

PERCEPTION AND PRODUCTION OF INTONATION AMONG ENGLISH-SPANISH BILINGUAL SPEAKERS
AT DIFFERENT PROFICIENCY LEVELS

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Spanish Linguistics

By

Germán A. Zárate-Sández, M.S.

Washington, DC
May 18, 2015

UMI Number: 3708036

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3708036

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Copyright 2015 by Germán A. Zárate-Sández
All Rights Reserved

PERCEPTION AND PRODUCTION OF INTONATION AMONG ENGLISH-SPANISH BILINGUAL SPEAKERS
AT DIFFERENT PROFICIENCY LEVELS

Germán A. Zárate-Sández, M.S.

Thesis Advisors: Cristina Sanz, Ph.D. & Alfonso Morales-Front, Ph.D.

ABSTRACT

This dissertation examined the perception and production of intonation among 55 English-native speakers of Spanish at three proficiency levels (low, high, and very high). Their performance was compared with monolingual speakers of Spanish ($n=17$) and English ($n=17$), and English-Spanish early bilinguals (heritage speakers, $n=16$). The target form was the intonational contour in neutral declarative utterances in Spanish, examined at two tonal events, namely prenuclear peak alignment and final boundary tone height. The study adhered to the theoretical principles of the Autosegmental-Metrical approach (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980, 2000). Participants completed an imitation task aimed at locating potential categorical shifts in the perception of both tonal events and two production tasks varying in speaking style (sentence reading and storytelling).

Results revealed a marked contrast between Spanish and English in perception and production of both tonal events. Spanish speakers generally preferred later alignment of prenuclear peaks and lower height of final boundary tone. In turn, second language (L2) and early bilingual speakers tended to produce values in the middle range between Spanish and English. Performance of low-proficiency speakers generally approximated English monolingual speakers, while L2 speakers of very high proficiency produced values at the same level of heritage speakers under most measures. As regards the role of speaking style in production, some minor effects were found in prenuclear alignment, while no effects were obtained in final boundary tone. A strong relationship between production and perception was also found for

prenuclear alignment but not for final boundary tone. Results also seemed to support some predictions made by the Speech Learning Model (Flege, 1995). Findings are discussed from the point of view of cross-linguistic influence, effects of high proficiency on L2 phonology, tonal representations for Spanish and English, and the link between production and perception in L2 prosody.

ACKNOWLEDGEMENTS

This dissertation is the culmination of perhaps the longest and hardest enterprise I have ever undertaken. As such, I could not have done it without many people who helped me during this journey. First and foremost, from the bottom of my heart, I want to thank Cristina Sanz and Alfonso Morales-Front, my wonderful co-advisors. Since my first year in Georgetown's Ph.D. program, I have spent countless hours as their student, research assistant, teaching assistant, and mentee. They are among the smartest, most knowledgeable, caring, and generous scholars and teachers I have ever known. Thank you, Cristina and Alfonso, for your incredible mentorship.

I also want to thank Joan Carles Mora for his help and advice during different phases of this dissertation. His dedication, comments, and guidance as a member of my dissertation committee greatly improved this manuscript and made me a better researcher. José Ignacio Hualde was involved at the initial stages of this dissertation and his suggestions greatly improved the research design.

My years of graduate coursework were very enjoyable thanks to many fine professors. In alphabetical order, I would like to thank Héctor Campos, Rusan Chen, Robert DeKeyser, Rod Ellis, Elena Herburger, Darleen Howard, Ron Leow, Alison Mackey, Rob Podesva, Tom Walsh, Matthew Wolf, and Elizabeth Zsiga. I want to extend a special thank you to Héctor Campos for spending so much time talking in his office about Argentina, Chile, and Linguistics; to Ron Leow for infusing in me his passion for SLA research and language program direction; and to Alison Mackey for giving me the chance to work closely with her on exciting research projects.

I will always have fond memories of the truly wonderful people I have met in the Department of Spanish and Portuguese at Georgetown University, most notably Francisco Cornejo, María Elvira Daza, Michael Ferreira, Kristen Hall, Gwen Kirkpatrick, Vivaldo Santos,

and María Wear. Thank you for your help and friendship. The camaraderie of graduate students from the departments of Spanish and Portuguese and Linguistics helped to ease the daily struggles associated with graduate life. Thanks to Missy Baralt, Annie Calderon, Marisa Filgueras, Julie Lake, Silvia Marijuan, Anthony Perry, Cecily Raynor, Jason Robinson, Carolina Rodríguez, and Nicole Ziegler. I want to especially thank my dear friends Ana María Ferreira, Ellen Johnson Serafini, Charlie Nagle, and Julio Torres for blessing me with their friendship and keeping me sane all these years. *Los quiero muchísimo*. I have also been fortunate to count on Tom Haran's friendship and hospitality all these years. Special thanks too to Veronica Salles-Reese for inviting me to discover other parts of the world. You are a wonderful and generous person.

Many people outside DC supported me as well. From Scranton, PA, my first residence in the United States, I am immensely grateful to friends at The University of Scranton: Bonnie Alco, Marie Karam, Joyce Knott, Linda Ledford-Miller, Robert Parsons, Virginia Picchietti, Yamile Silva, and Habib Zanzana.

In my native San Juan, Argentina, I want to express my sincerest appreciation to my friends and colleagues at the Universidad Nacional de San Juan: Myriam Arrabal, Marisel Bollati, Carmen Caturla, Rosita Cúneo, Ana María Laciár, Nancy Leánez, Andrea Leceta, Elvira Gargiulo, Marcela Morchio, María Rosa Marcovechio, Alejandra Muñoz, and Rosa María Sanou. They nurtured my passion for languages and linguistics and I am forever in their debt. Thank you, Graciela Albiñana, Claudia Castañeda, Cristina Castro, and Silvia Castro, at the Asociación de Cultura Saint John's, for fostering my love of foreign languages at an early age. Also, I wish to thank my best friends from San Juan, Cecilia Fernández, Cristián Gomez

Maldonado, Teté Manzur, Juan Saffe, and, above all, my beloved Celina Rodríguez. I am a much better person because of your friendship.

In terms of logistics, I owe a special note of thanks to the almost 140 people that participated in this study. Thank you to all undergraduate and graduate students, instructors, and language program coordinators at Georgetown University and The University of Scranton who helped recruit participants or actually participated in the study. Harriet Bowden and Lourdes Ortega shared the Elicited Imitation Task and suggestions on how to score it.

This dissertation and the path that brought me to it would not have been possible without the support and love of my family. My parents, Mario and Violeta, my brothers, Rubén and Mario, my sisters-in-law, Eugenia and Luli, my beautiful nieces and nephew, Valentina, Emilia, and Joaquín, have been tremendous sources of optimism when distance from Argentina or fatigue began to take their toll. My son, Tyler Julian, was a daily companion during the writing phase of the dissertation and mostly understood when I could not devote more time to him. Finally, to my wonderful husband, Lee Michael Penyak, for believing in me. Back in 2006 you thought it would be a good idea for me to stay in the US after finishing my Fulbright and to pursue graduate studies. I followed your advice. It was the best decision of my life, not only because of my academic achievements, but because staying here meant being with you. Thank you, honey, for all these years of unyielding love and support.

Germán

Para Violeta y Mario, mis padres y ejemplos de vida,

&

for Lee, my husband and best friend.

CONTENTS

Chapter 1: Introduction	1
Intonation in Spanish.....	2
Intonation in Spanish as a second language.....	2
Theoretical assumptions	3
The Autosegmental-Metrical model.....	3
Levels vs. configurations	4
Intonation as a phonetic and phonological unit of analysis	5
Intonation and SLA research	6
Perception and production in L2 intonation studies.....	8
Proficiency (or levels of bilingualism) in L2 intonation	10
Target form	11
Statement of the problem.....	12
Terminology and some conventions	13
Chapter 2: Literature Review.....	15
Overview	15
The Autosegmental-Metrical model.....	15
General description.....	15
Phonological and phonetic levels in the AM model	18
Alignment in the AM model	22
Spontaneous versus laboratory speech	26
Intonation of Spanish declaratives	30
Prenuclear pitch accents in Spanish	32
Production and perception of intonation	37
Speech production and perception	37
Production and perception of intonation	39
Second language intonation	43
Second language phonology	43
The study of L2 intonation.....	46
Perception studies.....	47
Production studies	55
Teaching intonation	66
Summary and conclusion.....	69

Research questions and hypotheses	70
Chapter 3: Perception	73
Overview	73
Methodology	73
Participants.....	73
Materials	80
Tasks.....	81
Procedure	85
Changes after pilot study	89
Data analysis	91
Results	94
Prenuclear alignment	95
Boundary tone	100
Discussion and Conclusion.....	106
Prenuclear alignment	106
Boundary tone	109
Conclusion	111
Chapter 4: Production	113
Overview	113
Methodology	113
Participants and materials	113
Tasks	113
Procedure	115
Data Analysis	117
Results	123
Prenuclear alignment	123
Boundary tone	134
Discussion and conclusion.....	139
Prenuclear alignment	139
Boundary tone	145
Conclusion	149
Chapter 5: Relationship between perception and production	150
Overview	150
Methodology and data analysis.....	150

Results	152
Discussion and conclusion.....	156
Conclusion	158
Chapter 6: Conclusion.....	160
Research questions and main findings	160
Theoretical and methodological implications.....	162
Role of L2 input	162
The bilingual and monolingual norms.....	164
Variation in intonation.....	165
Limitations and future research	166
Conclusion	168
Appendix A: Elicited Imitation Task	170
Appendix B: Scoring of Elicited Imitation Task	173
Appendix C: Background Questionnaire	174
Appendix D: Learning Context Questionnaire	175
Appendix E: Debriefing Questionnaire	178
Appendix F: Instructions for Imitation Task	179
Appendix G: Instructions for Sentence Reading Task	180
References	181

LIST OF TABLES

TABLE

1	Descriptive Statistics for Results in Elicited Imitation Task after Initial Recruitment	75
2	Group Size and Main Demographics.....	80
3	One-way ANOVA for Results in Prenuclear Alignment for All Groups	99
4	One-way ANOVA for Results in Final Boundary Tone for All Groups.....	104
5	Target Sentences for Production in Sentence Reading Task	115
6	Descriptive Statistics for Prenuclear Alignment in Sentence Reading and Storytelling Tasks.....	124
7	Post-hoc Results for Prenuclear Alignment of Final Boundary Tone in Sentence Reading Task.....	127
8	Post-hoc Results for Prenuclear Alignment of Final Boundary Tone in Storytelling Task	127
9	Results of <i>t</i> Tests for Differences in Prenuclear Alignment between Sentence Reading and Storytelling Tasks.....	129
10	Descriptive Statistics for Prenuclear Alignment in TwoIUS and ThreeIUS Utterances	130
11	Post-hoc Results for Prenuclear Alignment in TwoIUS Utterances	132
12	Post-hoc Results for Prenuclear Alignment in ThreeIUS Utterances	132
13	Results of <i>t</i> Tests for Differences in Prenuclear Alignment between TwoIUS and ThreeIUS Utterances	133
14	Descriptive Statistics for Final Boundary Tone Height in Sentence Reading and Storytelling Tasks	135
15	Post-hoc Results for Final Boundary Tone Height in Sentence Reading Task	138
16	Post-hoc Results for Final Boundary Tone Height in Storytelling Task	138
17	Results of <i>t</i> Tests for Differences in Final Boundary Tone Height between Sentence Reading and Storytelling Tasks.....	139

18	Results of Pearson's Correlations between Production and Perception of Prenuclear Alignment for All Groups	153
19	Results of Pearson's Correlations between Production and Perception of Final Boundary Tone for All Groups	154

LIST OF FIGURES

FIGURE

1	Phonological model for English intonation (based on Pierrehumbert, 1980, adapted from Botinis, Granström, & Möbius, 2001).....	16
2	Rendition of <i>Le dieron el número de vuelo</i> ‘They gave him/her the flight number.’	33
3	Boxplot for results in Elicited Imitation Task.....	76
4	Stimuli for prenuclear alignment.....	84
5	Stimuli for final boundary tone.	85
6	Example of typical response to prenuclear alignment stimuli (early t_1).....	93
7	Example of typical response to prenuclear alignment stimuli (late t_1).....	93
8	Results of perception of prenuclear alignment for LP group.....	95
9	Results of perception of prenuclear alignment for HP group.	96
10	Results of perception of prenuclear alignment for VHP group.....	96
11	Results of perception of prenuclear alignment for SNS group.	97
12	Results of perception of prenuclear alignment for ENS group.	97
13	Results of perception of prenuclear alignment for BS group.....	98
14	Results for post-hoc comparisons in prenuclear alignment stimuli.	99
15	Results of perception of final boundary tone for LP group.	101
16	Results of perception of final boundary tone for HP group.....	101
17	Results of perception of final boundary tone for VHP group.....	102
18	Results of perception of final boundary tone for SNS group.....	102
19	Results of perception of final boundary tone for ENS group.	103
20	Results of perception of final boundary tone for BS group.	103
21	Results for post-hoc comparisons in final boundary tone stimuli.	105

22	Shifts in perception in relation to segmental material	106
23	Example of coding and analysis for utterance produced in sentence reading task.	122
24	Boxplots for results of prenuclear alignment in sentence reading and storytelling tasks.	125
25	Comparisons between L2 learners and other groups in results of prenuclear alignment according to task.	126
26	Boxplots for results of prenuclear alignment for all groups in TwoIUS and ThreeIUS utterances.	131
27	Boxplots for results of final boundary tone height in sentence reading and storytelling tasks.....	136
28	Comparisons between L2 learners and other groups in results of final boundary tone according to task.	136
29	Comparison of prenuclear alignment between TwoIUS (top panel) and ThreeIUS (bottom panel) utterances.....	141
30	Scatterplots for relationship between perception and production data.....	153
31	Scatterplots for relationship between perception and production of prenuclear alignment for each group.	155

Chapter 1: Introduction

Intonation has been the object of study from an array of angles, which does justice to the many functions intonation can fulfill. Intonation is used linguistically to signal boundaries of syntactic units, determine the type of utterance (question, command, etc.), and present information as new or old. Intonation is also used pragmatically to convey personal attitudes, express emotions, and identify speakers' age, gender, and dialect (Prieto, 2003; Vaissière, 2005). This convergence of various functions on a single acoustic signal is one of the factors that have rendered intonation a relatively illusive object of study in the field of Linguistics and speech sciences. Nevertheless, the last two or three decades have witnessed a surge in the amount of research on intonation, probably fueled by a parallel consolidation of phonological theories that deal with it and the increasing complexity in the tools available to study intonation empirically.

Intonation and other prosodic features of pronunciation are known to be the main contributors of foreign accent among second language (L2) learners (e.g., Anderson-Hsieh et al., 1992; Pennington & Richards, 1986); however, the field of second language acquisition (SLA) has been rather slow to incorporate this growing interest into its own research. In its defense, one may argue that the study of SLA, especially in adult learners, is already quite complex in itself, with a plethora of variables both internal and external to the learner that affect the acquisition process. When the inherent complexity of SLA research is combined with an intrinsically complex unit of analysis, such as intonation, we are faced with research questions that are challenging to address empirically. Indeed, many SLA studies on intonation, especially early ones, suffer from methodological limitations in their

design, which makes it difficult to use these results to inform the field. Nonetheless, as I hope to show in this dissertation, investigating L2 intonation is possible and productive and far from a futile endeavor.

Intonation in Spanish

Important Spanish philologists such as Navarro-Tomás (1944) and Quilis (1975) laid the foundations for our modern understanding of intonation. From an Autosegmental-Metrical perspective, since the extensive publication on Spanish intonation by Sosa (1999) and a few seminal studies looking at key intonational features of Spanish (e.g., Llisterri, Marín, de la Mota, & Ríos, 1995; Prieto, 1998; Prieto, van Santen, & Hirschberg, 1995), the field has significantly contributed to our understanding of Spanish intonation. Several issues derived from an Autosegmental-Metrical analysis have been addressed for various dialects, such as the phonological status of the prenuclear delayed peak observed in declaratives across most dialects (Face, 2001, 2002; Face & Prieto, 2006; Hualde, 2002, Nibert, 2000) and the existence of an intermediate phrase in Spanish (Nibert, 2000). Furthermore, Spanish intonation has recently seen a steady increase in research that looks at how it compares to other languages, and, more importantly, how these cross-linguistic comparisons shed light on language-specific processes and phonological theories as a whole (e.g., Frota, D'Imperio, Elordieta, Prieto, & Vigário, 2007; Gabriel & Lleó, 2011; Simonet, 2010).

Intonation in Spanish as a second language

The field of Spanish as an L2, on the other hand, has produced only a handful of studies that address issues of acquisition and use of intonation (Henriksen, Geeslin, & Willis, 2010; Nibert, 2005, 2006). As is also the case with general L2 intonation, the

methodology and theoretical assumptions of these studies are too diverse to allow for discussion of any recurrent patterns or comparison of results. For example, Nibert (2005, 2006) gears her perception analysis of Spanish intonation toward the study of the intermediate phrase, an intonational unit whose applicability for Spanish has been debated (Hualde, 2002; Sosa, 1999).

Theoretical assumptions

The Autosegmental-Metrical model

The present study adheres to the Autosegmental-Metrical (AM) model of intonational analysis, which stemmed primarily from Pierrehumbert's (1980) seminal thesis, with later revisions by the same author and others (Beckman & Pierrehumbert, 1986; Ladd, 1983, 1996; Pierrehumbert, 2000).

One major hurdle in traditional studies of intonation has been the overwhelming number of intonational contours that appear to exist in languages. This posed a challenge for phonological approaches to intonation, where the organization of speech into a manageable system of finite units and contrasts is essential. Compared with the units at the segmental level, intonation was seen as a *half-tamed savage* (Bolinger, 1978) which resisted a straightforward analysis into discrete, phoneme-like units. AM phonology overcomes these apparent limitations by simplifying the structure of the intonational phrase. As proposed in Pierrehumbert (1980), intonation is composed of a binary system of high (H) and low (L) tones. A crucial premise for AM phonology is that melodies (the phonetic implementation) are generated from these underlying tones (the phonological representation). The phonetic realization, including details about pitch height and pitch register, is achieved by a series of

implementation rules that take into account factors such as the speaker's pitch baseline, downstepping rates, and the values of the tones to the left. In sum, the actual contours are the phonetic *by-product* of getting from one target to the next one. The resulting transitions between targets are responsible for the final—phonetic—shape of the melody (Gussenhoven, 2004; Ladd, 2008).

Levels vs. configurations

The AM model is normally included under analyses of intonation *by levels*, which are usually in opposition to analyses *by configurations*. This distinction comes mainly from differences in American and British approaches, respectively. The former goes back to the work by Pike (1945) and Trager and Smith (1951). The analysis by levels implies that the unit of representation is believed to be a tone (in the AM model, for example, L and H), which is an inherently static unit devoid of information about tonal excursions. For Spanish, the first applications of an analysis by levels can be traced to Stockwell, Bowen, and Silva-Fuenzalida (1956), followed later by Quilis (1975), while Sosa (1999) was perhaps the first comprehensive work on Spanish to use a by-levels analysis in the AM tradition. On the other hand, British approaches have favored an analysis by configurations (e.g., Brazil, 1981; Halliday, 1970), where the phonological representations of intonation are tonal movements (for example, a fall or a fall-rise). Various authors (e.g., Arvaniti, 2011; Ladd, 2008, Sosa, 1999) have now pinpointed strong arguments in favor of an analysis by levels over one by configurations, the details of which I discuss in the literature review. Another important observation for the goals of this research is that an analysis by levels, such as AM

phonology, appears to be a model with high explanatory and predictive adequacy to handle issues of intonation in SLA, including cross-linguistic influence (Mennen, 2008).

Intonation as a phonetic and phonological unit of analysis

For this dissertation, I adopt the position that intonation can be studied from a phonological and phonetic perspective, without the need to establish, at least not *a priori*, a semantic or pragmatic approach, as has been the case in many studies dealing with prosody. From a theoretical perspective, the field has recently developed a substantial understanding of phonological and phonetic issues of intonation. In turn, from a methodological point of view, we now have powerful technology that has increased the possibilities of implementing sophisticated research designs for the study of how intonation is perceived and produced. These constitute compelling reasons to treat intonation as a phonological and phonetic unit of analysis in its own right.

In this sense, I follow previous authors in the AM tradition who have defended the feasibility of—and need for—research on intonation independent of other levels of linguistic inquiry (Liberman, 1975; Pierrehumbert, 1980). The same position has been adopted for the study of Spanish intonation (Sosa, 1999). Other highly influential authors outside the AM approach have also supported the independence of intonation, especially from grammar (or *morphosyntax*, from a more modern interpretation of the term). Cruttenden (1981), for example, more than three decades ago noticed that there was a “new interest in the principles of intonational description for its own sake, and not simply as an awkward appendage to grammar” (p. 88).

Having said that, of course I believe that intonation interacts with other areas of linguistic analysis, but I think it is necessary to understand intonation as a separate part of the whole system—and to investigate whether it can actually stand as a coherent and self-sustained sub-system—before we look at how it interacts with other parts of the whole. This is particularly true for the relatively under-researched area of L2 intonation.

Intonation and SLA research

The field of L2 phonology has produced a number of models and theories that attempt to capture the processes—cognitive or otherwise—that characterize adult acquisition, development, and use of L2 phonology. For example, the Perceptual Assimilation Model (Best, 1995) and the Speech Learning Model (Flege, 1995) are among the most influential approaches to the study of L2 phonological perception and production. These models are often based on descriptions of parallel processes in first language (L1) speech perception, such as the Perceptual Magnet Effect (Kuhl, 2004). In SLA, the focus of several studies that have employed or tested these models has been the segmental level of analysis, with occasional explorations of larger units, such as the syllable, but this research makes no predictions about the acquisition of L2 intonation. Even though there have been some recent attempts to test aspects of these theories with the perception of lexical tones (e.g., Lee, Tao, & Bond, 2010) and the interaction of tone and intonation (Teague, 2011), in general we still lack a convincing and sound theoretical model to handle issues of L2 intonation.

Despite the aforementioned theoretical shortcomings and the fact that it has been significantly underrepresented in relation to other areas of interlanguage phonology, L2

intonation has been examined roughly since the inception of SLA as an independent field study. In Chun's opinion (2002), intonation was a key component of the audiolingual methods, in which the production of target-like pronunciation was a central goal of instruction. In contrast, the later emergence of more cognitive-based SLA studies did not initially prioritize intonation, probably due to the influence of communicative approaches to language teaching, where the emphasis lay on getting the message across rather than accurate pronunciation, let alone suprasegmental features such as intonation.

It appears that the field did not begin to take a serious interest in intonation until the late 1970s. Initially, intonation was touched upon tangentially in the prolific field of foreign accent (Flege, 1981; Piske, MacKay, & Flege, 2001; Purcell & Suter, 1980; Tahta, Wood, & Loewenthal, 1981), where the purpose has been to determine how intonation contributes to the level of perceived foreign accent among L2 learners (more recent discussions on the relationship between intonation and foreign accent can be found in Jilka, 2000, 2008; Moyer, 1999; and Munro, 1995). The explicit teaching of L2 intonation has been another popular area of research (Chun, 2002; de Bot & Mailfert, 1982; Lantolf, 1976; Levis, 1999), including some technology-assisted studies that employ visual aids (usually a graphic representation of the pitch track) as a means to increase learners' awareness of their own intonation (Chun, Hardison & Pennington, 2008; de Bot, 1983; James, 1977; León & Martin, 1972). Finally, the field of SLA has also followed the more traditional, discourse-based descriptions of intonation and its semantic and pragmatic functions (Pickering, 2009; Ramírez-Verdugo, 2005), including, for instance, its paralinguistic role as a cue to infer speakers' emotions (Holden & Hogan, 1993).

In sum, it would be unfair and inaccurate to suggest that SLA has not produced considerable studies dealing with intonation in an array of areas. However, this broad scope of coverage, ramification of interests, and the diverse definitions of intonation as an object of study may have prevented us from consolidating the field and achieving a greater understanding of L2 intonation.

Perception and production in L2 intonation studies

Intonation in general has been investigated in both perception and production studies. One goal of perception studies has been to provide a description of the relevant acoustic properties that actually play a role when perceiving (and processing) fundamental frequency (f_0) as intonation. This approach has been called a *listener's model of intonation* by, for example, 't Hart and associates ('t Hart et al., 1990). Categorical versus gradient perception of several intonational features (mostly tonal alignment and pitch height) has been another productive area. This research has sought to determine how stimuli are parsed into categories and has used this information to establish phonological units of intonation, or at least so the argument goes, since this purported relationship between perceptual categories and phonological units has not always been clear (e.g., Gussenhoven, 2004). As for production studies, research has looked at how speakers use intonation in an array of contexts and experimental conditions. Undoubtedly, the biggest challenge in production studies is the degree of control speakers have over their speech. It is probably desirable to use spontaneous and unplanned speech as much as possible, but the type of data that these designs yield can be too chaotic to analyze and they do not always contain answers to the research questions initially posed. On the other hand, highly controlled lab studies where

participants are asked, for example, to read sentences out loud allow the researcher to manipulate and observe specific features of intonation, but they have to deal with potential accusations that the language they examine may not reflect that of daily interaction. Indeed, research has demonstrated sizable differences between spontaneous and lab speech when examining intonation (e.g., Blaauw, 1994; Face, 2003; Howell & Kadi-Hanifi, 1991)

As expected, SLA has also looked at intonation from the point of view of perception and production. However, one major concern is that perception studies have been comparatively few, despite the repeated calls for additional research on perception prior to testing production, since it may be necessary to determine whether L2 learners can perceive a difference before making any claims about L2 intonation based on production data alone (Chun, 2002; Cruz-Ferreira, 1987). Another common and unfortunate misconception has been that transfer is more likely to occur in L2 production than perception. In fact, some early studies even dismissed the possibility of transfer in L2 perception (Neufeld, 1980). As stated by Broselow, Hurtig, & Ringen (1987), circumscribing transfer only to L2 production while excluding perception provides a limited and impoverished view of the phenomenon, since it assumes “the transfer of motor skills rather than of linguistic competence” (p. 251). Finally, to my knowledge, only one study (Grover, Jamieson, & Dobrovolsky, 1987) has looked at perception and production of L2 intonation simultaneously, which poses another problem when trying to identify major trends and generalizations in L2 intonation. Therefore, the present study investigates research questions that address both perception and production under comparable methodological designs.

Proficiency (or levels of bilingualism) in L2 intonation

Some studies have identified an improvement in intonation as learners receive more input in the L2 and as proficiency increases (Backman, 1978; Henriksen et al., 2010; Jun & Oh, 2000; Nibert, 2005, 2006), others have found no effect for proficiency (Lepetit, 1989), and still other studies suggest that higher proficiency relates to lower performance in some cases (Chen, 2009; Grosser, 1997). This scenario of mixed and even conflicting results warrants the investigation on how proficiency in the L2 interacts with L2 intonation use. Furthermore, recent studies such as Mennen (2004) and Simonet (2008, 2011) have shown that a very high proficiency level in the L2—learned during adulthood, as in the case of Mennen's study, or as a result of balanced bilingualism, as in Simonet (2008, 2011)—can reconfigure the intonational system of the native language (or languages, in the case of balanced bilinguals), with the result that the intonational systems in neither language function exactly as those of monolingual speakers of the languages in question. These findings seem to confirm the well-known claim that bilinguals are not the mere sum of two monolingual speakers and that functioning at such a high level in both languages often times involves reconfiguration of the linguistic competence bilinguals possess in each language (Grosjean, 1989). In addition, even though highly balanced bilinguals such as heritage speakers seem to have a clear advantage over L2 learners in phonology, they sometimes have their own *heritage accent*, which seems to be, when examined under close scrutiny, unlike the accent of native monolingual speakers (Benmamoun, Montrul, & Polinsky, 2011). In sum, the special behavior of bilinguals merits further investigation, especially in the field of intonation, which has been significantly underinvestigated. Furthermore, as L2 learners

make progress in their proficiency and level of bilingualism, we should expect their linguistic competence to resemble that of a balanced bilingual, and not necessarily, as stated before, that of monolingual speakers. That is why this study will use not only monolinguals as control groups—as has been traditionally done in L2 research—but also balanced bilingual speakers.

Target form

The target form for this study is the intonational contour of unmarked, non-emphatic declarative utterances in Spanish. There are various reasons for this choice. First, given the little knowledge we currently possess about L2 Spanish intonation, we need to posit research questions that address the most general and unmarked instances of discourse before moving to intonational patterns that occur in more restricted contexts. Second, there has been a good amount of research on Spanish declaratives that can be used to understand parallel processes in Spanish as an L2. Finally, there are reasons to believe speakers of American English have some difficulties in using target-like intonation in Spanish declarative sentences. While Spanish prefers a falling final intonation, a rising contour in declarative sentences seems to be quite common in American English (Bolinger, 1998), which English speakers may transfer to Spanish, as shown by Henriksen et al. (2010).

Another source of difficulty for Spanish learners is the relative complexity of Spanish declaratives. There is a general consensus that Spanish marks a distinction between nuclear (or final) and prenuclear (or prefinal) pitch accents in neutral declarative utterances. From the phonetic point of view, the nuclear pitch peak occurs within the boundaries of the

stressed syllable. For prenuclear pitch accents, however, the rise extends through the stressed syllable and reaches its peak on the posttonic syllable.

The existence of this delayed peak in prenuclear stressed syllables of declarative sentences is uncontroversial and seems to occur across most dialects of Spanish (Llisterri et al., 1995; O'Rourke, 2012; Prieto et al., 1995; Sosa, 1999). However, the phonological representation for this peak is still debated. Crucially, the intonation of declaratives in Spanish seems relatively complex at the phonetic and phonological levels, and as such it may prove difficult for the English-speaking learner who needs to hear it or utter it.

Statement of the problem

Intonation and other supra-segmental features seem to play a crucial role in the intelligibility of L2 speakers, and probably more so than the segmental make-up of L2 speech (e.g., Anderson-Hsieh et al., 1992; Pennington & Richards, 1986), but we know surprisingly little about this area of L2 phonology. Research on this issue has employed inconsistent and sometimes problematic designs. Moreover, recent theoretical and methodological advances in the study of intonation have—for the most part—not trickled down to inform SLA research. As a result, there exist few reliable and comparable findings upon which to build our understanding of L2 intonation.

The current study attempts to rectify these limitations in SLA research. By looking at a simple and highly-used structure of Spanish intonation, namely unmarked, non-emphatic declarative utterances in Spanish, it seeks to determine how adult English-speaking learners of Spanish of different proficiency levels use intonation as compared to Spanish-English bilinguals and Spanish monolingual speakers. Given the disconnect between perception and

production studies in the existing literature, another goal is to bring together these two areas in a single research design with the same underlying conception of intonation, L1-L2 pair, target form, and elements of intonation. Crucially, the same sample of participants will be included in both the perception and production sides of the study. Furthermore, the production of both controlled and quasispontaneous intonation will be investigated, which constitutes another unexplored area of L2 intonation. Finally, overall development in the L2 (operationalized as proficiency level) will be investigated in order to shed light on how and if intonation in the L2 develops, including the highest levels of proficiency in the L2.

Throughout this dissertation, I hope to show that the field of SLA is prepared to undertake further empirical research on Spanish L2 intonation. We have at our disposal a considerable amount of knowledge on at least three key areas: phonetic and phonological features of intonation in general, Spanish intonation, and the production and perception of L2 phonology. The overarching goal of this dissertation is to bring together these three areas and bridge the gap that presently exists among them.

Terminology and some conventions

Read speech: “Utterances produced in good recording conditions in response to the prompts provided by written scripts” (Beckman, 1997, p. 7)

Spontaneous speech: “speech that is not read to script” (Beckman, 1997, p. 7)

Proficiency: In order to provide a relatively stringent and thus more reliable measure, proficiency was operationalized as the combination of: a) L2 speakers' previous experience learning Spanish and b) results in an elicited imitation task (see Chapter 3).

Text: The term *text* is in this dissertation used here as some authors in the metrical tradition have done, that is, to designate the phonetic level, the spoken string of segments (e.g., Pierrehumbert, 1980).

Sequence of tones: Whenever a mere sequence of tones is presented without reference to the hierarchy or relationship among them, ellipses are used to separate them, e.g., H...L...H

Stressed syllables: When relevant, stressed syllables will be signaled by underlining the corresponding letters in the orthographic transcription. Example: *la mujer* ‘the woman’

Syllables in narrow focus: When relevant, syllables that appear in narrow focus will be capitalized in the orthographic transcription. Example: *la muJER, no el hombre* ‘the woman, not the man’

Chapter 2: Literature Review

Overview

The goal of this chapter is to explore the literature that informs and motivates the research questions of this dissertation. Previous studies, pertinent theories, and various methodological approaches are presented and discussed critically. This section also provides a detailed description of the linguistic target form under investigation, namely, the intonation of neutral declarative utterances in Spanish, along with the two specific tonal events under examination. Summaries of each subsection intend to highlight the gaps that this dissertation seeks to bridge. The chapter ends with the research questions that guided the study.

The Autosegmental-Metrical model

General description

Even though Pierrehumbert's (1980) thesis is undoubtedly the main foundation for the Autosegmental-Metrical¹ approach as presently conceived, she had at her disposal a solid theoretical framework to build upon. As early as Pike (1945) and Trager and Smith (1951), American scholars viewed intonation as a set of level tones that gave shape to the intonational contour (this is, as discussed above, in stark contrast with analyses by configurations). The same underlying concept was adopted by Pierrehumbert (1980) in a system with two tones: low (L) and high (H). L and H tones can be of three basic types, depending on their location in relation to elements of the phrase: an initial and a final boundary tone (T%); a pitch accent (T*), which is the most important element of the phrase

¹ The term *Autosegmental-Metrical* was coined by Ladd (1996).

and is associated with a stressed syllable; and a phrase accent (T–). The possibilities of well-formed sequences in English are presented in Figure 1.

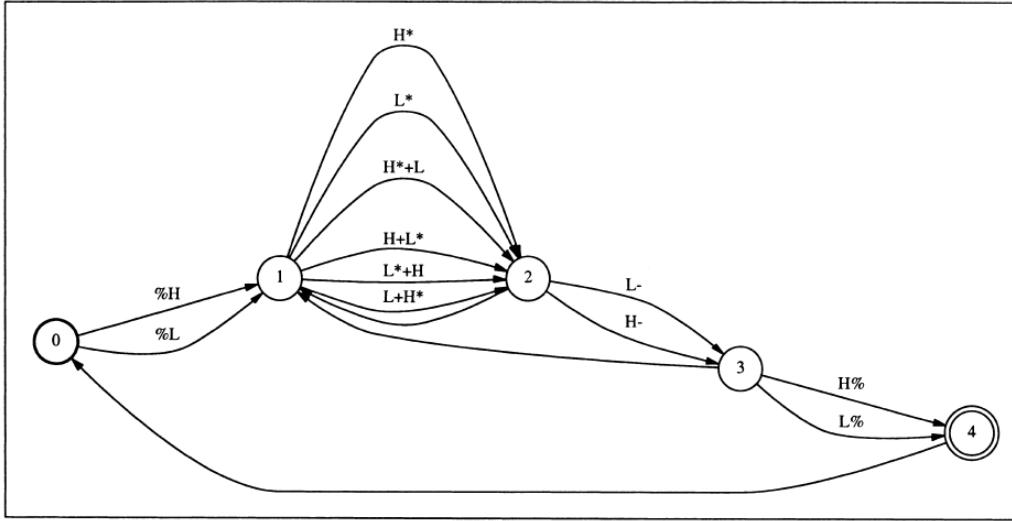


Figure 1. Phonological model for English intonation (based on Pierrehumbert, 1980, adapted from Botinis, Granström, & Möbius, 2001).

One key component for the assignments of tones to the text in the AM model is the stressed syllable. Pierrehumbert developed this aspect of her model from Liberman's (1975) thesis, where he proposed the existence of a metrical grid tier (where stress is assigned) which connects the text and the tunes: “text and tune are underlyingly separate. The metrical system [...] defines metrical patterns for text and tune independently, and then combines them (into a single abstract gesture) by the establishment of congruence” (p. 303, emphasis in original). In the AM model, the pitch accent is associated with a stressed syllable, thus making it the most important tone in the intonational phrase. Pitch accents in English and in many other languages can be monotonous (H^* or L^*) or bitonal (e.g., H^*+L , L^*+H). The starred tone in a bitonal pitch accent is the tone more strongly associated with the syllable,

and it can be accompanied by a *leading* tone (as L in L+H*) or a *trailing* tone (as L in H*+L).

Pierrehumbert most likely borrowed from Bruce (1977) the notion that tones can be basically described in a two-tone system of L and H, instead of a four-tone system as had been the prevailing approach (though Bruce did not fully develop this idea until Bruce & Gårding, 1978). Furthermore, Bruce's (1977) analysis of Swedish also added the central concept that tones (which he equated to *turning point* in the melodic curve) appear to be consistently synchronized with the segmental string, and that changes in this timing were the basic condition to produce phonological-like shifts, as is the case of Accent 1 and Accent 2 in Swedish. The strong version of the AM model takes f_0 maxima and minima as the phonetic implementation of a phonological tone. That is, phonological tones are phonetically realized as f_0 maxima and minima, though this version has been attacked on theoretical and empirical grounds (see Dilley & Brown, 2007, for discussion). What is crucial for the AM model is that the shape of the intonational phrase is not specified in the lexical entry. Only the locus for stress, where the pitch accent is anchored, is lexically specified. This feature differs from both *tone languages*, such as Chinese or Vietnamese, and *pitch accent languages*, such as Japanese, where not only location but also type of tone form part of the lexical entry for the word.

Another important characteristic in the AM model has been its treatment of downtrends. A general tendency for pitch to lower over the course of an utterance has been long noticed. This downturn has often been referred to as *declination*, which Gussenhoven defined as “the gradual, time-dependent downsloping of the fundamental frequency across

points that might be expected to be equal” (2004, p. 98). There is no agreement yet as to the extent to which declination is automatic (that is, the result of a physiological process such as the loss of subglottal air pressure) or controlled by the speaker. The grammaticalization of declination is called *downstep*. A crucial feature of L and H tones in the AM model is the fact that they do not carry information about pitch levels, and thus downstepping is generated by phonetic implementation rules. This feature allows the AM model to handle different phonetic pitch levels while keeping the phonological units to a minimum (Beckman & Pierrehumbert, 1986). The AM model (particularly in Liberman & Pierrehumbert, 1984) has described downstep as more or less stable and has proposed models that can predict the rate and pitch register of the decline (see Prieto, Shih, & Nibert, 1996, for an application to Spanish), though more recent analyses have expressed reservations about the potential for these models to work (see Ladd, 2008).

Phonological and phonetic levels in the AM model

In the field of intonation, the distinction between phonetics and phonology has not been clear-cut. In fact, some approaches disregard the possibility of these two levels of analysis, let alone any potential relationship between the two. On the side of those who support a phonological analysis, authors such as Rossi (2000) would argue that all approaches to intonation can be conceived as *model-builders*, and not simply as mere *measurers* of intonation and, given their common goal to characterize the abstract system that underlies intonation use, all theories can be considered more or less phonological in nature. However, other authors (e.g., Ladd, 2008) believe that certain descriptions of intonation are exclusively phonetic and they need not make attempts to pinpoint

phonological categories. These approaches seem to look for semantic or pragmatic regularities (such as the use of intonation in creating contrast, focus, etc.), but they connect these patterns directly to acoustic properties of intonation, such as duration of pauses and depths of valleys. Ladd (2008) suggests, for example, that the important work by authors like Xu (e.g., Xu, 2005) falls into this category. Ladd asserts that “by directly relating phonetic detail to categories of meaning, the approach taken in these studies presupposes that intonation is unlike the rest of language, because it has no place for a phonological level of description” (2008, p. 19).

Traditionally, there has been a more phonetic approach to intonation, and only with the advent of more recent approaches have some researchers started to move steadily towards an *intonational phonology* (Ladd, 1996, 2008). A promising feature of modern approaches to intonation, such as the AM model, is that they believe in the necessity of both areas of study. In other words, the study of the phonetic substance of intonation is crucial in the understanding of intonational phonology. Once we access the abstract representation by looking at how intonation is produced and perceived, we can proceed to tease apart what constitutes phonological representation and what features are added when this representation is actualized in speech.

