimuli as having been produced with different accents, original speech productions were presented to five native speakers of Spanish who did not participate in the main the experiment and who rated them for accent strength in a Likert scale from 1 *strong Spanish accent* to 4 *strong English accent*. Ratings were clearly different, with English-accented Spanish stimuli receiving a mean rating of 3.8 as compared to a mean rating of 1.0 for the native Spanish items.

Table 2. Speech Items in Experiment 1

Minimal Pair	Precursors
<i>Bata</i> ‘robe’	<i>Lo que yo te dije es</i> ‘what I told you is’
<i>Pata</i> ‘paw’	<i>Lo que él me dijo es</i> ‘what he told me is’
	<i>Esto que leo aquí es</i> ‘what I read here is’
	<i>Esto dice que es</i> ‘this is saying this’

The minimal pair consisted of two Spanish words (*bata* ‘robe’ and *pata* ‘paw’) differing in terms of the initial consonant (/b/ and /p/, respectively). Because these items were recorded in order to be used as a base sample for subsequent resynthesis, the speakers, both of whom had some phonetic training, were instructed to produce a range of VOTs. Each produced multiple repetitions of each word with the goal of obtaining at least one token with a strongly negative VOT token (VOT < -60 ms) of *bata* and one with a strongly positive VOT token of *pata* (VOT > 60 ms). To avoid semantic, segmental or supra-segmental bias, precursors were semantically neutral, of similar length, produced with a similar pitch contour, and ended in the same voiceless fricative consonant.

To reduce the opportunity of phonetic calibration (i.e., the establishment of expectations about the talker's bilabial stop categories), none of the precursors included exemplars of /b/ and /p/. They did, however, include a few coronal and velar stops, which were difficult to avoid. In particular, they included five voiceless velar stops (/k/) in the intervocalic position, one in each of the four tokens of the word *que* and one in the only token of the word *aquí*. They also included three voiced coronal stops (/d/), one in each of the tokens of *dije*, *dice*, and *dijo*, as well as three voiceless coronal stops (/t/), one in each of the three tokens of the word *esto*. Finally, each target token (*bata/pata*) included one instance of /t/.

Except for the two voiceless coronal stops /t/ in the word *esto*, all the contextual stops occurred in the intervocalic context, in which stop-consonant voicing differences are acoustically less prominent. For example, the VOT tends to disappear in Spanish in the intervocalic context because of the spirantization of voiced stops (Barlow, 2003; Martínez Celdrán & Planas, 2007; Ortega-Llebaria, 2003). In English, intervocalic stops tend to also undergo consonant lenition in the intervocalic context (Warner & Tucker, 2011). Also, English /t/ typically becomes unaspirated when it is immediately preceded by /s/, as in the English-accented Spanish word *esto*, thus limiting the existence of potential VOT contextual differences across accents.

Average VOT values for /k/, /t/, and /d/ in English-accented Spanish materials were 30.8, 23.0, and -44.6 ms respectively. Averaged VOT for intervocalic /k/ and /t/ in Spanish native materials were 21.2 and 16.2 ms, respectively (spirantized Spanish /d/ was excluded from the analyses). Statistical analyses of contextual VOT (a one-way ANOVA followed by a Tukey HSD post-hoc test) revealed a significant difference only between English-

accented /d/ and the rest of the stops (English accented /k/, English accented /t/, Spanish native /k/, and Spanish-native /t/). Thus, although the VOT values of the English speaker's voiceless stops were somewhat longer than those of the Spanish speaker, as might be expected, the difference was not statistically significant, probably due to the phonetic contexts in which these tokens appeared.

In summary, acoustic analyses suggested that the differences between the VOT values of the stop consonants in the different context sentences was not likely to bias listeners' expectations toward the English or Spanish phonetic norm, although of course many other phonetic aspects of the sentences (e.g. vowels) were clearly identifiable as English- or Spanish-like (respectively) by the five raters mentioned previously.

3.2.3 Speech resynthesis

Tokens were normalized to the root mean square intensity (RMS) of a 1000 Hz sine wave at 66 dB. Then, recorded tokens of *bata* and *pata* were examined to identify, for each talker, a baseline token of *bata* with a VOT value clearly lower than -60 ms and a baseline token of *pata* with a VOT clearly greater than 60 ms. Once these baseline tokens were identified, a series of VOT tokens, ranging from *bata* (-60 ms VOT) to *pata* (60 ms VOT) in terms of the first consonant, were resynthesized across 13 VOT steps of 10 ms each (Figure 8 shows some resynthesized tokens). This VOT range was selected to support the use of logistic regression analyses by including a significant amount of VOT tokens judged as /b/ and /p/ in both accents while still minimizing the total number of stimuli in the continuum to better constrain testing time (e.g., Llanos et al., 2013).

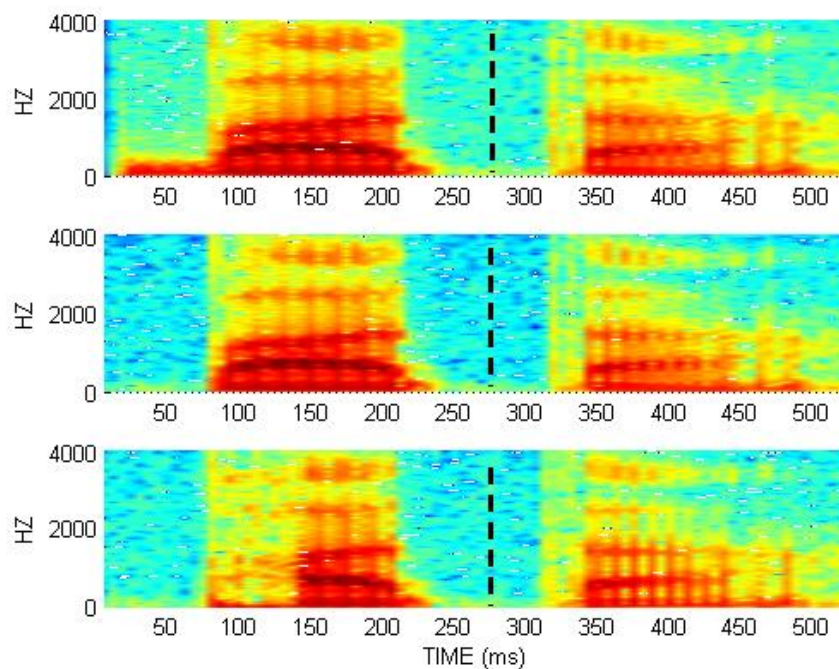


Figure 8. VOT Resynthesized Samples. Samples of edited tokens of *bata* and *pata* with voice onset time (VOT) from top to bottom of -60 , 0 , and 60 ms. VOT portions different from 0 ms are delimited by the low-frequency energy that can be observed at the beginning of the first spectrogram (VOT = -60 ms), or the aspiration shown right after the consonant release, at 80 ms, in the last spectrogram (VOT = 60 ms). Tokens are aligned by their consonant release. The dashed line indicates the syllable boundary

The seven voicing lead exemplars of the VOT series ($-60 \text{ ms} \leq \text{VOT} \leq 0 \text{ ms}$) were resynthesized by cutting out successively longer portions of prevoicing from the baseline token of *bata*. Previous research shows that the categorization of stop-consonant voicing may involve other cues than VOT, such as the onset of the first formant after the consonant release, or the value of F_0 at the onset of voicing (Holt, Lotto, & Kluender, 2001; Kingston, Diehl, Kirk, & Castleman, 2008; Kluender, 1991; Lisker, 1986; Raphael, 2005). To control for the effects of secondary cues to voicing, which were outside of the scope of the study, short lag and log lag tokens ($0 \text{ ms} < \text{VOT} \leq 60 \text{ ms}$) were resynthesized from the same 0 ms voicing lead token that resulted from the edition of the baseline version of *bata*. This

token was then cross-spliced with successively longer portions of aspiration, extracted from the baseline token of *pata* (of 60 ms VOT). Thus, portions of aspiration of the appropriate VOT length were inserted in the 0 ms token of *bata* to create six additional tokens with VOT values from 10 to 60 ms VOT (this cross-splicing method is illustrated in the left panel of Figure 9). The intensity of each cross-spliced portion was linearly attenuated in inverse proportion to its duration. To avoid acoustic artifacts, original and cross-spliced boundaries were smoothed by means of a cubic Hermite spline (see Figure 9, right panel).

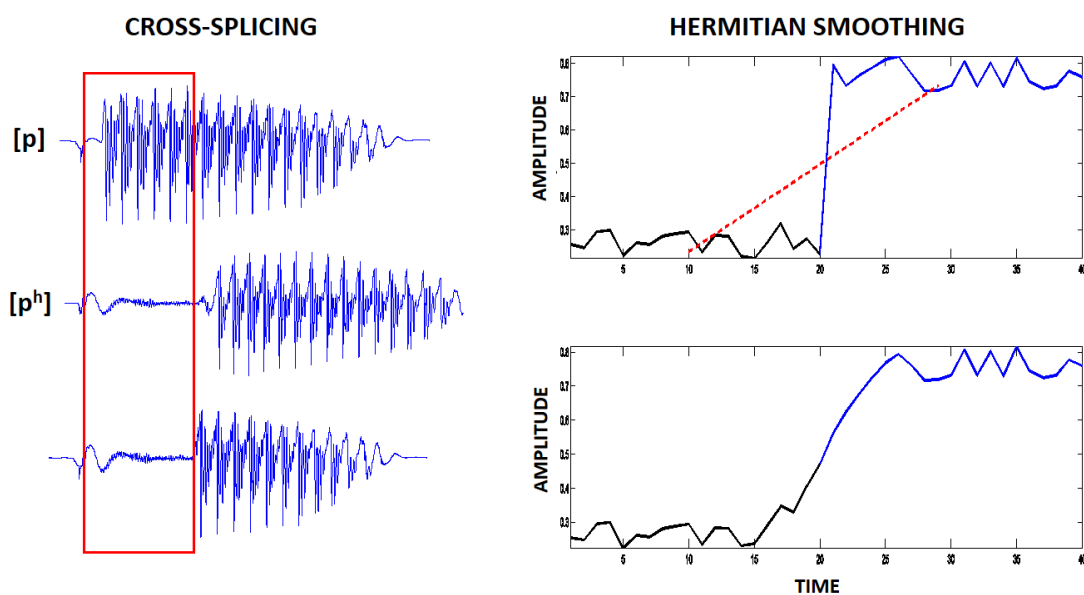


Table 9. Cross-splicing: from top to bottom, speech waveforms of a 0 ms VOT instance of [p], a 60 ms VOT instance of [p^h], and the 0 ms VOT instance of [p] cross-spliced with the aspirated portion of the 60 ms instance [p^h]. Hermitian smoothing: example of two cross-spliced, randomly generated waveforms before (top) and after the Hermitian interpolation that was applied to smooth cross-spliced junctures (bottom)

The resynthesis protocol explained in the previous paragraph was independently applied to speech samples from each talker. Therefore, experimental stimuli consisted of the four same-talker precursors combined with each of the 13 VOT resynthesized words from the corresponding talker; a total of 104 sentences across conditions: 52 sentences in the native-accented Spanish condition, and 52 sentences in the English-accented Spanish condition.

