Northwestern University

The Acquisition of English Focus Marking by Non-Native Speakers

A DISSERTATION
SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
for the degree
DOCTOR OF PHILOSOPHY
Field of Linguistics

By
Rachel Elizabeth Baker
EVANSTON, ILLINOIS
2010

ABSTRACT

The Acquisition of English Focus Marking by Non-Native Speakers

Rachel Elizabeth Baker

Second language learners experience difficulties mastering the various linguistic systems of their new language (L2), which may differ from the systems of their native language (L1). Correctly producing and understanding focus marking in a new language may be particularly challenging because it can require knowledge of several of these systems, including phonetics, phonology, syntax, semantics, and pragmatics. This study examines Mandarin and Korean speakers' acquisition of English prosodic focus marking.

In this study, 20 native English speakers, 20 native Mandarin speakers, and 20 native Korean speakers participated in four experiments: 1) a production experiment, in which they were recorded reading the answers to questions, 2) a perception experiment, in which they were asked to determine which word in a recording was the last prominent word, 3) an understanding experiment, in which they were asked whether the answers in recorded question-answer pairs had context-appropriate prosody, and 4) a pitch accent placement experiment, in which they were asked which word they would make prominent in a particular context. Finally, a new group of native English speakers listened to utterances produced in the production experiment, and determined whether the prosody of each utterance was appropriate for its context.

Based on the results of the five experiments, I propose a predictive framework for second language prosodic focus marking acquisition. This framework holds that both L1 transfer and features of the L2 itself affect language learners' acquisition of prosodic focus marking, so it includes two complementary models: the Transfer Model and the L2 Challenge Model. The Transfer Model predicts that prosodic structures in the L2 will be more easily acquired by

language learners that have similar structures in their L1 than those who do not, even if there are differences between the L1 and L2 in how the structures are realized. The L2 Challenge Model predicts that for hard tasks, language learners will rely on common prosodic patterns, making them more successful at prosodically marking broad focus than narrow focus. However, for easy tasks, language learners will more successfully mark information structures that have a more direct relationship between focus and accent placement.

ACKNOWLEDGEMENTS

Many people helped me to complete this dissertation. Ann Bradlow provided invaluable advice, feedback, and encouragement throughout this project. Brady Clark and Janet Pierrehumbert have also given me very helpful guidance on issues ranging from the questions that should be addressed to stimulus design to references in the literature. More generally, the faculty and staff of the Northwestern Linguistics Department have helped me reach this point though their dedicated teaching and approachability. I would particularly like to thank Chun Chan for his assistance with all manner of technical challenges, and Matt Goldrick for his statistical guidance. Of course, my fellow Ph.D. students have helped me both practically and emotionally for the last six years.

My family has also provided great support as I worked towards this goal. Patrick Quinan has seen me through all the highs and lows that make up the Ph.D. lifestyle. He has been endlessly supportive and loving. My parents, Rona and Gus Baker, have provided great role models and always supported my decisions. I would also like to thank my grandfather, Stan Baker, for his generous support of my education throughout my life.