Pierrehumbert (1980, 1990) has repeatedly acknowledged the need to look at the interaction (or *interface*) between both levels of analysis while still maintaining the autonomy of each. She clearly stated her two main goals in the introduction to her thesis: “to develop an abstract representation of English intonation” and “to investigate the rules which map these phonological representations into phonetic representations” (Pierrehumbert, 1980,

p. 10). Ladd (1983) also wrote a thorough analysis of dominant intonational theories which sheds light on how these theories treated phonetic and phonological analyses of intonation. In his opinion, theories that have contours as their unit of analysis (what he called *contour interaction* theories) focus on the functions of contours, overemphasize the generalizable aspect of intonation, and fail to capture the phonetic details of contours. On the other hand, theories that view intonation as consisting of atomistic individual tones (what he called *tone sequence* theories) place too much emphasis on the discrete units while overlooking native intuitions about similarities among melodic curves that may appear different phonetically. Ladd addressed these alleged shortcomings by proposing a set of three binary features while maintaining the basic principles of Pierrehumbert's model. This proposal did not gain popularity but its goal was legitimate and it is still valid today: to propose a model of intonation that achieves both "phonetic explicitness and phonological generality" (p. 732).

There has been a rather popular misconception that modern metrical models assume a straightforward relationship between the phonological representation of tones and their phonetic implementation. Pierrehumbert (1980) adopted the binary system proposed by Bruce but rejected the idea of locating phonological tones based on turning points in the contour. As she and other authors have underscored, two alternatives are possible: there can be turning points in the melodic contour that are not associated with any underlying tones and there can be underlying tones that do not generate any turning points in the contour. One example of the first possibility is the *sagging* contour—or dip—that usually occurs in English and other languages between two peaks that are far enough apart. The status of this dip has been the source of much analysis, since it is often just phonetic and not associated

with a tone (see, for example, discussion on Spanish below). As for the second possibility, the suspension (or level) intonation found at the end of some Spanish utterances is believed to be generated by the H*+H...L% sequence of tones, where the trailing H and the L% somehow neutralize one another and render the observed level intonation (Sosa, 1999).

As becomes clear, there are a few problems with this last point from the perspective of language learning. How are these tones acquired if there are no concrete representations in the contour? Any linguistic theory should make provisions for language learning, since it needs to prove that the abstract phonological rules or units it proposes are discoverable and learnable by the child exposed to the language. For example, Optimality Theory (McCarthy & Prince, 1993; Prince & Smolensky, 1993) has proposed a learning algorithm that explains how the learner is able to rank a set of markedness and faithfulness constraints (Tesar & Smolensky, 1998, 2000). From my perspective, the AM model of intonation lags behind in this respect.

The relationship between phonetics and phonology for intonation is more complex than I have outlined here. It is worth emphasizing, however, that the AM model should be seen as a phonological approach simultaneously grounded in phonetics. Despite some complications and unresolved issues, adopting a more comprehensive approach to intonation that brings together phonetics and phonology is, in my opinion, more fruitful than looking at the problem from a single point of view.

Alignment in the AM model

Many modern descriptions and analyses of intonation have focused on the empirical observation that f_0 changes tend to be timed with the segmental string. In the AM model, this has served as evidence supporting the relative *stability* of certain tones in several languages, as already observed by Bruce (1977) in his pioneering thesis on Swedish.

Some important production studies in the 1990s tested whether tonal alignment with segmental landmarks remains constant when other prosodic factors are manipulated experimentally. Silverman and Pierrehumbert (1990) analyzed the alignment of the prenuclear H* tone in English. They found that the peak was consistently aligned on or immediately after the accented syllable, but they also found systematic variation depending on speech rate, distance to the following word boundary, and number of intervening syllables between the prenuclear and nuclear syllables. In all cases, the prenuclear H* tone was proportionally aligned earlier whenever these factors reduced the duration of the utterance, either because it was uttered faster or because less segmental material existed between the peak and a prosodic boundary or an upcoming tone (that is, stress clash avoidance). Prieto, van Santen, and Hirschberg (1995) examined the H* tone in declarative sentences in Spanish as produced by two speakers of Mexican Spanish. They found that peak delay, as in English, is reduced as a result of adjacency to word boundaries, phrase boundaries, or another stressed syllable. Also, peak delay is positively correlated with vowel and onset duration, except for cases when vowel lengthening occurs as a result of an upcoming prosodic event. Another important finding is that valleys tend to align with onsets and show less displacement variation than peaks.

Building on these two studies, Arvaniti, Ladd, and Mennen (1998) investigated issues of alignment of prenuclear accents in Greek among five native speakers of the language. Greek prenuclear peaks are delayed with respect to the stressed syllable. As in the cases of English and Spanish discussed above, the authors found that the location of this delayed H tone is consistent when it does not appear immediately next to the end of the word and there are at least two intervening unaccented syllables. These three studies also provided evidence countering the notion that the slope and duration of a tonal movement were fixed, as had been supported by configurational analyses. Under this new view, tones are relatively stable, while the slope and duration of the rise vary and adjust when pressured by other constraints (stress clash, upcoming boundaries, etc., Ladd, 2008).

The rather unexpected finding from Arvaniti et al. (1998) that peaks are highly stable when not affected by prosodic factors became known as the *segmental anchoring hypothesis* (SAH), which suggests that peaks and valleys seem to be anchored to certain landmarks in the segmental string and that this stability is relatively unaffected by factors such as syllable structure or speaking rate. The SAH has been instrumental in questioning one major prediction made by the original AM theory, namely that there is a fixed temporal relationship between tones in a bitonal pitch accent. In the L+H* tone, for example, the starred tone is believed to be aligned—or perhaps *associated*, following Ladd's (2008) distinction—with the stressed syllable while the leading L tone will appear at a constant interval of time and unaffected by the segmental material. However, the studies by Ladd and Schepman (2003) and Dilley, Ladd, and Schepman (2005) examined valleys and peaks in English, respectively, and determined, based on results from both studies, that “the two

tones in a bitonal L+H* pitch accent are independently aligned with respect to the segmental string, and not with respect to each other” (Dilley et al., 2005, p. 118). These phonetic facts obviously pose an important challenge—which, to my knowledge, has not been successfully resolved—for the phonological status of bitonal pitch accent in the AM model.

In the area of speech perception, the manipulation of tonal alignment in the stimuli has also produced evidence supporting the stability of peaks and valleys in relation to segmental material. In what probably constitutes one of the seminal studies in this area, Pierrehumbert and Steele (1989) investigated the alignment of peaks relative to the onset of the stressed syllable *mi* in the phrase *only a millionaire*. Subjects were presented with 15 different versions of synthesized stimuli and asked to repeat them. The stimuli only differed in their peak alignment, with intervals of 20 ms between each version, which yielded an even continuum spanning over 300 ms. Despite the even distribution of the stimuli, subjects’ repetition clustered around two preferred peaks: one 150 ms and the other 300 ms after the release of the consonant [m]. The authors interpreted this distinction as evidence for a categorical, and thus phonological, status of peak alignment, in particular for the English pitch accents H*+L and L*+H.

The finding that manipulation of alignment can produce categorical-like changes in the perception of intonation has been a productive field of inquiry. Other categorical shifts have been reported in various languages and for a variety of functions: from statements to questions in Neapolitan Italian peak accents (D’Imperio & House, 1997; D’Imperio, Gili Fivela, & Niebuhr, 2010) and Russian (Makarova, 2007), and from contrastive focus to broad focus in peaks of Pisa Italian (D’Imperio et al., 2010). In a brief discussion on this

topic, Ladd (2008) concluded that “alignment can trigger essentially categorical responses from listeners, and [...] it is appropriate to regard the alignment of pitch movements relative to the segmental string as a phonetic dimension on which phonological distinctions can be based” (p. 172). Importantly, Ladd also notes that this observation is theory-neutral,² in the sense that it is consistent with an AM phonological approach to intonation, but it is not necessarily exclusive to it. In any case, Ladd asserts that these categorical distinctions constitute a “real phonetic phenomenon” for which any phonological characterization of intonation will have to account (p. 172).

There still remain, however, some issues to resolve in the field of tonal alignment. As for the SAH, we do not know whether tonal alignment can be related purely to syllable boundaries or if segmental factors such as vowel length should also be considered (see, Ladd, 2008, for discussion). Prieto and Torreira (2007) found that alignment in Spanish prenuclear peaks varies depending on syllable structure: the peak is located near the end of the accented vowel in CV syllables but toward the beginning of the sonorant coda in CVC syllables. The experiments suggest that the peak “is not anchored at acoustic segmental landmarks such as the vocalic or the syllabic offsets, and thus do not provide support for a strict view of the SAH” (p. 495). Another potential limitation seems to be that alignment studies have looked mostly at Indo-European languages, with comparably little research in other language families. In addition, from a methodological perspective, some scholars have recently begun to question some of the ways used to measure and locate peaks and valleys, which constitutes a crucial phase when making claims about alignment. Petrone and

² Ladd probably meant “theory-neutral” if considered only from a *phonological* perspective, since we cannot deny that processes such as categorical perception rest on rather well-established phonetic and psycholinguistic theories.

D'Imperio (2009), for example, analyzed the location of nuclear valleys in Neapolitan Italian as detected by three methods, namely manual, automatic, and semi-automatic. They concluded that the interpretation of results depended on the methodology used to detect tonal targets. They suggested that automatic procedures are probably more reliable than manual ones and call for a unification of methodologies in order to facilitate comparison of findings among alignment studies. Despite these shortcomings, research on the perception and production of tonal alignment is still a productive area of analysis. Given the key role it has played in understanding intonation thus far, there is every reason to believe that further empirical research on alignment will contribute substantially to current theoretical models of intonation (Xu, 2011).

Spontaneous versus laboratory speech

It has been long noted in the field of Linguistics that our speech varies tremendously depending on the communicative context. The language we produce is not the same if we address a child in the street or a superior at work, or if we have to improvise a speech or read it out loud, and this variation cuts across all linguistic domains, from pronunciation to pragmatics. This variable has been extensively investigated and is usually called *speaking style* (for a review of early work in this area, see Llisterri, 1992). This section will provide an overview of the *phonetic variation* observed in intonation research, when *speaking style* is roughly operationalized as the level of control and planning the speaker has over her production.

Researchers who embark on speech production analysis are usually confronted with methodological dilemmas regarding the degree of naturalness they expect to elicit in the

language samples under investigation. If we believe that one of the linguist's tasks is to describe language as it occurs in everyday interactions, then we should take into consideration the extent to which the language under study actually reflects spontaneous language use. In an insightful review of a book on Swedish intonation written more than 50 years ago, Lehiste (1963) astutely noted that a linguist needs to find a compromise between "naturalness" and "rigorousness in his experimental design" (p. 353). She added:

The balance between the two depends to a great extent upon the personal inclinations of the investigator. [...] The two requirements of naturalness and rigorousness appear to occupy the opposite ends of a continuous scale of gradual mutual exclusion. At one extreme we find complete naturalness coupled with complete lack of control; at the other end, complete control over the experimental design at the expense of naturalness and linguistic relevance.

(p. 353)

Though there is no reason to believe that Lehiste was favoring natural speech, authors such as Xu would probably disagree that "linguistic relevance" suffers in more controlled forms of speech elicitation. Xu (2010) wrote a compelling and clever article called "In Defense of Lab Speech," in which he attempted to dismiss the popular negative stereotypes about lab speech, such as its monotony and lack of communicative functions and emotions. He proposed that the frequently sought *naturalness* in speech might be a function of the degree of formality, which could also be studied in a laboratory setting. More importantly, he argued that "true progress in our understanding of speech has to rely heavily on lab speech" since "science progresses not by collecting more data" but through a process

of hypothesis testing (p. 334). In this sense, he believes that spontaneous speech does not easily allow the researcher to control the necessary variables for rigorous hypothesis testing.

In the field of intonation, Xu (2010) also rejected the idea that lab speech is impoverished in terms of its prosody, though he acknowledged that there are differences with spontaneous speech. What are some of these differences? A number of studies have identified critical ways in which read speech and spontaneous speech seem to diverge. Howell and Kadi-Hanifi (1991) asked six participants to describe freely a room of their choosing. Participants were asked later to read the orthographic transcription of their own description, along with the descriptions from other two participants. The most interesting findings showed that participants tended to produce more—and thus shorter—tone units in spontaneous speech than in read material, that the placement of stress (accented and deaccented vowels) varied in the two modes, and that this variation did not seem to correlate with speech rate (which was significantly faster in read speech). One possible confounding factor in this study, and to some extent one that was recognized by the authors themselves, is that punctuation in the transcription served as a cue for participants to produce tone groups while reading out loud.

In a perception study, Blaauw (1994) asked 42 native speakers of Dutch to classify a set of 109 utterances as *read* or *spontaneous*. The utterances had been previously recorded and contained instructions for building a cardboard object (the elicitation for the two versions followed the same procedure as in Howell & Kadi-Hanifi, 1991). Participants answered 77% correctly in the classification test. In a second phase of the analysis, the researcher looked for potentially relevant prosodic cues in the stimulus utterances and then

used these cues as predictor variables that could account for what listeners perceived as differences between the two modes. The results showed that variables such as presence of a falling contour, pre-boundary lengthening (within and at the end of the phrase), and presence of a pause accounted for 37% of the variance in the judgments.

In the field of Spanish intonation, Face (2003) also discussed several areas in which the intonational features under investigation could vary according to the elicitation method. He compared the most characteristic traits of intonation in Spanish declarative sentences as described in lab speech research against data from 150 utterances taken from an oral corpus of spontaneous speech. That is, Face did not analyze his own data for lab speech traits but, rather, based his discussion on features previously described in the literature on intonation in Spanish declaratives. Face's overall findings revealed differences in four major areas: pitch accents associated with stressed syllables (in spontaneous speech, there tend to be more deaccented syllables, and thus the otherwise expected rise in f_0 does not occur), peak alignment (early prenuclear peak alignment in spontaneous speech is not necessarily the result of the word being in narrow focus, as is often the case in more controlled lab speech, but perhaps a reflex of the speaker's attitudes and emotions), and final lowering (it seems to be more common in spontaneous speech). Downstepping across the declarative utterance seemed to be the only feature that remained more or less the same across both speaking styles. Despite some potential limitations, such as the lack of data for lab speech and the rather broad and at times vague analysis of the differences (there is little quantification and no statistical analysis), this study demonstrated that further empirical research using spontaneous speech is needed to test established assumptions about Spanish intonation. In a

more recent publication (Face, 2010), the author elaborated on these points and continued to underscore the necessity to look at naturally-occurring data.

In sum, research on prosodic differences between lab and spontaneous speech seems to warrant the investigation and comparison of both speaking styles. This is not just due to a mere methodological curiosity or to avoid the potential ‘mistake’ of preferring one method while overlooking the other. Rather, since both styles seem to tap into intonation differently, researchers can obtain a more complete and accurate picture of intonation when both styles are investigated.

Intonation of Spanish declaratives

Spanish phoneticians and phonologists have been drawn to the study of intonation for many decades. Prominent figures such as Navarro-Tomás (1932, and, more importantly, 1944) produced in-depth description of the intonation in Spanish and its dialects. Quilis (1975, 1993) also made a significant contribution to the field. His work adhered to the American analysis of intonation, that is, he viewed intonational primitives as a set of fixed levels that occurred at certain key points along the utterance. He was deeply concerned with demonstrating that intonation could be broken down into phonological units, thus highlighting its linguistic value (in contrast with other important scholars like Bolinger, who focused on the emotional and attitudinal functions of intonation). For Spanish, he suggested the following intonational units: three tonal levels (high, medium, and low), two terminal junctures (rising and falling), and two accent types (strong and weak). It is worth noticing how his system compared with other analyses of intonation by levels. Trager and Smith (1951), for example, proposed a system of four levels, while the modern AM model posits

the existence of two. Also, Quilis's system acknowledged the importance of the stressed syllable in the description of intonation, though it was not directly conceived as the locus of the occurrence of pitch accents, as in the case of more modern metrical models.

Moving forward in time, Sosa's (1999) detailed work constituted a landmark in our understanding of Spanish intonation. Apart from the considerable detail and scope of this work, he also provided strong theoretical support. He was the first scholar to use the AM model to describe Spanish intonation in a single volume, with a comprehensive analysis of different dialects and utterance types. He stated that a main goal of his work was to "show how different melodies depend on the sequences of tones that generate them, and how all intonational patterns in general Spanish can be explained by using these tones," thus clearly espousing the principles of the AM model (p. 93, my translation).

That said, Sosa proposed certain analyses that remain controversial to this day. Let us see two such analyses related to the declarative utterance. The first one has to do with the existence of a *phrase accent* in Spanish. Pierrehumbert (1980) used the phrase accent as an additional tone that occurs between the pitch accent and the final boundary tone, roughly at the end of the word carrying the pitch accent. In English, this accent is believed to control, for example, the type of fall from H* to L%, as either straight or gradual (that is, with a leveling phase in the slope). However, Sosa argued that Spanish can do without this tone, on the basis of the relatively brief stretch between pitch accent and boundary tone observed in the language. Sosa claimed that there can be only two syllables after the nuclear pitch accent and before the end of the utterance, and exceptionally three, so he concluded that there are not enough postnuclear syllables for English-like differences in tonal configurations to occur

(p. 95). This view was later contested, beginning with Nibert's (2000) dissertation, where the author looked at the disambiguating effect of phrase accents in the intonational phrase. The other rather contentious point of Sosa's is the differential treatment given to nuclear and prenuclear pitch accents, a topic that has received a great deal of attention, as reviewed in the following section.

Prenuclear pitch accents in Spanish

From a purely phonetic point of view, Spanish marks a distinction between nuclear (or final) and prenuclear (or prefinal) pitch accents in neutral declarative utterances: the nuclear peak occurs within the boundaries of the stressed syllable, while for prenuclear pitch accents, the rise usually extends through the stressed syllable and reaches its peak on the posttonic syllable. Figure 2 provides examples of both situations in the famous sentence by Sosa (1999) *Le dieron el número de vuelo* ‘They gave him/her the flight number.’ This displacement had been already noted by scholars such as Navarro-Tomás (1945) and investigated empirically in the 1990s (Garrido, Llisterri, de la Mota, Ríos, 1993; Llisterri et al., 1995). In a study based on a corpus of Peninsular Spanish, Llisterri et al. (1995) found that 79.5% of the peaks in paroxytone words appeared displaced to the right of the stressed syllable when they were not preceded by a pause, that is, in prenuclear position.

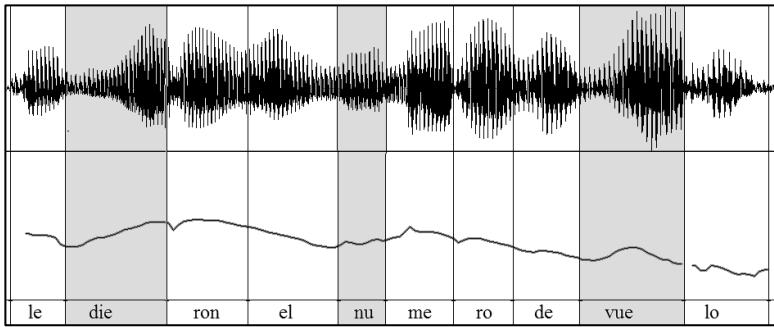


Figure 2. Rendition of *Le dieron el número de vuelo* ‘They gave him/her the flight number.’

The most straightforward explanation comes arguably from some of the studies on alignment reviewed above: nuclear peaks are constrained in time by an important upcoming prosodic boundary, namely the end of the intonation phrase. That is, in the example from Figure 2, the nuclear stressed syllable *vue* has only one syllable left to reach the boundary tone (L% in this case). For the prenuclear pitch accent *nu*, on the other hand, there are three unaccented syllables before the next stressed one. Indeed, studies like Prieto et al. (1995) have shown that when this space is reduced, the prenuclear delay also tends to shorten.

Although the existence of this prenuclear delayed peak is rather uncontroversial across most dialects of Spanish, the phonological representation for this peak is still debated. Approximately a decade ago, at least two different positions existed in this regard (I follow Hualde, 2002, in this synthesis): (1) the peak is the trailing H tone of a bitonal L*+H pitch accent, where the starred L tone is associated with the tonic syllable (Face, 2001; Sosa, 1999), and (2) the peak is the result of a monotonous H* associated with the tonic syllable (Nibert, 2000; Prieto, 1998). This is illustrated in (1) and (2) with the utterance *Le dieron el número de vuelo*.

- (1) Analysis (1)³
L*+H vs. L+H*

Le dieron el número de vuelo

L*+H	L*+H	L+H*	L%

- (2) Analysis (2)
All rising accents are H*

Le dieron el número de vuelo

H*	H*	H*	L%

Face's (2001) study provided evidence in support of Analysis (1). He tested peak alignment for prenuclear words in narrow focus. As he stated, he was not interested in words in nuclear position because of the general assumption that the peak occurs within the stressed syllable. His findings supported the analysis whereby there is a non-focal pitch accent L*+H which contrasts with a focal pitch accent L+H*.⁴ He also presented evidence that the valley is constantly aligned with the stressed syllable in both cases (more or less at the beginning). The peak, on the other hand, tends to occur on a posttonic syllable in broad focus (with some variation depending on distance with the following stressed syllable) while it is aligned within the syllable boundaries in instances of narrow focus. Put simply, Face's rationale for proposing two different pitch accents went as follows: a clear phonetic distinction triggers a clear change in the meaning of the utterance (narrow vs. broad focus), thus the distinction must also be phonological, which, following AM tenets, is the result of different pitch accents.

However, the difference between early and late peak alignment for syllables in narrow and broad foci, respectively, has also been accounted for by proposing that the following valley is an L- tone, which occurs at the end of the focused word and prevents the

³ Adapted from Hualde (2002, p. 104).

⁴ Face actually entertained the possibility of reanalyzing L+H* as (L+H)*, since both tones appeared to be aligned with the stressed syllable.

H^* pitch accent from displacing to the right (Nibert, 2000). Face (2001) refuted this analysis by showing that this valley sometimes occurred outside the focused word and, as a result, cannot be treated as a phrase accent. He found that this valley actually tended to align with the following stressed syllable, and thus should be considered part of the following pitch accent.⁵

Analysis (2), on the other hand, treats all pitch accents in the same way, and uses prosodic factors such as stress-clash avoidance and distance to a boundary to account for any differences in peak alignment. That is, peaks tend to displace to the right if they are relatively free to do so. Both Face (2001) and Hualde (2002), based on personal communication with Prieto, reported that, for instance, the prenuclear peak will tend to be timed within the syllable in oxytone words since the stressed syllable (and the associated pitch accent) occurs too close to the word boundary. Hualde (2002) provided an example of this in the utterance *Le darán el número de vuelo* ‘They will give him/her the flight number,’ where the peak of *darán* should tend to stay within the word instead of moving to the posttonic syllable.

The main problem with this analysis is its treatment of tonal valleys. The presence of a valley before each peak seems undeniable. This can be observed in the melody curve of Figure 2, where the three peaks are preceded by a valley. Once again, we are confronted with the dilemma of whether to assign a phonological status to what appears to be a phonetic regularity. Prieto (1998) found that tonal valleys were not stable enough—in their scaling or timing—to be classified as L tones. These findings contrasted with those, for example, of

⁵ In sum, Face (2001) was trying to dismiss the phrase accent L- and the existence of an intermediate phrase for Spanish. However, he later changed his position. In Face (2002), he proposed that Spanish resorts to multiple strategies to achieve narrow focus, including two that make use of the phrase accents H- and L-.

Greek (Arvaniti et al., 1998), where valleys seem to behave as L targets. In Prieto's (1998) analysis, valleys between H* would be the result of *dips*, a *sagging* transition between peaks. Pierrehumbert (1980) noticed this feature in English and wrestled with the problem of how to treat it phonologically, or if deserved to be addressed at all. Other studies on various Spanish dialects (Face, 2001; Prieto et al., 1995; Simonet, 2010; Sosa, 1999), however, have suggested that valleys are actually considerably stable and tend to occur at or near the syllable onset (as observed in Figure 2), even when the following peak may be displaced outside the stressed syllable as the result of, for example, alternations in focus, as in Face (2001). The need to treat both peaks and valleys as part of the same pitch accent led scholars like Hualde to propose a (L+H)* tone, thus reflecting the fact that both tones seem to be associated with the stressed syllable.⁶

Studies more or less contemporaneous with the beginning of this debate seemed to be inclined toward Analysis (1). Beckman, Díaz-Campos, McGory, and Morgan (2002), for example, adopted L+H* as the nuclear pitch accent of nuclear declaratives and L*+H as the prenuclear one. A few years later, however, authors that initially supported different analyses came together to propose a new phonological interpretation for prenuclear delayed peaks. Face and Prieto (2006) accepted that there is a difference in meaning between late and early peak alignment for prenuclear pitch accents, and that this difference translates into broad and narrow focus, respectively. However, they proposed L+H* as the pitch accent for both cases, the only difference being one of association: in narrow focus, the starred H tone has a secondary association with the syllable edge, represented as L+H*]σ, which forces it

⁶ Hualde follows Ladd's (1996, 2008) differentiation between tonal *association* and *alignment*. A tone may be associated with the stressed syllable (at a phonological level) but this does not necessarily mean that it will be aligned with it, as it can appear outside its acoustic boundaries.

to remain within the syllable boundaries, while in broad focus the H* tone has a primary association only and does not need to align with a metrical unit.

As can be seen from the previous discussion, even though the delayed peak in prenuclear position is quite consistent and can be considered a common feature among most dialects of Spanish (O'Rourke, 2012), the phonological representation is a matter of discussion. More recent analyses such as Face and Prieto (2006) seem promising, especially since parallel processes of secondary association between tones and other units in the metrical tree have been proposed for other languages (see Prieto, D'Imperio, & Gili-Fivela, 2005, for discussion on Romance languages). However, this analysis is relatively new and needs to be tested with more varieties of Spanish and in different methodological designs.

Production and perception of intonation

Speech production and perception

Most models of speech production assume that humans resort to multiple cognitive processes when uttering speech. At the very least, the following phases appear to be at play during speech production: (a) *conceptualization*, or the planning of what one wishes to say, (b) *formulation*, which involves encoding the message at different linguistic levels, (c) *articulation*, or the actual pronunciation of speech, and (d) *self-monitoring*, which refers to the process of making sure that the output is appropriate in view of communicative goals and context. Given the speed at which speech is initiated and produced in spontaneous communication, it is remarkable that all these processes occur almost simultaneously and with a high degree of efficiency. Levelt's (1989, 1999) modular approach is probably the

most well-known for speech production and the one most frequently applied to second language speech production. The model assumes that speech is planned and produced by different modules that operate in sequence and specialize in certain types of input. The *conceptual preparation* of speech is followed by the *grammatical encoding*, where the broad syntactic and lexical units are selected. At this point, the message has reached what Levelt calls *surface structure*. Before it can be articulated, it goes through *morpho-phonological encoding* and finally *phonetic encoding*. Three knowledge stores feed these different modules with the necessary information. Speech is also monitored simultaneously by a comprehension system which, crucially, parses and processes language from other speakers as well. For Levelt, therefore, the same critical modules are used in generating and perceiving speech.

The study of speech perception is a highly complex field, in part because speech perception relies on the human auditory mechanism, which also has the important task of perceiving non-speech stimuli. In general, the human sensory system seems to perceive and classify environment stimuli into groups and patterns and not just as a mere collection of dislocated individual stimuli, an idea which originated in Gestalt principles and which research has corroborated (see Remez, 2005, for review.). This tendency toward perceptual organization has been observed in non-speech stimuli (for example, beep-like sounds with different frequencies) and in speech (for example, syllables with different structures). Beyond this basic similarity between speech and non-speech, some theories propose that humans have a separate module that specializes in speech perception. This is the view held by the Motor Theory of speech perception (Liberman & Mattingly, 1985). The initial claims

for an independent speech perception module came from the observations that humans cannot parse spoken language in the same way they visually process systems of discrete units such as an alphabet. In this sense, much of the work in motor theory focused on how the speech perception module derives discrete linguistic units from continuous speech and coarticulation. This approach has also produced evidence that speech perception and production appear to share one system (e.g., Bell-Berti, Raphael, Pisoni, & Sawusch, 1978).

In sum, leading theories of speech perception and production suggest a strong connection between both areas. In a general discussion of this issue, Fowler & Galantucci (2005) concluded that “the review of behavioral and neuropsychological evidence within the speech domain [...] uncovers evidence for linkages between mechanisms that support motor performance and those that support perception” and they call for research that investigates more deeply “why perception-production linkages are so pervasive” (p. 647).

Production and perception of intonation

Prosody in language is composed of three major perceptual features, namely pitch, stress, and length, whose acoustic correlates are fundamental frequency (f_0), intensity, and duration, respectively. These features are superimposed on the segmental information, together with other features of less linguistic interest, such as voice quality. The specific weight of these prosodic features during speech perception depends on their position within the utterance and on language-specific rules (Vaissière, 2005). I will focus on f_0 as the most important element in the perception and production of intonation.⁷

⁷ There are acoustic cues other than f_0 that listeners use to process intonation, especially when setting boundaries of intonational units. For example, there is a cross-linguistic tendency to associate the strengthening of articulation with the beginning of an intonational unit and relaxation with the end. Also, nuclear accents are

I assume that perception and production of intonation function around a common system, as seems to be the case for other areas of language (as reviewed above). In this respect, I will adhere to the IPO (Institute for Perception Research, in Eindhoven) approach (Cohen & 't Hart, 1967; 't Hart, Collier, & Cohen, 1990), which argues that the study of listeners' perception and processing of intonation can provide us with information about the system of intonation. For the IPO approach, listeners process and make sense of pitch contours by *imposing* certain preexisting structures. This is usually called top-down processing. Clear evidence for top-down processing of intonation comes from cross-linguistic studies, where subjects of language A are tested on the perception of tones for language B, for which they have no previous knowledge, and the results show that subjects resort to the grammar of their own language to assign linguistic significance to the stimuli. For example, English speakers will interpret lexical pitch in Mandarin tones as intonation and not as tone, given the status that pitch has in English (So & Best, 2010). More interestingly, when pitch falls far outside any possible categories in the native language, it will be classified as non-speech. Crucially, participants in these studies are able to hear most acoustic differences (f_0 , intensity, or duration), but this does not amount to saying that they use all these differences linguistically. In fact, the speech signal contains many more acoustic features than are actually necessary for communication, such as minor changes and disruptions in the pitch contour, usually referred to as *microintonation* (Hermes, 2006).

The major goal of perception research within the IPO approach, therefore, is to provide a description of the relevant acoustic properties that actually play a role in

not characterized by pitch alone, but are usually accompanied by increased intensity and lengthening of stressed syllables, though the particulars of these mechanisms vary from language to language. For an example, see Ferreira (2008), who compares Spanish and Portuguese.

perceiving (and processing) f_0 as intonation. That is why this approach has been called a *listener's model* of intonation. This type of perception approach to intonation seems highly useful if we want to achieve a more accurate description of intonation as a linguistic system. Limiting ourselves to the study of production says little about two important questions that intonation research must address: (1) what acoustic features in the speech signal can the human ear actually distinguish? and, more importantly, (2) which of these cues are relevant for communication?

As for the first question, we have empirical evidence that the human ear is highly sensitive to changes in f_0 . 't Hart et al. (1990) presented a series of studies that showed that human beings can perceive differences as small as 5 Hz on the vertical axis of the pitch contour (that is, pitch range), as in a pair of tones of 1,000 Hz and 1,005 Hz, while on the horizontal axis, 30 ms is sufficient time to perceive a difference in pitch (though, as I have stated above, some research on categorical perception found much shorter spans). Needless to say, the threshold levels on both axes interact with one another in the sense that more time is needed to hear small changes in pitch, and vice versa. One caveat of these findings is that they come from psychoacoustic studies, where participants usually process tones or contours in isolation and devoid of speech. This scenario is very different from the one encountered when humans process speech for communicative purposes. This provides part of the answer to the second question. Pitch is just one more acoustic feature that listeners need to attend to in online communication, along with an array of other linguistic and non-linguistic stimuli (auditory and visual), often times in unfavorable contexts, where the speech signal is degraded to various degrees.

Under these circumstances, *listeners* are not expected—and are probably cognitively unable—to make use of minute pitch alterations that take place over a few milliseconds, such as the threshold levels described above. Crucially, the flip side of the coin is that *speakers* are not expected to encode these small f_0 movements. The intonation contour is highly complex, and vibrations generated at the vocal folds are greatly modified acoustically above the glottis. However, according to the IPO approach, the speaker only has control over major and relevant pitch movements, while the rest is composed of details beyond his voluntary control and are, actually, perceptually irrelevant from a linguistic perspective ('t Hart et al., 1990, p. 40).

Outside the IPO approach, perception has also been used to locate intonational units in studies of categorical perception. As in segmental phonology, differences perceived as categorical are believed to be phonological while gradient distinctions are phonetic. For example, VOT in English triggers a categorical effect, as in the distinction between /p/ and /b/. Categorical perception has also been studied in the field of intonation, with varying degrees of success (see section on alignment above and Gussenhoven, 1999, 2004).

In summary, perception studies can access the intonational system—believed to be shared by listeners and speakers—by presenting the listener with stimuli that aim to tease apart relevant pitch movements from irrelevant acoustic details. What the listener perceives as relevant changes can be assumed to be part of the intonational system. More importantly, this conception of intonation calls for research that integrates both production and perception. The field still lacks this type of approach, as expressed by Xu (2011):

Among the experimental approaches, there is often a preference for either a production- or perception-oriented strategy. From a functional perspective, an operable function by definition must be contrastively encoded through production, and reliably decoded through perception. The production and perception of a function therefore must have evolved in tandem [...] It is therefore beneficial for both perception- and production-oriented studies to take a comprehensive view in their interpretation of the empirical data. (p. 90)

Second language intonation

Second language phonology

The first attempts to explain L2 phonology systematically go back to the work of Lado (1957). He proposed that L2 errors (phonological and others) could not only be explained but also predicted by describing and contrasting the linguistic systems of the first language (L1) and L2. In this paradigm, errors would be expected in those areas where the systems differ, due to negative transfer from the L2, while shared features would positively transfer directly to the L2. In subsequent years, however, the predictive power of the Contrastive Analysis Hypothesis (CAH) proved problematic: there were errors attested in the L2 production that the CAH did not predict, and, conversely, some anticipated errors actually never occurred. This led to a reformulation of the hypothesis, in the form of a *moderate* (Oller & Ziahosseiny, 1970) or *weak* (Wardhaugh, 1970) versions of the CAH.

Some years later, however, Eckman (1977) claimed that the original and revised versions of the CAH were still flawed because the mere attempt to measure the level of difficulty between the L1 and L2 overlooked the fact that there are linguistic universals that

run across languages. Based on this observation, Eckman proposed his Markedness Differential Hypothesis (MDH), which predicted that more marked features in the L2 would be acquired later than the less marked version. For example, the distinction between voiced and voiceless obstruents (e.g., [d] vs. [t]) at the end of the word is more marked than the same distinction at the beginning of the word. Therefore, speakers of German (the unmarked case, with the distinction only made word-initially and medially) learning English (the marked case, where the distinction is made in all contexts) are expected to have problems acquiring the voiced/voiceless distinction for obstruents word-finally. Approaches like the MDH made a big leap in the understanding of interlanguage phonology in the sense that they attributed phonological development not only to L1 and L2 differences (causing *interference errors* due to the L1) but also to universal patterns of acquisition (causing *developmental errors*, similar to those made by native speakers when acquiring their L1).

Another valuable and more recent contribution in the field of L2 phonology comes from Flege's Speech Learning Model (SLM, Flege, 1995; Flege, Frieda, Walley, & Randazza, 1998). The SLM has been one of the most popular approaches to account for L2 phonology. Some of its strengths lie in its examination of age of acquisition, level of ultimate attainment, role of naturalistic versus classroom learning environments, and the link between production and perception of L2 phonology.

The SLM predicts that similar sounds between the L1 and the L2 are more difficult to acquire, since the difference is not salient enough for students to perceive it. If, on the other hand, the L2 does not contain a phonetically similar counterpart in the L1, the perception of the L2 sound as a new form will be enhanced, and thus facilitate its

acquisition. For example, Flege (1987) reports that English-speaking learners of French produced a native-like form for /ü/ (without a counterpart in English phonological system), while the pronunciation of French /u/ (perceived as a similar or equivalent sound in English) proved more problematic. One of the strongest predictions in this model is that L2 learners first need to attune their perception to the L2 system before they can create the appropriate *phonetics categories* in the L2. In simplified terms, once these categories are relatively stable, production that approximates the native-like performance can potentially follow.

Another major assumption in the SLM is that the learning mechanisms used to establish phonetic categories remain intact across the life span and are accessible for both L1 and, later in life, L2 acquisition. A closely related postulate is that L1 and L2 phonetic categories share a phonological space. In this model, this interaction between both phonological systems accounts for the influence of the L1 over the L2 (and of the L2 over the L1 under certain circumstances). When phonetic properties are similar between the two languages, sometimes speakers produce an intermediate resolution, a ‘merged’ category that falls in between the values of monolingual speakers of the language (e.g., Flege, Schirru, & MacKay, 2003). Notice, however, that these predictions were formulated on—and further tested with—L2 segments, while only recently has L2 prosody begun to examine claims made by the SLM (e.g., Teague, 2011).

The SLM’s prediction about phonological dissimilarity being a facilitator for L2 acquisition conflicts to some degree with the Markedness model, which predicts that different items will be harder to acquire, particularly if the L2 item is marked. In a reconciling fashion, Major drew from these two models to postulate his Similarity

Differential Rate Hypothesis (Major & Kim, 1996). He partially supports Flege's similar-is-harder position but in terms of rate of acquisition: similar sounds are slower to acquire, and he adds that marked items are acquired even more slowly. Later, the Similarity Corollary and the Markedness Corollary of his Ontogeny Phylogeny Model would refine the role of these two issues in L2 phonological acquisition (Major, 2001).

The study of L2 intonation

One persistent problem in the field of L2 intonation research has been the lack of consistency in the theoretical framework selected to approach this type of research. Part of the problem stems from the big divide that also exists outside L2 intonation, namely, the old debate between an analysis by levels and an analysis by configurations. Until the 1970s, most researchers of intonation in primary languages felt the need to adopt one position or the other, since most of the differences seemed irreconcilable. However, the influential AM model provided a new vantage point to look at intonation, which would result in a model that "is unquestionably the currently dominant approach to intonation description" (Ladd, 2000, p. 37). This dominance of the AM model does not imply that the analysis by configurations has disappeared. Rather, it has mostly specialized in different areas of intonational analysis, such as the role of intonation in the study of discourse analysis and in the teaching of English as a foreign language (García-Lecumberri, 2003).

Compared with research on primary languages, the field of SLA has been slow in adopting theoretical assumptions that inform L2 intonation research. Nevertheless, more recent L2 studies have recognized the importance of the AM approach, which is also adopted in this study and thus will be examined in depth in terms of contributions and

unresolved issues. As for methodological approaches, both production and perception studies will be discussed, since they are germane to my research. I will also address some of the pedagogical implications derived from instruction-based studies of intonation. I will use this discussion to explore major disconnects in the applicability of theory to the teaching and learning of L2 intonation.