3.2.4 Procedure

Participants were seated in a quiet room in front of an Acer Aspire 5830TG laptop computer showing an image of a rope (*bata*) and an animal paw (*pata*). No text was shown on the screen. All participants were tested in both conditions. Testing was conducted in two consecutive sessions of approximately 15 minutes each with a short break of less than 1 minute between sessions. The ordering of sessions was counterbalanced such that half of the participants in each group started with the Spanish-accented condition (ordering SE) and the other half started with the English-accented condition (ordering ES).

The experiment was controlled by a custom-written Psychophysics Toolbox 3.0.8 interface implemented in MATLAB R2011b. During the experiment, participants heard auditory stimuli presented at 66 dB via Sennheiser HD pro 280 headphones connected to a Dr. NANO USB external sound card. Stimuli in each session were presented in six randomized blocks of 52 experimental sentences each. After listening to each sentence, participants were asked to indicate the picture corresponding to the last word that they heard (*bata* ‘robe’ or *pata* ‘paw’) by pressing a button on a USTC RTBox 5.x response box (Li, Liang, Kleiner, & Lu, 2010). Participants were paid at a rate of US \$10 or € 8 per hour.

Throughout the experiment all participants were addressed only in Spanish by a native Speaker of Spanish, and all the background questionnaires and forms were written in Spanish.

3.3 Analysis

The location of the VOT boundary was estimated as the 50% cross-over point of the logistic regression curve. This point, commonly referred to as the median effective level, was calculated as $-\alpha/\beta$, in which α refers to the intercept and β refers to the first beta-coefficient of the logistic model (Agresti, 1996). The steepness of the VOT boundary was estimated as the slope of the logistic curve at the median effective level, calculated as 0.25β (Agresti, 1996).

To test the predictions made for each hypothesis of accommodation (see Table 3), individual boundary locations and slopes were analyzed using two independent mixed effects ANOVAs with location (or Slope) as the dependent variable, group (experienced, inexperienced) and ordering (SE, ES) as the between-subject factors, and Condition (English accented, native Spanish accented) as the within subject factor.

Table 3. Predictions in Experiment 1

Hypotheses	VOT Shift	Steepness Reduction
Phonetic relaxation	--	Yes
Phonetic calibration	No	No
Phonetic switching	Yes	No

3.4 Results

The mixed effects ANOVA for Slope did not reveal any significant effect of Group, Condition, Order, or any interactions. The mixed effects ANOVA for Location showed a main effect of group, $F(1, 28) = 7.92, p = 0.008, \eta^2 = 0.15$, and condition, $F(1, 28) = 4.92, p = 0.03, \eta^2 = 0.06$, and a significant three-way interaction of group, condition, and order of sessions, $F(1, 28) = 4.41, p = 0.04, \eta^2 = 0.05$.

The main effect of condition was examined by a two-sample t-test for Location across conditions. Results indicated that the average location of the VOT boundary in the English-accented condition ($M = 5.9$ ms) was significantly higher than in the native-Spanish condition ($M = 2.1$ ms), $t(62) = 2.19, p = 0.03, \eta^2 = 0.07$.

The main effect of Group was examined by a two-sample t-test for boundary location across groups. Results revealed that the experienced group relied on an average boundary location significantly higher ($M = 5.8$ ms) than the one of the inexperienced group ($M = 1.2$ ms, $SD = YYY$), $t(62) = 2.94, p < 0.001, \eta^2 = 0.12$.

The interaction of Group, Order, and Condition was explored by means of a battery of four independent two-sample t-tests (with p-values Bonferroni corrected to a threshold of 0.0125) to test for significant differences in VOT boundary location across condition for each possible combination of group and order (Figure 10).

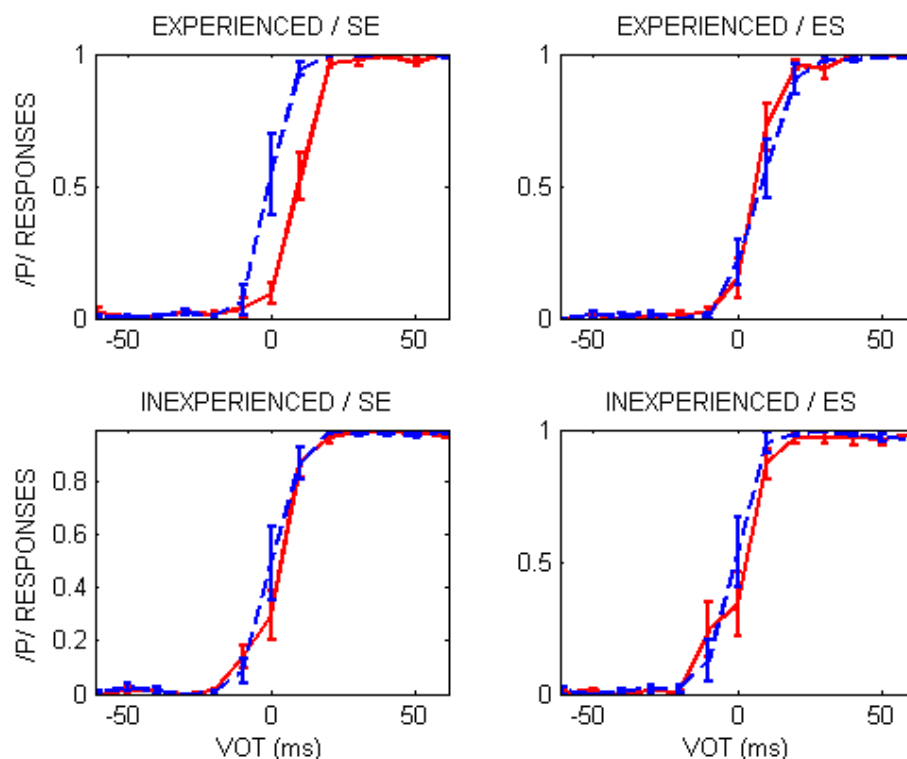


Figure 10. Voicing-identification Response Curves from Experiment 1. Averaged proportions of /p/ responses with standard errors (y-axis) for each target voice onset time (VOT) value (x-axis) in the English-accent condition (red solid line) and the native Spanish condition (blue dashed line). Top left: experienced listeners' performance in the SE order (Spanish accent first). Top right: experienced listeners' performance in the ES order (English accent first). Bottom left: inexperienced listeners' performance in the SE order. Bottom right: inexperienced listeners' performance in the ES order.

Results of the t-tests showed a significant difference between VOT boundaries only for the experienced listeners tested in the SE ordering (native Spanish first), such that the average VOT boundary in the English-accent condition was significantly higher than in the native-Spanish condition, $t(14) = -3.38, p = 0.004, \eta^2 = 0.45$. The other three comparisons did not reach significance (Table 4).

Table 4. VOT Boundaries from Experiment 1

	Ordering SE		Ordering ES	
	Session 1	Session 2	Session 1	Session 2
	Native	English	English	Native
	accented	accented	accented	accented
	Spanish	Spanish	Spanish	Spanish
Inexperienced group	[1.36 ms	1.64 ms]	[2.6 ms	-0.55 ms]
Experienced group	[-0.79 ms	7 ms] *	[8.67 ms	8.65 ms]

Note. VOT boundary locations averaged for each possible combination of group, order, and condition. Averaged boundaries (eight total) are bracketed into four groups of two means each to indicate the scope of each post hoc t-test. The only two significantly different VOT boundaries across accents were found in the experienced group in the ES order.

3.5 Discussion

The lack of a significant effect of Slope across accents indicates that listeners did not relax the VOT boundary across voicing categories. However, the main effects of Condition and Group both seem to support the phonetic switching hypothesis, which predicted that only experienced listeners would shift the VOT boundary across accents. The main effect of Condition shows that experienced listeners shifted the location of the VOT boundary across accents even in the absence of clear contextual cues to the pronunciation of target voicing. In addition, the main effect of Group indicates that the shift only occurred when listeners were highly familiar with English phonetic norms.

The phonetic switching hypothesis is further supported by the interaction of order, group, and condition. While inexperienced listeners relied on the same native Spanish boundary regardless of condition and order, the performance of the experienced group was more complex. In the SE order, in which they were first exposed to native Spanish, the

perceptual VOT boundary in English-accented Spanish was significantly higher than in native-accented Spanish. However, in the ES order, in which experienced listeners were exposed first to English-accented Spanish and then to native Spanish, they perseverated in holding an English-like boundary even into the second, native-accented Spanish, session.