TABLE OF CONTENTS

LIST OF TABLES 11

LIST OF FIGURES...17

Chapter 1: Introduction and Literature Review…………………………………………........20
1.1 Introduction………………………………………………………………………20
1.2 Significance of the Research…..……………………………………………….22
1.3 Review of the Literature…………………………………………………………..25
1.3.1 Typology of Focus Marking……………………………………………..25
1.3.2 English Prosody and Focus Marking……………………………………26
1.3.2.1 English Prosody………………………………………………26
1.3.2.2 English Focus Marking………………………………………..27
1.3.3 Mandarin Prosody and Focus Marking…………………………………..32
1.3.4 Korean Prosody and Focus Marking……………………………………..35
1.3.5 Do English, Mandarin, and Korean have the same Focus Marking System?..36
1.3.6 Discourse-Level Prosody Production in Non-Native English…………..40
1.3.7 Prosody Perception in Non-Native English……………………………...42
1.3.8 Relationship between Perception and Production………………………44
1.4 Previous Models of Second Language Acquisition………………………………46
1.4.1 Transfer Models of Phonological and Phonetic Acquisition……………46
1.4.2 Non-Transfer Models of Morpho-Syntactic Acquisition………………...50
1.5 Two Types of Models of English Prosodic Focus Marking Acquisition…………....51
1.5.1 Transfer Models…………………………………………………….....54
1.5.1.1 Comparisons between L1s and L2 for Transfer Models………....54
1.5.1.2 Transfer Model Predictions for Pitch Accent Perception and Realization………………………………………………………55
1.5.1.3 Transfer Model Predictions for Pitch Accent Understanding and Placement……………………………………………………...56
1.5.1.3.1 Opposing Transfer Models…………………………..56
1.5.1.3.2 Different Prominence Locations Hurt………………..58
1.5.1.3.3 Any Prominence Location Helps……………………59
1.5.2 L2 Challenge Models…………………………………………………60
1.5.2.1 L2 Challenge Model Predictions for Pitch Accent Perception and Realization………………………………………………………61
1.5.2.2 L2 Challenge Models' Predictions for Pitch Accent Understanding and Placement…………………………………………………62
1.5.3 Summary and Comparison of Model Predictions………………………..64
1.6 Native Perception of Non-Native Production……………………………………..65
1.6.1 Strict Native Perception Model…………………………………………..65
1.6.2 Relaxed Native Perception Model………………………………………66
1.6.3 Summary and Comparison of Model Predictions………………………..67

1.7 Relationship between Perception and Production……………………………………67
1.7.1 Perception/Production Dependence Model……………………………....68
1.7.2 Perception/Production Independence Model……………………………69
1.7.3 Summary and Comparison of Model Predictions…………………….....70

Chapter 2: Methods……………………………………………………………………………71
2.1 Introduction………………………………………………………………………….71
2.2 General Experiment Structure……………………………………………………….72
2.3 General Materials…………………………………………………………………….73
2.4 Participants…………………………………………………………………………...76
2.5 Experiment 1: Prominence Production Experiment…………………………………79
2.5.1 Experiment 1: Prominence Production Experiment Materials…………..79
2.5.2 Experiment 1: Prominence Production Experiment Design……………..79
2.6 Experiment 2: Prominence Perception Experiment……………………………….....80
2.6.1 Experiment 2: Prominence Perception Experiment Materials…………...80
2.6.2 Experiment 2: Prominence Perception Experiment Design……………..83
2.7 Experiment 3: Prominence Understanding Experiment……………………………..84
2.7.1 Experiment 3: Prominence Understanding Experiment Materials……....84
2.7.2 Experiment 3: Prominence Understanding Experiment Design…………85
2.8 Experiment 4: Prominence Placement Experiment…………………………………86
2.8.1 Experiment 4: Prominence Placement Experiment Materials…………...86
2.8.2 Experiment 4: Prominence Placement Experiment Design……………...86
2.9 Experiment 5: Native Perception of Non-Native Production Experiment……….....87
2.9.1 Experiment 5: Native Perception of Non-Native Production Participants…………………………………………………………………87
2.9.2 Experiment 5: Native Perception of Non-Native Production Experiment Materials……………………………………………………………….....88
2.9.3 Experiment 5: Native Perception of Non-Native Production Experiment Design…………………………………………………………………88
2.10 Statistical Analysis……………………………...……………………………..89

Chapter 3: Non-native Perception and Understanding of English Prosodic Prominence - Experiments 2 and 3…………………………………………………………………………...94
3.1 Introduction………………………………………………………………………….94
3.2 Experiment 2: Prominence Perception Experiment…………………………………96
3.2.1 Experiment 2: Prominence Perception Experiment Analyses…………...96
3.2.2 Experiment 2: Prominence Perception Native/Non-Native Regressions...98
3.2.3 Experiment 2: Prominence Perception Non-Native Regressions………100
3.2.4 Experiment 2: Prominence Perception Discussion……………………..101
3.3 Experiment 3: Prominence Understanding Experiment……………………………102
3.3.1 Experiment 3: Prominence Understanding Experiment Analyses……...102
3.3.2 Experiment 3: Prominence Understanding Context Regressions……....103
3.3.2.1 Experiment 3: Prominence Understanding Native/Non-Native Context Regressions……………………………………………103