Perception studies

Cruz-Ferreira's (1987) work was among the first cogent and well-conducted perception studies of L2 intonation. It also contributed substantially to our understanding of processes such as transfer and linguistic universals. A total of 60 university students participated in the study: 30 English speakers learning Portuguese and 30 Portuguese speakers learning English. In addition to being tested in the L2, all subjects also served as controls in their L1. Participants heard 60 pairs of sentences that differed in their intonational pattern (e.g., *Didn't John en joy it* vs. *Didn't John en,joy it*) and asked to say if the sentences were the same or different. Immediately after this stage, participants were provided with two glosses that they had to match with the sentences, either the same gloss if they thought the sentences were the same or different ones if they had judged them to be different. Cruz-Ferreira found that learners used at least three interpretive strategies when resolving the task. The first strategy was the *transfer* strategy, which learners used when the L1 and L2 had similar intonation structures and as a consequence the meaning from the L1 was assigned—correctly or not—to the L2. The second strategy was called the *pitch height* strategy, and was based on the observation that L2 learners identified L2 tones correctly when they broadly belonged to a certain category (for example, rise vs. fall), even if the L2

configuration was different from the L1. In her opinion, learners “seem to be sensitive to the gross phonetic shape of the pattern” (p. 112). This seems to be supported by scholars such as Cruttenden (1981) who assign the general universal meanings of *open statement* to rising contours and *closed statement* to falling ones. Finally, the *lexico-syntactic* strategy occurred when learners assigned the less marked interpretation to the sentences based only on their lexical or grammatical pattern.

Grabe, Rosner, García-Albea, and Zhou (2003) conducted two experiments aimed at testing the perception of falling and rising final intonation in English among adult speakers of Spanish ($n = 41$) and Chinese ($n = 40$), for whom the authors reported a low proficiency level in English. In addition, English speakers ($n = 42$) were used as controls. The stimuli were eleven synthesized versions of the phrase *Melanie Maloney*, which varied in the pitch movements and could be classified into two major groups: seven with a rising (LH) contour and four with a falling (LH) contour. The stimuli were presented in pairs, over earphones, and participants had to rate them on a scale from 1 (*very similar*) to 10 (*very different*). The second experiment was methodologically comparable but used frequency-modulated stimuli, that is, only contour movements without any speech. As predicted, all three groups clearly made a distinction between falling and rising contours, and this distinction was highly correlated between speech and non-speech stimuli. The authors suggested that this might be due to universal auditory mechanisms used in the perception of tone that do not seem to be language specific. The alleged universality in the meaning attributed to falling vs. rising tones has also been suggested in typological analyses of intonation (Bolinger, 1978; Cruttenden, 1981). However, the results also showed that the division between falling and

rising contours was not the same across the three languages, which the authors interpreted as the result of L1 influence, in particular for the case of Chinese.

Chen (2009) looked at how learners perceived the paralinguistic functions of emphasis and surprise as signaled by pitch range in L2 speech. The author used two groups of L2 learners: advanced English learners with Dutch as their L1 ($n = 17$), and beginning and intermediate Dutch learners with English as their L1 ($n = 17$). There were also control groups of native speakers of Dutch and English. The stimuli were generated from three pairs of utterances in each language, one utterance being a statement (e.g., *You asked them for a JOB interview*) and the other a question (e.g., *Did you ask her for a JOB interview?*). The original pitch of the source utterances was removed and assigned new pitch patterns that varied in peak height, pitch register, peak alignment, and end pitch. These specific variables were manipulated since they are known to trigger changes in emphasis and surprise. Participants were asked to judge the degree of surprise and emphasis they perceived in the stimuli. As initially hypothesized, the author found that L1 transfer accounted for some of the results in both groups. However, transfer played a bigger role among the group of English learners, who actually had a higher proficiency level in the L2 than Dutch learners. This unexpected difference might be due in part to the fact that Dutch learners received native input, while English learners received instruction primarily from non-native speakers. Besides transfer, the results revealed that both groups also used their L2 knowledge, which enabled them to judge the meaning of the stimuli like native speakers did. For example, English learners perceived the *surprise* meaning as signaled by end pitch, a strategy that

native speakers also used. The author claimed that participants easily learned this feature due to the salience of pitch in utterance-final position.

The combined studies by Nibert (2005, 2006) also looked at different proficiency levels. It is worth highlighting that, as far as I know, this is the only author who has dealt with the perception of intonation in Spanish as an L2. English-speaking learners of Spanish (3 proficiency levels: 22 beginners, 37 intermediate, and 18 advanced) were tested on 19 auditory stimuli consisting of five minimal groups. These groups—as minimal pairs do—differed only in one feature, namely their pitch contours, while the segmental information was the same. Specifically, Nibert tested the effect of phrase accents (H- and L-) in Spanish to disambiguate utterances like [[lilas]H- [y lirios amarillos]L-] ‘lilacs and yellow irises’, [[lilas y lirios]H- [amarillos]L-] ‘yellow lilacs and yellow irises’, and [[lilas y lirios amarillos]L-] where the lack of medial H- allows for either interpretation. Participants were required to assign meanings (or *glosses*) to pairs of utterances that differed only in their pitch configuration. Their performance was compared with that of a control group of native speakers (from Nibert, 2000). Results showed that all proficiency levels were able to perceive and interpret the presence of the disambiguating H- tone, though true beginners did so only in very limited cases. Overall, the results showed a clear positive effect of proficiency level on the perception of this particular tone. The data suggested that there can be a “gradual development or restructuring of L2 interlanguage grammar toward a more restrictive and native-like state” (2006, p. 146). In addition, the data for true beginners showed that L1 transfer was at play, though in a facilitative manner, since some instances of

target-like performance were due to the convergence of English and Spanish intonational systems.

The perception of pitch in languages with lexical tone, such as Chinese, has been another popular area of research. Participants in these studies usually have no prior knowledge of the L2 in question, but the findings are still useful since they provide information about how humans perceive and process pitch in natural languages, which certainly has more ecological validity than exposing participants to a series of artificially created tones. Also, these findings shed light on processes of L2 intonation, since participants in these studies are processing, first and foremost, differences in f_0 , which is the same acoustic property used in intonation.

So and Best's (2010) study is one example of research on the perception of tones. This work is particularly important in the field of L2 intonation because it constitutes one of the first attempts to apply the Perceptual Assimilation Model (PAM, Best, 1995; Best & Tyler, 2007), widely tested for segmental L2 phonology, to the realm of prosodic L2 phonology. The authors investigated how Cantonese, Japanese, and English speakers processed the four tones in Mandarin Chinese. Participants ($n = 30$, 10 per language group) had no previous experience with Mandarin. The hypotheses regarding the potential L1 influence on tone perception differed based on the status of pitch in each language: Cantonese is a lexical tone language (like Mandarin), Japanese is a pitch accent language, and English is an intonational language. After a short familiarization phase with the 4 tones, participants were asked to identify which tone they heard, in a total of 144 trials in two blocks. In general, results showed that Cantonese and Japanese speakers outperformed

English speakers in the identification of tones, most likely due to the special relevance that pitch has in these two languages. Also, listeners showed different levels of confusion depending on how much the L2 tones resembled L1 tones. If resemblance was high, it was likely that the L2 category was assimilated as an existing L1 phonological category, which is one of the predictions made by the PAM. On the other hand, all participants had the same difficulties differentiating Chinese tones that are perceptually similar (for example, tones 1 and 4 have a high pitch component). In sum, the overall results did support some PAM predictions, especially in relation to the areas that will be more susceptible to the influence of the L1 (or *linguistic experience*, as usually called in this model), while they also provided evidence which suggests that some perceptual processes, such as the generalized confusion of similar tones, are language-independent. One unfortunate limitation in this study is that the authors could not determine when English-speaking participants perceived Chinese tones as non-speech (which would turn the stimulus into a *non-assimilable* category in the PAM) or, if they did perceive it as speech, whether it was due to similar grammatical (such as questions vs. statements) or attitudinal (such as higher pitch for certain strong emotions) uses of intonation in English.

Despite the relative difficulty that speakers of non-tonal languages have when hearing tones, perception can be considerably enhanced after a training phase, as shown by Leather (1987). His study first exposed participants (L1 English, $n = 10$; L1 Dutch, $n = 10$; no previous knowledge of Mandarin) to a short familiarization phase with the Mandarin minimal pair /ŷ/ ('mud', with Tone 1, high level) and /ý/ ('fish', with Tone 2, high rise). Then participants were asked to identify the word they heard, based on nine stimuli

(synthetically generated) which differed only in pitch height, to form an even continuum between /ŷ/ (Tone 1) and /y/ (Tone 2), in a total of 90 trials. Participants from both groups performed as controls did when identifying the boundary between both tones, although they showed less categorical effects (that is, they showed more confusion for stimuli in middle of the continuum). The authors underscored the lack of L1 transfer effects in the study and the possibility of rapid perceptual familiarization with previously unknown tones.

In conclusion, perception studies of L2 intonation, though not abundant, do seem to indicate some interesting trends. First, they shed light on processes of L2 intonation that may be universal (Cruz-Ferreira, 1987; Grabe et al., 2003; So & Best, 2010). Isolating such processes has tremendous potential for language learning and teaching, as L2 practitioners can predict areas where learners, regardless of the L1, may have difficulty acquiring intonation. Second, these studies also show the pervasive effect of the L1 on L2 intonation learning and use. As in other areas of L2 acquisition, this influence can be facilitative (Cruz-Ferreira, 1987; Nibert, 2006) or inhibitory (So & Best, 2010).

One cause for concern in some of the perception studies described here is that they used participants with no knowledge of the L2. From a purely psychoacoustic or psycholinguistic perspective, this type of research is valid since it provides ‘clean’ measurements of perception in the L2: participants possess no previous knowledge of the L2 and thus the occasionally problematic variable *previous exposure to the L2* is not as much of a concern. However, restricting the sample to this type of participants may limit the possibilities of describing L2 intonation for learners who already have some proficiency in the L2. As we know from anecdotal and empirical evidence, proficiency is a crucial variable

in the development of L2 competence, with an array of other factors that are expected to covary with proficiency as L2 competence increases. For example, Grabe et al. (2003), at the beginning of the paper, warned the reader that their work “is *not* aimed at issues in second-language acquisition” (p. 376). As such, it suffers from a few methodological problems. First, Chinese- and Spanish-speaking participants clearly had some proficiency in English, which is acknowledged by the authors, yet not controlled as a potential intervening variable. Second, the authors mixed monolingual and bilingual speakers, as the Spanish group was actually composed of bilingual speakers of Spanish and Catalan, and there is evidence to suggest that intonation in bilingual speakers has specific features not found in monolinguals (Mennen, 2004; Simonet, 2008). Finally, the authors found clear evidence of universals of language acquisition and the effect of transfer, both of which have been widely explained in the L2 literature. The study could have been significantly enhanced if the results had been discussed in light of this literature.

By avoiding—intentionally or not—a deeper analysis of other variables that affect L2 learning, a few crucial questions remain unanswered. First, how do factors known to affect L2 perception (e.g., transfer) diminish as proficiency levels increase? In order to address this question, perception studies could adopt a longitudinal design and measure the same group of participants at different times or, more feasible from a methodological perspective, carry out a cross-sectional study where multiple groups with different proficiency levels are observed simultaneously (e.g., Nibert, 2005, 2006). I think either methodology would provide invaluable insight into the development of intonation, rather than only a static description of L2 intonation at one interlanguage stage. Second, what does

performance in perceptual studies tell us about the perception of tones/intonation during online communication? The seemingly effortless and accurate perception found in studies like Leather (1987) may deteriorate in real communication, when L2 learners have to pay attention to several other factors in their own speech and in the interlocutor's.

That being said, there are obvious advantages in these perception studies. As I mentioned previously, the use of participants who are naïve in the L2 can prove methodologically advantageous as it eliminates potential confounding effects of previous knowledge in the L2. Other methodological advantages include the possibility of synthesizing the stimuli. This process usually begins with a recording of natural speech from a native speaker of the language (e.g., Chen, 2009; Leather, 1987), which is then resynthesized so as to alter the desired variables (pitch height, end pitch, etc.). This procedure allows for control of other confounding variables that usually co-occur with changes in intonation, such as duration, intensity, and voice timber. All things considered, perception studies add to our comprehension of processes in L2 intonation, though it may be necessary to depart from the purely psychoacoustic design in order to arrive at conclusions which are more relevant for L2 research.

Production studies

Backman (1978) is often cited as one of the first studies that investigated empirically L2 production of intonation, though it suffered from a few methodological limitations. Backman studied the production of eight native speakers of Spanish, all males from Venezuela (age range: 17–29) who had just arrived in the U.S. to undertake university studies. Five native speakers of English were used as a control group. Subjects were

recorded while interacting with the researcher using a scripted dialogue. Recordings were made at the beginning and end of the semester, with a period of 3 months in between. No specific treatment was given to the subjects. The contours were visually inspected particularly for the pitch range used in declarative and interrogative sentences. Furthermore, eleven judges gave overall impressions about how target-like they perceived participants' intonation to be. For declarative utterances, the results suggested that learners' pitch range was not as wide as that of native speakers, terminal tones tended to be flat, unstressed words and syllables were generally produced with a relatively high tone (i.e., there was lack of vocalic reduction), and prominence placement sometimes occurred too early in the sentence. As for the effects of time, subjects in general improved by the end of the semester. For example, judges heard considerable improvement in 15% and partial improvement in 19% of the 184 sentences analyzed. Despite some interesting results and the breadth of target forms analyzed, the lack of sufficient quantitative and statistical analysis calls into question some of the results. Judges, for example, rated the data as either appropriate or inappropriate, which seems a rather subjective measure. The acoustic analysis did measure some frequencies (in Hz), but no inferential statistics were used.

The lack of sufficient and reliable quantitative data seems to be an unfortunate—and sometimes unavoidable—reality in studies of L2 intonation production. Huang and Jun's (2009) study of age effects is an example of this. They analyzed the production in English of 30 Mandarin speakers. Based on their age of arrival (AoA), participants were classified into child arrival (AoA: 5-9), adolescent arrival (AoA: 12-17), and adult arrival (AoA: 20-26). Length of residence for all the groups was at least five years long. Participants were asked to

read a short paragraph in English. The recording was analyzed for a series of prosodic variables. Some important findings were that children and adolescents behaved like the control group in the prosodic boundaries they used to segment speech. Also, adults used overall more H* tones than all other groups. They authors did not find any significant differences for use of pitch accents, phrase accents, or boundary tones among any of the groups.

With regards to the methodology used in the analysis, Huang and Jun had to deal with the fact that intonation is more often than not subject to a high degree of variation, even among native speakers of a language. For example, due to certain attitudinal overtones, speakers may choose to emphasize certain parts of the utterance, and thus produce different pitch peaks. This degree of variation can be more prominent among non-native speakers. Given this scenario, the authors used an analysis of *dominant patterns*, that is, “common patterns of intonation produced by the majority of the speakers within each group, instead of one absolute intonation pattern” (p. 11). First, the researchers transcribed all sentences for the 10 participants using the Mainstream American English Tones and Breaks Indices (MAE_ToBI), which is a standard prosodic transcription system developed out of the AM model. Recall that the sentences were the same across participants since it was based on a reading task. Second, the authors analyzed the annotation for each individual word across the 10 participants, and selected the label that was used by at least 5 participants (hence the *dominant pattern*). If no single label reached the critical 50%, no pattern was assigned for that word. The final and overall characterization for the group of learners was not based on means, but rather on patterns that seemed to be more prevalent. This is indeed an ingenious

way to deal with the problem of variation in the production of intonation but, as the authors acknowledged, it is always desirable to produce quantitative data in order to compare groups statistically.

Another early and frequently-cited study is Lepetit (1989). The author examined intonation during speech production among college learners of Canadian French from two L1 backgrounds: Canadian English ($n = 45$) and Japanese ($n = 30$). Participants of both L1s were evenly divided into three proficiency groups (15 per group for L1 English and 10 for L1 Japanese). There was no control group. Participants in the L1 English and L1 Japanese were asked to read a set of 25 and 22 French sentences, respectively.⁸ The target form was the intonation in neutral declarative sentences (e.g., *Jules, le grand blond, part à la guerre*, ‘Jules, the tall blond, goes off to war’). The melodic curve of the learners’ production was analyzed only in relation to the pitch used on the stressed syllable, and the researcher coded each instance as either correct or incorrect. The between-group analysis revealed that there were no significant differences among the three proficiency levels, for which the author did not provide a convincing analysis. He simply attributed it to a lack of explicit instruction and to the students’ “restricted intake” in the classroom (p. 404). Unfortunately, little was said about the participants’ background (age, sex, previous exposure, amount of instruction, etc.) and thus the reader cannot derive any alternative explanations for the lack of proficiency effects. Finally, the author did find effects for negative cross-linguistic influence in both

⁸ The author does not justify why different groups used different sets of French sentences. He simply states that the sentences “reflected the range of possible phonosyntactical structures for French sentences” (p. 402). But given the author’s initial hypothesis about the effects of cross-linguistic influence, the reader is led to believe that the sets of stimuli were tailored to each L1 to guarantee that such effects would emerge.

L1groups. For example, Japanese speaker's contours in French were rather flat, which was attributed to the influence of the L1 (Japanese has a relatively narrow pitch range).

Wennerstrom (1994) is arguably among the first studies to use the AM model consistently to analyze intonation in L2 speech production. She looked at the production of 30 adult intermediate learners of English from three language backgrounds, namely, Spanish, Japanese, and Thai (10 participants—5 males and 5 females—per language group). Ten English natives were used as a control group. Wennerstrom used a reading task and a free speech task (picture description), the goal being to analyze intonation in different speaking styles. The analysis measured key intonational components for the AM model: high pitch accents, low pitch accents, and boundary tones. Following more usage-based approaches within the AM model (Pierrehumbert & Hirschberg, 1990), one goal of the study was to assess how L2 learners used these intonational components to convey meaning in English, as in the contrast between new and old information. The results demonstrated that none of the L2 groups used the same pitch increase on new or contrastive information as did native speakers. Conversely, L2 learners reduced pitch less than native speakers for instances of old or redundant information. The author commented on the potential barriers in communication that can occur between native speakers and L2 learners if pitch and intonational boundaries that signal the information structure of the message are not used consistently. Note, however, that these were all intermediate learners, which begs the question: would observed non-native like patterns become more target-like as speakers' L2 competence increases?

Grosser (1997) examined how intonation and accentual features of L2 learners develop over time. The participants were eight native speakers of German, aged 10 to 12, learning English at an Austrian grammar school. The author recorded participants reading a passage in English and in German at three times: end of first school year (average age: 11), middle of second school year, and end of second school year. The data analysis adhered to the AM model and looked specifically at nuclear pitch accents. The results showed that L1 and L2 intonational systems tended to converge more, and deviate from the L2 norm, at year two rather than year one. In other words, interference from the L1 seemed to become a stronger factor as time passed. It would have been interesting to determine whether this apparent regression occurred as a result of a general fossilization pattern in L2 development, but the author did not report any measures of overall proficiency. There was, however, a progression in the use of the L+H tone. Specifically, its use in English was overgeneralized during the first year and it also occurred significantly more often than in German. By the second year its use in the L2 decreased significantly. The author interpreted this change as a preference for unmarked forms (of which L+H would be an example) during initial stages of development, while subsequently these unmarked forms could be dropped if they do not conform to L2 norms.

The study by Henriksen, Geeslin, and Willis (2010) is of particular interest for the present study since it constitutes, to the best of my knowledge, the only published study that addresses the production of intonation in Spanish as an L2. The authors sought to determine how the intonation of native speakers of English learning Spanish ($n = 4$) changed over a period of 7 weeks while studying abroad (in León, Spain). Participants were described as

highly proficient in Spanish and they had received most of their instruction in foreign language classrooms in the U.S. Furthermore, proficiency was tested by means of a grammar and vocabulary tests, administered during the first week of the program (Time 1) and at the end (Time 2). The target forms selected for the study were declarative sentences, pronominal interrogatives, and absolute interrogatives. The study followed the AM model in the selection of the relevant components of intonation (specifically, pitch accents and boundary tones). The elicitation task consisted of a contextualized sentence reading task, delivered by a computer, where participants were first given a context related to a story they had read before (e.g., *Contexto: Mamini te pregunta, “¿Qué hacía Papini ayer por la tarde?* ‘Context: Mamini asks you: “What was Papini doing yesterday afternoon?”’) and then asked to read the response (target item) out loud (e.g., *Respondes: “Mimaba a la nena”* ‘You respond: “He was spoiling the girl”’). The task was administered at Times 1 and 2. The analysis yielded a rather complicated picture, with lack of consistent patterns, and a good amount of variation among the four participants, especially for pronominal interrogatives. As for declaratives, two participants at Time 1 lacked the characteristic delayed prenuclear rise of Spanish declarative sentences. Only one of them showed a frequency increase in the use of this contour at Time 2. The second interesting finding for declaratives concerned the nature of the final boundary tone. Two speakers predominantly produced a final rise at Time 1 (47% and 41%) instead of the more typical fall for Spanish declaratives. The authors believed that such pattern could be the result of influence from English, the participants' L1. By Time 2, the frequency of rises decreased to 26% for one speaker and 22% for the other. In sum, this study provided valuable findings in an area that had been seriously neglected,

but the results should be viewed with caution due to the small number of participants upon which they were based.

The study of intonation production among highly advanced L2 speakers has produced some interesting results. In a sociophonetic study of the Spanish and Catalan spoken in Majorca, Simonet (2008, extended in 2011) found unexpected instances of mutual influence of the two intonational systems among bilingual speakers. Based on his own previous descriptions of both languages as spoken in Majorca, the author hypothesized that declarative sentences for Catalan-dominant bilinguals would have the final pitch contour H...L*...L% (fall), while Spanish-dominant bilinguals would produce a final pitch with the configuration L...H*...L% (rise-fall). A total of 20 speakers were analyzed, balanced between language dominance and between males and females. Both groups (Catalan- and Spanish-dominant bilinguals) read sentences in both Catalan and Spanish. The data showed that Catalan-dominant bilinguals generally kept both pitch accents apart in both languages, as predicted. However, Spanish-dominant bilinguals used only one system, which was an intermediate resolution between Catalan and Spanish. In other words, these speakers used one *converged* system in both languages.

Mennen (2004) found similar results in her study of native speakers of Dutch who were also highly proficient in Greek ($n = 5$). From a phonological point of view, both languages have a similar prenuclear pitch accent, consisting of a peak that usually occurs late in relation to the stressed syllable. However, both languages differ phonetically in that the peak in Dutch occurs earlier than in Greek. Participants were recorded while reading 20

declarative sentences in Greek and 40 in Dutch.⁹ Monolingual speakers of both Greek ($n = 5$) and Dutch ($n = 5$) were also recorded in their respective languages and served as control groups. Measurements (in ms) were obtained for syllable duration, rise duration, and alignment of peaks and valleys associated with the stressed syllable. Surprisingly, the results showed a bi-directional interference in the production of peak alignment in four out of the five participants: “not only does the L1 influence the L2, but the L2 also has an effect on the L1” (p. 558). Participants’ peak alignment in their native language differed substantially from the control group of monolinguals. Some of them developed a system more like the L2, while others lost the characteristic difference in Dutch peak alignment based on vowel duration. Unlike the study by Simonet (2008) described above, participants here did not develop a ‘merged’ system that could be described as falling ‘in the middle’ of both languages. Nevertheless, just like the results from Simonet, Mennen showed that a very high proficiency level in the L2—even if learned during adulthood, as in the case of this study—can reconfigure the intonational system of the native language, with the result that the intonational systems in neither the L1 nor the L2 function exactly as those of monolingual speakers of the languages in question. Note that similar processes have been documented for segmental interlanguage phonology (e.g., MacKay, Flege, Piske, & Schirru, 2001).

In conclusion, production studies of L2 intonation suggest that L2 learners can have serious difficulties using L2-like patterns at initial levels of proficiency, both in syntactic (e.g., Lepetit, 1989) and semantic (e.g., Wennerstrom, 1994) uses of intonation. Development towards more target-like stages is possible but apparently contingent upon

⁹ Dutch presents a further complication: there are both short and long vowels in the language and peak alignment can vary as a relation of vowel length. Therefore, the 40 sentences for Dutch actually comprised two subsets, one with short vowels and the other with long vowels.

factors such as type of input (native vs. nonnative, for example, as in Grosser, 1997; and Henriksen et al., 2010) or the perceptual salience of the target form. For instance, leaners may have fewer difficulties with boundary tones (e.g., Huang & Jun, 2009), due to their final position in the utterance, than with utterance-internal tones (e.g., Backman, 1978). As in the case of intonation perception, most production studies have shown that transfer from the L1 can be a major cause of departure from native patterns in the L2, though its influence varies according to target items and level of L2 competence.

When it comes to the role of proficiency, some of these studies have shown no development in intonation as proficiency increases (Grosser, 1997; Lepetit, 1989), though methodological limitations cast doubt on the validity of these findings. Moreover, some studies have found that a very high level of proficiency in an L2 acquired in adulthood produces changes in both L1 and L2 intonational systems which are comparable to those found in bilinguals.

Another important conclusion from production studies deals with the challenge of measuring intonational units. As we have noticed, there seems to be some degree of uncertainty about how intonation is measured and operationalized. Part of this decision will be largely constrained by the theoretical position adopted in the study. For example, a study within the British school may look at the shape of the contour on and after the last stressed syllable (the *nucleus* and the *tail* of the intonation group, using this model's terminology), while a study conducted within the AM model is likely to assign more importance not only to nuclear but also prenuclear material (e.g., Mennen, 2004). Also, the basic differences between these models will lead the former to provide a more holistic description of the tonal

movement (e.g., Lepetit, 1989), while the latter will probably use frequency (Hz) measures of more localized targets (e.g., peaks and valleys associated with the stressed syllable, as in Wennerstrom, 1994).

Nevertheless, theoretical hesitations in L2 intonation production may be a concern of the past. The most recent studies reviewed here followed the AM model (Henriksen et al., 2010; Huang & Jun, 2009; Mennen, 2004; Simonet, 2008, 2011) and they also measured intonational units that stem from this approach (pitch accents, boundary tones, etc.).

The problem of measurement also implies that production studies are usually smaller in scale than perception studies. As has been shown above, production studies tend to use relatively few participants (sometimes as few as four). For perception studies, it is generally less time-consuming to synthesize auditory stimuli, apply it with a large number of participants, and then obtain computer responses that can be easily turned into quantitative data. The analysis of production, on the other hand, usually relies on the visual inspection of pitch tracks, and, even when software can automatically obtain measurements of different prosodic variables, the location of the relevant points or stretches to be measured often needs to be done manually. Furthermore, production studies have to deal with the complication of variation as a consequence of speaking styles. If the researcher decides to use more than one elicitation task (e.g., Wennerstrom, 1994), the results might yield a more complete picture of L2 production, but they will also increase the amount of data and relevance of the analysis. The overall complexity of quantifying the variability in intonation has led some researchers to adopt a more qualitative description of tones (e.g., Huang, & Jun, 2009), or to reduce any quantitative analysis to basic descriptive statistics (e.g., Backman, 1978). Despite these

methodological challenges, production studies are valuable and, as some of the trends above suggest, necessary to understand L2 intonation more fully.

Teaching intonation

The teaching of L2 intonation has attracted the attention of some researchers and instructors who believe that intonation has been a rather neglected aspect in the second and foreign language classroom. Since the beginning of the communicative approach, scholars have called not only for more research into L2 intonation but also for ways to bridge the gap between this type of research and its application to language learning (Lantolf, 1976; León & Martin, 1972). Even though the current study was not conducted in a classroom setting, participants are L2 speakers who learned Spanish primarily in a classroom environment. Therefore it has broader implications for language teaching. Furthermore, despite some methodological issues, current research in this area is substantial and merits discussion. This section thus presents a few examples of such research and a brief critical view of this field.

James (1977) reports on experiences with English-speaking students learning French at Canadian universities. He used a speech visualizer which allowed students to see two melody curves on a screen. The curve at the top of the screen was the intonation to be imitated, while the curve on the bottom half of the screen was generated as participants recorded the target sentence. The underlying assumption was that participants could compare their own melodic curve with the model and adjust their output until they produced a more target-like rendition of the utterance. There were, however, quite a few problems with this approach. First, it implied mere imitation of an L2 sentence, with a primary focus on its melody, but sentences were not processed for meaning. It might be the case, for

example, that participants do not understand particular lexical items or syntactic structures. Second, this mere imitation cast doubt on the possibility of actually learning the pattern, abstracting generalizable rules, and applying them to novel contexts. In its defense, however, I think that the premise of providing a rudimentary type of feedback to students was valid and deserves further investigation.

Later research specifically examined the type of feedback that seems to be more beneficial when learning intonation. de Bot (1983) investigated the effects of the variables *type of feedback* (auditory vs. audiovisual) and *practice time* (45 vs. 90 minutes) on the intonation of 12 English sentences produced by 63 Dutch speakers learning English. The experiment consisted of a pretest-treatment-posttest design. The treatment used a computer program to train participants in the native-like production of intonation for 50 English sentences. In addition, participants were divided into four experimental groups and received feedback on their production, either auditory only (listening to the model sentence and their own production) or audiovisual (listening plus the possibility of seeing the melody curves). For each type of feedback, there were two practice times, either 45 or 90 minutes, for a total of four experimental groups. The results indicated that participants who received instruction and feedback achieved better results in the posttest than the control group, who received no instruction or feedback. Participants who received audiovisual feedback also outperformed the auditory-alone group, with no significant effects for amount of practice. The author concluded that an enhanced motivation caused by the use of technology and the provision of feedback in two sensory modalities, rather than one, was probably responsible for the benefits observed in the audiovisual group. I agree with the author that it is debatable how

these findings can be extrapolated to L2 classrooms throughout the world, especially in cases where the use of technology is not always available or not easily incorporated into traditional teaching curricula.

Traditionally, helping students achieve a more target-like intonation—through explicit teaching or through more implicit methods—has been largely neglected in the foreign and second language classroom (Chun, 2002). Furthermore, the way students perceive intonation will play a major role in the success of any pedagogical intervention. Unfortunately, some students will not see the potential benefits of making extra effort to improve their intonation. As stated by de Bot and Mailfert (1982), “teachers in these kinds of situations have to work out marketing strategies for selling intonation” (p. 76).

In addition, when intonation is taught, teachers’ knowledge about the subject matter is sometimes insufficient. Teaching materials do not necessarily improve this deficiency. For Levis (1999), the teaching of intonation in North America has failed to incorporate the recent theoretical advances in our understanding of intonation and thus current teaching practices are “still largely trapped in another era,” where the communicative power of intonation “is shackled by outdated descriptions and teaching practices that comfortably exist in the absence of communicative purpose” (p. 60). Levis provided a detailed and interesting discussion of the theoretical assumptions that the L2 teaching practice had generally overlooked or misinterpreted. It is worth noticing that, though not explicitly stated, much of the theoretical proposals that Levis made came from the AM model.

The lack of connection between research and practice often leads to serious mischaracterizations of intonation, as illustrated by the widespread misconceptions that exist

in teaching materials. For example, students of English are often told that pronominal questions and statements are uttered with falling intonation while yes/no questions carry a rising intonation. Research has shown that this is far from a valid generalization. For example, yes/no question in English can have virtually any intonational pattern (Bolinger, 1978; Gutknecht, 1979), probably because the syntax already marks them as a question.

This section provided a general picture of the needs and weaknesses in the field of L2 intonation teaching. The lack of a connection between theory and practice might be due to a ramification of theoretical positions in primary research or, more seriously, to the fact that instructors may not consider the teaching of intonation an endeavor worth undertaking. One hopes that with the recent growth of L2 research on intonation and the consolidation of theoretical frameworks such as the AM model, the findings will inform and benefit the teaching practice.

Summary and conclusion

Both perception and production studies reviewed in this section seem to suffer from certain methodological limitations. Taken separately, they could provide a partial view of issues pertaining to L2 intonation acquisition and use. Comparing existing studies from both research strands can be highly problematic since the theoretical foundations, definitions of intonation, measurements of intonation, participants' characteristics, and the languages involved (both L1 and L2) vary considerably from study to study. Furthermore, the amount of research in the field is relatively small to begin with, which complicates the possibility of pinpointing major trends and making generalizations. I believe we can overcome this hurdle

by designing studies whose research questions address both perception and production simultaneously, under comparable methodological designs.

As regards the theoretical basis supporting L2 intonation research, the present review has shown that focusing on the units of intonational analysis proposed in the AM model seems to yield an adequate level of description and explanation (though much less prediction) of issues in L2 intonation. This appears to be the case in the field of intonation production, while perception still lacks a dominant theoretical approach, with the concomitant difficulty of generalizing findings from this area.

Nonetheless, research from both areas has identified some crucial trends: the positive and negative effects of the L1, universal processes that run across L2 development and seem to be language-independent, and the relative difficulty of learning certain units of intonation, probably due to markedness or salience. These trends need to be confirmed and further explored by looking at underrepresented L2 populations (as is the case of Spanish).

Another major area worth highlighting is the role of proficiency in L2 intonation learning. As informative as they are for some aspects of L2 learning, studies using participants naïve in the L2 or a cross-sectional design with only one proficiency group say little about how intonation develops or how this development potentially interacts with other factors known to affect L2 intonation. In short, future studies need to look at proficiency as a critical variable in understanding L2 intonation.

Research questions and hypotheses

Since we possess almost no knowledge about L2 Spanish intonation, the research questions for this study were broad in scope. They did not seek to determine so much *if* L2

learners produce a certain intonational pattern or learn a target form under certain experimental conditions but, rather, *what* learners produce. With this in mind and considering previous findings on L2 intonation, the following research questions and hypotheses were formulated.

1. How do English-speaking learners of Spanish perceive the intonation of Spanish declarative utterances?
 - a. How does their performance compare to that of monolingual native speakers of Spanish, monolingual native speakers of English, and early Spanish-English bilinguals?
2. How do English-speaking learners of Spanish produce the intonation of Spanish declarative utterances?
 - a. How does their performance compare to that of monolingual native speakers of Spanish, monolingual native speakers of English, and early Spanish-English bilinguals?
 - b. How does L2 learners' production of intonation vary according to speaking style?

Hypotheses for RQs 1 and 2:

End tones might be relatively easier to produce and perceive, perhaps due to their salience in the intonational phrase (Broselow, Hurtig, & Ringen, 1987; Chen, 2009). However, transfer from English, and in particular American English, where declaratives may have a rising intonation, may intervene negatively, especially at lower proficiency levels. I also predict that alignment of prenuclear pitch accents will

be relatively difficult to perceive and produce, with proficiency making little or no difference (Henriksen et al., 2010; Mennen, 2004). In terms of differences in proficiency, some studies have found an improvement in intonation as learners' proficiency increases (Backman, 1978; Henriksen et al., 2010; Jun & Oh, 2000; Nibert, 2005, 2006), others have found no effect for proficiency (Lepetit, 1989), and other studies suggest that proficiency and performance may be negatively correlated (Chen, 2009; Grosser, 1997) . Given these conflicting results, I do not have specific predictions for this variable.

3. Is there a relationship between L2 learners' perception and production of intonation? If so, what is the nature of this relationship?

Hypothesis for RQ 3:

Even though research into the connection between production and perception of L2 intonation is practically nonexistent, I do expect to find a strong relationship between both areas. This is basically supported by the assumption that learners share one intonational grammar (or system) that governs both areas, as supported by a listener's model of intonation (Hermes, 2006; 't Hart, Collier, & Cohen, 1990). From this perspective, I expect patterns of transfer, use, and development in one area to account for patterns in the other.

Chapter 3: Perception

Overview

The purpose of this chapter is to report the findings of perception data aimed at answering the first research question. Perception was measured using an imitation task directed at the two tonal events of interest within the target form, namely prenuclear peak alignment and height of final boundary tone. This chapter also provides the methodological details that pertain to the study as a whole (both perception and production): description of participants, measurement of proficiency, and materials employed.

Methodology

Participants

L2 learner group

The group of L2 learners was initially comprised of 70 participants, all native speakers of American English. Fifty participants were recruited among students enrolled in third ($n = 26$) or sixth ($n = 24$) semester of Spanish at two universities in the northeastern United States. In addition, 20 participants of higher proficiency were recruited, with the following characteristics: (1) they used Spanish in their daily interactions, (2) their academic and/or work life was conducted in both Spanish and English, and (3) they learned Spanish primarily through formal instruction. In order to maintain some homogeneity in this group, all participants were graduate students enrolled in a Spanish Ph.D. program who also taught Spanish at the undergraduate level. The main reason for selecting graduate students was because their competence in the L2 is usually the highest among L2 speakers who started

learning the L2 after childhood. Although this is usually true for most university teachers, graduate students' age is generally closer to undergraduate students' age than the age of other types of university instructors. Similar populations have been examined successfully in studies of advanced L2 learners (e.g., Moyer, 1999, for L2 phonology).

L2 Proficiency

After initial recruitment, participants' proficiency was ultimately measured with the Elicited Imitation Task (EIT) test (Ortega, 2000, Ortega, Iwashita, Rabie, & Norris, 1999, see Appendix A). The EIT belongs to a family of imitation tests used in SLA research to assess participants' proficiency. It is quick to administer and it relies exclusively on aural stimuli and oral production, which is particularly appealing for studies dealing with spoken language only. It has been reported to correlate strongly with measures of proficiency such as the Simulated Oral Proficiency Interview (Ortega et al., 1999) and has been adapted and used for several languages (e.g., Iwashita, 2006, for Japanese; Tracy-Ventura, McManus, Norris, & Ortega, 2014, for French; and Wu & Ortega, 2013, for Chinese). The test produces a score on a scale from 0 to 120 (see Appendix B for scoring instructions and criteria). Table 1 shows the descriptive statistics for EIT scores for the three initial recruitment groups (third-semester, sixth-semester, and graduate students).

Table 1. *Descriptive Statistics for Results in Elicited Imitation Task after Initial Recruitment*

Group	<i>n</i>	Mean	<i>SD</i>	Range	
				Min.	Max.
Third-semester	26	57.96	12.24	30.00	82.00
Sixth-Semester	24	93.33	11.65	63.00	118.00
Graduate students	20	116.40	3.78	108.00	120.00
Total group	70	86.78	26.18	30.00	120.00

A one-way ANOVA and accompanying post-hoc tests (Bonferroni) demonstrated a significant difference among the EIT scores for the three groups, $F(2, 67) = 188.28, p < .001$ ($\eta^2 = .89$, observed power = 1.00), despite the rather large ranges of scores for the third- and sixth-semester groups. Nonetheless, in order to maintain relative homogeneity in each group, outliers (2 standard deviations above or below the group mean) were eliminated from the study. One subject was eliminated in the third-semester group for this reason, three in the sixth-semester group, and one in the group of graduate students. This process also eliminated the overlap in the ranges between groups (see whiskers in boxplot in Figure 3).

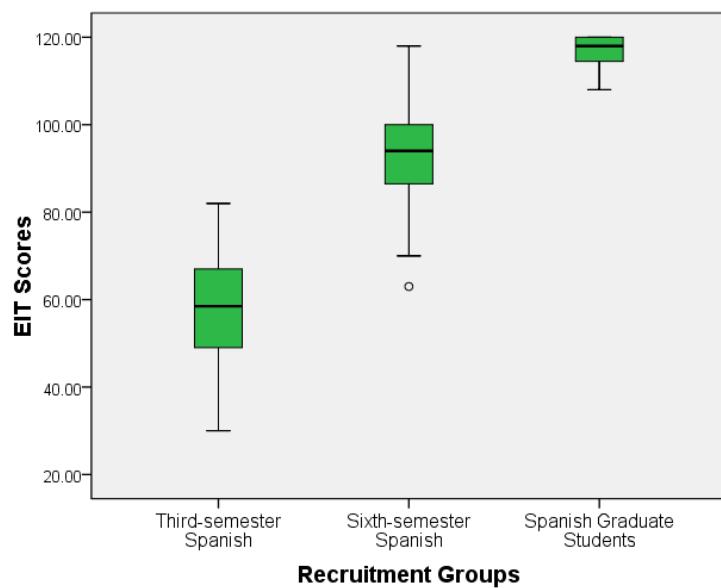


Figure 3. Boxplot for results in Elicited Imitation Task.

Final proficiency was hence operationalized based on a set of two and simultaneous criteria: previous experience and scores in the elicited imitation task. As a result, the following three proficiency levels were established: Low Proficiency (LP, $n = 25$), comprised of subjects enrolled in third-semester Spanish and with EIT scores in the range of 42-82, High proficiency (HP, $n = 21$), comprised of subjects enrolled in sixth-semester Spanish and with EIT scores in the range of 83-107, and Very High Proficiency (VHP, $n = 19$), comprised of graduate students and with EIT scores in the range of 109-120. Ranges include maximum and minimum scores obtained by participants in each group, after outliers were excluded.