The performance of the inexperienced group confirms that none of the acoustic differences between the English- and Spanish-accented precursors were sufficient to trigger a boundary shift across accents in listeners without English experience. This highlights the important role played by previous experience with English-accented norms in experienced listeners, who shifted the location of the VOT boundary despite the absence of clear contextual cues to the phonetic calibration of /b/ and /p/ in English.

In addition, the difficulties shown by experienced listeners in returning to their dominant language (Spanish) in the ordering ES (English-accented Spanish first) suggests that their mode of perceptual processing was highly independent of the acoustic information provided by the speech context; otherwise, they should have been able to shift to a Spanish-like VOT boundary in the native-accented Spanish context. Although this lack of return to the dominant background was unexpected, it could be a consequence of the higher amount of cognitive resources recruited by late bilinguals when processing speech from the non-dominant language (i.e., English), relative to the processing of speech from the dominant one (i.e., Spanish). While bilinguals tend to be better at discriminating sounds from their dominant language (Antoniou, Tyler, & Best, 2012), they seem to require more time to switch back from their non-dominant language to the dominant one than vice versa (Olson, 2013). One possible basis for this increased switching time is that bilinguals may invest greater processing resource in listening to their non-dominant language (perhaps

because they are actively inhibiting processing of their dominant language; cf. Green 1986, 1998), and therefore have fewer resources available to process competitors, even those that enable switching back to their native language. Thus, considering that the experienced listeners in the present study fit the criteria for being unbalanced bilinguals with Spanish as their dominant language, it is possible that they also recruited more cognitive resources for the processing of English-accented Spanish, thereby ending up with insufficient resources to switch back to the native competitor (i.e., Spanish). While previous studies have indeed shown that bilinguals invest more processing capacity in understanding speech in competing speech (Colzato et al., 2008; see also Bialystok, 2007; Brouwer, Van Engen, Calandruccio, & Bradlow, 2012), it is not yet clear whether they are actually investing proportionately more when listening to their non-dominant language than to their dominant one. Further research is necessary to distinguish between these possibilities.

Another aspect that remains unclear is the nature of the English-like VOT boundary observed in the experienced group. Previous research shows that Spanish-English bilinguals, and Spanish learners of English, may rely on an English-like VOT boundary that is significantly lower (i.e., more Spanish-like) than the boundary used by native speakers of English (Elman, Diehl, & Buchwald, 1977; Flege & Eefting, 1987; Garcia-Sierra, Diehl, & Champlin, 2009). This has been also observed in studies of code-switching production, in which the lower boundary was attributed to a linguistic interference of the native norm at the switching interface. For example, Bullock, Toribio, González, and Dalola (2006) showed that, when switching from Spanish to English, English-Spanish bilinguals tend to produce shorter, more Spanish-like VOT tokens, as compared to the VOT of their tokens produced in non-switched English.

Based on these and other results suggesting that bilinguals do not simply switch between two strictly monolingual modes (Grosjean, 2001), it is possible that the English-like VOT boundary reported for the experienced group does not correspond to the boundary expected in a truly native-accented English context. In other words, it is possible that the boundary used to encode English-accented Spanish voicing resulted from a mode of bilingual processing that is intermediate between the Spanish and English monolingual modes.

3.6 Summary and Specific Research Questions

Experiment 1 examined the categorization of /b/ and /p/ by two groups of native Spanish listeners with a low and high degree of experience with English-accent norms. They performed in a native- and English-accented Spanish context with no clear phonetic bias to the phonetic realization of target voicing in either Spanish or English. Only the group of English accent-experienced listeners was able to shift the location of the perceptual VOT boundary between /b/ and /p/ across accents, as was predicted by the phonetic switching hypothesis. This result lends support to the hypothesis that experienced listeners accommodated English-accented Spanish voicing by switching to a non-native VOT boundary that was not calibrated from the previous speech context. Results of Experiment 1 motivate two specific research questions:

SQ1: Do English-accent experienced listeners process English-accented Spanish voicing differences with an intermediate VOT boundary between the Spanish and English one?

SQ2: Can English-accent inexperienced listeners calibrate the location of the Spanish VOT boundary in a speech context with contextual cues biased to the phonetic realization of target voicing in English?

Given that experienced listeners in Experiment 1 were bilingual speakers of Spanish and English, the first specific research question (1) asks whether the perceptual VOT boundary exhibited by them in the English-accented Spanish context would be the same that they would exhibit in an English native-accented speech context. This question is addressed in a second experiment (Chapter 4). Also, given that the same group of listeners seemed to accommodate English-accented Spanish voicing by phonetic switching, and that the speech context in Experiment 1 did not facilitate clear cues for phonetic calibration, the second specific-research question (2) asks whether the phonetic calibration hypothesis would at least hold for those listeners that are more unfamiliar with the phonetic norms of the foreign accent. This question is addressed in Experiment 4 (Chapter 6).

CHAPTER 4. EXPERIMENT 2 - THE PROCESSING OF ENGLISH NATIVE-ACCENTED VOICING BY BILINGUAL SPEAKERS OF SPANISH AND ENGLISH

4.1 Motivation and Overall Design

The goal of Experiment 2 was to determine whether the VOT boundary shown by the group of experienced listeners in the English-accented Spanish context in Experiment 1 was an intermediate VOT boundary between the Spanish and English one or an English-like VOT boundary. To address this question, experienced listeners were tested on their processing of /b/ and /p/ in a native-accented English context in order to induce an English mode of phonetic processing.

4.2 Methods

Three months after Experiment 1, 10 participants from the group of experienced listeners were recruited. Six of the participants that returned to participate in Experiment 2 were tested in the ES order from Experiment 1 (English-accented Spanish first) and the other four participants were tested in the SE order (native-accented Spanish first).

A new set of stimuli were recorded and resynthesized following the methods described in Experiment 1 (section §3.2). Since the purpose of Experiment 2 was to determine whether the VOT boundary shown in English-accented Spanish was also used in a realistic English-monolingual context, speech materials were recorded by a native

speaker of a Midwestern dialect of American English speaking English (the same speaker who produced the English-accented Spanish in Experiment 1).

This new set of stimuli consisted of two English words (*barking*, *parking*) and four English sentential precursors (see Table 5). The two English words had the same trochaic structure as the two Spanish words recorded in Experiment 2 (*bata* ‘robe’, *pata* ‘paw’), but contained phonemes that were clearly English. Also, as in Experiment 1, precursors were semantically neutral, of similar length and prosodic structure, and ended in the same fricative consonant. Although speech materials included several non-bilabial stops (as in Experiment 1), they were naturally produced by a native speaker of English speaking native-accented English. This should guarantee the derivation of the English-like VOT boundary, such that listeners should perform as they would in a realistic English monolingual environment.

Table 5. Speech Items from Experiment 2

Minimal pair	Precursors
<i>Parking</i>	<i>This set of words lists</i>
<i>Barking</i>	<i>The smart young child writes</i>
	<i>He will need to stress</i>
	<i>I will try to guess</i>

Target VOT words were resynthesized and combined with all precursors as in Experiment 1, and presented in six randomized blocks of 52 experimental sentences using the procedure and equipment described in Experiment 1 (section §3.2). However, in

Experiment 2 participants were addressed and instructed only in English by a native speaker of American English (not the speaker who produced the stimuli).

4.3 Analysis, Results and Discussion

The location of the VOT boundary for each participant was estimated by the logistic regression modeling described in Experiment 1 (section §3.3). Then, individual VOT boundaries in native-accented English were compared with those identified for experienced listeners in native-accented Spanish and English-accented Spanish in Experiment 1.

As a result of the interaction of order, experienced listeners tested in the ES order in Experiment 1 relied on an English-like VOT boundary also in the native-accented Spanish context. Since Experiment 2 included participants who were tested in different orders in Experiment 1, the statistical comparison of VOT boundaries across accents was not based on a within-subject design (e.g., repeated measures ANOVA). To avoid order-based biases, individual VOT boundaries in native-accented English were thus compared via one-way ANOVA with the individual VOT boundaries reported for experienced listeners in Experiment 1 in each order (SE, ES) and accent (native Spanish, English-accented Spanish) (Table 6).

Table 6. VOT Boundaries in Experiments 1 and 2

	Experiment 1			Experiment 2	
	SE Order		ES Order		
	Native accented Spanish	English accented Spanish	English accented Spanish	Native accented Spanish	Native accented English
VOT boundary	-0.79 ms	7 ms	8.67 ms	8.65 ms	9.76 ms

The ANOVA revealed a main effect of boundary location, $F(4,38) = 5.19$, $p = 0.002$, $\eta^2 = 0.35$, which was examined by a Tukey HSD post-hoc test. Results of the post-hoc analysis ($p < 0.05$) revealed only one significant difference between the location of the VOT boundary in native-accented Spanish (Experiment 1, SE order) and the rest of the VOT boundaries included in the comparison. Thus, no significant difference was reported between the location of the VOT boundary in native English and English-accented Spanish (voicing-identification response curves for those boundaries are depicted in Figure 11). These results indicate that experienced listeners did not probably encode English-accented Spanish VOT with an intermediate VOT boundary between the English and the Spanish one.