3.3.2.2 Experiment 3: Prominence Understanding Non-Native Context Regressions……………………………………………………..105
3.3.3 Experiment 3: Prominence Understanding Accent Location Regressions…………………………………………………………..107
3.3.3.1 Experiment 3: Prominence Understanding Native/Non-Native Accent Location Regressions…………………………………...107
3.3.3.2 Experiment 3: Prominence Understanding Non-Native Accent Location Regressions………………………………………...109
3.3.4 Experiment 3: Prominence Understanding Discussion………………...112
3.4 Experiments 2 and 3: Prominence Perception and Understanding General Discussion………………………………………………………………118

Chapter 4: Non-native Placement and Realization of English Prosodic Prominence - Experiments 1, 4, and 5……………………………………………………………………………119
4.1 Introduction…………………………………………………………………...119
4.2 Experiment 4: Prominence Placement Experiment………………………………...120
4.2.1 Experiment 4: Prominence Placement Analysis………………………..120
4.2.2 Experiment 4: Prominence Placement Native/Non-Native Regressions.120
4.2.3 Experiment 4: Prominence Placement Non-Native Regressions……....123
4.2.4 Experiment 4: Prominence Placement Error Analysis………………....123
4.2.5 Experiment 4: Prominence Placement Discussion……………………..127
4.3 Experiments 1 and 5: Native Perception of Non-Native Production Experiment….131
4.3.1 Experiments 1 and 5: Native Perception of Non-Native Production Analysis………………………………………………………………...131
4.3.2 Experiments 1 and 5: Native Perception of Non-Native Production Native/Non-Native Regressions………………………………………...131
4.3.3 Experiments 1 and 5: Native Perception of Non-Native Production Non-Native Regressions………………………………………………...134
4.3.4 Experiments 1 and 5: Native Perception of Non-Native Production Discussion………………………………………………………………137
4.4 Prominence Placement and Realization General Discussion………………………139

Chapter 5: Acoustic Analysis of Production Recordings – Experiment 1……..………………143
5.1 Purpose of Acoustic Analysis……………………………………………………143
5.2 Selection of Recordings for Acoustic Analysis……………………………………144
5.3 Acoustic Labeling and Measurements…………………………………………...145
5.4 Non-Native SuNF Sentences with Inappropriate Prosody…………………………146
5.4.1 Phonological Features…………………………………………………..146
5.4.2 Phonetic Features………………………………………………………149
5.4.2.1 Statistical Analysis……………………………………………...149
5.4.2.2 Analysis of Duration……………………………………………150
5.4.2.3 Analysis of RMS Amplitude…………………………………...153
5.4.2.4 Analysis of F0 Maximum………………………………………156
5.4.2.5 Analysis of F0 Range…………………………………………..159

5.5 Acoustic Differences between Native and Non-Native Productions with Appropriate Prosody……………………………………………………………………………..163
5.5.1 Phonological Features……………………………………………………..163
5.5.2 Phonetic Features…………………………………………………………165
5.5.2.1 Statistical Analysis…………………………………………...165
5.5.2.2 Analysis of Duration…………………………………………....166
5.5.2.3 Analysis of RMS Amplitude…………………………………....172
5.5.2.4 Analysis of F0 Maximum………………………………………175
5.5.2.5 Analysis of F0 Range…………………………………………...179
5.6 Acoustic Analysis of Production Discussion………………………………………184
5.6.1 Discussion of Inappropriate Non-Native SuNF Productions…………...184
5.6.2 Discussion of Appropriate Native and Non-Native Productions……….185
5.6.3 Acoustic Analysis of Production Summary……………………………187