In order to control for L2 use, participants were screened for the amount of Spanish spoken outside the classroom, for both exposure and production. In addition, extended stays in a Spanish-speaking country were controlled as well. With this in mind, participants who

studied abroad for more than three months were excluded from the analysis, as shorter periods of time do not appear to translate into significant gains in L2 phonology (e.g., Díaz-Campos, 2004; Mora, 2007), though see Nagle, Morales-Front, Moorman, and Sanz (2015) for results suggesting that a short but intensive study abroad experience does yield certain gains in L2 segmental phonology. Notice that this requirement was only established for the LP and HP groups, since the standard profile of L2 learners in the VHP group typically includes extended visits to Spanish-speaking countries.

All 70 subjects were initially recruited for both the perception and production (see Chapter 4) components of this study. Since one of the research questions sought to investigate the nature of the relationship between perception and production (see Chapter 5), only subjects with complete and usable data in both components were included in the final analysis. Fifteen subjects were eliminated from both parts of the study for one or more of the following reasons: a) after recruitment and upon closer examination, subjects did not meet the criteria for previous language experience, b) more than 50% of the data for any one individual in either production or perception were missing, lost, or deemed poor in quality, c) subjects did not complete the online questionnaires, d) subjects did not meet the parameters for classification into one of the three proficiency groups (see above). The final size for each group was as follows: LP = 17, HP = 20, and VHP = 18. Table 2 contains additional details about the demographics for all groups.

Comparison groups

The perception of L2 speakers was compared with three groups: (1) Spanish native speakers (SNS), (2) English (monolingual) native speakers (ENS), and (3) English-Spanish

bilingual speakers (BS). Initially, 20 participants were recruited for the SNS, 22 for the ENS group, and 20 for the BS group, but due to the same reasons for exclusion (a), (b), and (c) explained above for L2 learners, the data from 12 participants were eliminated from the study. The final composition was as follows: SNS = 17, ENS = 17, and BS = 16.

Monolingual speakers of Spanish represented the following dialectal areas: Southern Cone ($n = 4$), Andean ($n = 4$), Mexican ($n = 3$), Central American ($n = 3$), and Peninsular ($n = 3$). Monolingual participants from various dialects were selected in order to represent as faithfully as possible the variety of dialects that L2 learners—in the U.S. in general and in this study in particular—are usually exposed to. Within the Southern Cone, speakers of *rioplatense* (or *porteño*) Spanish were not included, since this dialect has been reported to lack the typical prenuclear delayed peak and its nuclear fall also differs substantially from most other varieties (Colantoni & Gurlekian, 2004). Nine speakers were born and had lived permanently in their respective dialectal areas, while eight had been living in the U.S. for less than three months and reported little to no knowledge or use of English.

Monolingual speakers of English were recruited from undergraduate classes at the same two universities that L2 speakers attended. They were all monolingual native speakers of American English. No other language was spoken at home while growing up.

Balanced bilingual speakers were also recruited among students at the same universities. The inclusion of this group attempted to expand the comparison standards beyond the typical group of monolingual speakers (see, e.g., Ortega, 2009). The group was comprised of heritage speakers, that is, early bilinguals who were raised in families that speak a minority language (Kondo-Brown, 2006). The linguistic competence of this

population usually varies greatly due to their complex cultural, social, and linguistic backgrounds (see Benmamoun et al., 2011, for discussion). For the purposes of the present study, participants who met the following criteria were included in the analysis: (1) those who had been speakers and not only overhearers of Spanish before the age of seven,¹⁰ (2) those who had not acquired the *rioplatense* variety of Spanish (see above), and (3) those who were usually perceived as being native speakers of Spanish, even if their interlocutors could not pinpoint a specific dialect in their speech (it has been noticed that heritage speakers tend to level dialectal differences in pronunciation, e.g., Godson, 2004). This information was self-reported by the participants (in background questionnaire) and substantiated by an interview in Spanish with the researcher.

General considerations for all participants

Second language and bilingual speakers did not report high proficiency in any language other than English or Spanish. Subjects who declared substantial musical training beyond basic instruction at school were not included in the study since previous research on tone perception has shown that subjects with significant musical background outperform musically naïve participants (Alexander, Wong, & Bradlow, 2005; Burnham & Brooker, 2002; Dankovičová, House, Crooks, & Jones, 2007). All participants reported normal or corrected hearing and vision. Since L2 learners were university students and had thus completed secondary-level education, all other groups were required to meet the same educational level. Therefore, all groups had achieved secondary-level education, either in

¹⁰ Previous studies have demonstrated that adults who had only heard the language during childhood have a less native-like accent than those who also spoke it (Au, Knightly, Jun, & Oh, 2002; Au, Oh, Knightly, Jun, & Romo, 2008).

the United States (L2 learners, monolingual English speakers, and bilingual speakers) or in Spanish-speaking countries (monolingual Spanish speakers). Table 2 summarizes the size and main demographic characteristics of all groups.

Table 2. Group Size and Main Demographics

Group	n	Male – Female	Mean age (range)
LP	17	11 – 6	19.23 (18 – 21)
HP	20	8 – 12	20.40 (18 – 22)
VHP	18	8 – 10	27.39 (24 – 34)
SNS	17	7 – 10	28.41 (20 – 38)
ENS	17	8 – 9	20.75 (18 – 22)
BS	16	6 – 10	21.12 (18 – 25)
Total	105	48 – 57	22.87 (18 – 38)

Note. LP = Low Proficiency, HP = High Proficiency, VHP = Very High Proficiency, SNS = Spanish Native Speakers, ENS = English Native Speakers, BS = Bilingual Speakers.

Materials

Language background questionnaire

The language background questionnaire was designed to gather general demographic information about participants in all six groups (see Appendix C). Some sections targeted specific groups in order to determine whether they met the necessary criteria for each group. In order to facilitate data collection and subsequent analysis, the questionnaire was administered electronically using the online application Survey Monkey (www.surveymonkey.com).

Learning context questionnaire

A key feature of this study is that L2 speakers were tested primarily for Spanish acquired in the classroom setting. This means that participants' input included accented speech from peers and, in some cases, teachers who were native speakers of the language. Even though this complicated the analysis because it involved an additional variable to control, it did not necessarily pose an inherent limitation. Rather, it provided an opportunity to investigate the type of language students acquire when oral input from native speakers is scarce. This is, after all, the reality for a vast number of foreign language students in the United States. With this in mind, the three groups of L2 learners completed a learning context questionnaire to probe more deeply into the type of Spanish to which they had been exposed (see Appendix D).

Debriefing questionnaire

After completing the task, participants answered a short paper-and-pencil debriefing questionnaire for the imitation task (see Appendix E). Participants were informed about some of the research goals and asked about perceived difficulty of tasks and level of involvement, tiredness, and boredom while performing the tasks.

Tasks

Perception of prenuclear alignment and boundary tone was examined through a categorical perception (CP) test. Compared with CP of segments, research on CP of intonation is scant (Gussenhoven, 2004), and almost non-existent in the case of Spanish intonation. Given the paucity of studies that have looked at CP for intonation, scholars have tried different methods that trigger the clearest results in terms of either presence or absence

of CP effects. This section explains the creation of the stimuli, while the Procedure section will provide details about the CP task that was finally selected after piloting different possibilities.

The declarative sentence *La nena lloraba* ‘The girl was crying’ was used as the source utterance to create the stimuli. This sentence was selected for its simple lexical composition, which would make it appropriate for both monolingual and L2 speakers. Additionally, even though segmental accuracy was not within the scope of this study, the sentence did not pose a significant pronunciation challenge for L2 speakers, particularly considering that they would hear it multiple times and have opportunities to practice it before being asked to say it out loud. The following phonotactic and segmental features were taken into consideration when choosing the sentence: (1) all consonants are voiced, which is an important requirement to obtain an uninterrupted pitch track (2) only mid and low vowels (that is, [e, o, a]) are used (high vowels are known to produce small pitch changes by raising the larynx, and thus can potentially distort the pitch curve, Gussenhoven, 2004), (3) syllables have a simple onset and no coda, in order to avoid consonant clusters that could potentially pose a challenge for L2 speakers, while also maximizing sonority and thus improving pitch detection and (4) content words are paroxytone, which is the unmarked stress pattern in Spanish for words ending in a vowel.

The sentence was recorded by the researcher (native speaker of Western Argentinian Spanish), using a high-sensitivity microphone attached to a personal computer and the acoustic software Praat (Boersma & Weenink, 2010, version 5.1.25), at a sampling rate of 44 kHz. The file was resynthesized using the pitch-synchronous overlap and add (PSOLA)

method included in Praat. This technique allows for stylization of the pitch track by reducing it to critical pitch points while keeping the overall shape of the curve. The resulting curve still needs to be inspected acoustically and auditorily before manipulation begins, but usually no major problems are encountered if the quality of the source sound is good (Grabe et al., 2003).

Manipulation for prenuclear alignment consisted of the following steps (see Figure 4): (1) relevant pitch points were reduced to three: beginning of rise (A1), peak (A2), and end of fall (A3); (2) height (in Hz) of A2 was kept as in source utterance (138 Hz), while the timing (in ms) was displaced to make it equidistant from A1 and A3; (3) height and timing of A1 (118 Hz) were kept as in original, while height of A3 (114 Hz) was raised to match A1;¹¹ (4) the three relevant points were displaced eight times to the left of the original and once to the right, in 25-ms increments, creating 10 stimuli spanning over 225 ms. Point A1 of the left-most stimulus coincided roughly with the beginning of the sound [a], while the peak (A1) matched the segmental boundary between [n] and [e]. Similar studies on pitch alignment have used diverse durations for the increments in stimuli: 15 ms (D'Imperio et al., 2010), 20 ms (Pierrehumbert & Steele, 1989), 25 ms (Redi, 2003), and 35 ms (D'Imperio & House, 1997). Unfortunately, seldom have authors provided a justification for their decisions. In the current study, since the segmental stretch under examination (see above) spanned over roughly 230 ms, the authored considered that 10 stimuli in increments of 25 ms was a manageable number of stimuli and an increment duration that fell within parameters used in previous research.

¹¹ Manipulation in steps (2) and (3) resulted in equal absolute values for slopes of fall and rise.

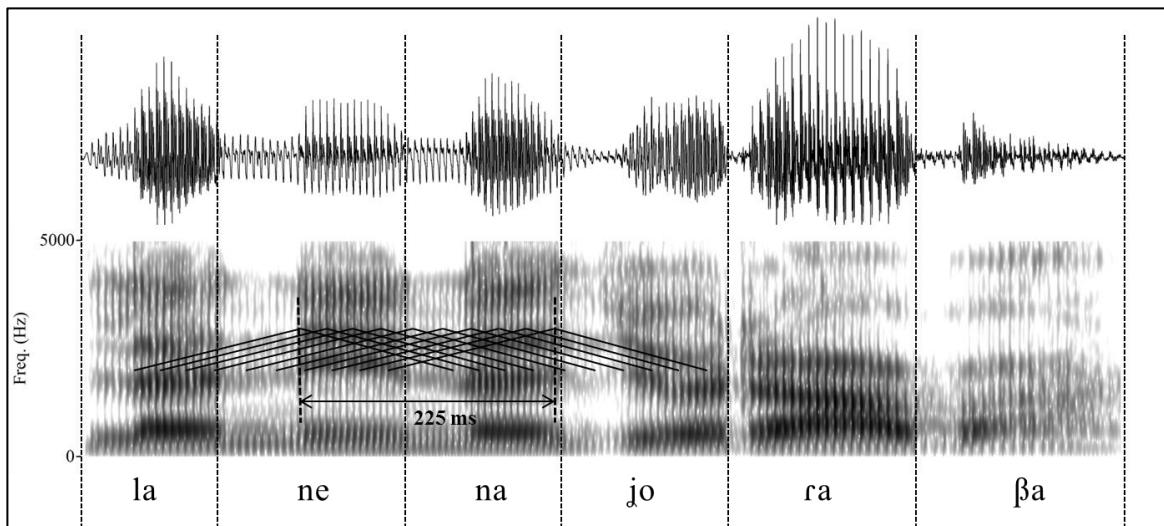


Figure 4. Stimuli for prenuclear alignment.

Manipulation for final boundary tone followed these steps (see Figure 5): (1) relevant pitch points were reduced to two: beginning of rise (B1) and end of the rise (B2), (2) B1 appeared in the middle of [a] (defined as the point with widest amplitude in waveform) in the syllable *ra* and its height (117 Hz) and timing were kept as in source utterance, (3) the temporal point for B2 was anchored at the last regular glottal pulse detected in the spectrogram, therefore excluding visible and audible creaky phonation at the end of the utterance. In terms of pitch height, B2 was moved upwards seven times and downwards twice, in 10-Hz increments, creating 10 stimuli spanning over 90 Hz.

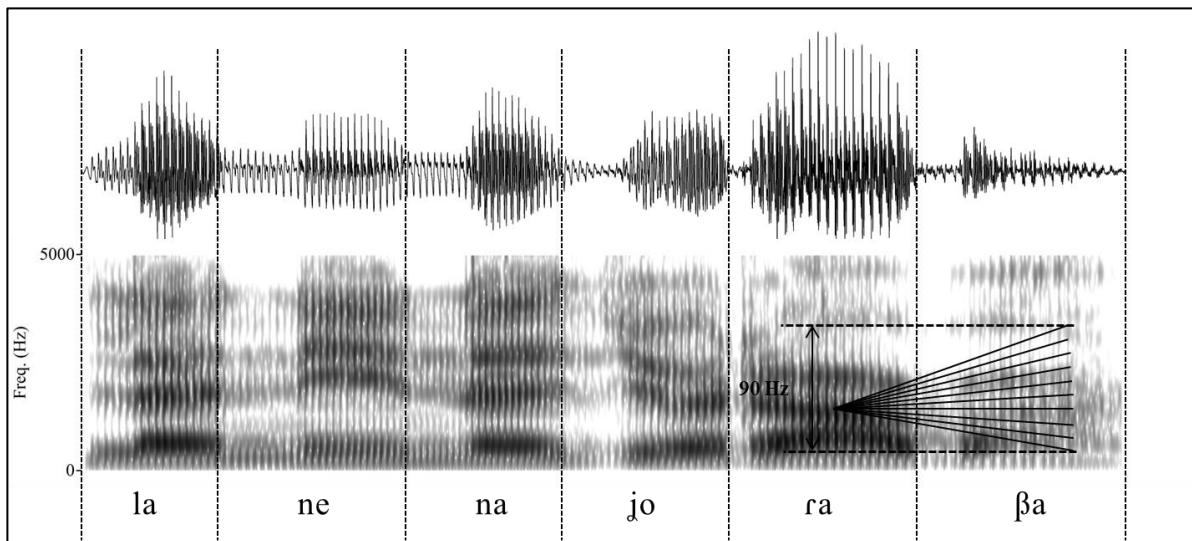


Figure 5. Stimuli for final boundary tone.

In total, 20 stimuli (10 for prenuclear alignment and 10 for boundary tone) were resynthesized, all of which were saved individually as WAV sound files, downsampled to 22 kHz, which was deemed sufficient for this type of perception task. Two native speakers of Spanish were asked to listen to all stimuli to check for naturalness. For this process, utterances were presented in random order and mixed with distractors that clearly sounded as if produced by a machine (15% of the stimuli). All stimuli were judged as natural-sounding.¹²

Procedure

As previously stated, scholars have shown a great deal of hesitation as to what tasks seem better suited to examine categorical perception for intonation. A *discrimination* task

¹² An initial round of stimuli contained 12 boundary tones (instead of 10), but the two highest ones were judged unnatural by at least one of the judges. The very strict manipulation and resynthesis criteria left little room to fix each stimulus without altering all others as well. Therefore, the researcher eliminated these two stimuli altogether. In any case, shifts in perception noticed during pilot testing seem to occur in the middle range of the stimuli, which also justified the elimination of these two highest points.

has probably been the canonical way to test categorical versus gradient perception. It was used extensively in early research conducted at the Haskins Laboratories in Connecticut, USA (e.g., Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Harris, Kinney, & Lane, 1961). This task involves asking subjects to judge whether a pair of stimuli adjacent in a continuum are the same or different. This technique has been applied to intonation (e.g., Ladd & Morton, 1997) as well, but has not yielded satisfactory results (see Gussenhoven, 2006, for discussion). Also, in particular at the segmental level, some authors have shown that participants in an identification task are frequently biased towards one of the answers (either *same* or *different*) and this greatly skews results and yields categorical effects where other tasks show such effects do not hold (Gerrits & Schouten, 2004; Schouten, Gerrits, & van Hessen, 2003). In an *identification* task, subjects listen to one stimulus at a time and are asked to label it. In studies that deal with voice onset time (VOT), for instance, subjects need to decide whether the stimulus approximates more closely English [p] or [b], for example, which are easily labeled with the corresponding letters *p* and *b*, respectively. In the case of melodic contours, however, it is not always easy to assign labels that will make sense to participants, as in, for example, tests of narrow versus broad focus (although others differences are certainly more intuitive, like questions versus statements, as in D'Impeiro & House, 1997). In order to steer away from labels, intonational research on CP has sometimes used an ABX task as an alternative to the traditional identification task format. Participants in this task are presented with the stimuli at both ends of the continuum (stimuli A and B) and then with one from the middle (stimulus X). Participants are asked to judge if stimulus X is more similar to A or to B (e.g., Barnes, Veilleux, Brugos, & Shattuck-Hufnagel, 2010).

One potential problem with this task stems from short-term memory limitations: subjects need to make at least two comparisons (A and X and B and X) before they make a final judgment, and by the time subjects hear stimulus X, they may not recall what stimulus A sounded like. This is particularly true of intonation research, where stimuli are usually stretches of language that span over a few syllables, if not complete sentences.

As a viable solution to overcome these problems, more recent studies have used an *imitation* task, where participants listen to one stimulus at a time and have to reproduce it out loud. Some prominent scholars have argued it is the best way to examine categoriality in intonation (e.g., Gussenhoven, 1999, 2004). Pierrehumbert and Steele (1989) were among the first to use this task. If there is a categorical distinction, participants will not reproduce the entire continuum they hear but, rather, utterances tend to group in a bimodal distribution. The technique has been employed to investigate pitch alignment (e.g., Pierrehumbert & Steele, 1989; Redi, 2003) and pitch height (e.g., Dilley, 2005; Dilley & Brown, 2007). The underlying assumption is that a purported perceptual system of intonation will restrict various renditions of intonation (the stimuli) to the unit or units that are linguistically meaningful. When hearers are asked to repeat the utterances they hear, production thus reveals what these units—if any—are. In the words of D’Imperio, Cavone, and Petrone (2014), “when reproducing randomly variable intonation contours, speakers appear to be able to extract linguistically meaningful (i.e., phonological) contrast by concentrating their productions around a limited number of prototypical patterns or ‘attractors.’” In other words, [...] speakers can only reproduce $f0$ contours which are grammatically meaningful, just as in segmental imitation speakers might only be able to imitate phonemes of their own language”

(p. 2). If production yields a bimodal distribution (e.g., Pierrehumbert & Steele, 1989), results have been normally interpreted as indicative of categorical shifts in perception. In regards to purported advantages of the imitation task over other perception tasks, it has been argued that, since subjects have to listen to only one stimulus before they are asked to repeat it, this task does not present the type of challenges for memory that discrimination or identification tasks do (e.g., Gussenhoven, 2006).

Consequently, in view of potential shortcomings associated with more traditional perception tasks and recent evidence that suggests imitation tasks are better suited to capture categorical effects in intonation, the researcher decided to adopt the latter in this present study. Participants were presented with the 10 prenuclear alignment stimuli twice, in two randomized blocks. Two additional blocks contained the 10 boundary tone stimuli, for a total of 40 stimuli to be repeated by each subject. Participants were instructed to reproduce as faithfully as possible the utterance they heard (see Appendix F with instructions). Since the stimulus was always the same, no beep sound was necessary to alert participants that they were ready to repeat. They were asked to focus on the pronunciation of the sentence and encouraged to imitate it within a comfortable pitch range (D'Imperio et al., 2014; German, 2012; Redi, 2003). Participants first listened to a block of five practice utterances, which could be repeated until participants felt comfortable with the task. Two of the five practice utterances were drawn from the stimuli, while the others were unrelated declarative utterances, similar in length (Redi, 2003). For the trial blocks, participants had the possibility of saying the utterance again if they hesitated, paused, or considered that their output was not faithful to the stimulus they had heard. The stimuli were delivered over high-

fidelity headphones and presented through E-prime software. Recordings were made in a sound-proof room, with a high-sensitivity, head-mounted microphone attached to a personal computer and using the Audacity software (Audacity Team, 2011, version 1.3.14), at a sampling rate of 32 kHz, which provided sufficient quality for acoustic analysis while keeping digital storage at a reasonable size. Finally, when completing the debriefing questionnaire, participants were asked to explain how they thought the stimuli they heard were different from one another.¹³

Changes after pilot study

As regards participants, the initial stages of piloting also included participants with basic proficiency in Spanish (only one semester of college Spanish), in an attempt to capture a more comprehensive picture of language development. In the imitation and production tasks (see Chapter 4), however, they produced highly disfluent speech (extended hesitations and pauses, unfinished utterances, and false starts) even when provided with opportunities to repeat and amend their production. As a result, the researcher decided not to recruit participants from this level for any parts of the current study.

In terms of stimuli manipulation, prenuclear alignment was originally tested with both plateaus and peaks. Plateaus are sometimes preferred in speech resynthesis since they usually correspond more closely to the source utterance, where pitch maxima appear flanked by smooth elbows that transition into slopes. Regardless of the exact shape, pitch tracks of naturally-occurring language rarely have an inverted V shape like the one in Figure 4.

¹³ Some participants provided highly unexpected answers, especially in the LP or HP groups. For example, three participants thought alignment stimuli were different in terms of duration, that is, some utterances were “said more slowly,” even though the stimuli were of the exact same duration.

Considering the methodological requirements for categorical perception, however, plateaus raise the problem of locating category shifts with precision. Approaches that consider a tonal center of gravity would predict that the shift occurs in the middle of the plateau (e.g., Veilleux, Barnes, Shattuck-Hufnagel, & Brugos, 2009). However, a more recent study (D'Imperio et al., 2010) examined this issue in German and two varieties of Italian (Naples and Pisa) by comparing alignment of peaks and plateaus. The results indicated that categorical shifts in both languages tended to occur toward the plateau offset rather than in the middle or at the onset. The results of the pilot for this present study did not produce sufficient data to make conclusive comparisons as in D'Imperio et al. (2010) but, in general, sharp peaks did tend to produce more categorical-like responses. Therefore, the researcher decided to utilize peaks rather than plateaus. The rise, peak, and fall of the resynthesized stimuli respected the overall shape of the original stylized contour, though rises and falls were slightly adjusted to render the same rising and falling slopes in order to exclude slope differences as a potential confounding factor. For the same reason, preceding and following valleys were kept at the same height. Similar stylization and manipulation decisions are customary in research on perception of tonal alignment (D'Imperio & House, 1997; D'Imperio et al., 2010; Pierrehumbert & Steele, 1989).

In the case of boundary tone, a different set of stimuli was originally tested as well. The increments in upper frequencies were made increasingly larger, since it has been noted that sometimes the higher the f_0 values for two points are, the wider the pitch span that is needed to perceive the difference between these points (e.g., Chen, Gussenhoven, & Rietveld, 2004). Five participants were tested with these stimuli and the results, even though

insufficient to conduct inferential statistics, did not produce any differences with the stimuli containing equal f_0 increments.

Regarding the imitation task, Dilley (2010) noted that some subjects are just poor imitators, which was the case among participants in the pilot for this present study. Toward the end of pilot testing, the researcher decided to increase the number of practice tokens, which seemed to help participants understand the goal of the task and perform it more naturally.

Data analysis

The quality of the 4,200 data points (40 utterances x 105 subjects) was both auditorily evaluated and visually inspected using spectrograms and intonational curves produced by Praat. Excessive creaky voice, hesitations, and uncommonly flat global pitch contours were the main reasons for excluding data points. For these reasons, 353 utterances were eliminated from the analysis of prenuclear alignment while 555 were removed from the analysis of boundary tone. In the latter case, final creaky, especially among male speakers, voicing was the main reason for elimination. Two male speakers were removed from the analysis altogether as more than 50% of their utterances contained excessive creaky voice (see description of participants). In total, 1,747 utterances were ultimately analyzed for prenuclear alignment (279 for LP group, 336 for HP, 304 for VHP, 286 for SNS, 279 for ENS, and 264 for BS) and 1,545 for final boundary tone (249 for LP group, 300 for HP, 276 for VHP, 252 for SNS, 231 for ENS, and 237 for BS).

Analysis of prenuclear peak alignment was carried out by locating f_0 extrema in the segmental stretch comprised between the onset of [e] and the end of [a] in *nena*, since *a*

priori predictions estimated that most if not all peaks, both early- or late-aligned, should fall within these boundaries. This prediction was based on previous studies on Spanish declarative utterances and, in order not to rule out any potential explanations, even conflicting views on the tonal representation of prenuclear tones were considered when selecting these boundaries (Face, 2001; Llisterri et al., 1995; Prieto, 1998; Sosa, 1999). The right edge of [a] could not be clearly located since the following sound—the initial sound in *lloraba*—was normally realized as the approximant [j]. Therefore, the right boundary for the stretch under analysis was redefined as the middle point of [j] (or [j]) and located acoustically at the point of minimum amplitude between the vowels [a] and [o] in [ajo]. These boundaries were manually marked and annotated using a Praat text grid. A special Praat script automatically extracted the time of f_0 maxima occurrence between these boundaries. In order to control for differences in speaking rates among participants, this value was normalized using the formula in (1).

$$(1) \quad T_N = \frac{t_1 - t_0}{d}$$

T_N is the normalized time for f_0 maximum, t_1 is the real time for f_0 maximum as extracted by the Praat script, t_0 is the onset of [e] in *nena*, and d is the total duration of the stretch under analysis, that is, from the onset of [e] to the midpoint of [j] (or [j]). Since all speakers produced t_1 within the boundaries of d , T_N in all cases consisted of a value in the range between 0 and 1.0. Similar normalization procedures have been used in prior studies dealing with location and timing of tonal events (e.g., Redi, 2003; Dilley & Brown, 2007). Figure 6 and Figure 7 show examples of analysis for a participant in the HP group, with a relatively early and late aligned t_1 , respectively.

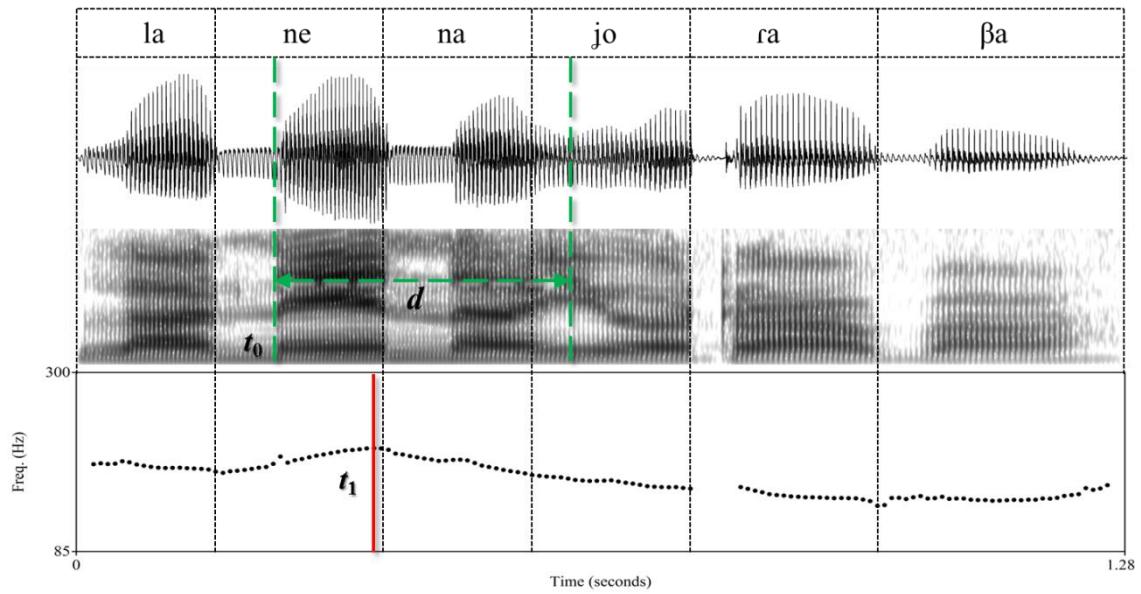


Figure 6. Example of typical response to prenuclear alignment stimuli (early t_1).

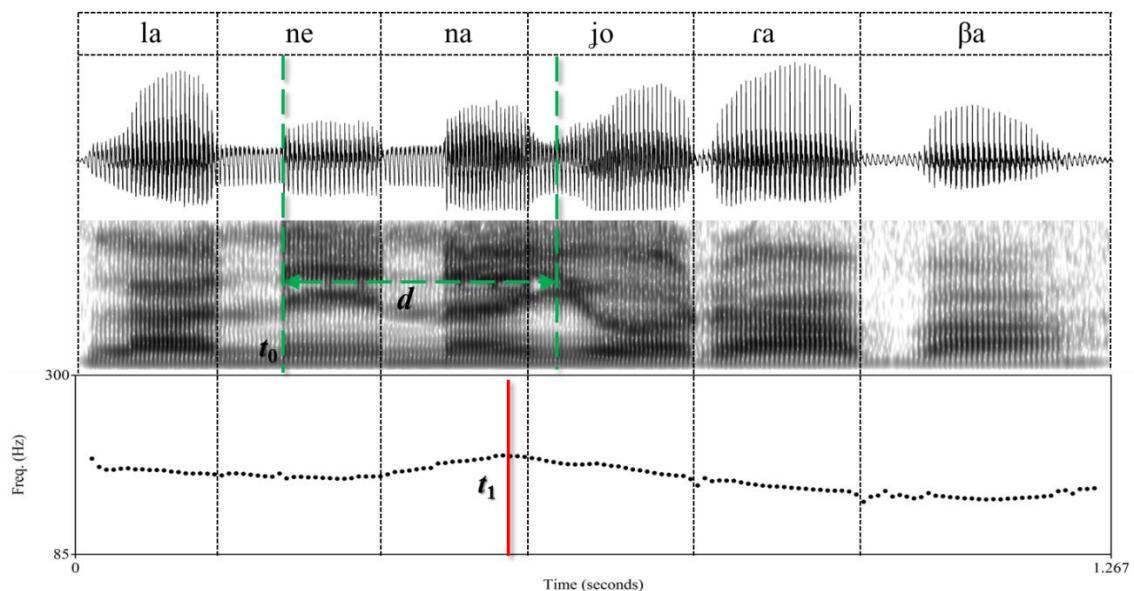


Figure 7. Example of typical response to prenuclear alignment stimuli (late t_1).

Final boundary tone, in turn, was defined as the last non-spurious f_0 point in the pitch track generated by Praat (see Arvaniti & Ladd, 2009; Henriksen, 2014, for similar procedures). If this point fell to the left boundary of syllable *ba* in *lloraba* due to pitch detection errors or excessive creaky voice, utterances were eliminated from the analysis.¹⁴ Twenty-three utterances were eliminated for this reason. Other utterances were excluded for the reasons mentioned earlier, creaky voice—especially among male speakers—being the main one. Extracted f_0 values were converted to Equivalent Rectangular Bandwidth (ERB) units in order to normalize for differences among speakers, in particular between males and females. This normalization procedure describes more accurately the relationship between f_0 and perceived pitch and has been increasingly used in recent research (e.g., Arvaniti & Ladd, 2009; Colantoni & Gurlekian, 2004; Simonet, 2008, 2009).

Results

Results for prenuclear alignment and boundary tone are presented separately. In both cases, individual boxplots were produced for each of the six groups. Potential effects for categorical shifts in perception were tested using one-way ANOVAs and post-hoc tests, with an alpha level set at 0.05 throughout. All graphs and statistical tests were done using the Statistical Package for the Social Sciences (SPSS), Version 22.

¹⁴ For the purpose of this study, the left boundary of this syllable was defined as the beginning of the closure of /b/, realized acoustically as an abrupt drop in amplitude in the waveform if realized as [ba], and as the point of minimum amplitude in the waveform if realized as [βa]. This procedure was done visually and was only needed in 10 cases. In all other cases, it was visually clear in the spectrogram and pitch track that the final non-spurious f_0 point occurred well into the syllable /ba/ or well before its beginning.

Prenuclear alignment

Boxplots (Figure 8 through Figure 13) for each group present the descriptive statistics for the prenuclear alignment perception test. The X axis represents the 10 points in the stimuli (see description of tasks above), while the Y axis represents the normalized time scores for each group (T_N , see description of analysis above).

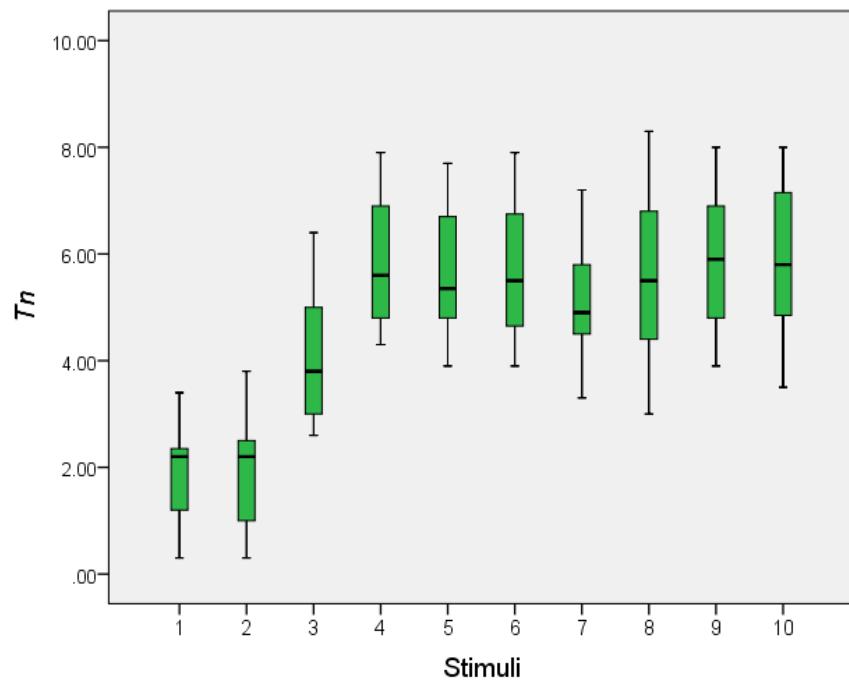


Figure 8. Results of perception of prenuclear alignment for LP group.

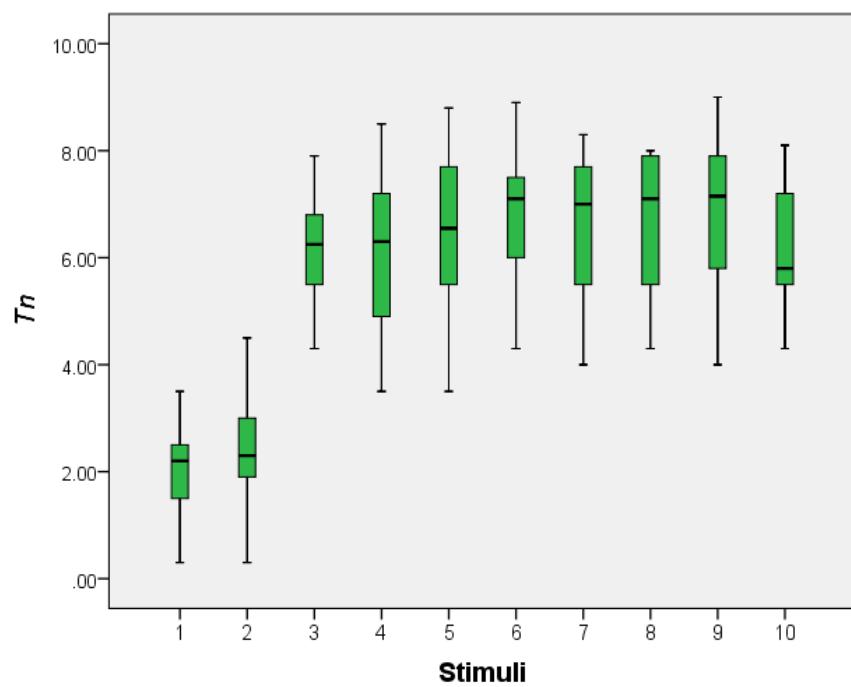


Figure 9. Results of perception of prenuclear alignment for HP group.

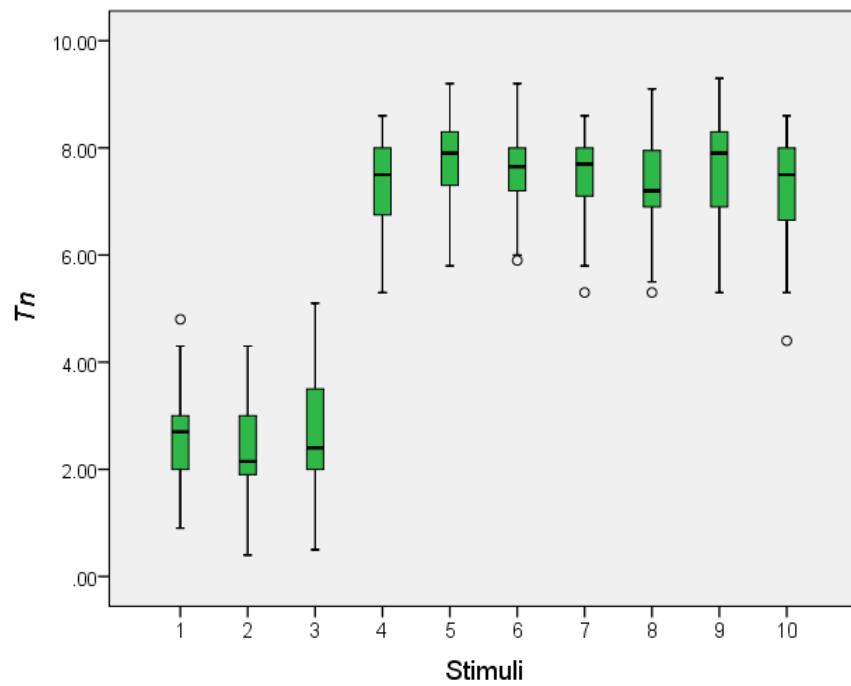


Figure 10. Results of perception of prenuclear alignment for VHP group.

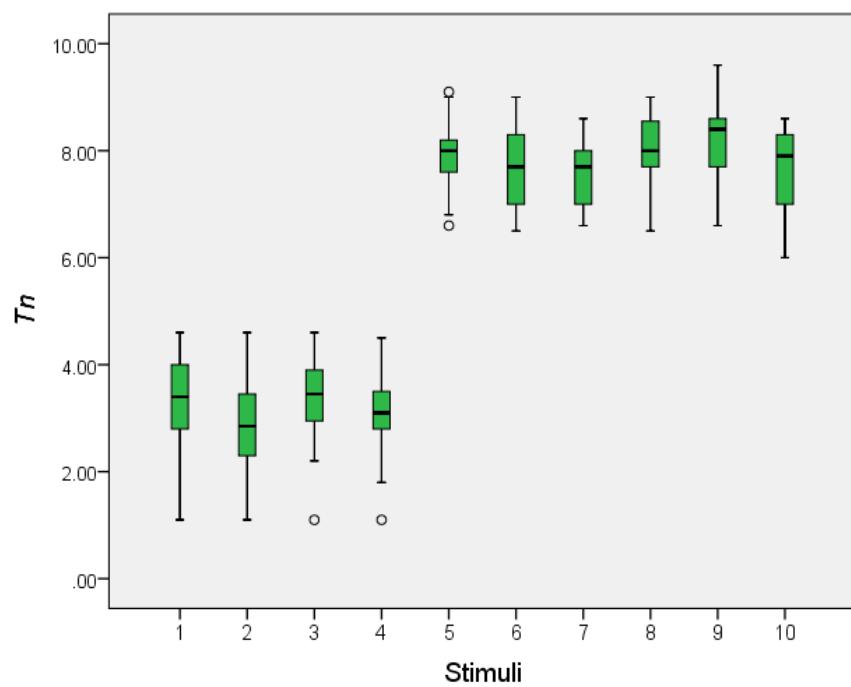


Figure 11. Results of perception of prenuclear alignment for SNS group.