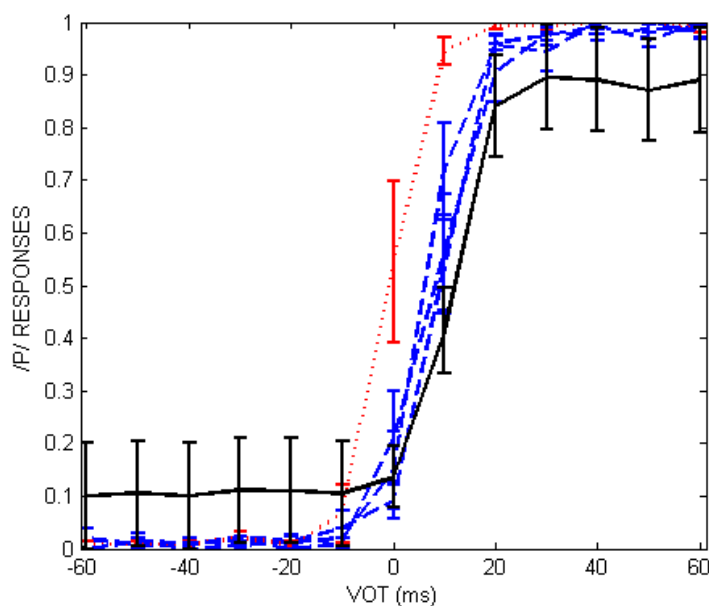


Figure 11. Voicing-identification Response Curves from Experiments 1 and 2. Proportions of /p/ responses with standard errors (y-axis) for the experienced group in native English (black solid line) and native Spanish –SE order (red dotted line). Curves for the rest of the contexts (English-accented Spanish – orderings 1 and 2, and native-accented Spanish –ES order) are plotted with blue dashed lines.

4.4 Summary and further research questions

In summary, the results of Experiments 1 and 2 indicated that Spanish listeners processed English-accented Spanish voicing differently depending on the degree of familiarization with English-accented speech. While inexperienced listeners relied on a Spanish-like VOT boundary across accents, experienced listeners relied on either a Spanish- or English-like VOT boundary predictable from the language mode of phonetic processing (i.e., Spanish or English). Specifically, they accommodated native-accented English voicing and English-accented Spanish voicing with the same English-like VOT boundary. This suggests that they were not processing English-accented Spanish voicing with an intermediate VOT boundary between the Spanish and English one. Given that English-accent experienced listeners were processing English-accented Spanish VOT with an English-like VOT boundary that was not significantly different from the one that they used in a native-accented English context, we wondered whether bilingual experience with the native-accented Spanish and native-accented English could facilitate a phonetic switching in the absence of experience with English-accented Spanish.

SQ3: Can Spanish-English bilinguals perform a phonetic switch based solely on their experience with native-accented English norms?

This research question was addressed in a third experiment (Chapter 5), in which a group of bilingual speakers of Spanish and English with little or no experience with English-accented Spanish was tested in their accommodation of /b/ and /p/ in the same speech contexts and accents that were used in Experiment

CHAPTER 5. EXPERIMENT 3 - THE EFFECTS OF BILINGUAL EXPERIENCE WITH NATIVE-ACCENTED ENGLISH ON THE PROCESSING OF ENGLISH- ACCENTED SPANISH VOICING

5.1 Motivation and Overall Design

The goal of Experiment 3 was to determine whether Spanish-English bilinguals were able to switch to an English-like VOT boundary in an English-accented Spanish context based on their experience with native-accented English and not with English-accented Spanish. A group of bilingual speakers of Spanish and English with little or no experience with English-accented Spanish was recruited.

5.2 Methods

Eight new bilingual speakers of Spanish and English (3 men, 5 women) were recruited and tested in Lafayette, IN ($M = 28.6$ years, $R = 21 - 35$ years, $SD = 2.1$ years). As the group of Spanish-English bilinguals that participated in Experiment 1 (Chapters 3 and 4), all of the participants included in Experiment 3 were older than 15 years when they first get immersed in an English environment. In particular, there were 3 participants from Colombia, 1 from Argentina, 2 from Spain, 1 from Chile and 1 from 1 from Mexico. In contrast to bilinguals in Experiment 1, bilinguals in Experiment 3 had no prior experience

with Spanish instruction at Purdue University or any other educational institution in the USA.

The relative amount of contact with English and with English-accented Spanish were measured as in Experiment 1, by two independent Likert scales from 1 to 5. Two independent two-sample t-tests with English contact and English-accented Spanish contact (EASC) as the dependent variables, and group as the independent one, revealed that bilinguals in experiment 3 ($M = 1.25$ EASC) had a significantly lower amount of contact with English-accented Spanish than bilinguals in experiment 1 ($M = 4.8$ EASC), $t(22) = 15.7, p < 0.001$. However, they did not differ in terms of the amount of contact with native-accented English.

In Experiment 3, participants were tested in the accommodation of /b/ and /p/ in the same speech contexts that were used in Experiment 1 (section §3.2.2). However, to avoid the ordering effect reported in Experiment 1 (sections §3.5 and §3.6), participants were tested first in the native-accented Spanish session, followed by another session of English-accented Spanish (SE order). During experiment, they performed the task through the same experimental apparatus and control that was used in Experiment 1 (section §3.2.4).

5.3 Analysis, Results and Discussion

Individual VOT boundaries were logistically modeled as described in Experiment 1 (section §3.3). However, in Experiment 3, individual VOT boundary locations and slopes were analyzed by means of two independent two-sample t-test with VOT location (or slope) as the dependent variable, and condition (native Spanish and English-accented Spanish) as the independent variable.

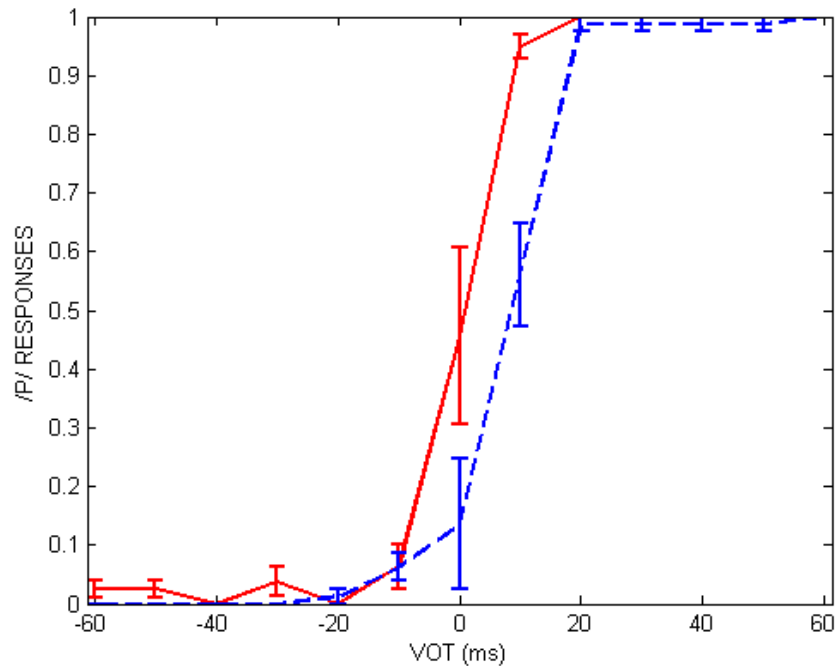


Figure 12. Voicing-identification Response Curves from Experiment 3. Averaged proportions of /p/ responses with standard errors (y-axis) for each target voice onset time (VOT) value (x-axis) in the experimental condition (English-accented Spanish context; blue dashed line) and the control condition (native-accented Spanish context; red solid line)

As in Experiment 1, statistical analyses revealed no significant differences across accents in terms of boundary slope. The VOT boundary location in the English-accented Spanish context (VOT boundary location = -0.63 ms average) was however significantly higher than the one in the native-accented Spanish context ($M = 8.07$ ms VOT), $t(14) = -6.06$, < 0.01 (Figure 12). This result supports the phonetic switching hypothesis for Spanish-English bilinguals with a low degree of experience with English accented Spanish.

CHAPTER 6. EXPERIMENT 4 - PHONETIC CALIBRATION OF STOP
CONSONANT VOICING BY NATIVE SPANISH LISTENERS WITH LITTLE
EXPERIENCE WITH ENGLISH ACCENTED SPEECH

6.1 Motivation and Overall Design

Given that experienced listeners in Experiment 1 seemed to accommodate English-accented Spanish voicing by means of phonetic switching rather than phonetic calibration, Experiment 4 investigated whether inexperienced listeners were able to use phonetic calibration to accommodate a boundary change such as is observed in English-accented Spanish. This research question could not be properly addressed in Experiment 1, in which speech contexts did not provide clear contextual cues to the phonetic realization of target voicing in either Spanish or English. Therefore, in Experiment 4, a new group of Spanish listeners with little or no experience with English-accented speech classified bilabial stop consonants varying in VOT. These sentences all had Spanish sentence precursors that included several exemplars of /b/ and /p/ that were artificially manipulated for a prototypical English or Spanish VOT realization, depending on the condition: the control condition (Spanish VOT) and the experimental condition (English V

6.2 Methods

6.2.1 Participants

A group of 16 native speakers of Spanish (10 women, 6 men) were recruited and tested in Madrid, Spain ($M = 31.7$ years, $R = 20$ to 50 years, $SD = 5.1$ years). The amount of experience with English for each participant was quantified by the variables used in Experiment 1: English background ($M = 0.5$ years), English contact ($M = 1.3$ Likert), and English-accented Spanish contact ($M = 1.8$ Likert). Three two-sample t-tests revealed no significant differences ($p < 0.05$) between this group and the group of English-accent inexperienced listeners tested in Experiment 1 across all of these variables.

6.2.2 Stimuli

Recordings were made following the protocol described in Experiment 1. However, since the goal was specifically to manipulate exposure to certain VOT values in the context, the speech context was distinct from that of Experiment 1. The speech context here was made up of a series of words including a specific number of exemplars of /b/ and /p/ that were then systematically manipulated to exhibit a Spanish- or English-like VOT, depending upon the condition. This type of speech context provided a higher degree of control over the phonetic properties of contextual materials. For example, this decision made it possible to include up to four words beginning with a bilabial stop in a given trial, and to counter-balance the VOT values of those words across trials (Table 8) in a manner that would not have been possible using actual sentences or phrases.