Chapter 6: Relationship between Perception and Production in the Prosodic Prominence of Non-Native English Speakers……………………………………………………………….….189
6.1 Introduction……………………………………………………………………...189
6.2 Correlations between Prominence Placement (Exp. 4) and Understanding(Exp. 3).190
6.3 Correlations between Prominence Production (Exp. 1 and 5) and Understanding (Exp. 3)……………………………………………………………………………...192
6.4 Correlations between Prominence Production (Exp. 1 and 5) and Perception (Exp. 2)…………………………………………………………………………194
6.5 Discussion of Prominence Placement, Production, Understanding, and Perception Correlations………………………………………………………………………195

Chapter 7: Conclusions and Applications……………………………………………………200
7.1 Introduction……………………………………………………………………...200
7.2 Summary of Results………………………………………………………..………200
7.2.1 Summary of Results for Experiment 2: The Prominence Perception Experiment…………………………………………………………………...200
7.2.2 Summary of Results for Experiment 3: The Prominence Understanding Experiment…………………………………………………………………...201
7.2.3 Summary of Results for Experiment 4: The Prominence Placement Experiment…………………………………………………………………...201
7.2.4 Summary of Results for Experiment1: The Prominence Production Experiment…………………………………………………………………...202
7.2.5 Summary of Perception/Production Correlations……………………...203
7.3 Predictive Framework for Studying Prosodic Prominence Acquisition……………203
7.3.1 Predictive Framework for Studying Prosodic Prominence Acquisition: Perception and Realization……………………………………………203
7.3.2 Predictive Framework for Studying Prosodic Prominence Acquisition: Understanding and Placement………………………………………….205
7.4 Native Perception of Non-Native Pitch Accent Productions………………………212
7.5 Perception and Production in Prosodic Prominence Acquisition…………………..213
7.6 General Implications for English Prosodic Prominence Education………………..214

7.6.1 The Role of L1 in Prosodic Prominence Education……………………….213
7.6.2 The Relationship between Perception and Production in English Prominence Education…………………………………………………….216
7.6.3 Teaching English Prominence Understanding……………………………216
7.6.4 Teaching English Prominence Production………………………………..217
7.7 Conclusions……………………………………………………………………217

References……………………………………………………………………………219

Appendix A: Experiment Stimulus Sentences...………………………………………………..226

Appendix B: Experiment Instructions……..……………………………………………………..228

LIST OF TABLES

Table 1.1. Model predictions for pitch accent perception and realization

Table 1.2. Model predictions for pitch accent understanding and placement

Table 1.3. Model predictions for native perception of non-native production of English prosodic focus marking

Table 1.4. Model predictions for the relationship between perception and production in second language acquisition of prosodic focus marking

Table 2.1. Segmental features that were eliminated or controlled across the twelve target word sets

Table 2.2. Mean and range of participants' age-of-arrival to an English-speaking country (in years), and length of time spent in English-speaking countries (in months), broken down by native language

Table 3.1. Parameter values for fixed variables in the prominence perception native/non-native regression, with English as the language group baseline

Table 3.2. Parameter values for fixed variables in the prominence perception non-native regression, with Korean as the language group baseline

Table 3.3. Parameter values for fixed variables in the prominence understanding native/non-native context regressions, with English as the language group baseline and matched as the match baseline

Table 3.4. Parameter values for fixed variables in the prominence understanding non-native context regressions, with matched as the match baseline

Table 3.5. Parameter values for fixed variables in the prominence understanding non-native context regressions, with mismatched as the match baseline

Table 3.6. Parameter values for fixed variables in the prominence understanding native/non-native accent location regression, with matched as the match baseline, English as the language group baseline, and object as the accent location baseline

Table 3.7. Parameter values for fixed variables in the prominence understanding non-native accent location regression, with matched as the match baseline and object as the accent location baseline