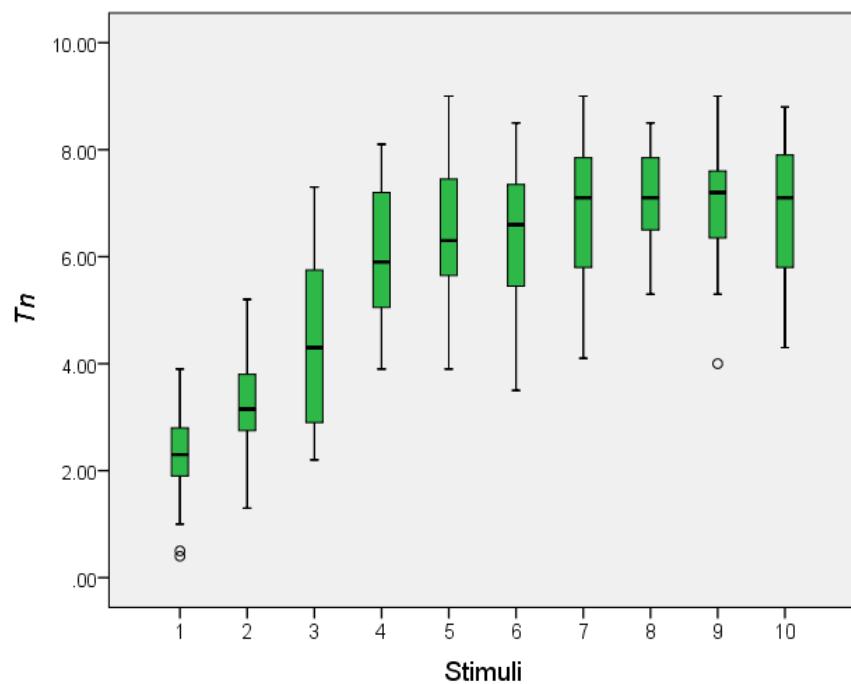


Figure 12. Results of perception of prenuclear alignment for ENS group.

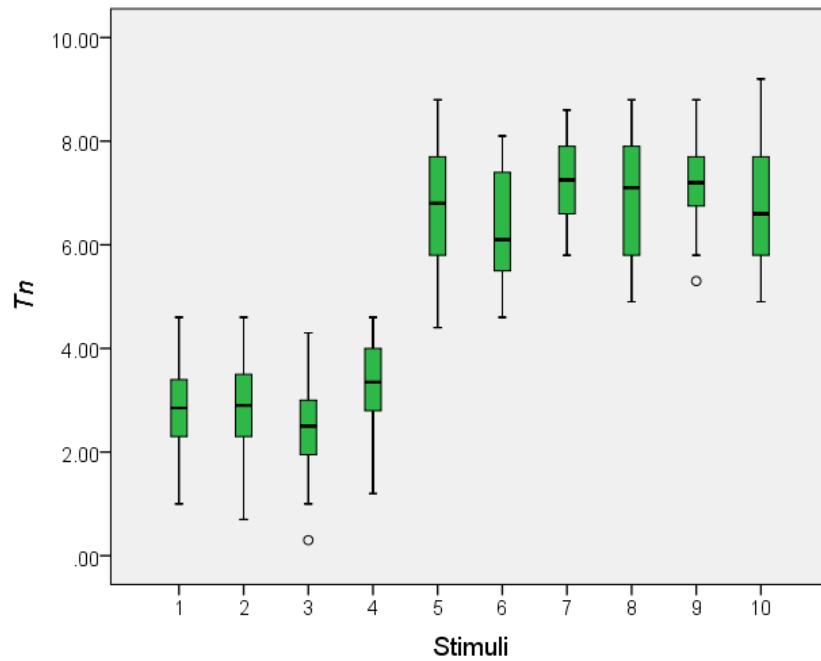


Figure 13. Results of perception of prenuclear alignment for BS group.

All graphs generated some level of clustering of data points around two sets of stimuli, except those for LP and ENS groups. The cut-off line between the clusters, however, seems to vary considerably from one group to the other. These preliminary results show that participants seemed to perceive prenuclear stimuli quite differently. In order to verify the statistical significance of these observations, a one-way omnibus ANOVA was run for each group, with Stimuli as the independent variable and T_N as the dependent variable. Results of one-way ANOVAs, effect size (partial Eta squared, η^2), and observed power are reported in Table 3.

Table 3. One-way ANOVA for Results in Prenuclear Alignment for All Groups

Group	<i>F</i>	<i>df</i>	Sig.	Observed			power
				Partial η^2			
LP	52.49	9 – 269	.000	.64			1.00
HP	84.76	9 – 326	.000	.70			1.00
VHP	168.88	9 – 294	.000	.85			1.00
SNS	257.66	9 – 276	.000	.89			1.00
ENS	62.19	9 – 269	.000	.67			1.00
BS	107.33	9 – 253	.000	.79			1.00

Note. LP = Low Proficiency, HP = High Proficiency, VHP = Very High Proficiency, SNS = Spanish Native Speakers, ENS = English Native Speakers, BS = Bilingual Speakers.

LP	1.91	1.96	3.97	5.82	5.69	5.64	5.25	5.71	5.89	5.96
HP	1.97	2.23	6.14	6.17	6.40	6.65	6.68	6.81	6.87	6.15
VHP	2.72	2.36	2.60	7.29	7.78	7.54	7.40	7.28	7.60	7.22
SNS	3.23	2.90	3.33	3.11	7.89	7.71	7.59	7.93	8.11	7.72
ENS	2.23	3.11	4.38	6.02	6.48	6.37	6.87	6.95	6.96	6.83
BS	2.87	2.93	2.42	3.25	6.70	6.36	7.24	6.91	7.24	6.68
Group	1	2	3	4	5	6	7	8	9	10
Stimuli for prenuclear alignment										

Figure 14. Results for post-hoc comparisons in prenuclear alignment stimuli.

Note. Numbers represent mean normalized times (T_N) produced for each stimulus, by group. Mean T_N enclosed in same color indicate homogenous groups. That is, means in each group are statistically the same (Tukey's test, $p > .05$).

Post-hoc analyses were conducted in order to determine where significant differences actually lay. Due to the relatively high number of groups that needed to be entered for pairwise comparisons, Tukey's test was used. This test controls for Type I errors

well and allows for comparisons among unequal sample sizes by using the harmonic mean of the sample sizes (Howell , 2002; Larson-Hall, 2010). Results of post-hoc analyses are presented graphically in Figure 14 where statistically homogenous stimuli appear in the same color.

Post-hoc results showed that for all but the LP and ENS groups there are clear categorical shifts in perception and that the stimuli were perceived in two distinctive groups. Information from the debriefing questionnaire revealed that, as expected, participants perceived an unmarked, non-emphatic version and an emphatic (narrowly focused) version. They tended to describe the latter as ‘said with more emphasis/accent on *nena*’. The SNS and BS groups had statistically comparable perception patterns, with categorical shifts between stimuli 4 and 5, while the cutoff point for the VHP group appeared between stimuli 3 and 4 and between stimuli 2 and 3 for the HP group.

Boundary tone

The results for boundary tone in this section follow the same organization as those for prenuclear alignment. Boxplots (Figure 15 through Figure 20) for each group present the descriptive statistics of the imitation task. The X axis represents the 10 points in the stimuli, while the Y axis represents the normalized f_0 scaling of final tone, expressed in ERB units (see description of analysis above).

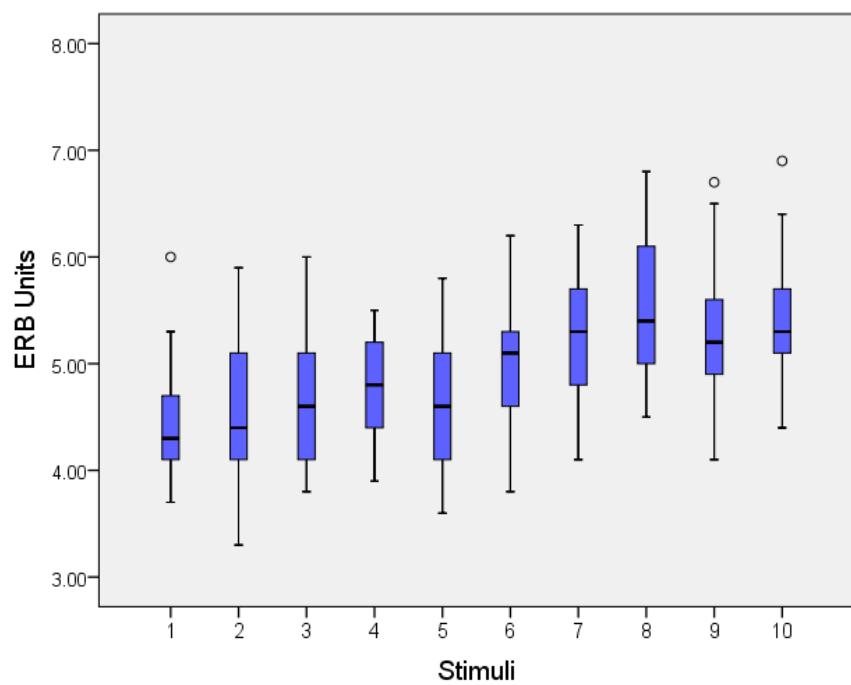


Figure 15. Results of perception of final boundary tone for LP group.

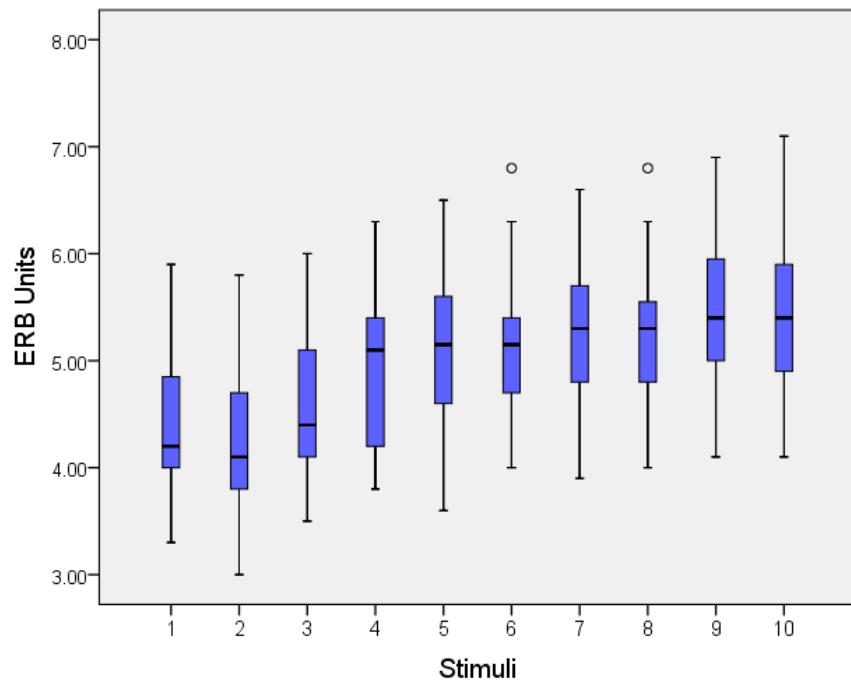


Figure 16. Results of perception of final boundary tone for HP group.

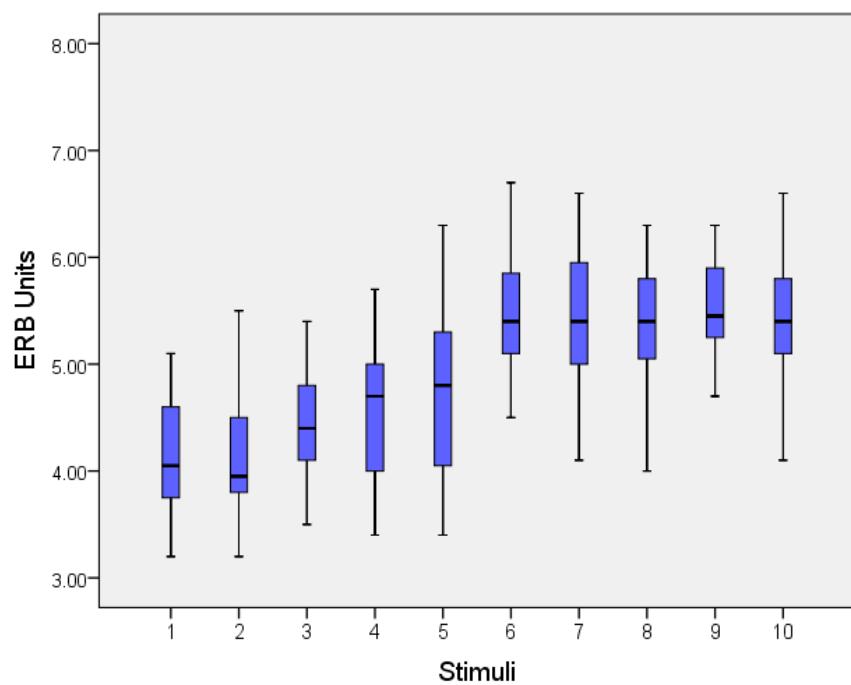


Figure 17. Results of perception of final boundary tone for VHP group.

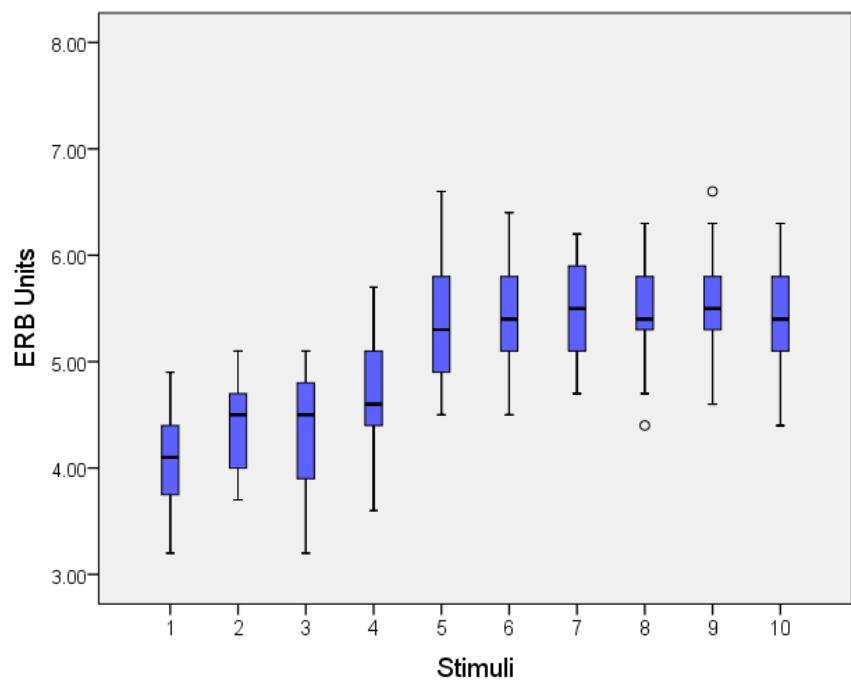


Figure 18. Results of perception of final boundary tone for SNS group.

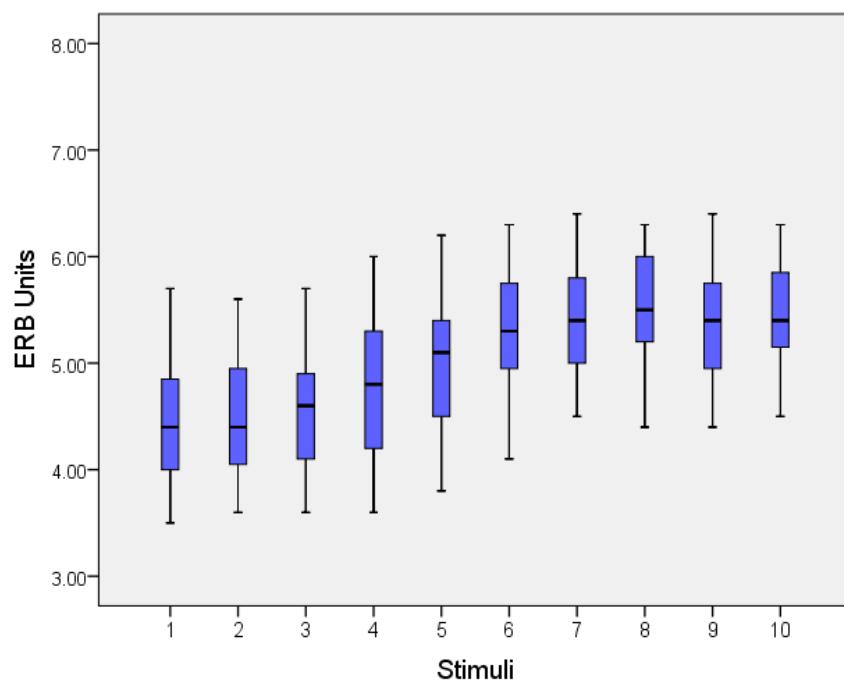


Figure 19. Results of perception of final boundary tone for ENS group.

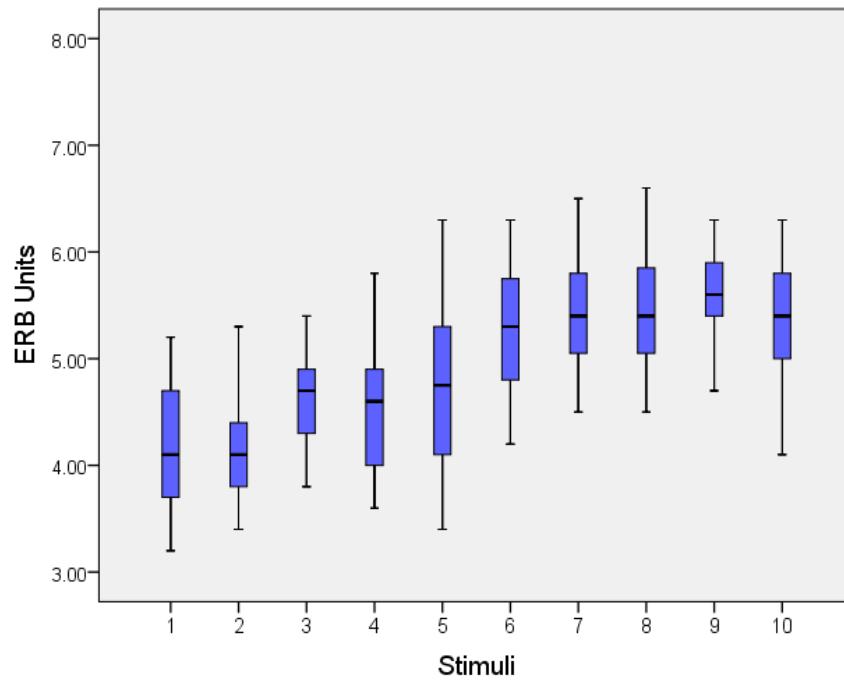


Figure 20. Results of perception of final boundary tone for BS group.

In general terms, boxplots for perception of final boundary tones suggest that participants perceived increments in final f_0 in a gradient manner. That is, in contrast with prenuclear peak alignment, descriptive statistics do not suggest clear categorical differences in the perception of boundary tone stimuli. A one-way ANOVA was run for each group in order to explore further any potential categorical effects in the data. The 10 points in the stimuli served as the independent variable, and ERB units served as the dependent variable. Results of one-way ANOVAs, effect size, and observed power are reported in Table 4.

Table 4. *One-way ANOVA for Results in Final Boundary Tone for All Groups*

Group	<i>F</i>	<i>df</i>	Sig.	Partial η^2	power	Observed
LP	11.09	9 - 239	.000	.29	1.00	
HP	12.48	9 - 290	.000	.28	1.00	
VHP	27.81	9 - 266	.000	.48	1.00	
SNS	30.85	9 - 242	.000	.53	1.00	
ENS	13.65	9 - 221	.000	.36	1.00	
BS	23.34	9 - 227	.000	.48	1.00	

Results of the omnibus one-way ANOVA yielded significant difference in all six cases. Notice, however, the reduced effect sizes, reported in η^2 , for all groups as compared with inferential statistics for prenuclear alignment data. This could be attributed to a smaller sample due to greater elimination of data points. Nonetheless, since all tests found significant differences among stimuli, a post-hoc analysis (Tukey's) was run in order to determine the precise location of these differences. Results are presented graphically in Figure 21.

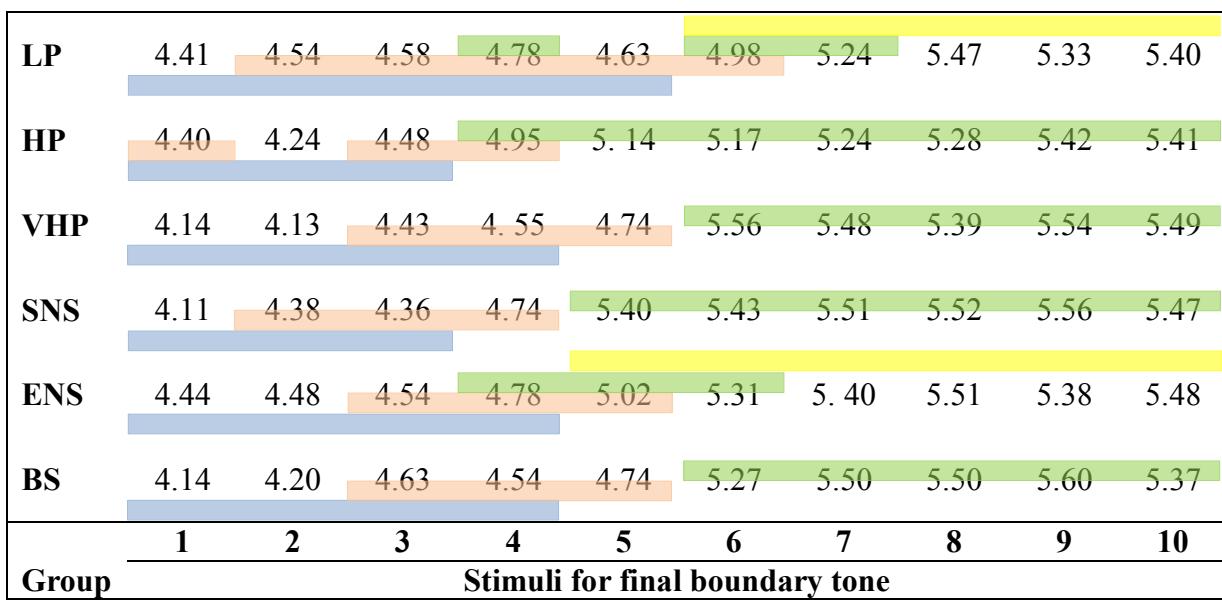


Figure 21. Results for post-hoc comparisons in final boundary tone stimuli.

Note. Numbers represent mean normalized Equivalent Rectangular Bandwidth (ERB) units produced for each stimulus, by group. Mean ERB units enclosed in same color indicate homogenous groups. That is, means in each group are statistically the same (Tukey's test, $p > .05$).

The results of post-hoc comparisons for boundary tone stimuli produced quite a different picture compared with prenuclear alignment. In general, no group achieved categorical effects in the perception of the stimuli as traditionally defined, since there was some degree of overlap in stimuli clustering for all 6 groups. Interestingly, however, if we are willing to consider initial overlapping groups (blue and red in Figure 21) as one, the VHP, SNS, and BS groups did have one clear cut off point between stimuli 4 and 5 for SNS, and between stimuli 5 and 6 for VHP learners and BS. Otherwise, the stimuli for all groups were arranged from falling (non-emphatic declarative) to rising (questions), which corresponded to the declarative versus question interpretations, respectively, that the vast majority of participants reported in the debriefing questionnaire.

Discussion and Conclusion

Prenuclear alignment

Despite the lack of clear categorical effects in the group of English native speakers, the difference with Spanish native speakers in the overall parsing of perceptual stimuli seems to reproduce the differences in prenuclear alignment between English and Spanish. That is, as reviewed in Chapter 2, English tends to align prenuclear pitch accents within the limits of the stressed syllable (H^* , in AM notation¹⁵), while Spanish prefers a late-aligned prenuclear pitch accent (L^*+H).

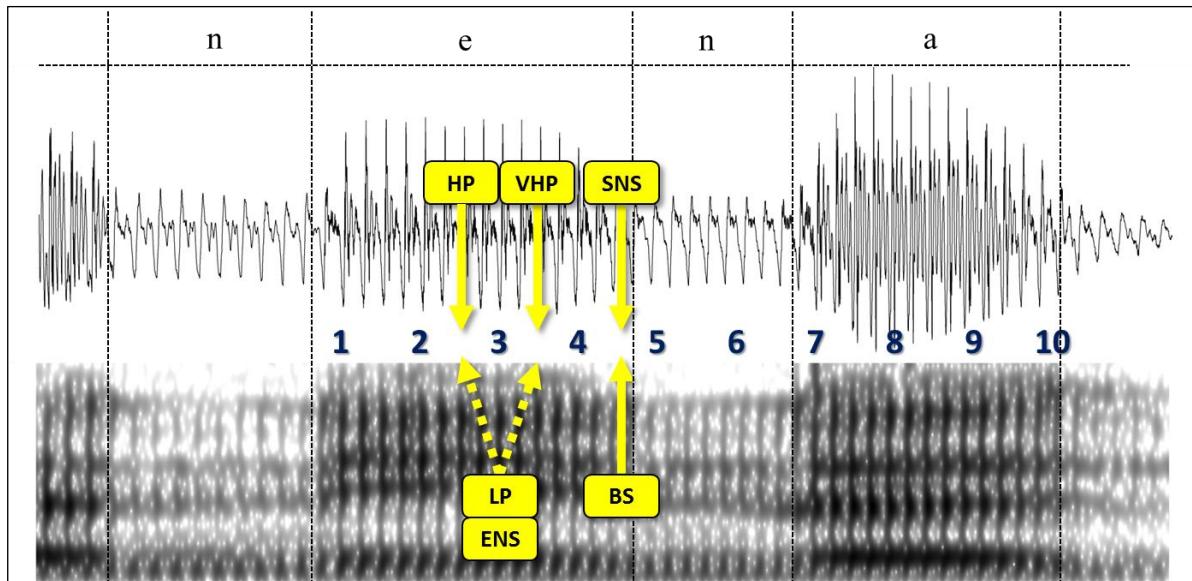


Figure 22. Shifts in perception in relation to segmental material.

Note. Numbers indicate points where pitch peak for each stimulus occurred (see description of tasks). Solid arrows show perception shifts for groups with a single shift location ($p < .001$). Dotted arrows show shifts for groups with multiple shift locations ($p < .001$).

¹⁵ Scholars still debate whether this pitch accent in English is mono- (H^*) or bitonal ($L+H^*$). The leading L tone assumes there is a noticeable valley that precedes the rise and usually occurs in the preceding syllable. This valley, however, is not always easy to detect reliably, which has led some scholars and ToBI transcribers to opt for H^* in neutral declarative utterances (see full discussion in Ladd, 2008). Since the main interest of the current study is whether the peak is aligned with the accented syllable or not, the presence or absence of a leading tone is not relevant at this point.

As shown in Figure 22, the shift in the perception of a Spanish declarative utterance as either emphatic (narrow focus) or non-emphatic occurs precisely at the end of the prenuclear syllable (between points 4 and 5) for Spanish-native and bilingual speakers. For monolingual English speakers, on the other hand, anchoring the f_0 maximum at points 3 and possibly 4 is early enough for the utterance to be perceived as non-emphatic, but too early for Spanish speakers. In other words, English speakers seem to allow prenuclear pitch to be aligned earlier and closer to the midpoint of the stressed syllable (consistent with a H* pitch accent) while Spanish-native and bilingual speakers prefer a later aligned peak (consistent with a L*+H pitch accent).

Although outside the scope of the present study, whose main interest is the examination of second language learners, these findings may provide valuable data to the dilemma of how to treat phonologically prenuclear peaks in Spanish, which most production studies have found to appear in a delayed position, usually at the posttonic syllable. As reviewed in Chapter 2, even though this displacement is uncontroversial in most dialects of Spanish, its phonological representation within an AM model is still debated. Interestingly, the perception data in this study showed that shifts between narrow and broad focus occur inside the tonic syllable, just before the right boundary, and certainly quite far from the midpoint or right boundary of the posttonic syllable, where most production studies have located this delayed peak. Alternative analyses to the L*+H tone have proposed that the prenuclear pitch accent is a monotonal H* (Nibert, 2000; Prieto, 1998), a bitonal L+H* (Face & Prieto, 2006), or even an associated (L+H)* tone for Hualde (2002). In all these proposals, the phonological representation for Spanish prenuclear pitch accents would

consist of an H anchored within the stressed syllable, while the observed delay is a phonetic process, governed by phonetic factors related to timing, such as tonal crowding, avoidance of stress clash, or distance with upcoming prosodic boundaries. The perception data for Spanish speakers in this study seem to support this analysis.

In regards to the group of L2 learners, results show a surprisingly clear development of perception from an L1 state toward a more target-like system. Though not clearly categorical, the results for the low-proficiency group matched exactly those of monolingual English speakers. Let us recall that participants in the LP group had had at least two semesters of Spanish instruction, yet their perception of Spanish intonation was still heavily—if not totally—determined by their L1. Similar stagnant patterns in initial perception of intonation have been found for L2 Spanish (Nibert, 2005, 2006). The two more proficient groups in the current study, however, did progressively depart from the L1 and approached L2-perception, but still fell slightly short of completely native- or bilingual-like performance. Previous studies have indeed found that, compared with other components of L2 intonation, target-like prenuclear alignment appears to be particularly difficult to attain (e.g. Henriksen et al., 2010), perhaps due to its relatively poor perceptual salience in the utterance. In sum, perception of prenuclear pitch accents seems to be heavily influenced by the L1 at initial proficiency. Furthermore, this L1-governed stage appears to hold well into the first stages of instruction in Spanish and probably lags behind the comparatively faster development of segmental L2 phonological perception (see, for instance, Kissling, 2012; Nagle, 2014). This initial stage is overcome at higher proficiency levels, however, and

development of perception of prenuclear tones does come closer to—though not completely like—perception of native or bilingual speakers.

Boundary tone

The results for perception of boundary tone showed that all participants were able to perceive broad increments in pitch in an overall similar manner. The data from debriefing interviews also determined that these differences in pitch have at least two interpretations: questions (for rising tones) or declaratives (for falling tones). These findings are consistent with previous claims for a universal tendency to perceive a falling intonation as a closed statement and a rising contour as an open statement (Cruttenden, 1981). Indeed, regardless of proficiency, participants perceived overall pitch height and excursions accurately, a perception process that Cruz-Ferreira (1987) noticed in L2 speakers and called *pitch height strategy*. This rudimentary strategy may constitute the first phase in perceiving and shaping the L2 intonational system. In addition, the fact that this pattern was also observed in monolingual English speakers may speak to the robustness of this perceptual strategy: it can be activated successfully even if the hearer does not possess familiarity with the L2. That is, since English speakers in this study had no previous knowledge of Spanish, it could be argued that they were processing the auditory stimuli not as intonation of Spanish but merely as changes in pitch excursions. Yet, these speakers perceived global falls and rises like all other participants in this study.

Upon closer examination, however, there were differences between monolingual English speakers and the other groups. The results of the debriefing questionnaire revealed that question and statement were the only two possible interpretations that low-proficiency,

high-proficiency, and bilingual speakers assigned to boundary tone stimuli. However, some very high proficiency L2 speakers and Spanish native speakers reported that certain utterances also seemed to be ‘unfinished’ or ‘as if the speaker wanted to say something else’. We could hypothesize that this third interpretation may be attributed to a level final tone, pragmatically interpreted as a suspension tone. This possibility appears to be confirmed by the perception data: these three groups of participants actually perceived three distinct groups of stimuli (see Figure 21), where the middle group, represented in red and spanning stimuli 2-5, could be linked to this purported suspension tone. This is an appealing possibility and would suggest that, as proficiency in Spanish increases, perception of final tone is attuned to perceive three basic types of utterances (declarative, suspended, and questions). Though not reported in the debriefing questionnaire, the same pattern was observed in the BS group. Unfortunately, given the lack of clear shifts in categorical perception, these promising results should be taken with caution and certainly call for further investigation.

Previous research has found that perception of end pitch, though universal to certain extent, is also marked by differences in where exactly languages draw these differences (Grabe et al., 2003). Such differences have indeed emerged in the data for Spanish and English in the current study. That is, English and Spanish speakers (or hearers) parse final pitch scaling differently. The data from L2 learners also point in this direction, as low-proficiency speakers’ perception resembled that of English monolingual speakers, but then approximated the L2 as proficiency increased. It is also worth noticing that only the SNS, VHP, and BS groups—undoubtedly the groups with the highest proficiency in Spanish—

achieved some level of categorical perception for boundary tone. Even more interestingly, the pattern of clustering for the VHP groups was exactly the same as that of BS (see Figure 21). These findings are particularly important as they contribute to the growing body of research that suggests high proficiency in a foreign language may result in native-like processing (syntactic processing in particular, e.g., Bowden, Steinhauer, Sanz, & Ullman, 2013). To the researcher's knowledge, this is the first study that obtains such results for perception of L2 prosody.

Conclusion

The first research question for this study sought to determine how L2 learners perceive intonation in Spanish and how their perception compares to groups of native and bilingual speakers. In terms of prenuclear alignment, it was hypothesized that learners would have difficulty perceiving this tonal event in declarative utterances due to its relative low salience in the utterance. The results of this study proved this hypothesis to be incorrect. Learners of Spanish as a second language can progressively fine-tune their perception of prenuclear pitch alignment and approximate it to that of native and bilingual speakers. That being said, none of the L2 groups achieved perception mechanisms at the exact same level of native or bilingual speakers.

Though clear signs of categorical effects were not found for final boundary tone, there was not only development toward more target-like perception for this tonal event but also performance at the same level of bilingual speakers among participants at the highest level of proficiency. In this sense, relative salience of a final tone as compared with the

rather obscured position of prenuclear pitch accent may help learners achieve target-like perception.

The gradation from completely L1-like to an approximation to L2 perception systems in both prenuclear alignment and boundary tone has shown a clear effect of cross-linguistic influence at play. The influence of the L1 on the L2 has been consistently noticed in the literature of L1 transfer. The vast majority of these effects, however, were examined in studies of L2 production of segmental phonology. The results of this study prove that similar processes of initial heavy L1 transfer and subsequent development toward L2-like performance can also apply in the perception of L2 prosody.

The findings in this chapter shed light on how learners perceive Spanish intonation in declarative utterances at different stages of their second language development. The overall results suggest that perception is heavily influenced by the L1 at initial stages of proficiency and then gradually moves toward perception patterns of monolingual and bilingual speakers. It appears to reach native-like performance, however, only for final boundary tone but not for prenuclear alignment. The remaining pieces of the puzzle include how these speakers produce Spanish intonation in declarative utterances and, more importantly, whether there is a connection between perception and production. The following two chapters attempt to answer these questions.

Chapter 4: Production

Overview

The goal of this chapter is to address the second set of research questions in this dissertation, namely, how English-speaking learners of Spanish produce intonation in declarative utterances, how their production compares with bilingual, Spanish monolingual, and English monolingual speakers, and how learners' production of intonation varies according to task. As was the case for perception, production of intonation was examined in two tonal events: prenuclear peak alignment and final boundary tone height. A sentence reading task (controlled speech) and a storytelling task (quasispontaneous speech) were used to elicit different speaking styles. A brief description of relevant methodology is given, followed by sections on results and discussion.

Methodology

Participants and materials

Participants for the production tasks were the same as those in the perception tasks. If data for one given subject in either production or perception were missing, significantly incomplete or of poor quality, the participant was eliminated from the study altogether. As far as questionnaires and surveys, the materials described in the previous chapter also served the purpose of collecting data relevant for the production tasks.

Tasks

Production data were obtained from all participants via two tasks: sentence reading (controlled speech) and storytelling (quasispontaneous speech).

For the sentence reading task, four stimuli sentences were used for Spanish and four for English (see Table 5). The English set was used for the ENS group and served the purpose of providing a baseline for comparison with all other groups. The segmental characteristic and syllabic structure of *critical* syllables in the Spanish sentences were controlled in order to facilitate acoustic analysis and comparison across sentences and tasks. Given the goals of this study, critical syllables comprised (1) the prenuclear stressed syllable, (2) the prenuclear posttonic syllable, and (3) the last syllable in the sentence. These critical syllables were composed of a single-consonant onset and had no coda so that consonant clusters that could potentially hinder reading were avoided. Furthermore, the onset was a sonorant consonant in order to maximize sonority and voicing, which would improve pitch detection. All critical syllables contained a mid or low vowel as the nucleus and no diphthongs (see description of perception stimulus in previous chapter). Finally, since previous research has shown that pitch alignment may vary as a function of number of unstressed syllables between pitch accents (e.g., Prieto et al., 1995), half the sentences contained two intervening unstressed syllables between the nuclear and prenuclear syllables, while the other half contained three such syllables. The four sentences for English followed the same criteria as far as possible. The condition of syllables with no codas, however, could not be applied in some cases, due to the more stringent nature of English phonotactics in this area.

Table 5. Target Sentences for Production in Sentence Reading Task

Language	Number of intervening syllables between stressed syllables			
	Two		Three	
Spanish	<i>EI mono la lava.</i> ‘The monkey washes it/her.’	<i>Elena me llama.</i> ‘Elena calls me.’	<i>Manolo la mimaba.</i> ‘Manolo pampers her.’	<i>Amaba la moneda.</i> ‘He/She loved the coin.’
English	<i>Her manners are modern.</i>	<i>They remember a lemon.</i>	<i>The dilemma was Alana.</i>	<i>They manage a momentum.</i>

For the storytelling task, participants used the book *Frog, where are you?* (Mayer, 1969). The story consists of a sequence of 24 wordless pictures and has been widely used in studies of first and second language acquisition, among children and adults, and in an array of languages. This story is popular in language research because it elicits natural-occurring language under minimal instructions and has been used to examine a broad scope of issues in L2 and bilingual acquisition, from strategies in narrative discourse (Pearson, 2002) to syntax, (Sánchez, 2003) and phonology (Gallardo del Puerto, Gómez Lacabex, & García Lecumberri, 2007), including a recent study on intonation among Turkish-German bilinguals (Queens, 2012). Participants in the ENS group completed the task in English, while all other groups did so in Spanish.

Procedure

For the sentence reading task, the four stimuli were presented on individual slides on a computer screen in order to avoid the list-reading intonation that usually occurs if sentences appear together. Four sentences of other types (commands and questions) were also included as distractors and to avoid monotony. Participants first received printed cards

with all sentences and were given a few minutes to review them silently. They were allowed to ask questions about lexical items they did not understand or were unsure how to pronounce. The goal of this silent reading phase was to allow participants to become familiar with the stimuli and thus minimize disfluent speech while reading (Huang & Jun, 2009).

Participants received written instructions for the recording process (see Appendix G). The focus of the instructions was to elicit natural, fluent, and loud reading. All sentences were presented twice, in two randomized blocks (8 items x 2 repetitions = 16 renditions), preceded by a practice block. The recording process was monitored by the researcher and at the end of each block participants were asked to repeat any utterances with disfluency or hesitations.