Table 8. Speech Items from Experiment 4

Targets		<i>vaso</i> ‘glass’, <i>paso</i> ‘step’
Precursors	b-precursors	<i>vado</i> ‘no parking area’, <i>barro</i> ‘clay’, <i>vale</i> ‘ticket’, <i>valle</i> ‘valley’, <i>valla</i> ‘fence’, <i>baza</i> ‘trick’
	p-precursors	<i>parra</i> ‘vine’, <i>palo</i> ‘stick’, <i>paro</i> ‘unemployment’, <i>paja</i> ‘straw’, <i>pana</i> ‘clothing’, <i>pala</i> ‘shovel’
Foils		<i>sala</i> ‘room’, <i>mesa</i> ‘table’, <i>silla</i> ‘chair’, <i>mano</i> ‘hand’, <i>loro</i> ‘parrot’, <i>lana</i> ‘wool’, <i>faro</i> ‘lighthouse’, <i>cera</i> ‘wax’

Speech materials included three types of Spanish CVCV trochaic nouns: 1 Spanish minimal pair, 12 precursors and 8 foils (Table 8). Both contextual and target VOT tokens consisted of stop consonants in word-initial position, followed by the vowel /a/ and sharing the same place of articulation (bilabial). The minimal pair consisted of two Spanish words (*vaso* ‘glass’ and *paso* ‘step’; in contemporary Spanish, letters *v* and *b* represent the same phoneme /b/; Tomás, 1990) differing in terms of the initial consonant (/b/ and /p/, respectively).

As in Experiment 1, the speaker was asked to pronounce exemplars of *vaso* with a very long prevoicing (strongly negative VOT) and exemplars of *paso* with a very long lag (strongly positive VOT) in order to provide an adequate base for resynthesis. Precursors included the exemplars of /b/ and /p/ that were resynthesized to have either a prototypical English or Spanish VOT value, depending on the condition. They consisted of six Spanish words starting with /b/ (b-precursors) and six Spanish words starting with /p/ (p-precursors). Because the intended manipulation would include shifting the VOT values of the b-precursors from a strongly Spanish /b/-like negative VOT to a more English-like slightly positive VOT, which is simultaneously similar to a Spanish /p/, it was important

to ensure that the b-precursors could not have a corresponding Spanish word that began with /p/. Foils consisted of eight Spanish words with no exemplars of stop consonants.

Prior to the beginning of the experiment, each participant's level of lexical familiarity with each word was assessed using a Likert scale from 1 (*I have never heard this word*) to 5 (*it is one of the most frequent words that I know*). Results indicated that participants were very familiar with all words ($M = 3.65$, $SD = 0.25$). Also, no significant difference between b- and p-words was found in terms of familiarity, $t(12) = 0.4$, $p = 0.6$.

6.2.3 Stimuli creation

Experimental stimuli were created by editing from original recordings. Before editing, tokens were normalized to the RMS of a 1000 Hz sine wave at 66 dB. Recorded tokens of *vaso* and *paso* were then examined to identify two tokens with VOT values clearly lower than -60 ms and greater than 60 ms, respectively. Once these tokens were identified, a series of target VOT tokens ranging from *vaso* (-60 ms VOT) to *paso* (60 ms VOT) was resynthesized in 13 VOT steps of 10 ms each by following the resynthesis method detailed in Experiment 1.

Precursor words were similarly edited using the same cutting and cross-splicing methods. In the control condition (Spanish-like contextual VOT), b- and p-precursors were modified to match typical Spanish VOT values of -90 ms and 10 ms, respectively (Dmitrieva et al., 2015). In the experimental condition (English-like contextual VOT), b- and p-precursors were edited to match typical English VOT values of 10 and 60 ms, respectively. To make the acoustic quality of prevoicing and aspiration in precursors more

like those in targets, precursors were cross-spliced with corresponding portions (of prevoicing or aspiration) from the target words (*vaso* and *paso*) as needed.

To assess the perceptual quality of the editing and cross-splicing method across targets and precursors, newly generated tokens with typical Spanish VOT (b-words of –60 and –90 ms VOT, and p-words of 10 ms VOT) were rated on naturalness by five Spanish speakers not included in the experiment. Both targets and precursors were rated on an ordinal scale from 1 *very unnatural* to 4 *very natural* as being well within the ‘natural’ range, with both groups of words receiving the same mean score of 3.8.

6.2.4 Procedure

Listeners were paid at a rate of €8 for their participation in 35 minutes of testing. In the experiment, participants were seated in a quiet room in front of an Acer Aspire 5830TG laptop showing an image of a glass (*vaso*) and a step (*paso*). No words were shown on the screen. Half of the participants (8 participants) were tested in the Spanish-like condition, in which all precursors were realized with Spanish VOT, and the other half were tested in the experimental condition, in which all precursors were realized with prototypical English VOT. This between-subject design was meant to avoid the effect of order observed in Experiment 1, in which participants’ performance in the second session was affected by their performance in the first session. Thus, in Experiment 4 participants were tested in either the English- or Spanish-like VOT condition. However, in each condition participants were selected from the same linguistically homogeneous population of Castilian speakers.

As in Experiment 1, each session was controlled by a custom-written Psychophysics Toolbox 3.0.8 interface implemented in MATLAB R2011b. Stimuli were

presented at 66 dB via Sennheiser HDpro 280 headphones connected to a Dr. NANO USB external sound card. In each trial, listeners heard a different series of six words, including one target word among different precursors and foils according to the scheme shown in Table 9. Participants were told that every trial would consist of six words, one of which would be one of the two words pictured on the screen. Their task was to decide which of the two words was said, regardless of where it appeared in the list and to indicate the picture corresponding to the target word by pressing the corresponding button on a USTC RTBox 5.x box (cf. Li et al., 2010).

In target trials, target words with VOT ranging from -60 to 60 ms were located at the third or fifth position of the series to allow it to be preceded by an equal numbers of randomly selected /b/- and /p/-precursors according to the scheme shown in Table 9. During the experiment, listeners heard a total of 52 randomized target trials per block (4 target trials \times 13 VOT target words). The purpose of the foil trials was to encourage listeners to attend to all the VOT tokens included in the context, including precursors in target trials, by making the location of target words more unpredictable. Target words in foil trials were thus located at the second and fourth position and combined with a random selection of experimental precursors and foils.

Table 9. Target and Foil Trials in Experiment 3

Type	Target Trials						Foil Trials					
1	b-	p-	t	f	f	f	p-	f	b-	t	f	p-
2	p-	b-	t	f	f	f	b-	p-	f	t	f	b-
3	b-	p-	p-	b-	t	f	b-	t	p-	b-	f	f
4	p-	b-	b-	p-	t	f	p-	t	b	p-	f	f

Note. Each trial consisted of a fixed sequence of targets (t), b-precursors (b-), p-precursors (p-) and foils (f), with items from each category randomly selected

To avoid the total duration of the experiment exceeding a limit of 40 minutes, target words in foil trials were constrained to an extreme VOT value of either -60 or 60 ms. Extreme VOT values were chosen to avoid other VOT acoustic biases than those that were provided in b- and p-precursors. During the experiment, participants heard a total of 16 randomized foil trials per block (4 foil trials \times 2 extreme VOT values \times 2 repetitions of each extreme VOT value).

In each condition, participants listened to a total of 260 target trials (5 blocks \times 52 target trials per block) and 80 foil trials (5 blocks \times 16 foil trials per block). Words within each trial were separated by an inter-stimulus interval (ISI) of 200 ms of silence. From the point of recruitment through the entire experiment, participants were addressed only in Spanish by a native speaker of Spanish. Background questionnaires and forms were also provided only in Spanish.

6.3 Analysis

Target-trial responses for each participant in each condition were logistically modeled as in Experiments 1. The location of the VOT boundary for each participant was estimated as the VOT value corresponding to the median effective level. The steepness of the VOT boundary for each participant was also estimated, like in Experiment 1, as the slope of the logistic curve at the median effective level. Individual boundary locations and slopes were submitted to two independent two-sample t-tests with condition as the independent variable, and location (or slope) as the dependent variable.

6.4 Results and Discussion

Results of the t-test for boundary location indicated that the VOT boundary in the experimental condition (English-like contextual VOT) was significantly higher ($M = 6.2$ ms VOT) than the one in the control condition (Spanish-like contextual VOT), $M = -4.3$ ms VOT, $t(14) = -5.15$, $p < 0.001$, $\eta^2 = 0.65$). In contrast, the t-test for boundary slope did not reveal any significant difference across conditions. These results suggest that listeners did not relax their VOT boundary to accommodate English-accented Spanish voicing. However, the VOT boundary shift observed across conditions suggests that the location of the VOT boundary was actually calibrated from the speech context (Figure 13).

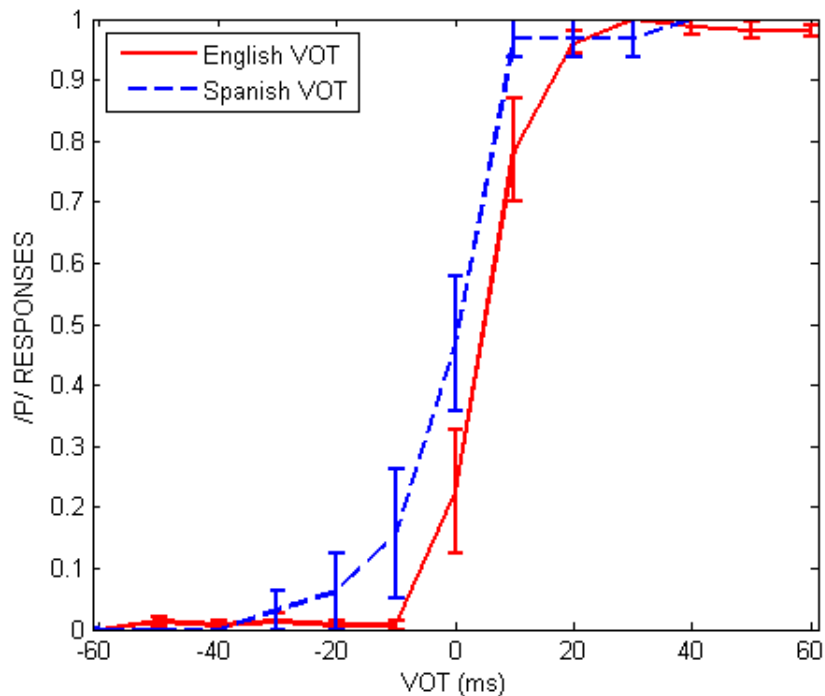


Figure 13. Voicing-identification Response Curves from Experiment 4. Averaged proportions of /p/ responses with standard errors (y-axis) for each target voice onset time (VOT) value (x-axis) in the experimental condition (English-like contextual VOT; red solid line) and the control condition (Spanish-like contextual VOT; blue dashed line)

The effects of block and target-trial type in the location of the VOT boundary were subsequently tested by a two-way ANOVA with VOT as the dependent variable and type and block as the independent variables. Results revealed no significant effects of block or target-trial type in the location of the VOT boundary, which indicates that accommodation was quite fast across all contexts (Figure 14 illustrates participants' performance across the five blocks).