Table 3.8. Parameter values for fixed variables in the prominence understanding non-native accent location regression, with mismatched as the match baseline and object as the accent location baseline

Table 3.9. Parameter values for fixed variables in the prominence understanding non-native accent location regression, with matched as the match baseline and subject as the accent location baseline

Table 4.1. Parameter values for fixed variables in the prominence placement native/non-native regressions, with English as the language group baseline

Table 4.2. Parameter values for fixed variables in the prominence placement non-native regressions, with Korean as the language group baseline

Table 4.3. Parameter values for fixed variables in the prominence placement non-native regressions, with Mandarin as the language group baseline

Table 4.4. Mean percentages of responses by language group and context. The shaded squares indicate the 'correct' answer for each context.

Table 4.5. Parameter values for fixed variables in the native perception of non-native production native/non-native regressions, with English as the language group baseline

Table 4.6. Parameter values for fixed variables in the native perception of non-native production non-native regressions, with Korean as the language group baseline

Table 4.7. Parameter values for fixed variables in the native perception of non-native production non-native regressions, with Mandarin as the language group baseline

Table 5.1. Number of productions classified as prosodically appropriate, by language group and context

Table 5.2. Number of productions classified as prosodically inappropriate, by language group and context

Table 5.3. Parameter values for fixed variables in the subject duration regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.4. Parameter values for fixed variables in the verb duration regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.5. Parameter values for fixed variables in the object duration regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.6. Parameter values for fixed variables in the subject RMS amplitude regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.7. Parameter values for fixed variables in the verb RMS amplitude regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.8. Parameter values for fixed variables in the object RMS amplitude regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.9. Parameter values for fixed variables in the subject F0 maximum regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.10. Parameter values for fixed variables in the verb F0 maximum regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.11. Parameter values for fixed variables in the object F0 maximum regression for ‘ appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.12. Parameter values for fixed variables in the subject F0 range regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.13. Parameter values for fixed variables in the verb F0 range regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.14. Parameter values for fixed variables in the object F0 range regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

Table 5.15. Parameter values for fixed variables in the subject duration regressions on prosodically appropriate productions

Table 5.16. Parameter values for fixed variables in the verb duration regressions on prosodically appropriate productions, with nativeness as the native baseline

Table 5.17. Parameter values for fixed variables in the object duration regressions on prosodically appropriate productions, with nativeness as the native baseline

Table 5.18. Parameter values for fixed variables in the subject RMS amplitude regressions on prosodically appropriate productions

Table 5.19. Parameter values for fixed variables in the verb RMS amplitude regressions on prosodically appropriate productions

Table 5.20. Parameter values for fixed variables in the object RMS amplitude regressions on prosodically appropriate productions, with native as the nativeness baseline

Table 5.21. Parameter values for fixed variables in the subject F0 maximum regressions on prosodically appropriate productions

Table 5.22. Parameter values for fixed variables in the verb F0 maximum regressions on prosodically appropriate productions, with native as the nativeness baseline

Table 5.23. Parameter values for fixed variables in the object F0 maximum regressions on prosodically appropriate productions, with native as the nativeness baseline

Table 5.23. Parameter values for fixed variables in the subject F0 range regressions on prosodically appropriate productions, with native as the nativeness baseline

Table 5.24. Parameter values for fixed variables in the subject F0 range regressions on prosodically appropriate productions, with native as the nativeness baseline

Table 5.25. Parameter values for fixed variables in the verb F0 range regressions on prosodically appropriate productions, with native as the nativeness baseline

Table 5.26. Parameter values for fixed variables in the object F0 range regressions on prosodically appropriate productions, with native as the nativeness baseline

LIST OF FIGURES

Figure 1.1. Chart demonstrating the three levels of prosodic focus marking and the relationships between them

Figure 2.1. Boxplot of overall Versant scores for Korean and Mandarin participants

Figure 2.2. Pitch contour labeled with word boundary and pitch accent locations for the SuNF version of a typical stimulus sentence used in the prominence perception experiment (Exp. 2)