For the storytelling task, participants were interviewed individually by the researcher. Participants were asked to show the book and tell the story to the researcher, who was supposedly unfamiliar with the story. Prior to recording, participants were given a few minutes to review the entire book. A list with key vocabulary (e.g., *rana* ‘frog’, *ciervo* ‘deer’, *abeja* ‘bee’) was also provided, and used mainly by participants in the LP and HP groups. All participants were encouraged to ask questions about vocabulary before being recorded, but emphasis was placed on natural speech. They were told they could skip some pages if they could not understand exactly what was happening or thought they lacked sufficient vocabulary to describe a certain scene. Nonetheless, participants were asked to tell a coherent story. When they felt ready, recording started. The researcher would ask

questions to elicit further language in cases of extended hesitations. In rare cases when participants spoke for more than five minutes, they were politely asked to stop.

With the exception of Spanish monolinguals, participants received written and oral instructions—as deemed necessary by the researcher—only in English. Recordings for both tasks were made in a sound-proof room, with a high-sensitivity, head-mounted microphone attached to a personal computer. Recordings for each participant were done with the software Audacity and saved individually as WAV files, at a sampling rate of 32 kHz.

Both production and perception tasks were completed in one session, whenever possible. Production always preceded perception, but within production, the order of tasks was counterbalanced: half of the subjects were randomly chosen to complete the storytelling task first while the other half started with the sentence reading task.

Data Analysis

For the storytelling task, 90 seconds extracted from the recording were used for analysis, beginning after the first 60 seconds, when participants usually felt more relaxed and comfortable with the task. The selection of the utterances that ultimately were part of the analysis (henceforth referred to as *analyzable utterances*) was done in three major phases. These data-trimming criteria were selected in order to maintain some homogeneity between sentences in this task and the sentence reading task. The goal was to obtain eight analyzable utterances per participant, half of them with two and the other half with three intervening syllables between nuclear and prenuclear pitch accents.

First, the 90 seconds under examination were transcribed orthographically by the researcher. This transcription was used to eliminate sentences based on their syntactic and

morphological features. Only sentences that fulfilled the following criteria were kept: (1) the sentence was syntactically complete, that is, it contained a subject (overt or tacit) and a predicate, and any obligatory arguments within the predicate, (2) the sentence was syntactically simple (no subordination or coordination), (3) it was a declarative sentence, and (4) it did not contain more than five content words.

Second, the utterances that met the previous criteria were examined with Praat to check for their general acoustic and auditory properties. Utterances with the following characteristics were excluded from the analysis: (1) any part of the utterance was emphasized or narrowly focused, (2) the utterance contained filled or unfilled pauses or hesitations, (3) it was produced in a very soft voice and/or contained excessive creaky voice in critical areas for analysis (i.e., prenuclear pitch accents and final boundary tone), (4) it was uttered unnaturally slow or fast, (5) the last word was not paroxytone, and (6) there were only two or three syllables between the nuclear stressed syllable and the preceding prenuclear syllable.¹⁶ This last requirement was the most difficult to meet, and the researcher had to go back to the raw recording for 18 participants and perform a quick scan to find additional utterances that met all the above criteria. Recall that, in order to allow for comparison with the sentence reading task, four utterances needed three intervening syllables between prenuclear and nuclear accents, and an additional four needed two intervening syllables. Nonetheless, nine participants (three in the LP group, two in the HP group, two in the BS group, and two in the ENS group) ended with fewer than four utterances of each kind, but never fewer than two.

¹⁶ Some utterances had two or even three prenuclear pitch accents. These utterances were included if they met all other requirements, but only the last prenuclear pitch accent was examined for the current study.

After the first two sets of data-trimming criteria were applied, an average of 13 utterances remained for each subject. The differences were due mainly to speakers' fluency and syntactic complexity of sentences. As for fluency, LP participants tended to produce fewer fluent utterances than more advanced participants, including native speakers and bilinguals. Regarding syntactic complexity, however, proficient participants produced utterances that tended to be emphatic/focused and overall more complex (more use of subordination, for example) than lower-proficiency groups. The opposite effects of fluency and complexity, therefore, compensated one another to yield a relatively even number of analyzable utterances across the different groups.

Finally, the remaining utterances were examined for the particular acoustic requirements necessary for intonational analysis. Unlike the controlled sentences designed for the sentence reading task, the pitch track in this quasispontaneous task was often interrupted or distorted by excessive stop and fricative consonants. As is the case in this type of research, methodological compromises had to be made in selecting the final eight analyzable utterances for each participant. In order to minimize any potential bias in the selection of these utterances, the researcher asked a trained phonetician to examine acoustically the remaining utterances and select the eight utterances that had the least amount of disruptions and detection errors in the pitch track. This phonetician was aware that the utterances needed to be acoustically suitable for intonational analysis but was otherwise naïve to the specific goals of the study. The only condition was that the eight choices needed be balanced (whenever possible) between two and three syllables intervening between nuclear and prenuclear stressed syllables. In total, 823 utterances were

ultimately analyzed for both prenuclear alignment and final boundary tone in the storytelling task (131 for LP group, 156 for HP, 144 for VHP, 136 for SNS, 132 for ENS, and 124 for BS). Notice that the goal was to obtain and analyze a total of 840 utterances (105 participants x 8 utterances), but recall that even a thorough examination of nine participants' entire production could not find eight analyzable utterances.

Regarding sentences from the reading task, given the strict conditions under which they were designed and recorded, these data were easier to analyze and showed less variation. They were also inspected with Praat to verify their acoustic and auditory features. Twenty-six utterances were discarded due to poor pitch track detection. In total, 814 utterances were ultimately analyzed in this task (129 for LP group, 155 for HP, 142 for VHP, 134 for SNS, 132 for ENS, and 122 for BS).

All analyzable data from both production tasks were examined for prenuclear peak alignment. The analysis consisted of locating f_0 extrema in the segmental stretch comprised between the onset of the stressed syllable and the offset of the last unstressed syllable before the nuclear syllable (e.g., from the onset of [no] to the offset of [ma] in *Manolo la mimaba*). These landmarks were manually marked and annotated using a Praat text grid. In the case of the sentence reading task, the main researcher and a second trained phonetician coded 20% of the utterances independently. Then, both phoneticians got together, examined differences in coding, agreed on a single coding and on specific criteria to mark segmental landmarks, and finally proceeded to code the rest of the data separately (around 40% each). For the storytelling task, both phoneticians coded 100% of the data independently, using the same criteria established before, and then got together to discuss differences in coding and agree

on a final coding for each utterance. Once all relevant boundaries were marked, a special Praat script automatically extracted the time of f_0 maxima occurrence between these boundaries. In order to control for differences in speaking rates among participants, this value was normalized using the formula in (1).

$$(1) T_N = \frac{t_1 - t_0}{d}$$

T_N is the normalized time for f_0 maximum, t_1 is the real time for f_0 maximum as extracted by the Praat script, t_0 is the onset of the prenuclear stressed syllable, and d is the total duration of the stretch under analysis, that is, from the onset of prenuclear stressed syllable to the offset of the last unstressed syllable before the nuclear syllable. Since all speakers produced t_1 within the boundaries of d , T_N always consisted of a value in the range between 0 and 1.0 (see Dilley & Brown, 2007; Redi, 2003, for similar procedures). Figure 23 shows an example of analysis and coding for an utterance produced by a male participant in the LP group.

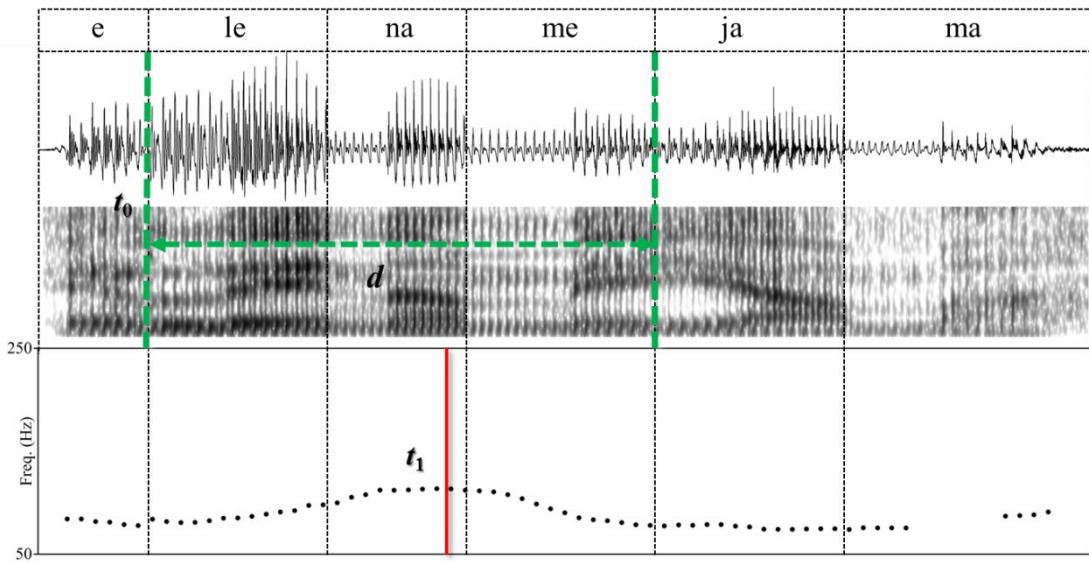


Figure 23. Example of coding and analysis for utterance produced in sentence reading task.

Note. Male speaker in LP group.

Final boundary tone, in turn, was defined as the last non-spurious f_0 point in the pitch track generated by Praat (see Arvaniti & Ladd, 2009; Henriksen, 2014 for similar procedures). First, an interval tier was created to delimit the analysis to a window between the onset of the last stressed syllable and the end of voicing for the entire utterance (defined as cessation of glottal pulses in the waveform). Segmentation was done manually via simultaneous inspection of spectrogram and waveform. Values for end pitch were extracted automatically using a script with instructions to search for end of pitch in the interval and return a value in Hz. As was done in the study of perception (see Chapter 3), extracted f_0 values were converted to Equivalent Rectangular Bandwidth (ERB) units in order to normalize for differences among speakers, in particular between males and females.

Normalized values were averaged so that each participant contributed to the statistical analysis with four aggregate scores (2 tonal events \times 2 tasks): (1) prenuclear peak

alignment in sentence reading task (expressed in T_N), (2) prenuclear peak alignment in storytelling task (expressed in T_N), (3) final boundary tone in sentence reading task (in ERB units), and (4) final boundary tone in storytelling task (in ERB units).

Results

This section presents results for both prenuclear alignment and final boundary tone in both tasks. Descriptive statistics are followed by inferential analyses using Multivariate Analysis of Variance (MANOVA) and post-hoc comparisons.

Prenuclear alignment

Descriptive statistics for all groups are given in Table 6 and Figure 24. All means fell between T_N .44 and .74, that is, all groups seemed to place prenuclear pitch accents rather late and probably to the right boundary of the stressed syllable.

Descriptive statistics, however, also show a great deal of variation among groups. L2 learners seemed to align prenuclear pitch accents progressively later in relation to proficiency and thus approached performance of the SNS group. The ENS group yielded the earlier alignment of all groups, and did not seem to behave like any of the other five groups. In turn, alignment for the BS group fell in between that of SNS and ENS groups, as did the VHP group. A clearer contrast between groups of L2 learners and the three comparison groups is given in Figure 25. As for results in task type, scores for sentence reading and storytelling appear to be quite similar within each group.

Table 6. Descriptive Statistics for Prenuclear Alignment in Sentence Reading and Storytelling Tasks

Group	Task	n	Mean	SD	Min.	Max.
LP	SR	17	.49	.037	.43	.54
	ST	17	.52	.036	.45	.60
HP	SR	20	.50	.023	.45	.54
	ST	20	.53	.039	.46	.59
VHP	SR	18	.61	.043	.55	.68
	ST	18	.61	.040	.55	.69
SNS	SR	17	.70	.046	.62	.78
	ST	17	.74	.056	.65	.84
ENS	SR	17	.44	.042	.37	.54
	ST	17	.41	.042	.34	.51
BS	SR	16	.62	.041	.54	.69
	ST	16	.57	.051	.49	.66

Note. Means represent T_N scores.

LP = Low Proficiency, HP = High Proficiency, VHP = Very High Proficiency, SNS = Spanish Native Speakers, ENS = English Native Speakers, BS = Bilingual Speakers, SR = Sentence Reading Task, ST = Storytelling task.

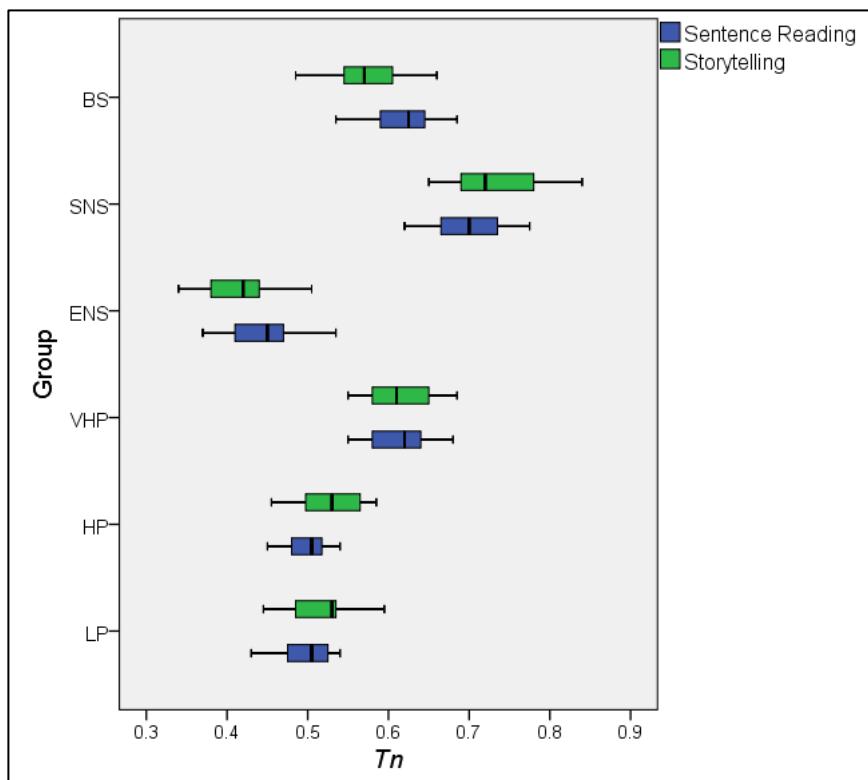


Figure 24. Boxplots for results of prenuclear alignment in sentence reading and storytelling tasks.

Note. T_N = Normalized time, LP = Low Proficiency, HP = High Proficiency, VHP = Very High Proficiency, SNS = Spanish Native Speakers, ENS = English Native Speakers, BS = Bilingual Speakers.

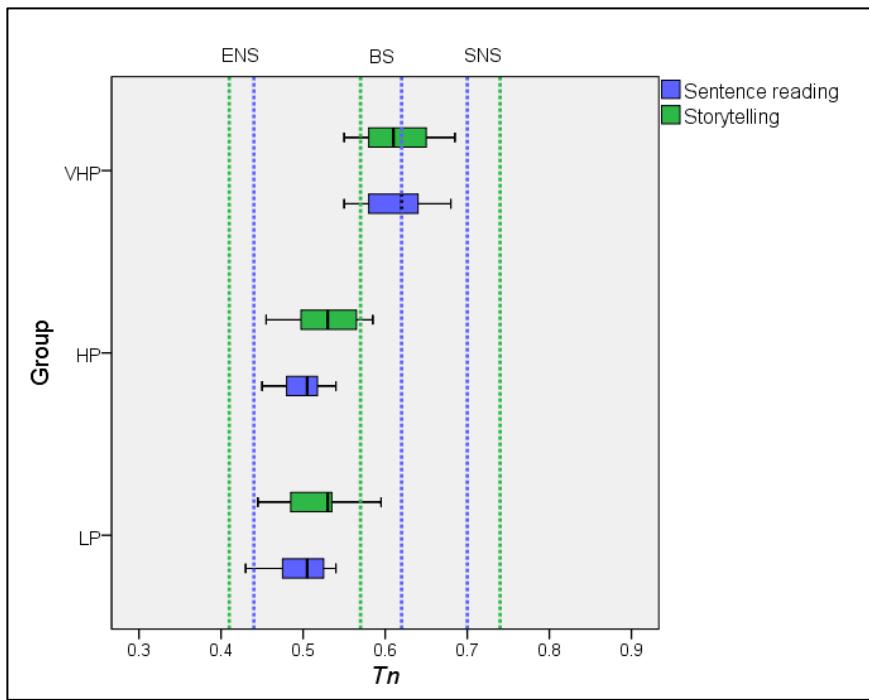


Figure 25. Comparisons between L2 learners and other groups in results of prenuclear alignment according to task.

Note. Vertical dotted lines represent mean scores for the group indicated above. Color legends apply to boxplots and dotted lines.

Results from prenuclear alignment were entered into a one-way MANOVA, with Groups as the independent variable and values in sentence reading and storytelling as the outcome scores. There was a statistically significant difference among groups, $F(10, 196) = 59.56, p < .001$, partial $\eta^2 = .75$, observed power = 1.00. In addition, Groups had a statistically significant effect on both dependent variables: sentence reading, $F(5, 99) = 104.34, p < .001$, partial $\eta^2 = .84$, observed power = 1.00; and storytelling, $F(5, 99) = 100.13, p < .001$, partial $\eta^2 = .83$, observed power = 1.00. Post-hoc comparisons (Tukey) revealed differences among groups as well, in both tasks. Results of post-hoc comparisons are given in Table 7 and Table 8, where subsets contain statistically homogenous groups.

Table 7. Post-hoc Results for Prenuclear Alignment of Final Boundary Tone in Sentence Reading Task

Group	<i>n</i>	Subset			
		1	2	3	4
ENS	17	.45			
LP	17		.50		
HP	20			.50	
VHP	18				.61
BS	16				.62
SNS	17				.70
Sig. (<i>p</i>)		1.00	1.00	1.00	1.00

Note. Subsets indicate groups of homogenous means ($p < .05$). Since sample sizes are unequal, harmonic mean is used.

Table 8. Post-hoc Results for Prenuclear Alignment of Final Boundary Tone in Storytelling Task

Group	<i>n</i>	Subset				
		1	2	3	4	5
ENS	17	.42				
LP	17		.52			
HP	20			.53	.53	
BS	16				.57	.57
VHP	18					.61
SNS	17					.74
Sig. (<i>p</i>)		1.00	1.00	.051	.088	1.00

Note. Subsets indicate groups of homogenous means ($p < .05$). Since sample sizes are unequal, harmonic mean is used.

Some interesting trends emerged from the post-hoc analyses presented in Table 7 and Table 8. First, the ENS and SNS groups appeared separated from all other groups in both tasks. That is, the ENS group produced significantly earlier peaks than all other groups,

while the SNS group produced significantly later peaks than all other groups. Second, the production of BS and VHP participants was statistically the same in both tasks, with alignment times that fell roughly in the middle between the SNS and ENS groups. Lastly, LP and HP participants' alignment in both tasks was statistically the same and, in turn, later than the alignment of ENS participants and earlier than VHP and SNS participants. Notice that the only difference between production tasks is that the post hoc analysis for storytelling task produced one more homogeneous group—identified as subset 3 in Table 8—comprised of participants in the HP and BS groups. This result seems odd given the general patterns obtained for all groups in both tasks. Furthermore, the non-significance at $p = .051$ in the difference between means seems marginal. Therefore, this subset is noted now but will be ignored in the general discussions of results.

The following part of the analysis examined whether there is a difference between tasks. As discussed above (Table 6 and Figure 24), descriptive statistics showed relatively little differences among tasks for all six groups. This observation was tested using six paired sample t tests, the results of which are presented in Table 9.

Results of paired t tests showed that tasks were statistically different for all groups except for the HP group. The direction of the difference, however, varied among groups. LP, VHP and SNS produced significantly later prenuclear peaks in the storytelling task, while ENS and BS produced later peaks in the sentence reading task. These results do seem puzzling and, though no clear pattern emerges in these data, an explanation will be attempted in the discussion section.

Table 9. Results of *t* Tests for Differences in Prenuclear Alignment between Sentence Reading and Storytelling Tasks

Group	Differences		<i>t</i>	<i>df</i>	Correlation
	Mean	<i>SD</i>			
LP	-0.026	0.025	-4.21**	16	0.77***
HP	-0.029	0.028	-4.69	19	0.71**
VHP	-0.002	0.019	-0.38***	17	0.90***
SNS	-0.038	0.032	-4.92***	16	0.82***
ENS	0.030	0.003	44.08***	16	1.00***
BS	0.045	0.031	5.88***	15	0.80***

Note. Differences are between T_N scores in sentence reading and storytelling tasks. A correlation coefficient is also provided and shows the strength and significance of the correlation between tasks. Due to multiple *t* tests performed, alpha level was adjusted following Bonferroni's procedure.

** $p < .01$, *** $p < .001$

The final part of the analysis for prenuclear pitch alignment looked at the effect of distance between stressed syllables on prenuclear alignment. As discussed in the review of the literature, some studies on Spanish intonation have suggested that the distance between stressed syllables may be related to the degree of displacement of prenuclear pitch accents. In simplified terms, these studies have observed that peak displacement increases as the temporal distance between stressed syllables becomes larger, normally as a result of additional unstressed syllables. To test this effect, aggregate T_N scores were reorganized by (1) separating scores based on number of intervening unstressed syllables (two, as in *El mono la lava*, or three as in *Manolo la mimaba*) and (2) collapsing scores between sentence reading and storytelling tasks. Therefore, each participant's production was entered into the analysis with two T_N aggregate scores, one for pitch alignment in utterances with two intervening unstressed syllables between stressed syllables (henceforth referred to as

TwoIUS utterances), and another one for pitch alignment in utterances with three intervening unstressed syllables between stressed syllables (henceforth referred to as ThreeIUS utterances).

Table 10. *Descriptive Statistics for Prenuclear Alignment in TwoIUS and ThreeIUS Utterances*

Group	Utterance type	<i>n</i>	Mean	<i>SD</i>	Min.	Max.
LP	TwoIUS	17	.53	.060	.41	.62
	ThreeIUS	17	.49	.028	.45	.54
HP	TwoIUS	20	.53	.037	.46	.59
	ThreeIUS	20	.49	.024	.45	.54
VHP	TwoIUS	18	.64	.042	.58	.71
	ThreeIUS	18	.59	.040	.53	.66
SNS	TwoIUS	17	.74	.046	.67	.83
	ThreeIUS	17	.69	.051	.61	.78
ENS	TwoIUS	17	.44	.044	.38	.54
	ThreeIUS	17	.42	.043	.33	.50
BS	TwoIUS	16	.62	.046	.53	.70
	ThreeIUS	16	.57	.042	.50	.65

Note. Means represent T_N scores.

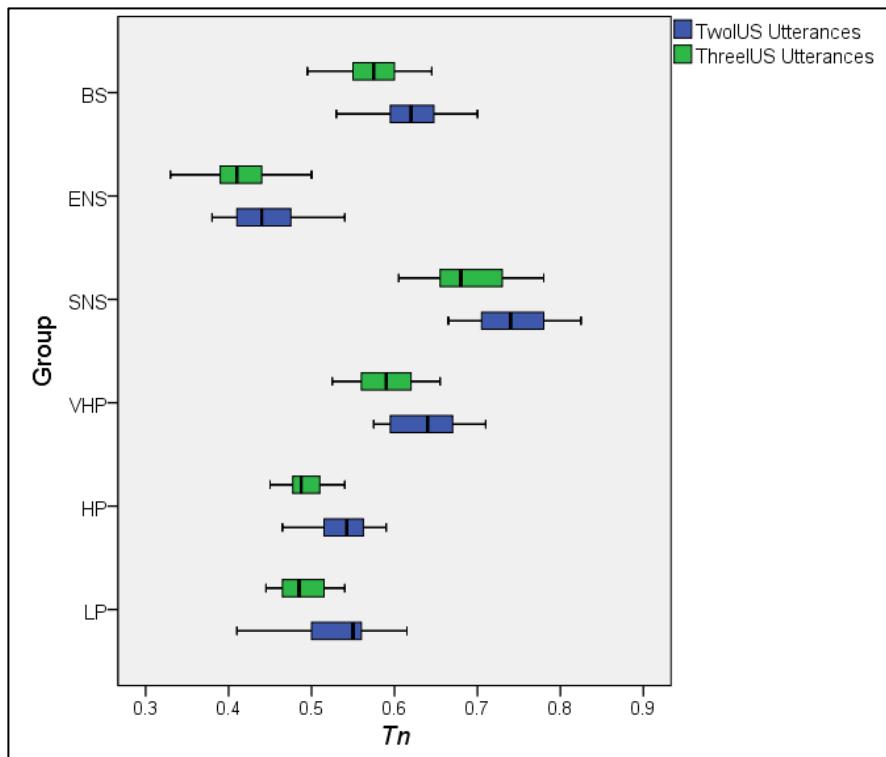


Figure 26. Boxplots for results of prenuclear alignment for all groups in TwoIUS and ThreeIUS utterances.

The general trends for distance between stressed syllables are quite similar to the results obtained in task type. In principle, this should not be completely surprising since the underlying data are the same but grouped into different dependent variables. In order to test for the statistical effect of Groups on distance between syllables, a one-way omnibus MANOVA was performed, with Groups (six) as the independent variable and aggregate scores for TwoIUS and ThreeIUS utterances as the two dependent values. There was a statistically significant difference among groups, $F(10, 196) = 31.29, p < .001$, partial $\eta^2 = .61$, observed power = 1.00. In addition, Groups had a statistically significant effect on both dependent variables: TwoIUS utterances, $F(5, 99) = 87.59, p < .001$, partial $\eta^2 = .82$, observed power = 1.00; and ThreeIUS utterances, $F(5, 99) = 107.11, p < .001$, partial $\eta^2 = .82$, observed power = 1.00.

.84, observed power = 1.00. Post-hoc comparisons (Tukey) revealed differences among groups as well, in both types of utterances. Results of post-hoc comparisons are given in Table 11 and Table 12.

Table 11. Post-hoc Results for Prenuclear Alignment in TwoIUS Utterances

Group	<i>n</i>	Subset			
		1	2	3	4
ENS	17	.44			
LP	17		.53		
HP	20		.53		
BS	16			.61	
VHP	18			.64	
SNS	17				.74
Sig. (<i>p</i>)		1.00	1.00	.82	1.00

Note. Subsets indicate groups of homogenous means ($p < .05$). Since sample sizes are unequal, harmonic mean is used.

Table 12. Post-hoc Results for Prenuclear Alignment in ThreeIUS Utterances

Group	<i>n</i>	Subset			
		1	2	3	4
ENS	17	.42			
LP	17		.49		
HP	20		.49		
BS	16			.57	
VHP	18			.59	
SNS	17				.69
Sig. (<i>p</i>)		1.00	1.00	.83	1.00

Note. Subsets indicate groups of homogenous means ($p < .05$). Since sample sizes are unequal, harmonic mean is used.

Results from post-hoc analyses yielded a clear picture that was actually the same for both TwoIUS and ThreeIUS utterances. Monolingual speakers of English and Spanish behaved statistically different from all other groups, producing the earliest and latest T_N values, respectively. In turn, alignment scores for LH and HP participants were statistically equivalent and aligned later than ENS values, while VHP and BS participants formed their own homogenous group and their alignment was later than ENS, LP, and HP scores.

The final step of prenuclear alignment analysis tested whether there was a statistically significant difference in alignment between utterances with two and three intervening unstressed syllables. A set of six t tests was run in order to test this effect. Results are presented in Table 13.

Table 13. Results of t Tests for Differences in Prenuclear Alignment between TwoIUS and ThreeIUS Utterances

Group	Differences		t	df	Correlation
	Mean	SD			
LP	.041	.064	2.68*	16	.105
HP	.044	.023	8.45***	19	.788***
VHP	.048	.013	15.57***	17	.951***
SNS	.052	.008	27.22***	16	.992***
ENS	.028	.023	4.98***	16	.859***
BS	.045	.005	32.86***	15	.996***

Note. Differences are between T_N scores in TwoIUS and ThreeIUS utterances. A correlation coefficient is also provided and shows the strength and significance of the correlation between tasks. Due to multiple t tests performed, alpha level was adjusted following Bonferroni's procedure.

* $p < .05$, *** $p < .001$

All t -tests results for differences between TwoIUS and ThreeIUS utterances reached significance. However, the LP group did not after the alpha level was adjusted using

Bonferroni's correction. As shown in Figure 26, the data for TwoIUS utterances in this group were quite dispersed, which is also confirmed by the relatively large standard deviation (.60, while standard deviations for all other groups fell within .24 and .46). It is worth highlighting that significance was reached in all but five groups despite the relatively small difference in T_N scores between TwoIUS and ThreeIUS utterances. This could be attributed to the very strong correlation between both outcome variables in each group and to the relatively homogeneity of scores within groups. A final important observation is that that T_N values for ThreeIUS utterances in all groups were aligned relatively earlier than in TwoIUS utterances.

Boundary tone

Results for final boundary tone followed the same statistical procedure described above for prenuclear alignment, though only two outcome scores were tested for each group, one aggregate score from the sentence reading task and one aggregate score from the storytelling task, both of which expressed in ERB units. Descriptive statistics are provided in Table 14, and Figure 27.

Table 14. Descriptive Statistics for Final Boundary Tone Height in Sentence Reading and Storytelling Tasks

Group	Task	n	Mean	SD	Min.	Max.
LP	SR	17	3.73	0.35	3.13	4.27
	ST	17	3.74	0.41	2.95	4.47
HP	SR	20	3.55	0.34	2.80	4.00
	ST	20	3.61	0.38	2.98	4.30
VHP	SR	18	3.52	0.37	2.70	3.93
	ST	18	3.60	0.47	2.80	4.30
SNS	SR	17	3.26	0.39	2.55	3.83
	ST	17	3.31	0.43	2.56	4.30
ENS	SR	17	3.75	0.35	3.00	4.34
	ST	17	3.84	0.39	3.20	4.60
BS	SR	16	3.50	0.32	2.95	4.10
	ST	16	3.55	0.40	2.75	4.18

Note. Means represent ERB Units.

LP = Low Proficiency, HP = High Proficiency, VHP = Very High Proficiency, SNS = Spanish Native Speakers, ENS = English Native Speakers, BS = Bilingual Speakers, SR = Sentence Reading Task, ST = Storytelling task.

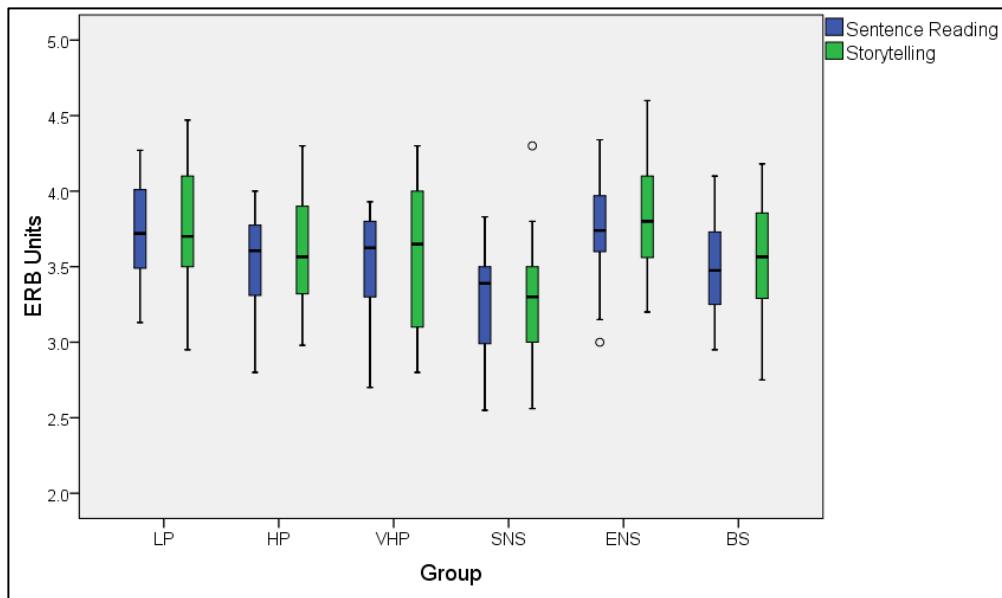


Figure 27. Boxplots for results of final boundary tone height in sentence reading and storytelling tasks.

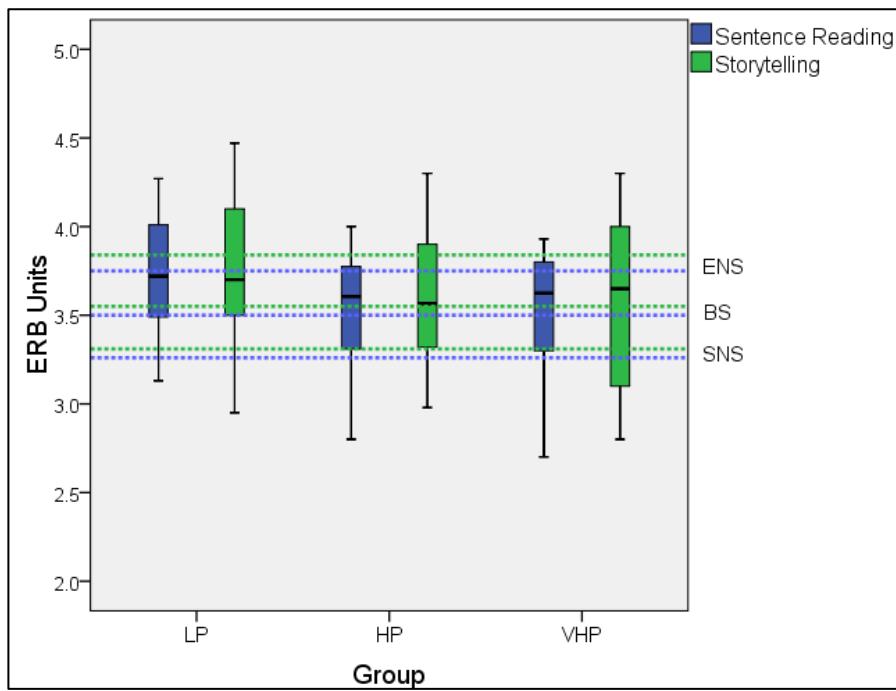


Figure 28. Comparisons between L2 learners and other groups in results of final boundary tone according to task.

Note. Horizontal dotted lines represent mean scores for the group indicated on the right. Color legends apply to boxplots and dotted lines.

Interestingly, overall results of descriptive statistics show similar trends as those obtained in prenuclear alignment. The ENS and SNS groups are at different extremes, with the former obtaining the highest mean and the latter the lowest mean for final boundary scaling. All other groups seem to appear in the middle of the monolingual groups.

Comparisons between the three groups of L2 learners and all other groups appear more clearly in Figure 28. In order to test for the statistical effect of Groups on type of task, a one-way MANOVA was run, with Groups (six) as the independent variable and aggregate ERB unit scores for sentence reading and storytelling as the two outcome values. The test yielded a statistically significant difference among groups, $F(10, 196) = 2.67, p < .05$, partial $\eta^2 = .10$, observed power = .92. Groups also had a statistically significant effect on both dependent variables: sentence reading, $F(5, 99) = 4.31, p < .05$, partial $\eta^2 = .18$, observed power = .96; and storytelling, $F(5, 99) = 3.25, p < .05$, partial $\eta^2 = .14$, observed power = .87. Post-hoc tests (Tukey) also produced significant differences among groups as well, in both tasks. Post-hoc comparisons are presented in Table 15 and Table 16.

Table 15. Post-hoc Results for Final Boundary Tone Height in Sentence Reading Task

Group	<i>n</i>	Subset	
		1	2
SNS	17	3.26	
BS	16	3.50	3.50
VHP	18	3.52	3.52
HP	20	3.55	3.55
LP	17		3.73
ENS	17		3.75
Sig. (<i>p</i>)		.16	.31

Note. Means represent ERB Units. Subsets indicate groups of homogenous means (*p* < .05). Since sample sizes are unequal, harmonic mean is used.

Table 16. Post-hoc Results for Final Boundary Tone Height in Storytelling Task

Group	<i>n</i>	Subset	
		1	2
SNS	17	3.31	
BS	16	3.55	3.55
VHP	18	3.60	3.60
HP	20	3.61	3.61
LP	17		3.74
ENS	17		3.84
Sig. (<i>p</i>)		.25	.32

Note. Means represent ERB Units. Subsets indicate groups of homogenous means (*p* < .05). Since sample sizes are unequal, harmonic mean is used.

The final part of analysis investigated whether there was a difference between tasks for production of final boundary tone. Six dependent sample *t* tests were run, the results of which are presented in Table 17.

Table 17. *Results of t Tests for Differences in Final Boundary Tone Height between Sentence Reading and Storytelling Tasks*

Group	Differences		<i>t</i>	Sig.	df	Correlation
	Mean	SD				
LP	-.007	.24	-0.13	.90	16	.82***
HP	-.066	.36	-0.82	.42	19	.49*
VHP	-.079	.48	-0.71	.49	17	.37
SNS	-.048	.29	-0.68	.50	16	.75***
ENS	-.095	.29	-1.33	.20	16	.70*
BS	-.054	.25	-0.88	.39	15	.78***

Note. Differences are between ERB Unit scores in sentence reading and storytelling tasks. A correlation coefficient is also provided and shows the strength and significance of the correlation between tasks. Due to multiple *t* tests performed, alpha level was adjusted following Bonferroni's procedure.

* $p < .05$, *** $p < .001$

No *t* test reached significance, which means that both elicitation tasks were statistically the same for all groups. Correlations were relatively strong (except for the VHP group), which meant that in general participants behaved consistently across both tasks. However, the fact that absolute differences between tasks were in general minimal, along with the relatively large standard deviations for all groups, probably resulted in the statistical nonsignificance observed in all six differences.

Discussion and conclusion

Prenuclear alignment

The results of prenuclear alignment clearly and consistently put monolingual speakers of English and Spanish at two ends of the alignment spectrum. Recall that the segmental text under analysis spanned from the left boundary of the prenuclear stressed syllable to the right boundary of the last unstressed syllable before the nuclear syllable (e.g.,

from the onset of [no] to the offset of [ma] in *Manolo la mimaba*). For English speakers, peaks were aligned relatively early in this string and thus closer to the prenuclear stressed syllable, with mean T_N scores in the .42–.45 range.¹⁷ In contrast, mean prenuclear peaks for Spanish monolinguals fell within the T_N range of .69–.74. These results support previous findings that prenuclear peak alignment occurs relatively later in Spanish than in English (Garrido et al., 1993; Gussenhoven, 2004; Llisterri et al., 1995). These claims, however, were based on results from studies that examined these languages separately, with divergent research goals and methodologies. To my knowledge, this dissertation is the first empirical study that specifically set out to compare Spanish and English prenuclear alignment. The results in the current study provide clear empirical evidence for claims that up to now have been mostly based on assumptions.

Despite the differences in relative timing of prenuclear peaks between Spanish and English, the results also revealed a remarkable similarity between both groups, and among all other groups, for that matter. When intervening syllables between stressed syllables were increased from two to three, alignment shifted to the left but not in the same incremental proportion. That is, when absolute duration between stressed syllables was increased by roughly 50% (assuming equal duration across syllables), alignment retracted by only .028 T_N for English monolinguals and .052 T_N for Spanish monolinguals (see Table 13). Figure 29 shows this observation graphically. When the total duration under analysis (d) increased, peak alignment (expressed in T_N values) occurred earlier, but not with the same magnitude as the increase in duration.

¹⁷ Recall that T_N is a ratio value that indicates the relative position of the aligned peak within the segmental string under analysis. Possible T_N values range from 0 to 1, where values closer to 0 represent earlier peak alignment (that is, closer to the onset of the prenuclear stressed syllable).