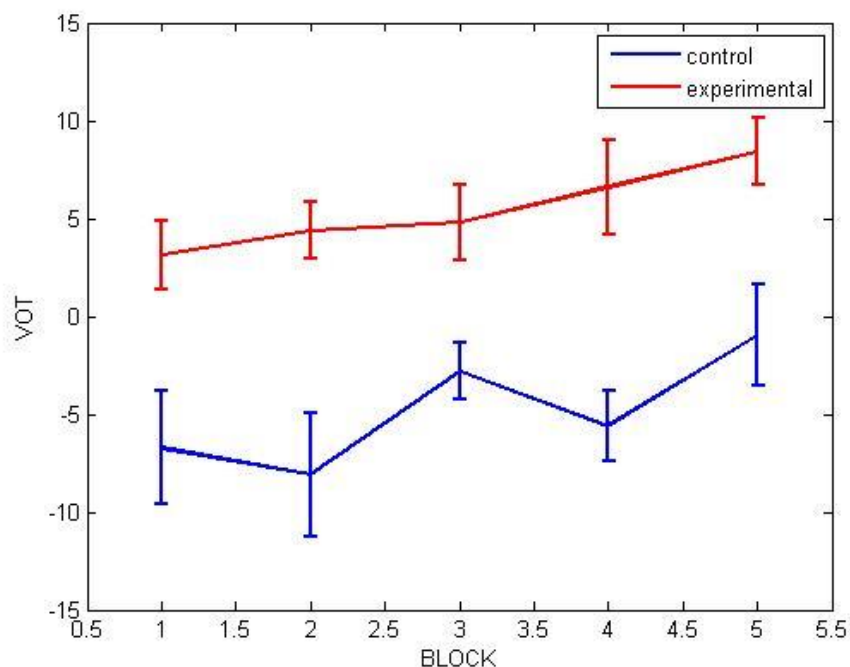


Figure 14. Performance by Block in Experiment 4. VOT boundary location means and standard errors (y-axis) across the five blocks (x-axis) for the control group (bottom line, Spanish-like contextual VOT) and the experimental group (top line, English-like contextual VOT)

In summary, results of Experiment 4 show that inexperienced listeners were able to calibrate the location of the VOT boundary solely based on contextual VOT. Interestingly, the magnitude of the VOT boundary shift reported for inexperienced listeners in

Experiment 4 (of approximately 10 ms) was similar to the one reported in previous experiments for experienced listeners performing in a native English versus a native Spanish context (ordering 1). This magnitude was also greater than the amount of VOT variability that would be expected among Castilian speakers with no significant experience with English based on prior results. In Experiment 1, the standard deviation of the VOT boundary among inexperienced listeners performing in native Spanish was 5.6 ms VOT, barely half of the boundary shift reported in Experiment 4. This indicates that the amount of phonetic calibration observed across the two groups of inexperienced listeners tested in Experiment 4 is better predicted by the manipulation of the context than by the distribution of participants across conditions.

CHAPTER 7. GENERAL DISCUSSION

7.1 Summary

The last four chapters presented four experiments investigating the accommodation of /b/ and /p/ in different English and Spanish accented contexts by native Spanish listeners with a low and high degree of familiarity with English-accented Spanish and/or native-accented English. In Experiment 1 (Chapter 3), native Spanish listeners accommodated target voicing in a native- and English-accented Spanish context with no clear contextual cues to the phonetic realization of /b/ and /p/ in either the Spanish or English. English-accent inexperienced listeners relied on the same Spanish-like VOT boundary across accents. In contrast, English accent-experienced listeners processed English-accented Spanish by switching to an English-like VOT boundary that was not predictable from the phonetic realization of /b/ and /p/ in the speech context. None of the groups tested relaxed the slope of the VOT boundary in the English-accented Spanish context.

In Experiment 2, English-accent experience listeners accommodated /b/ and /p/ in an native-accented English context, in which they switched to an English-like VOT boundary that was not significantly different from the English-like VOT boundary that they used in the English-accented Spanish context in Experiment 1.

In Experiment 3, a new group of listeners with experience with native-accented Spanish and native-accented English, but not with English-accented Spanish,

accommodated /b/ and /p/ in the speech contexts that were used in Experiment 1 (native Spanish and English-accented Spanish). Despite the lack of experience with the English-accented Spanish norm, in the English-accented Spanish context, this group of listeners was shown to switch to an English-like VOT boundary that was significantly different from the VOT boundary that they used in the native Spanish context, in a similar way to which the group of experienced listeners switched their phonetic boundaries in Experiment 1.

In Experiment 4, a group of English-accent inexperienced listeners was tested in the accommodation of target voicing in a Spanish context with clear contextual cues to the phonetic calibration of target voicing in either Spanish or English. Results of Experiment 4 indicated that English-accent inexperienced listeners were able to calibrate the location of the VOT boundary based on the phonetic realization of target voicing in the previous speech context.

7.2 Findings and implications

This section discusses the findings and implications of all of the experiments summarized in the previous section in the light of the research questions raised in the previous chapters.

7.2.1 Overarching question

The overarching question (OQ) addressed in the present study was whether listeners accommodate foreign-accented sounds differently, depending on their degree of familiarization with the phonetic norm of the foreign accent.

OQ: Does accent-specific experience changes the phonetic accommodation of sounds from that accent?

Results of experiments 1 and 4 support the hypothesis that English-accent experienced listeners and English-accent inexperienced listeners rely on different mechanisms of accommodation of foreign-accented sounds. However, the specific nature of these mechanisms, as well as the amount of experience responsible for these differences, were further examined in a series of three general research questions (GQs):

GQ1: Do accent-experienced listeners relax the slope of perceptual boundaries in listening to foreign-accented speech?

GQ2: Do accent-experienced listeners calibrate the location of perceptual boundaries in listening to foreign-accented speech?

GQ3: Do accent-experienced listeners switch to a non-native set of phonetic boundaries in listening to foreign-accented speech?

7.2.2 General Research Questions

Overall, the results of Experiment 1 suggest that English-accent experienced listeners did not accommodate English-accented Spanish voicing by relaxing the slope of the phonetic boundary. This provides support against the phonetic relaxation hypothesis (GQ1). In addition, they seemed to process English-accented Spanish voicing by means of a perceptual boundary that was not likely calibrated from the previous speech context, because the speech context used in Experiment 1 did not provide significant cues to the

pronunciation of target voicing in English. This result thus lends support to the phonetic switching hypothesis for the group of English-accent experienced listeners (GQs 2-3).

7.2.3 Specific research questions

The general research questions discussed in the previous section were further investigated across a series of three specific research questions, which were motivated by the results of Experiments 1 and 2.

SQ1: Do English-accent experienced listeners process English-accented Spanish voicing differences with an intermediate VOT boundary between the Spanish and English one?

SQ2: Can English-accent inexperienced listeners calibrate the location of the Spanish VOT boundary in a speech context with contextual cues biased to the phonetic realization of target voicing in English?

SQ3: Can Spanish-English bilinguals perform a phonetic switch based solely on their experience with native-accented English norms?

Results of Experiment 2 (Chapter 4) indicate that English accent-experienced listeners processed English-accented Spanish voicing by switching to the same English-like VOT boundary that they used in a native-accented English speech context. This result contradicts the hypothesis of an intermediate mode of phonetic processing postulated in question SQ1. In Experiment 3 (Chapter 5), Spanish-English bilinguals with little or no experience with English-accented Spanish were able to perform a phonetic switching in an

English-accented Spanish context, thus supporting the proposal that Spanish-English bilinguals are processing English-accented Spanish sounds and native-accented English sounds in the same mode of phonetic processing (question SQ3). With respect to the group of English-accent inexperienced listeners, results of Experiment 4 (Chapter 6) suggest that they may be able to recalibrate phonetic boundaries in a speech context with enough opportunities for phonetic calibration (question SQ2). This result supports the proposal that accent-inexperience listeners may accommodate to foreign-accented speech by means of the mechanism of phonetic recalibration.

7.2.4 Conclusions

Overall, the experimental findings discussed in the previous sections (§7.2.2 - §7.2.3) suggest that native listeners may not accommodate foreign-accented speech sounds by relaxing the phonetic boundary between native speech categories. They rather seem to accommodate foreign-accented speech sounds differently, depending on the amount and type of experience with the phonetic norm of the foreign-accent. For example, when Spanish listeners had little or no experience with English-accented speech norms, they encoded English-accented Spanish voicing with an English-like VOT boundary only when contextual VOT was clearly biased toward the English norm. However, when they were very familiar with the English-accented speech norm, they processed English-accented Spanish voicing with an English-like VOT boundary even in the absence of clear contextual cues to the prototypical pronunciation of /b/ and /p/ in English. Therefore, while inexperienced listeners seem to accommodate foreign-accented speech categories by adjusting the short-term representation of those categories to their pronunciation in the

previous speech context, as predicted by bottom-up models of phonetic processing (Dahan et al., 2008; Nygaard et al., 1994), experienced listeners seem to accommodate to the same speech categories by switching the language-specific mode of phonetic processing, as predicted by top-down models of phonetic processing (García-Sierra et al. 2012; Gonzales and Lotto, 2013).