Figure 2.3. Pitch contour labeled with word boundary and pitch accent locations for the VPBF version of a typical stimulus sentence used in the prominence perception experiment (Exp. 2)

Figure 2.4. Pitch contour labeled with word boundary and pitch accent locations for the SBF version of a typical stimulus sentence used in the prominence perception experiment (Exp. 2)

Figure 3.1 Boxplots showing proportions of correctly identified accent locations by language group, match, and context

Figure 3.2. Boxplots showing proportions of correctly interpreted accents by language group, match, and context

Figure 3.3. Boxplots showing proportions of correctly interpreted items by language group, match, and context

Figure 3.4. Scatterplots showing relationship between accuracy and Versant score for prominence understanding non-native subject and object items

Figure 4.1. Boxplots showing proportions of correct predictions of accent placement by language group and context

Figure 4.2. Scatterplot of participants' proportions of correct accent placement by their Versant scores

Figure 4.3 Boxplots of proportion of productions perceived to have context-appropriate prosody, by talker native language and context

Figure 5.1. Barplots showing proportions of non-native SuNF sentences perceived to have context-inappropriate prosody and non-native SuNF, VPBF, and SBF sentences perceived to have context-appropriate prosody. Bars in dark gray represent accent patterns that are predicted to be acceptable for VPBF contexts, bars in light gray represent accent patterns that are predicted to be acceptable for SuNF contexts.

Figure 5.2. Line graph showing the mean durations in seconds for the subject, verb, and object, in inappropriate SuNF productions and appropriate SuNF, VPBF, and SBF productions by non-native speakers

Figure 5.3. Line graph showing the mean RMS amplitudes in dB for the subject, verb, and object in inappropriate SuNF productions and appropriate SuNF, VPBF, and SBF productions by non-native speakers

Figure 5.4. Line graph showing the mean F0 maxima in Hz for the subject, verb, and object in inappropriate SuNF productions and appropriate SuNF, VPBF, and SBF productions by non-native speakers

Figure 5.5. Line graph showing the mean F0 ranges in Hz for the subject, verb, and object in inappropriate SuNF productions and appropriate SuNF, VPBF, and SBF productions by non-native speakers

Figure 5.6. Barplots showing proportions of prosodically appropriate sentences in each context

produced with each accent pattern by native and non-native English speakers. Bars in dark gray represent accent patterns that are predicted to be acceptable for VPBF contexts, bars in light gray represent accent patterns that are predicted to be acceptable for SuNF contexts, bars in black represent patterns that are not predicted to be acceptable in any of these three contexts.

Figure 5.7. Boxplots showing the object durations in seconds in appropriate SuNF, VPBF and SBF productions by native and non-native speakers

Figure 5.8. Boxplots showing the object F0 range values in Hz in appropriate SuNF, VPBF and SBF productions by native and non-native speakers

Figure 6.1. Scatterplots comparing individual English learners' accuracy at placing prominence (Exp. 4) to their accuracy at understanding the meaning of prominence location (Exp. 3), both overall and for the SuNF, VPBF, and SBF contexts

Figure 6.2. Scatterplots comparing individual English learners' accuracy at producing context-appropriate prosody (Exp. 1 and 5) to their accuracy understanding the meaning of prominence location (Exp. 3), both overall and for the SuNF, VPBF, and SBF contexts

Figure 6.3. Scatterplot comparing individual English learners' overall accuracy at producing context-appropriate prosody (Exp. 1 and 5) to their accuracy at perceiving the location of prominence (Exp. 2)

Chapter 1

Introduction and Literature Review

1.1 Introduction

Second language learners experience difficulties mastering the various linguistic systems of their new language, from syntax to phonetics, which may differ from the systems of their native language. Correctly producing and understanding focus marking in a new language may be particularly challenging because it can require knowledge of several of these systems, including phonetics, phonology, syntax, and pragmatics. Focus allows speakers to highlight words and phrases that are new or otherwise informative. Focus is part of the information structure of a sentence, which describes how information in the sentence relates to the discourse as a whole. This study examines Mandarin and Korean speakers' acquisition of English prosodic focus marking.