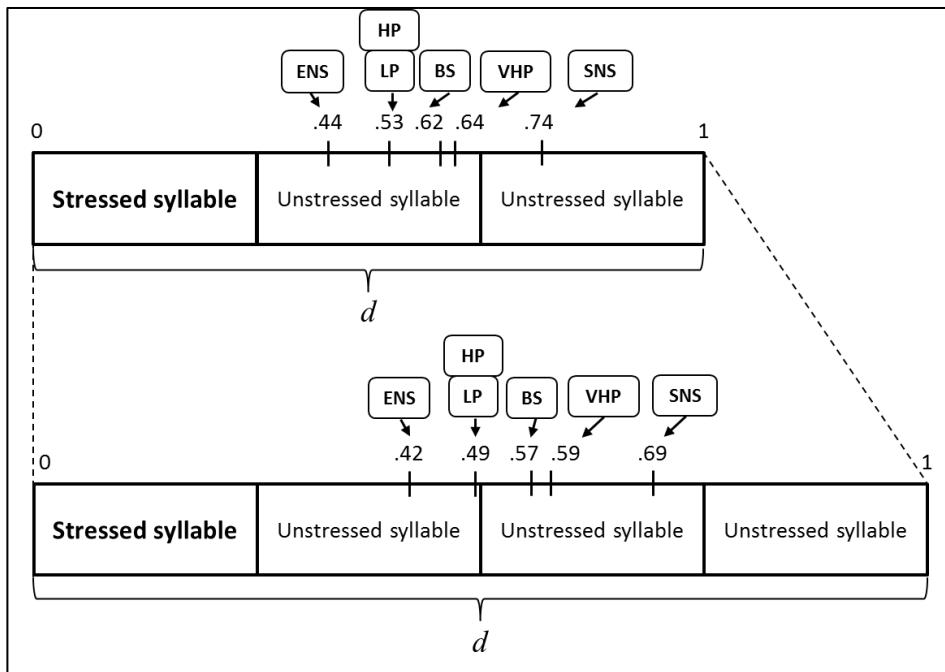


Figure 29. Comparison of prenuclear alignment between TwoIUS (top panel) and ThreeIUS (bottom panel) utterances

Note. Boxes represent idealized duration of syllables. Ticks and numbers represent T_N values for the group labeled right above. Duration of the entire stretch under analysis is represented by d (see data analysis section).

These findings have multiple implications for theories of alignment. First, the results do not lend support to a segmental anchoring hypothesis (SAH, e.g., Arvaniti et al., 1998). A strict interpretation of the SAH would predict that peaks, and probably most tonal events, appear anchored to segmental landmarks (e.g., syllable codas, vowel midpoints, etc.) and are thus unaffected by other factors such as speaking rate, distance between stressed syllables, etc. This was not the case in the data for the current study. Peaks tended to appear regularly distributed in the tonal space available between stressed syllables. In the case of Spanish monolingual speakers, for example, prenuclear alignment appeared on the last unstressed syllable when d was shorter, but on the penultimate syllable when d was longer. The

idealized representation in Figure 29 shows that there was no preference for peaks to appear anchored to any particular syllable onset or offset. These findings are consistent with previous studies on Spanish pitch alignment (e.g., Henriksen, 2014, Prieto & Torreira, 2007). That said, there is certainly some degree of tonal stability in peak alignment, represented by the fact that English speakers consistently produced earlier peaks than Spanish speakers, and within small T_N variations based on distance between stressed syllables in each group. This stability did not stem from segmental landmarks, as the SAH would predict, but from the relative space available between stressed syllables. If the distance is shorter, peaks will be aligned relatively earlier within that space, but always respecting language-specific preferences for an overall early (English) or late (Spanish) peak alignment. Phonetic and phonological processes seem to be at play. The pressure from reduced time will displace peaks to the right (phonetic process) but within the limits of preferred alignment for each language (phonological process). So, should a phonological H* tone be proposed as the most common prenuclear tone for English and a L*+H tone for Spanish? The answer to this question is beyond the scope of this study, but results add very interesting considerations to the debate. Note that English speakers also produced peaks that aligned to the right boundary of the prenuclear syllable, a phenomenon noted and discussed by Gussenhoven (2004). In this sense, as many authors have suggested, displacement could be just a phonetic factor: peaks will be aligned later if they have sufficient space to do so (e.g., Hualde, 2002; Prieto et al., 1995). So we could propose that the same H* pitch accent has different phonetic realizations in English and Spanish, though that would bring us to the more problematic field of language-specific phonetic constraints. At this point, therefore, I

prefer to assume that differences in alignment should be represented by different tonal configurations. In sum, results in this study have shown there are clear differences in prenuclear alignment between Spanish and English and that, beyond these language-specific differences, both languages behave similarly in terms of how the available time affects the degree of displacement.

Where does L2 learners' production fall in relation to monolingual speakers? In straightforward terms, it falls in the middle. The two lower-proficiency groups had statistically equivalent results and aligned peaks significantly later than monolingual English speakers but earlier than very high proficiency participants, who in turn produced earlier pitch alignment than monolingual Spanish speakers. These distances in T_N scores among groups are, moreover, practically equidistant, if scores are examined either according to task (see Figure 25) or according to number of intervening syllables (see Figure 26 and Figure 29). These findings offer a clear and unprecedented look at language development in the area of L2 prosody. First, results of production of the lower-proficiency groups showed the bias towards the L1 when producing the L2. Values in prenuclear alignment for these two groups approximated English values and were still far from bilingual or Spanish-native speakers. VHP participants, in turn, approximated but did not reach values of Spanish native speakers. They did, however, produce prenuclear pitch alignment within the same values as bilingual speakers. This particular point is discussed in the next section along with results for final boundary tone.

From a theoretical standpoint, the gradual development in the production of L2 intonation attested in this dissertation also provides serious support for the use of the

Autosegmental-Metrical (AM, Beckman & Pierrehumbert, 1986; Ladd, 1983, 1996; Pierrehumbert, 1980, 2000) approach for the study of L2 intonation. This support is extended to an analysis of intonation using levels instead of configurations (e.g., Brazil, 1981; Halliday, 1970). As reviewed in Chapters 1 and 2, the treatment of tones (H and L) by the AM approach as individual units has significantly increased the sophistication of intonation analysis. The detailed acoustic measurement of peak alignment, the isolation of peaks (H tones) from valleys (L tones), the interplay between phonological and phonetic levels of analysis, among other factors, are powerful theoretical and methodological tools in the AM model that have allowed the current study to examine in such great detail L2 intonation use and potential development. Such findings would have been greatly obscured, if not altogether impossible, if the intonational curve had been treated as a complete tonal movement, as is the case in analyses by configurations.

Finally, no clear results were obtained for the effect of task type on prenuclear alignment. Compared with values from the sentence reading task, results from the storytelling task showed that monolingual English speakers produced significantly earlier peaks while Spanish speakers produced later peaks. In other words, it seems that the more spontaneous nature of the storytelling task only accentuated the difference already noted between English and Spanish alignment. With regards to groups of L2 learners and bilingual speakers, the picture is not as clear. There was no significant difference between tasks for the group of VHP speakers, while LP group produced later peaks (like Spanish monolinguals) and BS produced earlier peaks (like English monolinguals). In general, differences, though statistically significant, are relatively small and the effect of task type

does not support or reject a pattern of language development, understood as an approximation to monolingual or bilingual norms. As noted by Face (2003, 2010), quasispontaneous data pose significant problems for analysis and comparison with speech elicited under more controlled conditions. However, it is worth highlighting that the lack of clear results occurred when the tasks were compared within each group. At the level of comparisons across groups, however, quasispontaneous speech produced the same results as those in the sentence reading task. This fact speaks to the robustness of peak alignment as a tonal event regardless of speaking style conditions.

Boundary tone

The examination of final boundary tone scaling in this study has produced a number of interesting results. First, the two groups of monolingual participants produced final tone scaling quite differently. Spanish native speakers produced the lowest boundary tone (L%), while English native speakers produced the highest. It is a relatively well-known fact that some English dialects produce a rising intonation in what otherwise is a declarative statement. This phenomenon, usually referred to as high rising terminal (HRT) or ‘uptalk,’ has been long observed in New Zealand and Australian English (Britain, 1992; Fletcher, Grabe, & Warren, 2005; Horvath, 1985) and is becoming increasingly common in American English (Ching, 1982; Ladd, 2008), probably driven by younger generations. The contexts for this tone are normally narratives, invitations to acknowledgement from the listener, transaction openers, and answers to wh-questions (Ladd, 2008). Even though English-speaking participants in this study did not produce HRT in the strict sense of the phenomenon (characterized as H*H-H%), arguably because most of the contexts for this

tone were not present in the tasks, we could hypothesize that an overall higher L% tone was produced in English as a result of the existence of the HRT tone in this language. However, the data from the storytelling task may suggest that English speakers were using this tone to some degree: the difference between this task and the sentence reading was the widest among all groups (see Table 17), and the narrative nature of the task may have matched one of the contexts described by Ladd where the HRT tone occurs. Although this difference between tasks was not significant and the goal of this dissertation was not to examine English intonation in depth, the general trends and the anecdotal observations made by the researcher, who did notice cases of HRT tones during data collection, could offer a plausible explanation for the relatively high final tone observed among monolingual speakers of English.

Second, the production of final L% tone for L2 learners produced mixed results. On the one hand, low-proficiency speakers produced tones at the same level of English monolingual speakers. These results are in line with previous studies that found a strong effect of transfer from English to Spanish in the production of final tones. Henriksen et al. (2010), for example, also observed predominance for final rises (i.e., the HRT tone) among participants who were beginning a study abroad program in Spain. On the other hand, all other L2 learners in the current study produced final pitch height at the same level of native and bilingual speakers. In fact, as shown in post-hoc analyses (Table 15 and Table 16), boundary tone values for HP, VHP, and BS participants did not differ significantly from either the ENS or the SNS groups. However, comparisons also showed that values for ENS and SNS groups did differ significantly. Taken together, these results suggest that end pitch

height for HP, VHP, and BS participants fell in between Spanish and English. This was the phenomenon noticed before for prenuclear alignment, save the fact that the HP group clustered with LP participants.

The results for production of both tonal events examined in this study, therefore, revealed two important facts: (1) advanced L2 learners produced intonational patterns equivalent to those of Spanish-English bilinguals (heritage speakers) and (2) these patterns appeared in between values of English and Spanish monolingual speakers. The first observation has implications for issues of development and ultimate attainment in second language acquisition. There has been a growing impetus in the field of SLA to drop the monolingual native norm as a benchmark for SLA (e.g., Ortega, 2009) and adopt, instead, bilingualism (and early bilinguals) as the target against which L2 development and use are measured. Of course the data in this study do not rule out the possibility that production for L2 learners could change ('develop') in the future to match monolingual native standards. However, the relatively stable and exceptionally high level of proficiency of the most advanced group in this study suggests that their intonational patterns should not change substantially beyond their current state. The second observation provides strong support for the 'merger' category predicted by Flege's Speech Learning Model (SLM, 1995). Recall that this hypothesis predicts that certain phonological aspects in bilingual speech are composed of categories taken from both languages, thus creating a sort of single, mixed resolution in the middle of both languages. This is exactly what the current study found in both tonal events under investigation. Though the evidence is scant for prosody as compared to segmental phonology, similar results have been obtained, for instance, by Mennen,

Schaeffler, and Dickie (2014) in a study of pitch range among advanced German-speaking learners of English. Other studies, however, have tested this prediction and the merger effect was not found, such as in Teague's investigation of Mandarin-English bilinguals (2011). The current study did not set out to test the merger prediction (a full account would have required to test bilingual speakers and L2 learners in English as well) but the results still showed that highly advanced proficiency in the L2 can result in bilingual-like (and also native-like in the case of boundary tone) production of intonation.

The last part of this section will briefly discuss the lack of difference within subjects for the effect of task type. Production among all groups followed the exact same pattern in both controlled and quasispontaneous speech. Previous research on L2 intonation has found differences in length of tone units (Howell & Kadi-Hanifi, 1991), preboundary lengthening (Blaauw, 1994), and number of deaccented syllables (Face, 2003) between controlled and spontaneous speech, but such differences were not found for final boundary tone height. From a methodological perspective, Xu (2010) suggested that not all phonological features of speech vary to the same degree according to speaking styles. The results in this study suggest that final pitch height could be one feature that does not vary from spontaneous to controlled speech. This constitutes only a hypothesis, restricted in scope to these two tonal events and to the neutral declarative utterance in Spanish. It is perhaps not surprising that a short and syntactically simple declarative utterance, that carries information but otherwise usually lacks the expressiveness of other sentence types, will not vary substantially across different speaking styles. It is possible that differences in speaking styles do arise in more

marked utterance types, such as declaratives in narrow focus, or in questions or exclamations.

Conclusion

The second research question in this dissertation sought to investigate what type of intonation L2 learners of Spanish produced in declarative utterances, how their production compared to that of bilingual speakers and monolingual speakers of Spanish and English.

The hypothesis that prenuclear alignment would be impervious to development had to be rejected. Results showed that there is a clear development toward, and achievement of, bilingual standards as proficiency increased. The hypothesis that final boundary tone would be higher for L2 learners as a result of transfer from English was partially correct; it was found to hold only among learners with the lowest proficiency level.

There was a general approximation to bilingual values in both tonal events under investigation. These results showed that production of intonation in declarative sentences can not only develop toward a more L2-like state but also reach bilingual-like production. Additionally, the highest proficiency of L2 learners under analysis matched the production of balanced bilinguals. Both groups, in turn, produced values intermediate between English and Spanish, suggesting that one merged system could be utilized in the production of L2 intonation.

Finally, a secondary research question addressed the effects of task type in the production of intonation of declarative utterances. Some minor though unclear effects were found for prenuclear alignment, while no effects were obtained in final boundary tone.

Chapter 5: Relationship between perception and production

Overview

This chapter examines whether there is a relationship between the findings in perception and those obtained in production. Statistical analyses and examination of global trends will be used to compare results in both areas. Findings are discussed as they pertain to the research goals of this study and also as they shed light on methodological and theoretical considerations in the study of the relationship between perception and production in L2 speech.

Methodology and data analysis

Data for the analysis in this chapter were drawn from the same studies on perception and production described in previous chapters. The comparison was done by performing correlational analyses between perception and production data for both tonal events under analysis, namely, prenuclear peak alignment and final boundary tone height.

In the case of perception, recall that the analysis sought to determine whether there were categorical shifts in the perception of both tonal events. Locations of such shifts, when they occurred, were compared across groups. Presence of categorical effects was defined as the division of the 10 perception stimuli into non-overlapping clusters of statistically equal means. The location of the shift was defined as the middle point between these clusters (for example, the shift was marked between points 3 and 4 if one cluster was comprised of stimuli 1, 2, and 3 while the other cluster contained stimuli 4, 5, 6, 7, 8, 9, and 10). Since each participant needed to have one data point for the correlation analysis in the current

chapter, the shift in categorical perception for each participant needed to be determined. For this purpose, two one-way ANOVA's were run for each participant (one for prenuclear alignment and one for final boundary tone), followed by a post-hoc analysis if the ANOVA was significant. The ANOVA compared the means for the 10 perception stimuli (dependent variable). As expected, the ANOVA did not reach significance for some participants, and other times it did reach significance but post-hoc analyses showed that clusters were overlapping, and thus claims of categorical shifts could not be made. When this was the case, participants were eliminated from the correlation analysis. For this reason, a total of 29 (out of 105) participants were eliminated from the analysis of prenuclear alignment and 59 (out of 105) from the analysis of final boundary tone. These results, though deleterious for the goals of the comparison, are not surprising if we consider the fact that boundary tone did not produce clear categorical effects (see Chapter 3). That is, since categorical effects were not found in a certain group, it was expected that performance of individual participants in that group would follow similar trends. Participants from the LP and ENS groups tended not to achieve categorical perception, and thus these groups suffered the greatest amount of participant exclusion. For the sake of simplicity of reference, scores were computed as the stimulus immediately preceding the purported categorical shift (for example, if the shift occurred between stimuli 3 and 4, a score of 3 was selected as the data point to be entered in the correlation analysis).

As for production, the data discussed in chapter 4 were rearranged so that each participant would have one aggregate score for prenuclear alignment, expressed in T_N , and

one score for final boundary tone, expressed in ERB units. Aggregate scores were obtained by collapsing data from both speaking styles.¹⁸

In sum, scores from 76 participants were used to test the correlation of perception and production of prenuclear alignment while scores from 46 participants were used to analyze the correlation between perception and production of final boundary tone.

Results

A Pearson's correlation for each tonal event was performed with data points from all participants. This process would first indicate whether there was a general association between perception and production regardless of group belonging. In addition, since a large percentage of data points were eliminated in the data for final boundary tone, an overall correlation would gain statistical power. The correlation between perception and production of prenuclear alignment was statistically significant ($r = .43, p < .001, n = 76, R^2 = .18$) while it was not significant for final boundary tone ($r = -.08, p = .58, n = 46, R^2 = .007$). Effect size—as reported by R^2 —was between medium and large¹⁹ for the correlation in prenuclear alignment. The positive correlation implies that the later categorical shifts in perception occur, the later peaks are aligned in production. Scatterplots in Figure 30 present the association between production and perception for both tonal events. A quick comparison between graphs shows how the nature of the association for each tonal event differed substantially.

¹⁸ Analyzing any other covariates or intervening variables was beyond the scope of the analysis in this study, which was meant to tap into the general nature of the relationship—or lack thereof—between perception and production of intonation in neutral declarative utterances. Furthermore, the variable ‘speaking style’ produced mixed and marginal differences in the case of prenuclear alignment, and no difference at all in the case of final boundary tone.

¹⁹ Larson-Hall (2010) and Cohen (1992) define R^2 effect sizes as small for $R^2 = .01$, medium for $R^2 = .09$, and large for $R^2 = .25$.

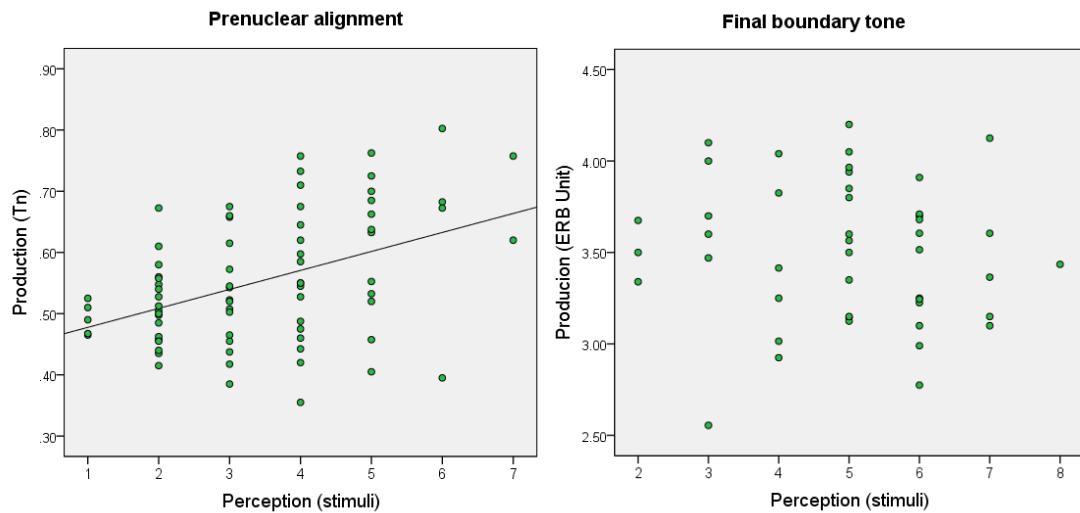


Figure 30. Scatterplots for relationship between perception and production data.

Note. $n = 76$ for prenuclear alignment, $n = 46$ for final boundary tone. Regression line is provided ($R^2 = .18$) when correlation was statistically significant.

Table 18. *Results of Pearson's Correlations between Production and Perception of Prenuclear Alignment for All Groups*

Group	<i>n</i>	<i>r</i>	Effect size - R^2
LP	12	.091	(.01)
HP	15	.689**	.47
VHP	13	.605*	.37
SNS	13	.699**	.49
ENS	11	-.486	(.24)
BS	12	.732**	.53

Note. Correlation coefficients (*r*) significant at * $p < .05$, ** $p < .01$
Effect sizes appear in parentheses when correlation coefficient did not reach significance, thus rendering them irrelevant.

The next step in the analysis involved looking at the association for each individual group. A considerable amount of power was lost as correlations were now performed with a

much smaller N . Extreme cases were the LP ($n = 4$) and ENS groups ($n = 5$) in final boundary tone. Even though significant correlations were not expected in these cases, for the sake of consistency all results are reported (see Table 18 and Table 19).

Table 19. Results of Pearson's Correlations between Production and Perception of Final Boundary Tone for All Groups

Group	<i>n</i>	<i>r</i>	Effect size (R^2)
LP	4	.326	(.10)
HP	9	-.379	(.14)
VHP	9	-.622	(.39)
SNS	10	.121	(.01)
ENS	5	.236	(.05)
BS	9	.082	(.00)

Note. Effect sizes appear in parentheses since correlation coefficient did not reach significance, which renders them irrelevant.

Results from individual groups are consistent with results from all groups considered together. For prenuclear alignment, perception of categorical shifts correlated positively with production of peak alignment, for all but the LP and ENS groups, for whom correlations did not reach significance. It is worth highlighting that effect sizes for the significant correlations increased considerably as compared with the effect size reported for the analysis of collapsed groups. This can be interpreted as a strong relationship between perception and production of prenuclear alignment for the HP, VHP, SNS, and BS, whereby one variable explained roughly half of the variance in the other variable (47% for HP, 37% for VHP, 49% for SNS, and 57% for BS). Individual scatterplots are presented in Figure 31.

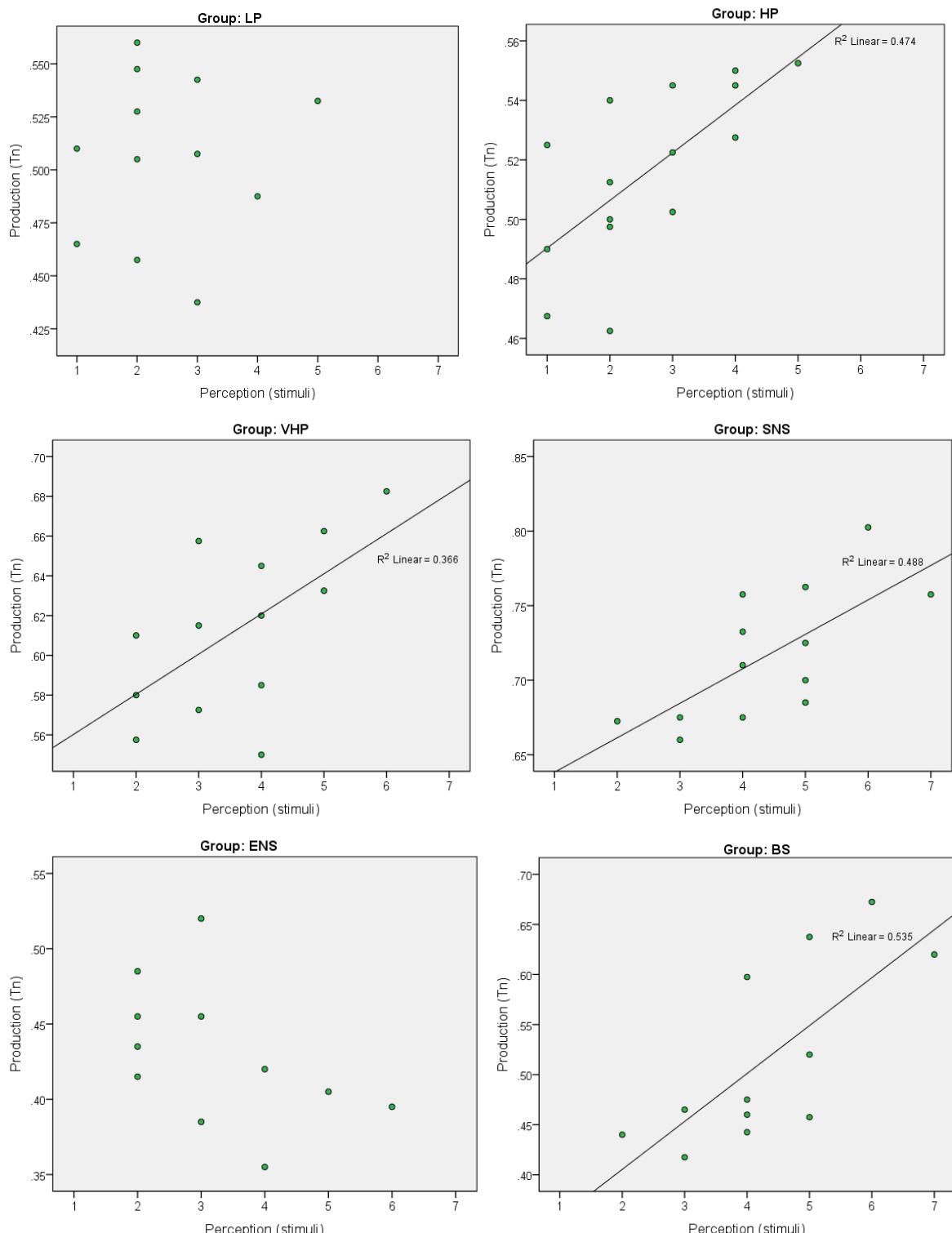


Figure 31. Scatterplots for relationship between perception and production of prenuclear alignment for each group.

As for final boundary tone, no group achieved a statically significant correlation. As seen in Table 19, results yielded a mixture of positive and negative correlations, which could be regarded as an anomaly in the data due to the small sample size in each test. Since no significant results were obtained for the relationship between production and perception of final boundary tone, no further statistical analyses were pursued. Interpretation of these results will be provided in the discussion section.

Discussion and conclusion

Results from this chapter suggest there is a strong relationship between production and perception of prenuclear pitch alignment in declarative utterances in Spanish. The only two groups for whom this relationship did not hold were, perhaps not surprisingly, low-proficiency L2 speakers and monolingual English speakers.

To my knowledge, this type of relationship has not been established before for L2 Spanish intonation. As discussed in Chapter 2, claims about any purported relationship between perception and production of L2 speech come almost exclusively from the field of segmental phonology, along with some rare incursions into other areas, such as syllable structure (e.g., Huensch, 2013) or stress placement (e.g., Archibald, 1997). For L2 intonation, however, such relationship has been neglected, which makes the comparison of current findings with previous research particularly challenging. The comparison is further blurred by the rather mixed contributions that come from L2 segmental phonology. In general, models of L2 perception such as Flege's Speech Learning Model (SLM) would predict a relatively straightforward association between L2 perception and production, whereby attunement of the former is a prerequisite for development of the latter. The results

for prenuclear alignment in the current study showed that there is indeed a correlation between production and perception, but no clear conclusions can be drawn as to whether perception leads production, as predicted by the SLM. In fact, if we examine the separate results of perception and production for the VHP group, for example, we could conclude that production leads perception: this group achieved bilingual-like measures in production but fell short of showing bilingual- or native-like perception of prenuclear peaks. Similar results have been obtained in studies of segmental L2 phonology (e.g. Baker, Trofimovich, Flege, Mack, & Halter, 2008).

We can, however, extract a general pattern of development when we compare the three groups of L2 learners. Neither monolingual English speakers nor low-proficiency L2 learners achieved significant correlations between production and perception. We could hypothesize that this was not totally unexpected for English speakers since they were actually being tested on different languages (perception of Spanish but production of English). Low-proficiency L2 learners, however, were tested only in Spanish, yet they did not show a relationship between perception and production either. On the other hand, correlations for groups of high and very high proficiencies proved to be not only significant but quite strong. These findings are in line with predictions of the SLM that at least moderate correlations between production and perception usually occur among more experienced L2 learners (Flege, 2003). Lack of correlations at lower proficiency levels could be attributed to the effect of transfer of L1 categories into the L2. How transfer affects perception is still a matter of debate, though we do have a clearer understanding of how L1 transfer affects production (see Kormos, 2006, for a brief review).

The lack of a statistical relationship between perception and production of final boundary tone may have been the result of a methodological limitation, and thus it cannot be taken as evidence that such relationship does not actually exist. In this regard, Llisterri (1995) points to the “inadequacy of comparing results from tests in speech production with results derived from speech perception tests, since there are important differences in the nature of the techniques used to assess these activities” (p. 98). Some recent studies in the realm of segmental phonology have also reported that the absence of a perception-production relationship could be due to the methodology used to tap each construct (e.g., Kissling, 2012; Peperkamp & Bouchon, 2011). The test of categorical perception used in the current study did not produce clear results; hence it is not surprising that when these data were entered into a correlation, no association between perception and production was found. Moreover, a significant amount of statistical power was lost since most data points had to be eliminated.

Conclusion

The goal of this chapter was to answer the third and last research question formulated for this dissertation: Is there a relationship between L2 learners’ perception and production of intonation? If such relationship was found, an ancillary question examined the nature of this relationship. The hypothesis that there would a strong relationship between production and perception was confirmed for prenuclear alignment but rejected for final boundary tone.

Positive correlations in prenuclear alignment among the highest proficiency levels seemed to support the SLM’s prediction that production and perception of L2 phonology

begin to act in tandem as proficiency increases. However, the prediction that perception leads production could not be tested. With regards to final boundary tone, no statistical correlation was obtained for any group. Rather than discarding the possibility that such association actually exists, the results speak to a potential limitation in the original measurement of perception, which was carried over to the correlation analysis.

Significant results for prenuclear alignment offer a novel and promising look into the association between perception and production of intonation in Spanish. However, interpretations should be taken with caution as they are informed mostly by findings in—and models of—segmental L2 phonology.

Chapter 6: Conclusion

Research questions and main findings

This section revisits the research questions that guided this study and, by answering them, also provides a brief summary of main findings.

RQ1. How do English-speaking learners of Spanish perceive intonation of Spanish declarative utterances?

Results showed that learners of Spanish as a second language can progressively fine-tune their perception of prenuclear pitch alignment. As for boundary tone, no clear categorical effects were found, but results suggested that there was some level of development toward more target-like perception. The gradation from completely L1-like to an approximation to L2 perception systems in both prenuclear alignment and boundary tone has shown the clear effect of cross-linguistic influence at play.

RQ1a. How does their performance compare to that of monolingual native speakers of Spanish, monolingual native speakers of English, and early Spanish-English bilinguals?

Low-proficiency participants resembled English monolingual speakers in the perception of both tonal events. The overall results suggest that perception is heavily influenced by the L1 at initial stages of proficiency and then gradually moves toward perception patterns of monolingual and bilingual speakers. It appears to reach native-like performance, however, only for final boundary tone but not for prenuclear alignment. In this sense, relative salience of a final tone as compared with the rather obscured position of prenuclear pitch accent may help learners achieve target-like perception.

RQ2. How do English-speaking learners of Spanish produce intonation of Spanish declarative utterances?

Results for prenuclear alignment showed that there was a clear development toward and achievement of bilingual standards as proficiency increased. The hypothesis that final boundary tone would be higher for L2 learners as a result of transfer from English was partially correct, as it was confirmed only among learners with the lowest proficiency level.

RQ2a. How does their performance compare to that of monolingual native speakers of Spanish, monolingual native speakers of English, and early Spanish-English bilinguals?

Values in both tonal events for monolingual Spanish speakers were drastically different from those of monolingual English speakers. All other groups produced values in the middle range between English and Spanish. For L2 learners, there was generally an approximation to bilingual scores in both tonal events. These results showed that production of intonation in declarative utterances can not only develop toward a more L2-like state but also reach bilingual-like production. Additionally, the highest proficiency of L2 learners matched the production of balanced bilinguals. Both groups, in turn, produced values intermediate between English and Spanish, suggesting that one merged system could be utilized in the production of L2 intonation.

RQ2b. How does L2 learners' production of intonation vary according to speaking style?

Some minor though unclear effects were found in prenuclear alignment, while no effects at all were obtained in final boundary tone. The general lack of effect of speaking

style was attributed to the relatively little expressiveness that a neutral declarative utterance inherently carries.

RQ3. Is there a relationship between L2 learners' perception and production of intonation? If so, what is the nature of this relationship?

A strong relationship between production and perception was found for prenuclear alignment but not for final boundary tone. The relationship in prenuclear alignment was in line with the Speech Learning Model's prediction that production and perception of L2 phonology are related as proficiency increases. Regarding final boundary tone, no statistical correlation was obtained for any group. This was attributed to a potential limitation in the original measurement of perception which was carried over to the correlation analysis.

Theoretical and methodological implications

Role of L2 input

As described in Chapter 3, quantity and quality of input that participants received was controlled to the extent possible, particularly among learners with low and high proficiency. Participants were excluded if they reported they received unusually high amounts of native input outside the classroom setting (extended study abroad periods, interactions with Spanish-speaking caretakers during childhood, etc.). Given the general profile of participants in the two lower proficiency levels, it is thus safe to assume that most of the L2 input these participants received came from an instructional setting. In this type of scenario, learners are expected to receive input from non-native speakers (classmates and non-native instructors, for example) who can potentially reinforce learners' foreign-accented speech (Flege, 2009). The analysis of learning context questionnaires revealed that in fact all

L2 participants received substantial input from non-native speakers of Spanish. We can assume that participants in the highest proficiency group compensated for this with ample opportunities to interact with native speakers. However, this was not the case for the two lower proficiency groups. Yet, these participants showed significant development in many of the measures in this study. This was particularly true for the high proficiency group. In other words, these regular college-level students of Spanish seem to have improved their perception and production of intonation *despite* being exposed to a good amount of non-native L2 input.

These findings are certainly unexpected given prior studies that highlight the importance of receiving native speaker input for the development of L2 phonology (e.g., Flege & Liu, 2001). The importance of native speaker input is surely uncontested but its scope should probably be accepted with caution. The findings in this study suggest, for example, that native speaker input may not be as critical as it is for segmental areas of L2 phonology. We could hypothesize that non-native teachers probably acquire certain prosodic features more easily than they acquire segmental features, and thus their speech is not ‘accented’ in this regard, or at least not as accented as for segments. It also possible that L2 learners were maximizing exposure to native input, in the form of audio classroom material, interactions with native speakers outside the classroom, etc. In sum, findings in this study suggest that development of L2 intonation, at least for certain unmarked tonal events, may necessitate less native input than segmental phonology does. Future studies will need to probe more deeply and more systemically into the type of input that L2 learners receive. In

any case, these results certainly provide a positive outlook for the role of L2 classroom instruction in developing intonation in a foreign language.

The bilingual and monolingual norms

Findings in this dissertation provide evidence in favor of reconsidering the monolingual norm as the target for ultimate L2 attainment and the yardstick for measuring L2 development in SLA research. L2 users and learners are not—or will not become—two monolingual speakers in one, following the well-known claim made by Grosjean (1989). The current study demonstrated first that bilingual and monolingual speakers are not alike under many measurements of intonation, and second, and more importantly, that highly experienced L2 speakers resembled bilingual speakers and not monolingual Spanish speakers.

A related topic brings us to what constitutes *accuracy* in the perception and production of L2 phonology. Many studies in this area have shown a big concern for measuring this construct (e.g. Piske, Flege, MacKay, & Meador, 2011; Waltmunson, 2006). While arguably a valid concept used to assess development and use of L2 segmental phonology, *accuracy* is normally not clearly defined in most studies that deal with it. There is possibly a tacit understanding that accuracy means performance at the level of native speakers (i.e., degree of ‘nativelikeness’) if such speakers are used as control or reference groups. This dissertation, however, has shown that bilingual speakers may constitute a better benchmark for comparison. If such speakers are not available or the interest of the study actually lies in comparing L2 learners with monolingual speakers, studies on L2 phonology

at the very least should provide a definition of *accuracy* and a justification for the selection of control groups and baselines for comparison.

Variation in intonation

The results of the present dissertation shed light on the variation of intonation across languages and within Spanish. In regards to the differences between English and Spanish, overall results corroborated previous assumptions on contrasts between these languages in neutral declarative utterances. The sharp difference found in this study between Spanish and English not only provided a solid framework to understand perception and production of L2 learners but also highlighted basic and relatively unexplored contrasts between these two languages. Sometimes the seemingly least interesting target form may be overlooked, and in so doing basic typological differences among languages are also overlooked.

Spanish, in turn, presents a great deal of dialectal variation in intonation (Hualde, 2003; Sosa, 1999). As I have argued previously, this could be one of the reasons why researchers have been discouraged from further investigating L2 intonation. However, results in the current study seem to suggest that dialectal variation may not be a major concern at least for the tonal events under consideration. Of course dialectal and idiolectal variation did exist among Spanish speakers, as attested by anecdotal observations during data collection and some trends in the statistical analysis,²⁰ but this variation did not seem to be sufficiently large to skew or hinder comparisons with other groups.

²⁰ For example, monolingual Spanish speakers had a greater variance than any other group in the production of prenuclear alignment in the storytelling task.

Limitations and future research

As described in Chapter 3, the imitation task used for the perception portion of this dissertation was chosen after contemplating alternative approaches to test categorical perception. Prominent researchers in the field have supported the advantages of this task after weighing the pros and cons of other methodologies (Gussenhoven, 2004; Ladd, 2008). Laura Dilley has extensively used the technique and has concluded that it is the “optimal task” to test categorical perception in intonation (2010, p. 68). Despite being arguably the best methodological compromise we have to date, there are limitations in its application and the findings that it produces. Empirical research relies on *observable* behavior or data. What constitutes observable data is a major challenge in the investigation of speech perception. In the specific case of an imitation task, the observable behavior is speech *production*, which may prove problematic if we are testing and making claims about speech *perception*. One hypothesis, for instance, claims that imitation is constrained by the speakers’ grammar, that is, speakers can only reproduce what is permitted in their grammar. This *selective imitation* hypothesis has been tested in the realm of segmental phonology (Kraljic & Samuel, 2006; Nielsen, 2007). For this hypothesis, therefore, imitation taps into production but perhaps not perception, which is not necessarily a problem for models that assume the same underlying system for both domains (e.g., the *listener’s model* of intonation, Collier, & Cohen, 1990; ’t Hart et al., 1990). The selective imitation hypothesis, however, has not been widely tested in segmental phonology and has not held when applied in perception of intonation (D’Imperio et. al., 2014). All this being said, the results in the present study yielded significant and categorical effects for most groups in at least one tonal event under examination. Moreover,

these results were far from random: they depicted coherent contrasts between Spanish and English and clear patterns of development in second language learning. Given the task used and the scarcity of studies in this area, these findings call for replication research and further testing.

Proficiency among the three groups of L2 learners undoubtedly proved to be a highly important and informative variable in the current study. However, very advanced L2 users were, in many respects, significantly different from the other two proficiency levels. First, the scores for this group in the elicited imitation task—used to determine proficiency—were almost at ceiling and, in turn, much higher than those in the low and high proficiency groups. Levels of proficiency in this study, therefore, did not increase in a gradual progression. Second, the context of learning in the most advanced group differed considerably from the type of language learning experienced by the two lower-proficiency groups. Graduate students in general have a higher motivation to learn the language, which in turn leads them to seek more opportunities to learn and use Spanish, inside and outside the classroom. The best example of this is the significantly higher contact with native speakers that graduate students reported, both in the United States and in Spanish-speaking countries. These experiences were quite different in the other two groups of learners, which brings about many potentially intervening variables that remained uncontrolled for in the study. This is certainly an important limitation in the interpretation of the findings but is also a necessary methodological compromise needed in order to probe into the most advanced proficiencies while also examining language development.

Given the paucity of studies that have looked at intonation in Spanish as a second language, the current study proposed crucial research questions aimed at understanding how Spanish intonation is perceived and produced by L2 speakers in the most common utterance type, namely, unmarked declarative utterances. Future research should address other utterance types (questions, commands, exclamations) and other functions normally attributed to intonation (expression of sarcasm, surprise, etc.). The interplay between production and perception should still be at the core of any future analysis of intonation to gain a more comprehensive understanding of L2 speech at the prosodic level.

Finally, we should wonder to what extent the differences between L2 users/learners and comparison groups would actually impact comprehension and intelligibility of L2 speech. Many scholars in the field of L2 accent and pronunciation (e.g., Derwing & Munro, 2009) have called for studies that examine the features that contribute to a foreign accent and also whether these features actually affect communication. While the present study lacks a measure of how deviations from bilingual or monolingual intonation standards contribute to a perceived foreign accent among L2 learners, it constitutes a solid starting point for such research.