This interpretation of the experimental findings is compatible with a model of perceptual adaptation to foreign-accented speech in which foreign-accented sounds are accommodated in three consecutive steps modulated by the degree of linguistic experience with the accent speech norm. First, when listeners are very unfamiliar with the phonetic norm of the foreign accent, they may infer the intended pronunciation of foreign-accented sounds by means of the lexical, semantic, or pragmatic information provided by the context. For example, without any experience with English-accented speech, native Spanish listeners listening to an English-accented pronunciation of the Spanish word *bata* as [paʔa] could infer the intended pronunciation [baʔa] based on the semantic or pragmatic expectations provided by the conversational context. This might correspond to a phase in which listeners exhibit a higher degree of processing cost, as a result of the conflict of expectations provided by phonetic, lexical, semantic, and pragmatic features.

Then, listeners with a little bit more experience with a foreign-accented norm may be able to re-calibrate the decoding of foreign accented sounds based on the pronunciation of these sounds by the talker that they are listening to. For example, a particular English speaker may systematically pronounce Spanish /b/ as English-accented Spanish [p]. A native Spanish speaker listening to this talker may recalibrate the decoding of [p] as Spanish /b/ based on the talker's repeated productions of Spanish /b/ as [p] in utterances

preceding a particular ambiguous token, in a similar way to which the same native speaker of Spanish would accommodate the speech idiolect of an unfamiliar but native-accented talker of Spanish.

Finally, when listeners are very familiar with the speech norm of the foreign accent that they are listening to, they may rather switch to the sound system that better accounts for the type of speech variability that they are listening to, in a similar way in which bilinguals accommodate speech variability from different languages.

7.3 Limitations and future directions

There are a few aspects of the present study that remain unclear and would be worthwhile to explore in future work. First, while it seems clear that the amount of language experience facilitates the ability to switch across different phonetic norms, the role played the type of experience is not yet clear. Given that the group of English accent-experienced listeners that participated in Experiments 1 - 3 were bilingual speakers of Spanish and English, and therefore had considerable experience with both English and English-accented Spanish, it is difficult to determine whether the phonetic switching reported in Experiment 1 (Chapter 3) was more strongly facilitated by experience with English-accented Spanish or with native-accented English. On the one hand, the mechanism of phonetic switching might be facilitated by long-term experience with the English-accented norm in any of its variants, independently of whether this experience comes from contact with native-accented English, English-accented Spanish, or both. Alternatively, it might be necessary that listeners first have well-formed lexical representations in their second language, ones previously develop through bilingual

experience, in order to maintain a non-native system of accented speech representations in long-term memory. To this perspective, long-term experience with a foreign accent might not be enough to induce the mechanism of phonetic switching.

Similarly, there are factors related to language dominance that may also play an important role in the accommodation of foreign-accented speech sounds by bilinguals, but that are not deeply explored in the present study. For example, while the Spanish-English bilinguals tested in Experiments 1 to 3 were Spanish dominant, the ability to switch sound systems across different language and accented contexts (e.g. Spanish and English) in early bilinguals and long-term immigrants might be also modulated by their degree of attrition in L1 (e.g. Spanish).

Another aspect that might be interesting to explore in future research is the role played by the type of speech context that is being used in the perceptual task. While in Experiment 1 the speech context consisted of natural sentences, in Experiment 4 it consisted of lists of words because of the need for a higher degree of phonetic control in developing the contextual materials. Thus, although results of Experiment 4 suggest that Spanish listeners should be able to adjust the processing of target voicing based on contextual VOT, the amount of adjustment might vary in a more realistic speech context depending on different variables, such as the density of relevant phonetic cues in the previous speech context (e.g. the number of bilabial stop-consonants, as in Experiment 4) or the distribution of these cues across different phonetic or lexical contexts. While in some models of perceptual accommodation, the mechanism of phonetic calibration is established in terms of auditory biases that are independent of the lexical or phonetic environment in which contextual cues are provided, other models suggest that listeners' perceptual

calibration could be also sensitive to the information provided by the lexical or phonetic context (cf. Norris et al, 2003; Reinisch & Holt, 2014; McMurray & Jongman, 2011). These are aspects that would be worthwhile to explore in future research.

In the same vein, it would be interesting to investigate the role played by secondary cues in the accommodation of foreign-accented speech contrasts. Speech sounds are processed across multiple acoustic dimensions that tend to interact in perception. The neutralization of one of these dimensions by synthesis (or resynthesis) might reduce the number of cues involved in the contrast and thus weaken the perceptual difference across speech categories. The resynthesis method used in Experiments 1 to 4 (which only manipulated the VOT dimension) could actually explain the unexpectedly short VOT boundary shift, of approximately 10 ms, reported for Spanish listeners across the Spanish and English phonetic norms. However, these are also aspects that were not explored in the present study and that require of further investigation.

REFERENCES

REFERENCES

- Abramson, A. S., & Lisker, L. (1964). A cross-language study of voicing in initial stops: Acoustical measurements.
- Abramson, A. S., & Lisker, L. (1972). Voice-timing perception in Spanish word-initial stops. *Haskins Laboratories Status Report on Speech Research*, 29(30), 15-25.
- Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: two activation likelihood estimation (ALE) meta-analyses. *Brain and language*, 122(1), 42-54.
- Agresti, A. (1996). *An introduction to categorical data analysis* (Vol. 135). New York: Wiley.
- Alexander, J. M., & Kluender, K. R. (2010). Temporal properties of perceptual calibration to local and broad spectral characteristics of a listening context. *The Journal of the Acoustical Society of America*, 128(6), 3597-3613.
- Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of phonetics*, 40(4), 582-594.

- Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. *Handbook of child psychology: formerly Carmichael's Manual of child psychology/Paul H. Mussen, editor.*
- Auer, P. (Ed.). (2013). *Code-switching in conversation: Language, interaction and identity.* Routledge.
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America, 137*(5), 2823-2833.
- Barlow, J. A. (2003). The stop-spirant alternation in Spanish: Converging evidence for a fortition account. *Southwest Journal of Linguistics, 22*(1), 51
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America, 114*(3), 1600-1610.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege, 13-34.*
- Bilmes, J. A. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute, 4*(510), 126.
- Blom, J. P., & Gumperz, J. J. (2000). Social meaning in linguistic structure: Code-switching in Norway. *The bilingualism reader, 111-136.*
- Blumstein, S. E. (1998). The mapping from acoustic structure to the phonetic categories of speech: The invariance problem. *Behavioral and Brain Sciences, 21*(02), 260-260.
- Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic inquiry, 32*(1), 45-86.

- Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in second language acquisition, 19*(04), 447-465.
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707-729.
- Brookes, M. (2009). Voicebox. *Speech Processing Toolbox for Matlab, Department of Electrical & Electronic Engineering, Imperial College.*
- Brouwer, S., & Bradlow, A. R. (2014). Contextual variability during speech-in-speech recognition. *The Journal of the Acoustical Society of America, 136*(1), EL26-EL32.
- Bullock, B. E., & Toribio, A. J. (2009). *Themes in the study of code-switching*. NA.
- Bullock, B. E., Toribio, A. J., González, V., & Dalola, A. (2006). Language dominance and performance outcomes in bilingual pronunciation. In *Proceedings of the 8th generative approaches to second language acquisition: The Banff conference* (pp. 9-16).
- Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics, 41*(6), 520-533.
- Chiba, R., Matsuura, H., & Yamamoto, A. (1995). Japanese attitudes toward English accents. *World Englishes, 14*(1), 77-86.
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647-3658.
- Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in psychology, 3*.

- Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations?. *Cognition*, 108(3), 710-718.
- Delattre, P. (1964). Comparing the vocalic features of English, German, Spanish and French. *IRAL-International Review of Applied Linguistics in Language Teaching*, 2(1), 71-98.
- Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27(4), 769-773.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in second language acquisition*, 19(01), 1-16.
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, 1(2), 121-144.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annu. Rev. Psychol.*, 55, 149-179.
- Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, 49, 77-95.
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 914.
- Eddington, D., & Elzinga, D. (2008). The phonetic context of American English flapping: Quantitative evidence. *Language and speech*, 51(3), 245-266.

- Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America*, 62(4), 971-974.
- Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America*, 62(4), 971-974.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.
- Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Netherlands Graduate School of Linguistics.
- Escudero, P., & Boersma, P. (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. *University of Pennsylvania Working Papers in Linguistics*, 8(1), 7.
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of phonetics*, 15(1), 47-65.
- Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, 84(1), 70-79.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press
- Flege, J. E. (2009). Give input a chance. *Input matters in SLA*, 175-190.

- Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, 6(3), 185-202.
- Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, 83(2), 729-740.
- Flege, J. E., & Liu, S. (2001). THE EFFECT OF EXPERIENCE ON ADULTS' ACQUISITION OF A SECOND LANGUAGE. *Studies in second language acquisition*, 23(04), 527-552.
- Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, 24(2), 125-146.
- Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, 34(2), 153-175.
- Flege, J. E., & Hammond, R. M. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition*, 5(01), 1-17.
- Flege, J. E., MacKay, I. R., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, 23(4), 567-598.
- Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16(1), 1-26.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of memory and language*, 41(1), 78-104.

- Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt?. *Journal of Psycholinguistic Research*, 38(4), 379-412.
- Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing?. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1276.
- Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing?. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1276.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3-28.
- Fowler, C. A., & Magnuson, J. S. (2012). Speech perception. *The Cambridge Handbook of Psycholinguistics*, 3-26.
- Gantz, B. J., & Turner, C. W. (2003). Combining acoustic and electrical hearing. *The Laryngoscope*, 113(10), 1726-1730.
- Garcia-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech communication*, 51(4), 369-378.
- García-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J., & Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and language*, 121(3), 194-205.
- Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language learning*, 34(1), 65-87.