Conclusion

The present study constituted, to the best of my knowledge, the first examination of both production and perception of intonation in Spanish as a second language. Findings showed sharp differences between Spanish and English for both tonal events examined in neutral declarative utterances. In addition, results revealed that intonation in the second language can develop as a function of proficiency in terms of perception and production.

Unresolved issues in this dissertation should lend themselves to further investigation of intonation, in an attempt to come to terms with what Dwight Bolinger wisely called a ‘half-tamed savage.’

Appendix A: Elicited Imitation Task

Repetition task.

For this task you will be asked to first repeat some sentences in English and then some sentences in Spanish. Please pay careful attention to the instructions on the recording. Please do not take any notes during this exercise. Now, let's begin.

You are going to hear several sentences in English. After each sentence, there will be a short pause, followed by a tone sound {TONE}. Your task is to try to repeat exactly what you hear in English. You will be given sufficient time after the tone to repeat the sentence.
REPEAT AS MUCH AS YOU CAN. Remember, DON'T START REPEATING THE SENTENCE UNTIL YOU HEAR THE TONE SOUND {TONE}. Now let's begin.

We drove to the park

I'll call her tomorrow night

You can buy meat at the butcher shop

My brother just bought a brand new computer

Sometimes they take their dog for a walk in the park

We're going to play volleyball at the gym that I told you about

That was the last English sentence

Now, you are going to hear a number of sentences in Spanish. Once again, after each sentence, there will be a short pause, followed by a tone sound {TONE}. Your task is to try to repeat exactly what you hear in Spanish. You will be given sufficient time after the tone to repeat the sentence. REPEAT AS MUCH AS YOU CAN. Remember, DON'T START REPEATING THE SENTENCE UNTIL YOU HEAR THE TONE SOUND {TONE}. Now let's begin.

[30 Spanish sentences, see below]

1. Quiero cortarme el pelo.
2. El libro está en la mesa.
3. El carro lo tiene Pedro.

4. Él se ducha cada mañana.
5. ¿Qué dice usted que va a hacer hoy?
6. Dudo que sepa manejar muy bien.
7. Las calles de esta ciudad son muy anchas.
8. Puede que llueva mañana todo el día.
9. Las casas son muy bonitas pero caras.
10. Me gustan las películas que acaban bien.
12. El chico con el que yo salgo es español.
11. Después de cenar me fui a dormir tranquilo.
13. Quiero una casa en la que vivan mis animales.
14. A vosotros os fascinan las fiestas grandiosas.
15. Ella sólo bebe cerveza y no come nada.
16. Me gustaría que el precio de las casas bajara.
17. Cruza a la derecha y despues sigue todo recto.
18. Ella ha terminado de pintar su apartamento.
19. Me gustaría que empezara a hacer más calor pronto.
20. El niño al que se le murió el gato está triste.
21. Una amiga mía cuida a los niños de mi vecino.
22. El gato que era negro fue perseguido por el perro.
23. Antes de poder salir él tiene que limpiar su cuarto.
24. La cantidad de personas que fuman ha disminuido.
25. Después de llegar a casa del trabajo tomé la cena.

26. El ladrón al que atrapó la policía era famoso.
27. Le pedí a un amigo que me ayudara con la tarea.
28. El examen no fue tan difícil como me habían dicho.
29. ¿Serías tan amable de darme el libro que está en la mesa?
30. Hay mucha gente que no toma nada para el desayuno.

This is the end of the repetition task. Thank you.

Appendix B: Scoring of Elicited Imitation Task

- *Score 0:** Nothing (silence); • Garbled (unintelligible, usually transcribed as XXX); • Minimal repetition, then item abandoned: - Only 1 word repeated; - Only 1 content word plus function word(s), - Only function word(s) repeated, - Only 1 or 2 content words out of order plus extraneous words that weren't in the original stimulus
- Score 1:** • When only about half of idea units are represented in the string but a lot of important information in the original stimulus is left out; sometimes the resulting meaning is unrelated (or opposed) to stimulus; • Or when string doesn't in itself constitute a self-standing sentence with some (related or not to stimulus) meaning (This may happen when only 2 of 3 content words are repeated and no grammatical relation between them is attempted)
- Score 2:** When content of string preserves at least more than half of the idea units in the original stimulus; string is meaningful, and the meaning is close or related to original, but it departs from it in some slight changes in content, which makes content inexact, incomplete, or ambiguous
- Score 3:** • Original, complete meaning is preserved as in the stimulus. Strings which are ungrammatical can get a 3 score, as long as exact meaning is preserved. Some synonymous substitutions are acceptable: -Anything with or without "muy" (very) should be considered synonymous. -Substitutions of 'y'/pero' (and & but) are acceptable. • Changes in grammar that don't affect meaning should be scored as 3. (Ambiguous changes in grammar that could be interpreted as meaning changes from a NS perspective should be scored as 2. That is, as a general principle in case of doubt about whether meaning has changed or not, score 2.)
- Score 4:** • Exact repetition: String matches stimulus exactly. Both form and meaning are correct without exception or doubt.

Appendix C: Background Questionnaire

BACKGROUND QUESTIONNAIRE

Please read the following questions and answer them to the best of your knowledge. As stated in the consent form, this information will be kept confidential, and your name shall never appear in any part of this project.

- (1) Name: _____
- (2) Gender: _____
- (3) Age: _____
- (4) First language: _____
- (5) Do you speak any other language(s) besides English and/or Spanish? Yes No
If your answer is 'yes,' please specify: _____

If you are studying Spanish

- (6) Spanish class: _____
- (7) Number of semesters of formal college instruction in Spanish: _____

Appendix D: Learning Context Questionnaire

Please answer the following questions to the best of your knowledge. They are all related to your learning and use of Spanish, inside and outside a classroom. Some of them require retrieving knowledge that may have faded in your memory, so if you don't remember specific information, you can give an estimate or write "I don't know/remember" if you don't feel comfortable giving an estimate.

Remember that the goal is to obtain as much information as possible about your contact with Spanish, so if you notice that the questions do not describe your specific situation, please let the researcher know about this. Also, the researcher may contact you at a later time and ask you to elaborate on some answers if he believes that the questions have not fully addressed a specific experience you had.

Childhood

1. **People who spoke Spanish around you.** Consider your childhood before the age of 12 and think about all the native Spanish speakers with whom you regularly interacted in Spanish, even if either of you used Spanish only occasionally in these interactions. You may want to think about friends, classmates, neighbors, relatives, nannies, babysitters, etc.
 - a. Where was he/she from (nationality)? _____
 - b. What dialect of Spanish did he/she speak? _____
 - c. How long was he/she approximately your nanny/friend/etc. for?
_____ years, _____ months
 - d. How old were you during this period (indicate age range if applicable)?

 - e. During this period, how many hours a week did you approximately spend with him/her? _____
 - f. During this time, how often did he/she use Spanish when talking to you?
Never Sometimes Usually Always
 - g. During this time, how often did you use Spanish when talking to him/her?
Never Sometimes Usually Always
 - h. How much Spanish do you feel you "learned" from this person? Explain.

(repeated as necessary)

Precollege/adolescence

2. (same questions as in 1, but within the age range of 12-17)
3. **Spanish you had at school.** Did you have Spanish classes at school (or in any other formal instructional setting)? If so, please answer the following questions. The same

set of questions will apply each time the instructional situations changes (you start a new school year, new teacher, etc.)

- a. Grade: _____
- b. What level of Spanish was the class (name of the class if you are not sure about the level)? _____
- c. How many hours of Spanish lessons did you have a week? _____ hours
- d. Was your teacher a native speaker of Spanish? _____. If so,
 - i. Where was he/she from (nationality)?

 - ii. What dialect of Spanish did he/she speak?

- e. Think about the amount of speaking the teacher did in Spanish and English. Choose the option that more closely approximates the percentage of Spanish he/she used in class (as opposed to English).
10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
(repeated as necessary)

College

4. List all the college Spanish courses you have taken, even if you did not complete them. Be sure you indicate the level if it is not obvious from the course name. For each class, answer the following questions.
 - a. Did you finish this course? If not, how many classes did you attend?

 - b. How many hours of Spanish classes did you have a week? _____ hours
 - c. Was your teacher a native speaker of Spanish? _____. If so,
 - i. Where was he/she from (nationality)?

 - ii. What dialect of Spanish did he/she speak?

 - d. Think about the amount of speaking the teacher did in Spanish and English. Choose the option that more closely approximates the percentage of Spanish he/she used in class (as opposed to English).
10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Visits to Spanish-speaking countries

5. Think about your visits to Spanish-speaking countries. For each one, answer the following questions:
 - a. What country did you visit? _____
 - b. For how long? _____

- c. What was the main reason of the visit (tourism, study abroad, etc.)?

- d. Who did you go with? _____
- e. Where did you stay (hostel, host family, friends' house, etc.)?

- f. What was your Spanish level when you visited the country?
i. I didn't speak any Spanish.
ii. Poor
iii. Good
iv. Very good
- g. Where your daily interactions mostly in English or Spanish?
- h. How often did you interact with native speakers? Please explain the main type of interactions you had (for example, with your host family, with a tour guide, etc.).

(repeated as necessary)

Spanish outside the classroom

6. For this question, please consider all opportunities you have outside the classroom to speak and listen to Spanish. Please do not take into account any school assignments (for example, watching a film for a class).
- a. Do you have regular interactions in Spanish with native speakers of Spanish?
If so,
i. Where is the person from? _____
ii. How many hours a week do you interact with him/her?

- (repeated as necessary)
- b. Do you have any other opportunities to listen to Spanish? If so,
i. What type of activity is it (watching movies, listening to music, etc.)?
ii. What dialect of Spanish do you mostly listen to in this activity?

- iii. How many hours a week do you do this activity? _____ hours
(repeated as necessary)

Appendix E: Debriefing Questionnaire

Debriefing questionnaire administered after imitation task (perception)

Answer these questions to the best of your knowledge. Try to remember what your reactions were while doing the task.

1. Do you think all the sentences sounded the same?
2. If you think they were different, how do you think they differed?
3. Did sentences in general sound natural to you?
4. Can you think of a context where someone would say these sentences, as you heard them?
5. Was the task difficult for you?
6. Did you have problems recalling the sentence before you had to say it out loud?
7. Were you tired, bored, or distracted while doing the task?

8. Thank you for your answers! Now please give this sheet to the researcher and feel free to ask him any questions you want about this research project.

Thank you for participating!

Appendix F: Instructions for Imitation Task

Slide 1

In this task, you will hear the same sentence said in different ways. This is the sentence:

La nena lloraba.

Your task is to listen to the sentence carefully and repeat it just as you heard it. Pay particular attention to the overall pronunciation.

It's OK if the sentences sound the same to you. Just try to imitate each sentence individually.

Slide 2

You will always see the sentence on the slide. Click on the audio button and then repeat it. You can listen to it a second time if you need to. You can also take a second and say it in your head first.

Let's practice first.

[practice sentences, 3 slides]

Slide 6

We will start with the task now.

Any questions?

Appendix G: Instructions for Sentence Reading Task

Slide 1

In this task, you will be recorded while reading some sentences in Spanish.

Slide 2

First, you will have the chance to read the sentences without being recorded, so that you become familiar with them. In this practice part, you can read the sentences out loud or to yourself. You can also ask the researcher if you are not sure how to pronounce the sentences.

Slide 3

The next three slides contain the target sentences.

Take a few minutes to review them.

[three slides with target sentences and distractors]

Slide 7

Now you will see one sentence per slide. Read it out loud, as naturally as possible. You can read the sentence again if you hesitate or make pauses in the middle of it. You can move to the next slide when you are ready.

Some sentences will appear more than once.

The researcher may also ask you to reread a sentence if you speak too softly.

Slide 8

Please, follow the researcher's instructions to set up the microphone.

Do you have any questions?

References

- Aguilar, L., De-la-Mota, C., & Prieto, P. (Eds., 2009). Sp_ToBI Training Materials. Web page: http://prosodia.upf.edu/sp_tobi/
- Aliaga-García, C. & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista, (Eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 2–31). Newcastle upon Tyne, UK: Cambridge Scholars Publisher.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42, 529–555.
- Archibald, J. (1997). The acquisition of English stress by speakers of tone languages: Lexical storage versus computation. *Linguistics*, 35, 167–181.
- Arvaniti, A. (2011). The representation of intonation. In M. van Oostendorp, C. J. Ewen, E. V. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology: Vol. 2. Suprasegmental and prosodic phonology* (pp. 757–780). Malden, MA: Wiley-Blackwell.
- Arvaniti, A., & Ladd, D. R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, 26(1), 43–74.
- Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, 26, 3–25.
- Au, T. K., Knightly, L. M., Jun, S. A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science*, 13, 238–243.

- Au, T. K., Oh, J. S., Knightly, L. M., Jun, S.-A., & Romo, L. F. (2008). Salvaging a childhood language. *Journal of Memory and Language*, 58, 998–1011.
- Audacity Team (2011). Audacity (Version 1.3.14) [Computer program]. Retrieved from <http://audacity.sourceforge.net/>
- Backman, N. (1978). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage Studies Bulletin*, 4, 239–266.
- Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, 51(4), 317–342.
- Barnes, J., Veilleux, N., Brugos, A., Shattuck-Hufnagel, S. (2010). The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. *Speech Prosody 2010*, 100445, 1–4. Retrieved from <http://speechprosody2010.illinois.edu/papers/100445.pdf>
- Beckman, M. E. (1997). A typology of spontaneous speech. In Y. Sagisaka, N. Campbell, & N. Higuchi, (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 7–26). New York: Springer.
- Beckman, M. E., Díaz-Campos, M., McGory, J. T., & Morgan, T. A. (2002). Intonation across Spanish, in the tones and break indices framework. *Probus*, 14(1), 9–36.
- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.
- Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. *Phonetica*, 36, 373–383.

- Benmamoun, E., Montrul, S., & Polinsky, M. (2011). White paper: Prolegomena to heritage linguistics. National Heritage Language Resource Center. Retrieved from <http://nhlrc.ucla.edu/news/article.asp?parentid=122865>
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore: York Press.
- Best, C., & Tyler, M. (2007). Non-native and second language speech perception. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 15–34). Amsterdam: John Benjamins.
- Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, 14(4), 359–375.
- Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.25) [Computer program]. Retrieved from <http://www.praat.org/>
- Bolinger, D. (1978). Intonation across languages. In J. P. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language. Volume 2: Phonology* (pp. 471–524). Stanford: Stanford University Press.
- Bolinger, D. (1998). Intonation in American English. In D. Hirst, & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 45–55). Cambridge, U.K.: Cambridge University Press.

- Borràs-Comes, J., Vanrell, M. d. M., Prieto, P. (2010). The role of pitch range in establishing intonational contrasts in Catalan. *Speech Prosody 2010*, 100103:1-4
<http://speechprosody2010.illinois.edu/papers/100103.pdf>
- Botinis, A., Granström, B., & Möbius, B. (2001). Developments and paradigms in intonation research. *Speech Communication*, 33(4), 263–296.
- Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, 51(13), 2492–2511.
- Braun, B., Kochanski, G., Grabe, E., & Rosner, B. S. (2006). Evidence for attractors in English intonation. *Journal of the Acoustical Society of America*, 119(6), 4006–4015.
- Britain, D. (1992). Linguistic change in intonation: The use of high rising terminals in New Zealand English. *Language Variation and Change*, 4, 77-104.
- Broslow, E., Hurtig, K. K., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup, & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 350–361). Cambridge, MA: Newbury House.
- Bruce, G. (1977). *Swedish word accents in a sentence perspective*. Lund: Gleerup.
- Bruce, G., & Gårding, E., (1978). A prosodic typology for Swedish dialects. In: E. Gårding, G. Bruce & R. Bannert, R. (Eds.), *Nordic Prosody* (pp. 219–228). Lund: Gleerup.
- Burge, S., Laskowski, K., & Wölfel, M. (May, 2008). Comparative cross-domain study of the occurrence of laughter in meeting and seminar corpora. In Proceedings of the Sixth International Conference on Language Resources and Evaluation, Marrakech,

- Morocco. Retrieved from http://www.lrec-conf.org/proceedings/lrec2008/pdf/815_paper.pdf
- Carlisle, R. S. (2001). Syllable structure universals and second language acquisition. *International Journal of English Studies*, 1(1), 1–19.
- Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning*, 59(2), 367–409.
- Chen, A., Gussenhoven, C., & Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech*, 47(4), 311–350.
- Ching, M. K. L. (1982). The question intonation in assertions. *American Speech*, 57(2), 95–107.
- Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: Benjamins.
- Chun, D. M., Hardison, D. M., & Pennington, M. C. (2008). Technologies for prosody in context: Past and future of L2 research and practice. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 323–346). Amsterdam: J. Benjamins Publishing.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Cohen, A., & 't Hart, J. (1967). On the anatomy of intonation. *Lingua*, 19, 177–192.
- Cruttenden, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics*, 17(1), 77–91.
- Cruttenden, A. (1997). *Intonation*. Cambridge, U.K.: Cambridge University Press.

- Cruz-Ferreira, M. (1987). Non-native strategies for intonational meaning: An experimental study. In A. James & J. Leather, J. (Eds.), *Sound patterns in second language acquisition* (pp. 103–120). Dordrecht: Foris.
- Dankovičová, J., House, J., Crooks, A., & Jones, K. (2007). The relationship between musical skills, music training, and intonation analysis skills. *Language and Speech*, 50, 177–225.
- de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language & Speech*, 26(4), 331–350.
- de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, 16(1), 71–77.
- Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490.
- Dilley, L. C. (2005). *The phonetics and phonology of tonal systems* (Doctoral dissertation). Massachusetts Institute of Technology.
- Dilley, L. C. (2010). Pitch range variation in English tonal contrasts: Continuous or categorical? *Phonetica*, 67, 63–81.
- Dilley, L. C., & Brown, M. (2007). Effects of pitch range variation on f0 extrema in an imitation task. *Journal of Phonetics*, 35, 523–551.
- Dilley, L. C., Ladd, D. R., & Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics*, 33, 115–119.
- D'Imperio, M., Gili Fivela, B., & Niebuhr, O. (2010). Alignment perception of high intonational plateaux in Italian and German. In International Conference on Speech

- Prosody, Chicago, USA. In <http://hal.archives-ouvertes.fr/docs/00/54/19/71/PDF/4340.pdf>
- D'Imperio, M., & House, D. (1997). Perception of questions and statements in Neapolitan Italian. In G. Kokkinakis, N. Fakotakis, & E. Dermatas (Eds.), *Proceedings of Eurospeech*. Rhodes, Greece. Retrieved from http://www.iscspeech.org/archive/archive_papers/eurospeech_1997/e97_0251.pdf
- D'Imperio, M., Cavone, R., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.01226
- Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27, 315–330.
- E-Prime version 1.1 [Computer software]. Pittsburgh, PA: Psychology Software Tools.
- Face, T. L. (2001). Focus and early peak alignment in Spanish intonation. *Probus*, 13, 223–246.
- Face, T. L. (2002). Local intonational marking of Spanish contrastive focus. *Probus*, 14, 71–92.
- Face, T. L. (2003). Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics*, 2, 115–131.
- Face, T. (2010). The necessity of both naturally-occurring and elicited data in Spanish intonational phonology. *Studies in Hispanic & Lusophone Linguistics*, 3, 485–499.
- Face, T. L., & Prieto, P. (2006). Rising accents in Castilian Spanish: A revision of Sp_ToBI. *Journal of Portuguese Linguistics*, 5, 117–146.

- Ferreira, L. (2008). *High initial tones and plateaux in Spanish and Portuguese neutral declaratives: Consequences to the relevance of F0, duration and vowel quality as stress correlates* (Doctoral dissertation). University of Illinois, Champaign.
- Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly*, 15, 443–455.
- 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, 45–65.
- 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). Baltimore: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319–355). Berlin: Mouton de Gruyter.
- Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Briston, England: Multilingual Matters.
- Flege, J. E., Frieda, E. M., Walley, A. C., & Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition*, 20, 155–188.
- Flege, J. E., & Liu, S. (2001). The effect of experience on adults’ acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527–552.

- Flege, J. E., Schirru, C., & MacKay, I. R. A., (2003). Interaction between the native and second language subsystems. *Speech Communication*, 40, 467–491.
- Fletcher, J., Grabe, E., & Warren, P. (2005). Intonational variation in four dialects of English: The high rising tune. In S.-A. Jun (Ed.). Prosodic typology: *The phonology of intonation and phrasing*. Oxford: Oxford University Press.
- Fowler, C. A., & Galantucci, B. (2005). The relationship of speech perception and production. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 633–652). Malden, MA: Blackwell Pub.
- Frota, S., D'Imperio, M., Elordieta, G., Prieto, P., & Vigário, M. (2007). The phonetics and phonology of intonational phrasing in Romance. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.). *Segmental and prosodic issues in Romance phonology* (pp. 131–153). Amsterdam: John Benjamins Publishing.
- Gabriel, C., & Lleó, C. (2011). *Intonational phrasing in Romance and Germanic: Cross-linguistic and bilingual studies*. Amsterdam: John Benjamins Publishing Co.
- Gallardo del Puerto, F., Gómez Lacabex, E., & García Lecumberri, M. L. (August, 2007). The assessment of foreign accent by native and non-native judges. In *Proceedings of the Phonetics Teaching & Learning Conference, London, England*. Retrieved from http://www.phon.ucl.ac.uk/ptlc/proceedings/ptlcpaper_20e.pdf
- García-Lecumberri, M. L. (2003). Análisis por configuraciones: La escuela británica. In P. Prieto (Ed.), *Teorías de la entonación* (pp. 35–61). Barcelona: Ariel.
- Garrido, J. M., Llisterri, J., de la Mota, C., Ríos, A. (1993). Prosodic differences in reading style: Isolated vs. Contextualized Sentences. In *Eurospeech'93. 3rd European*

- Conference on Speech Communication and Technology, Berlin, Germany* (pp. 573–576). Retrieved from http://www.isca-speech.org/archive/archive_papers/eurospeech_1993/e93_0573.pdf
- German J. (2012). Dialect adaptation and two dimensions of tune. In *Proceedings of Speech Prosody*. Shanghai.
- Gerrits, E., & Schouten M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychology*, 66(3), 363–376.
- Godson, L. (2004). Vowel production in the speech of Western Armenian heritage speakers. *Heritage Language Journal*, 2. Retrieved from <http://hlj.ucla.edu/ViewPaper.ashx?ID=SdjoekCuvTH8%2fA8cm3E8Bw%3d%3d>
- Grabe, E., Roser, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, 46, 375–401.
- Grice, M., Baumann, S., & Jagdfeld, N. (2009). Tonal association and derived nuclear accents: The case of downstepping contours in German. *Lingua*, 119, 881–905.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3–15.
- Grosser, W. (1997). On the acquisition of tonal and accentual features of English by Austrian learners. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 211–228). Berlin: Mouton de Gruyter.
- Grover, C., Jamieson, D. G., & Dobrovolsky, M. B. (1987). Intonation in English, French and German: Perception and production. *Language & Speech*, 30(3), 277–295.

- Gussenhoven, C. (1999). Discreteness and gradience in intonational contrasts. *Language & Speech*, 42(2), 283–305.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. New York: Cambridge University Press.
- Gussenhoven, C. (2006). Experimental approaches to establishing discreteness in intonational contrasts. In S. Sudhoff et al. (Eds.), *Methods in empirical prosody research* (pp. 321–334). Berlin: Walter de Gruyter.
- Gutknecht, C. (1979). Intonation and language learning: The necessity for an integrative approach. *Studies in Second Language Acquisition*, 1(2), 25–36.
- Halliday, M. A. K. (1970). *A course in spoken English: Intonation*. Oxford: Oxford University Press.
- Henriksen, N. C., Geeslin, K. L., & Willis, E. W. (2010). The development of L2 Spanish intonation during a study abroad immersion program in León, Spain: Global contours and final boundary movements. *Studies in Hispanic and Lusophone Linguistics*, 3, 113–162.
- Henriksen, N. (2014). Initial peaks and final falls in the intonation of Manchego Spanish wh-questions. *Probus* 26(1), 83–133.
- Hermes, D. J. (2006). Stylization of speech contours. In S. Sudhoff et al. (Eds.), *Methods in empirical prosody research* (pp. 29–61). Berlin: Walter de Gruyter.
- Holden, K. T., & Hogan, J. T. (1993). The emotive impact of foreign intonation: An experiment in switching English and Russian intonation. *Language & Speech*, 36(1), 67–88.

- Horvath, B. (1986). *Variation in Australian English: The sociolects of Sydney*. Cambridge: Cambridge University Press.
- Howell, D. C. (2002). *Statistical methods for psychology*. Pacific Grove, CA: Duxbury/Thomson Learning.
- Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, 10(2), 163–169.
- Hualde, J. I. (2002). Intonation in Spanish and the other Ibero-Romance languages: Overview and status quaestionis. In C. R. Wiltshire & J. Camps, J. (Eds.), *Romance Phonology and Variation: Selected Papers from the 30th Linguistic Symposium on Romance Languages, Gainesville, Florida, February 2000* (pp. 101–115). Amsterdam: John Benjamins Pub.
- Hualde, J. I. (2003). El modelo Métrico y Autosegmental. In P. Prieto (Ed.), *Teorías de la entonación* (pp. 155–184). Barcelona: Ariel.
- Huang, B. H., & Jun S.-A. (2009). Age effect on the acquisition of second language prosody. In J. Chandlee, M. Franchini, S. Lord, & M. Rheiner (Eds.), *Online Supplement to Proceedings of the 33rd Boston University Conference on Language Development*. Retrieved from <http://www.bu.edu/linguistics/BUCLD/supplement33/Huang.pdf>
- Huensch, A. R. (2013). *The perception and production of palatal codas by Korean L2 learners of English* (Doctoral dissertation). University of Illinois, Champaign.

- Iwashita, N. (2006). Syntactic complexity measures and their relation to oral proficiency in Japanese as a foreign language. *Language Assessment Quarterly*, 2(4), 151–170.
- James, E. (1977). The acquisition of second-language intonation using a visualizer. *Canadian Modern Language Review*, 33(4), 503–506.
- Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent* (Doctoral dissertation). University of Stuttgart.
- Jilka, M. (2008). Different manifestations and perceptions of foreign accent in intonation. In J. Trouvain & U. Gut (Eds.), *Non-native prosody: Phonetic description and teaching practice* (pp. 77–90). Berlin: Mouton de Gruyter.
- Jun, S.-A., & Oh, M. (2000). Acquisition of second language intonation. In *Proceedings of International Conference on Spoken Language Processing*, Beijing, China, vol. 4 (pp. 76–79). Retrieved from www.linguistics.ucla.edu/people/jun/sun-ah.htm
- Kissling, E. (2012). *The effect of phonetics instruction on adult learners' perception and production of L2 sounds* (Doctoral dissertation). Georgetown University.
- Kondo-Brown , K. (Ed.). (2006). *Heritage language development: Focus on East Asian immigrants*. Amsterdam: Benjamins.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Kraljic, T., & Samuel, A.G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, 13(2), 262–268.

- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843.
- Ladd, D. R. (1983). Phonological features of intonation peaks. *Language*, 59(4), 721–759.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge; New York: Cambridge University Press.
- Ladd, D. R. (2000). Bruce, Pierrehumbert, and the elements of intonational phonology. In M. Horne (Ed.), *Prosody: Theory and experiment: Studies presented to Gösta Bruce* (pp. 37–50). Dordrecht: Kluwer Academic Publishers.
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, 25, 313–342.
- Ladd, D. R., & Schepman, A. (2003). “Sagging transitions” between high accent peaks in English: Experimental evidence. *Journal of Phonetics*, 31, 81–112.
- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.
- Lantolf, J. P. (1976). On teaching intonation. *Modern Language Journal*, 60(5), 267–274.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
- Leather, J. (1987). F0 pattern inference in the perceptual acquisition of second language tone. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 59–80). Dordrecht: Foris.

- Lee, C., Tao, L., & Bond, Z. S. (2010). Identification of acoustically modified Mandarin tones by non-native listeners. *Language & Speech*, 53(2), 217–243.
- Lehiste, I. (1963). [Review of the book *Acoustico-phonetic studies in the intonation of southern Swedish*, by K. Hadding-Koch]. *Language*, 39, 352–360.
- León, P. R., & Martin, P. (1972). Applied linguistics and the teaching of intonation. *Modern Language Journal*, 56(3), 139–144.
- Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning*, 39, 397–413.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3, 223–232.
- Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, 33(1), 37–63.
- Liberman, A. M., Harris, K. S., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and non-speech patterns. *Journal of Experimental Psychology*, 61, 379–388.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.

- Liberman, M. Y. (1975). *The intonational system of English* (Doctoral dissertation). Retrieved from <http://dspace.mit.edu/handle/1721.1/27376>
- Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length . In M. Aronoff & R. Oehrle (Eds.), *Language sound structure* (pp. 157–233). Cambridge, MA: MIT Press.
- Llisterri, J. (1992, July). Speaking styles in speech research. Paper presented at ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language, Dublin, Ireland. Retrieved from http://liceu.uab.es/~joaquim/publicacions/SpeakingStyles_92.pdf
- Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In *ICPhS 1995 Proceedings of the 13th International Congress of Phonetic Sciences, Vol. 4* (pp. 92–99).
- Llisterri, J., Marín, R., de la Mota, C., & Ríos, A. (1995). Factors affecting F0 peak displacement in Spanish. In *Eurospeech'95. 4th European Conference on Speech Communication and Technology, Madrid, Spain* (pp. 2061–2064). Retrieved from http://www.isca-speech.org/archive/archive_papers/eurospeech_1995/e95_2061.pdf
- MacKay, I., Flege, J., Piske, T., & Schirru, C. (2001). Category restructuring during second-language (L2) speech acquisition. *Journal of the Acoustical Society of America*, 110, 516–528.
- Major, R. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Major, R. C., & Kim, E (1996). The Similarity Differential Rate Hypothesis. *Language Learning*, 43, 465–496.
- Makarova, V. (2007). The effect of pitch peak alignment on sentence type identification in Russian. *Language and Speech*, 50, 385–422.
- Mayer, M. (1969). *Frog, where are you?* New York: Dial Press.
- McCarthy, J., & Prince, A. (1993). Prosodic morphology I: Constraint interaction and satisfaction. Unpublished manuscript, University of Massachusetts at Amherst and Rutgers University, New Brunswick, NJ.
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32(4), 543–563.
- Mennen, I. (2008). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.), *Non-native prosody: Phonetic description and teaching practice* (pp. 53–76). Berlin: Mouton de Gruyter.
- Mennen, I., Schaeffler, F., & Dickie C. (2014). Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition*, 36, 303–329.
- Mora, J. C. (2007). Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau, & A. Bel (Eds.), *Child language and child development: A portrait of the young in Multilingual Spain* (pp. 241–263). Clevedon, London: Multilingual Matters.
- Moyer, A. (1999) Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, 21, 81–108.

- Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, 17, 17–34.
- Nagle, C. (2014). *A longitudinal study on the role of lexical stress and motivation in the perception and production of L2 Spanish stop consonants* (Doctoral dissertation). Georgetown University.
- Nagle, C., Morales-Front, A., Moorman, C., & Sanz, C. (2015). Disentangling research on study abroad and pronunciation: Methodological and programmatic consideration. Manuscript submitted for publication.
- Navarro-Tomás, T. (1932). *Manual de pronunciación española*. Madrid: Consejo superior de investigaciones científicas.
- Navarro-Tomás, T. (1944). *Manual de entonación española*. New York: Hispanic Institute in the United States.
- Nibert, H. J. (2000). *Phonetic and phonological evidence for intermediate phrasing in Spanish intonation* (Doctoral dissertation). University of Illinois, Champaign.
- Nibert, H. (2005). The acquisition of the phrase accent by intermediate and advanced adult learners of Spanish as a second language. In D. Eddington (Ed.), *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 108–122). Somerville, MA: Cascadilla Proceedings Project.
- Nibert, H. (2006). The acquisition of the phrase accent by beginning adult learners of Spanish as a second language. In M. Díaz-Campos (Ed.), *Selected Proceedings of the*

- 2nd *Laboratory Approaches to Spanish Phonetics and Phonology* (pp. 131–148) Somerville, MA: Cascadilla Proceedings Project.
- Nielsen, K. (2007). The interaction between spontaneous imitation and linguistic knowledge. *UCLA Working Papers in Phonetics*, 105, 125–137.
- Ohala, J. J. (2010). The relation between phonetics and phonology. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2nd ed., pp. 653–677). West Sussex, UK: Blackwell Publishing.
- Oller, J. W., & Ziahosseiny (1970). The contrastive analysis hypothesis and spelling errors. *Language Learning*, 20, 183–189.
- O'Rourke, E. (2012). Intonation in Spanish. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *The handbook of Hispanic linguistics* (pp. 173–192). Malden, MA: Wiley.
- Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners* (Doctoral dissertation). University of Hawaii at Manoa.
- Ortega, L. (2009). *Understanding Second Language Acquisition*. London: Routledge.
- Ortega, L., Iwashita, N., Rabie, S., & Norris, J. (1999, August). *A multilanguage comparison of L2 syntactic complexity measures*. Paper presented at the 12th AILA (World Congress of International Association of Applied Linguistics) Conference, Tokyo.
- Ortega-Llebaria, M., & Colantoni, L. (2014). L2 English intonation: Relations between form-meaning associations, access to meaning, and L1 transfer. *Studies in Second Language Acquisition*, 36, 331–353.

- Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon, UK: Multilingual Matters.
- Pennington, M., & Richards, J. (1986). Pronunciation revisited. *TESOL Quarterly*, 20, 207–225.
- Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association* (Interspeech 2011) (pp. 161-164). Rundle Mall: Causal Productions.
- Petrone, C., & D'Imperio, M. (2009): Is tonal alignment interpretation independent of methodology? *INTERSPEECH-2009*, 2459-2462. Retrieved from http://www.isca-speech.org/archive/interspeech_2009/i09_2459.html
- Pickering, L. (2009). Intonation as a pragmatic resource in ELF interaction. *Intercultural Pragmatics*, 6(2), 235–255.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation). Retrieved from dspace.mit.edu/handle/1721.1/16065
- Pierrehumbert, J. (1990). Phonological and phonetic representations. *Journal of Phonetics*, 18, 375–394.
- Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and experiment: Studies presented to Gösta Bruce* (pp. 11–36). Dordrecht: Kluwer Academic Publishers.

- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning and intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.
- Pierrehumbert, J., & Steele, S. (1989). Categories of tonal alignment in English. *Phonetica*, 46, 181–196.
- Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Piske, T., MacKay, I. R. A., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215.
- Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. (2011). Investigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010 Vol. II* (pp. 195–205). Bern, Switzerland: Peter Lang.
- Prieto, P. (1998). The scaling of the L values in Spanish downstepping contours. *Journal of Phonetics*, 26(3), 261–282.
- Prieto, P. (2003). Teorías lingüísticas de la entonación. In P. Prieto (Ed.), *Teorías de la entonación* (pp. 13–33). Barcelona: Ariel.
- Prieto, P., D'Imperio, M., & Gili-Fivela, B. (2005). Pitch accent alignment in Romance: Primary and secondary associations with metrical structure. *Language and Speech*, 48, 359–396.

- Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, 24(4), 445–473.
- Prieto, P., & Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics*, 35(4), 473–500.
- Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23(4), 429–451.
- Prince, A., & Smolensky, P. (1993). Optimality Theory: Constraint interaction in a generative grammar. Unpublished manuscript. New Brunswick, NJ: Rutgers University Center for Cognitive Science.
- Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, 30, 271–287.
- Queens, R. (2012). Turkish-German bilinguals and their intonation: Triangulating evidence about contact-induced language change. *Language*, 88(4), 791–816.
- Quilis, A. (1975). Las unidades de entonación. *Revista Española de Lingüística*, 5(2), 261–280.
- Quilis, A. (1993). *Tratado de fonología y fonética españolas*. Biblioteca románica hispánica, 74. Madrid: Editorial Gredos.
- Ramírez Verdugo, D. (2005). The nature and patterning of native and non-native intonation in the expression of certainty and uncertainty: Pragmatic effects. *Journal of Pragmatics*, 37(12), 2086–2115.

- Redi, L. (2003). Categorical effects in production of pitch contours in English. In *Proceedings of the 15th International Congress of the Phonetic Sciences* (pp. 2921–2924). Barcelona.
- Remez, R. E. (2005). Perceptual organization of speech. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 28–50). Malden, MA: Blackwell Pub.
- Rossi, M. (2000). Intonation: Past, present, future. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology* (pp. 13–52). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Sánchez, L. (2003), *Quechua/Spanish bilingualism. Interference and convergence in functional categories*. Amsterdam: John Benjamins.
- Schouten, B., Gerrits, E., & van Hessen (2003). The end of categorical perception as we know it. *Speech Communication*, 41, 71–80.
- Silverman , K. E. A., & Pierrehumbert , J. B. (1990). The timing of pre-nuclear high accents in English. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I* (pp. 72–106). Cambridge: Cambridge University Press.
- Simonet, M. (2008). *Language contact in Majorca: An experimental sociophonetic approach* (Doctoral dissertation). University of Illinois, Champaign.
- Simonet, M. (2009) Nuclear pitch accents in Majorcan Catalan declaratives: Phonology, phonetics, diachrony. *Studies in Hispanic and Lusophone Linguistics*, 2, 77–115.
- Simonet, M. (2010). A contrastive study of Catalan and Spanish declarative intonation: Focus on Majorcan Spanish. *Probus*, 22, 117–148.

- Simonet, M. (2011). Intonational convergence in language contact: Utterance-final contours in Catalan-Spanish bilinguals. *Journal of the International Phonetic Association*, 41, 185–205.
- So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language & Speech*, 53(2), 273–293.
- Sosa, J. M. (1999). *La entonación del español: Su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra.
- Sosa, J. M. (2003). La notación tonal del español en el modelo Sp-ToBI. In P. Prieto (Ed.), *Teorías de la entonación* (pp. 185–208). Barcelona: Ariel.
- Stockwell, R. P., Bowen, J. D., & Silva-Fuenzalida, I. (1956). Spanish juncture and intonation. *Language*, 32(4), pp. 641-665.
- 't Hart, J., Collier, R., & Cohen, A. (1990). A perceptual study of intonation: An experimental-phonetic approach to speech melody. Cambridge, UK: Cambridge University Press.
- Teague, K. (2011). *The interaction of lexical and discourse-level categories in second language phonetics* (Doctoral Dissertation). Georgetown University.
- Tesar, B., & Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, 29(2), 229–268.
- Tesar, B., & Smolensky, P. (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

- Tracy-Ventura, N., McManus, K., Norris, J. M., & Ortega, L. (2014). ‘Repeat as much as you can’: Elicited imitation as a measure of global proficiency in L2 French. In L. P. Lelercq, A. Edmonds, & H. Hilton (Eds.). *Proficiency assessment issues in SLA research: Measures and practices* (pp. 143–166). Bristol: Multilingual Matters.
- Trager, G. L., & Smith, H. L. (1951). *An outline of English structure*. Norman, Oklahoma: Battenburg Press.
- Vaissière, J. (2005). Perception of intonation. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 236–263). Malden, MA: Blackwell Pub.
- Veilleux, N., Barnes, J., Shattuck-Hufnagel, S., & Brugos, A. (2009, September). *Perceptual robustness of the tonal center of gravity for contour classification*. Poster presented at the Workshop on Prosody and Meaning, Univeristy Pompeu Fabra, Barcelona.
- Vilaplana, E. E., & Prieto Vives, P. (2008). La notación prosódica del español: Una revisión del Sp_ToBI. *Estudios de Fonética Experimental, XVII*, 263–283. Retrieved from <http://www.ub.edu/labfon/XVII-15.pdf>
- Waltmunson, J. (2006). *The relative degree of difficulty of L2 Spanish /d, t/, trill and tap by L1 English speakers: Auditory and acoustic methods of defining pronunciation accuracy* (Doctoral dissertation). University of Washington.
- Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, 4, 123-130.
- Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, 15(4), 399–420.
- Wu, S.-L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, 46(4), 680–704.

- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46, 220–251.
- Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, 38, 329–336.
- Xu, Y. (2011). Speech prosody: A methodological review. *Journal of Speech Sciences*, 1(1), 85–115.