- Gertken, L. M., Amengual, M., Birdsong, D., & Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. *Measuring L2 proficiency: Perspectives from SLA*, 208-225.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, 105(2), 251.
- Gonzales, K., & Lotto, A. J. (2013). A Bafri, un Pafri Bilinguals' Pseudoword Identifications Support Language-Specific Phonetic Systems. *Psychological science*, 24(11), 2135-2142.
- Grosjean, F. (2001). The bilingual's language modes. In Nicol, J. (Ed.). *One Mind, Two Languages: Bilingual Language Processing* (pp. 1-22). Oxford: Blackwell. Also in Li Wei (Ed.). *The Bilingual Reader* (2nd edition). London: Routledge, 2007.
- Hammond, R. M. (2001). *The Sounds of Spanish: Analysis and Application (with Special Reference to American English)*. Cascadilla Press, PO Box 440355, Somerville, MA 02144.
- Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of phonetics*, 36(4), 664-679.
- Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1), 283.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, 97(5), 3099-3111.

- Hintzman, D., L. (1986) "Schema Abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411–428.
- Holt, L. L. (2005). Temporally non-adjacent non-linguistic sounds affect speech categorization. *Psychological Science*, 16, 305-312.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *The Journal of the Acoustical Society of America*, 108(2), 710-722.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?. *Department of Psychology*, 167.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, 21(1), 60-99.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of phonetics*, 34(4), 485-499.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive psychology*, 29(1), 1-23.
- Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*.
- Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory phonology*, 2(1), 125-156.

- Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of phonetics*, 36(1), 28-54.
- Klatt, D. H. (1987). Review of text-to-speech conversion for English. *The Journal of the Acoustical Society of America*, 82(3), 737-793.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237(4819), 1195-1197.
- Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*.
- Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, 124(6), 3959-3971.
- Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of phonetics*, 38(4), 569-587.
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal?. *Cognitive psychology*, 51(2), 141-178.
- 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.

- Kuhl, P. K., & Iverson, P. (1995). Chapter 4: Linguistic Experience and the “Perceptual Magnet Effect,”. *Speech perception and linguistic experience: Issues in cross-language research*, 121-154.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1), 98-104.
- Lenneberg, E.H. (1967). *Biological foundations of language*. New York: Wiley.
- Lev-Ari, S. (2014). Comprehending non-native speakers: theory and evidence for adjustment in manner of processing. *Frontiers in psychology*, 5.
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093-1096.
- Li, N., & Loizou, P. C. (2008). A glimpsing account for the benefit of simulated combined acoustic and electric hearing. *The Journal of the Acoustical Society of America*, 123(4), 2287-2294.
- Li, N., & Loizou, P. C. (2008). A glimpsing account for the benefit of simulated combined acoustic and electric hearing. *The Journal of the Acoustical Society of America*, 123(4), 2287-2294.
- Li, X., Liang, Z., Kleiner, M., & Lu, Z. L. (2010). RTbox: A device for highly accurate response time measurements. *Behavior Research Methods*, 42(1), 212-225.
- Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.
- Lipski, J. M. (1982). Spanish-English language switching in speech and literature: theories and models. *Bilingual Review/La Revista Bilingüe*, 191-212.

- Lisker, L. (1986). "Voicing" in English: a catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and speech*, 29(1), 3-11.
- Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America*, 134(3), 2213-2224.
- Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, 68(2), 178-183.
- Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60(4), 602-619.
- Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *The Journal of the Acoustical Society of America*, 102(2), 1134-1140.
- Mackay, I. R., Flege, J. E., & Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*, 27(02), 157-183.
- Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 391.
- Martínez Celdrán, E. M. C., & Planas, A. M. F. (2007). *Manual de fonética española: articulaciones y sonidos de español*. Editorial Ariel.

- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953-978.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7-8), 953-978.
- Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science, 32*(3), 543-562.
- Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental science, 11*(1), 122-134.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-B111.
- Mazzaro, N., Cuza, A., & Colantoni, L. (2016, February). Age effects and the discrimination of consonantal and vocalic contrasts in heritage and native Spanish. In *Romance Linguistics 2013: Selected papers from the 43rd Linguistic Symposium on Romance Languages (LSRL), New York, 17-19 April, 2013* (Vol. 9, p. 277). John Benjamins Publishing Company.
- McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech, Language, and Hearing Research, 26*(3), 451-458.
- McGowan, K. B. (2011). *The role of socioindexical expectation in speech perception* (Doctoral dissertation, University of Minnesota).

- Mckenzie, R. M. (2008, June). The complex and rapidly changing sociolinguistic position of the English language in Japan: A summary of English language contact and use. In *Japan Forum* (Vol. 20, No. 2, pp. 267-286). Taylor & Francis Group.
- McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2), 219.
- McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental science*, 12(3), 369-378.
- McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic bulletin & review*, 15(6), 1064-1071.
- Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *The Journal of the Acoustical Society of America*, 73(5), 1751-1755.
- Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 369.
- Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PloS one*, 4(11), e7785.
- Morrison, G. S. (2009). L1-Spanish Speakers' Acquisition of the English /i/—//Contrast II: Perception of Vowel Inherent Spectral Change¹. *Language and Speech*, 52(4), 437-462.

- Mullennix J, Pisoni DB, Martin C. Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*. 1989;85:365–378.
- Munro, M. J. (2008). Foreign accent and speech intelligibility. *Phonology and second language acquisition*, 193-218.
- Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech*, 38(3), 289-306.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, 47(2), 204-238.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, 115(1), 39.
- Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. *Talker variability in speech processing*, 109-132.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42-46.
- Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41(6), 407-420.
- Ortega-LLebaria, M. (2003). Effects of Phonetic and Inventory Constraints in the Spirantization of Intervocalic Voiced Stops: Comparing two Different Measurements of Energy Change. *Stress*, 7(6), 5.
- Pérez-Leroux, A. T., Cuza, A., & Thomas, D. (2011). From parental attitudes to input conditions: Spanish-English bilingual development in Toronto. *Bilingual youth: Spanish in English-speaking societies*, 149-176.

- Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. *Science*, 333(6042), 595-595.
- Peterson, G. E. and Barney, H. I. (1952). "Control methods used in the study of vowels", *Journal of the Acoustical Society of America*, 24, 75-184.
- Pfaff, C. W. (1979). Constraints on language mixing: intrasentential code-switching and borrowing in Spanish/English. *Language*, 291-318.
- Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, 46(2-3), 115-154.
- Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics*, 29(2), 191-215.
- Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic biology*, 53(5), 793-808.
- Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. *The handbook of speech perception*, 182-206.
- Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 75.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212(4497), 947-949.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(4), 511-531.

- Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, *14*(3), 337-353.
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207-1218.
- Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of phonetics*, *2*(3), 181.
- Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of phonetics*, *2*(3), 181.
- Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Developmental Science*, *12*(4), 583-601.
- Schmid, M. S. (2013). First language attrition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 117-123.
- Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *27*(4), 395-408.
- Sidas, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306-3316.
- Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America*, *130*(6), 4020-4031.

- Stephens, J. D., & Holt, L. L. (2003). Preceding phonetic context affects perception of nonspeech (L). *The Journal of the Acoustical Society of America*, 114(6), 3036-3039.
- Stephens, J. D., & Holt, L. L. (2003). Preceding phonetic context affects perception of nonspeech (L). *The Journal of the Acoustical Society of America*, 114(6), 3036-3039.
- Story, B. H. (2002). An overview of the physiology, physics and modeling of the sound source for vowels. *Acoustical Science and Technology*, 23(4), 195-206.
- Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7(5), 1074.
- Toribio, A. J. (2004). Convergence as an optimization strategy in bilingual speech: Evidence from code-switching. *Bilingualism: Language and Cognition*, 7(02), 165-173.
- Toribio, A. J., Bullock, B. E., Botero, C. G., & Davis, K. A. (2005). Perseverative phonetic effects in bilingual code-switching. *Amsterdam studies in the theory and history of linguistic science series. 4*, 272, 291.
- Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive science*, 34(3), 434-464.
- Toscano, J. C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics*, 74(6), 1284-1301.

- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(04), 905-916.
- Trude, A. M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations on adaptation to foreign accents. *Journal of memory and language*, 69(3), 349-367.
- van Wijngaarden, S. J., Steeneken, H. J., & Houtgast, T. (2001). The effect of a non-native accent in Dutch on speech intelligibility. *The Journal of the Acoustical Society of America*, 109(5), 2473-2473.
- Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of phonetics*, 46, 34-51.
- Weil, S. A. (2001). *Foreign accented speech: Adaptation and generalization* Doctoral dissertation, The Ohio State University.
- White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press.
- Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., & Merfeld, D. M. (2012). *Sensation & perception*. Sinauer.
- Xie, Z., Yi, H. G., & Chandrasekaran, B. (2014). Nonnative Audiovisual Speech Perception in Noise: Dissociable Effects of the Speaker and Listener. *PloS one*, 9(12), e114439.

VITA

VITA

Fernando Llanos is a Ph.D. Candidate at Purdue University with specialization in Hispanic linguistics and speech sciences. Prior to joining the graduate program at Purdue, he obtained his B.A. in Spanish language and literature from the Universidad Complutense de Madrid, and his M.A. in speech sciences from the Spanish National Research Council (CSIC), where he worked under the supervision of Professors Juana Gil Fernandez and Rafael Martinez Olalla.

Throughout his research program, he seeks to understand how speech sounds are efficiently heard, learned and produced. His scientific approach combines linguistic analysis with behavioral research and computational modeling. In his dissertation, under the supervision of Professors Alexander L. Francis and Robert M. Hammond, he investigated the effects of experience in processing foreign accented speech. He is also interested in the acquisition of novel speech categories, and the role played by speech redundancy in the encoding and decoding of speech signal, topics that he is currently investigating with Professor Keith R. Kluender