Different languages mark focus in different ways, through prosodic, syntactic, and morphological methods (Büring, 2009). Even if two languages both use prosody to mark focus, they can differ in how they mark focus on larger constituents, such as verb phrases and sentences. In these cases, some part of the focused constituent may be prosodically prominent, or focus may not be marked at all. Different languages can also make different words prominent within these larger focused constituents. Finally, different languages can use different phonological and phonetic cues to mark a word as prominent. For instance, English uses pitch accents, Korean puts focused words at the beginnings of accentual phrases and removes following phrase boundaries, and Mandarin expands the pitch range of focused words and decreases the pitch range of following words. English focus marking is a complex system because, although there is a relationship between focus and pitch accent placement, words that are in focus do not always receive a pitch accent, and some words that receive pitch accents are not in focus (Ladd, 1996). Non-native speakers of English must learn this complex system largely implicitly, as very little ESL instruction targets discourse-level accentuation (Celik 2005). The acquisition of prosodic focus marking in a second language is a particularly difficult challenge because it requires the learner to discover both which words to emphasize in a particular context and what acoustic cues to use to mark emphasis.

This study seeks to answer the following three questions:

I. What factors affect the relative difficulty that English learners have with accurately producing and perceiving English focus marking?

II. Can an English learner deviate from native-like pitch accent realization and still have his pitch accents accurately perceived by native listeners?

III. Does an English learner's ability to accent the appropriate word for a particular information structure and to produce this accent in a native-like manner depend on the ability to accurately perceive and understand such accents?

The remainder of Chapter 1 includes a review of the relevant literature and a discussion of the frameworks in which these questions will be studied. Section 1.2 describes the significance of the proposed research. Section 1.3 reviews the literature on the typology of focus marking, the prosody and focus marking systems of English, Mandarin, and Korean, English prosody production and perception by non-native English speakers, and the relationship between production and perception in language acquisition. Section 1.4 discusses language acquisition models that have been proposed in the literature. Section 1.5 lays out a framework for testing models of prosodic prominence acquisition and describes the predictions of opposing models for the acquisition of prosodic focus marking. Section 1.6 discusses two models that make opposing predictions regarding native perception of non-native prosodic focus marking production. Section 1.7 describes two models that make opposing predictions about the relationship between perception and production in acquisition.

The dissertation is organized as follows. Chapter 2 describes the five experiments used to test the models of prosodic focus marking acquisition. Chapter 3 presents the results of the prosodic focus marking perception and understanding experiments, and discusses the implications of these results. Chapter 4 presents and discusses the results of the pitch accent placement and prosodic focus marking production experiments. Chapter 5 explains the acoustic analyses of the non-native speakers' productions. Chapter 6 presents and discusses data on the relationship between perception and production in prosodic focus marking acquisition. Chapter 7 contains a novel predictive model of second language prosodic focus marking acquisition based on the experimental results. It also discusses the results of the analyses on the perception of non-native prosodic focus marking by native English listeners and the perception/production relationship. Finally it describes the pedagogical implications of these results.

1.2 Significance of the Research

Second language learners experience difficulties achieving native-like comprehension and pronunciation in their new language. While the factors leading to problems perceiving and producing the segments of a new language have been well-studied (e.g. Hammarberg 1990; Flege 1995; Flege, Munro et al. 1995), less attention has been devoted to non-natives' use of prosodic features, which include pitch, duration, and amplitude. In English, these features are used to mark certain syllables as more prominent than others. At the word level, prosodic stress can be used to distinguish between two different words (such as the noun *record* – with first syllable stress, and the verb *record* – with second syllable stress). At the sentence and discourse levels, pitch accents play an important role in providing information about how an utterance fits into the larger context of a conversation. For instance, the appropriate response to the question "Who ate the pizza?" is "JOHN ate the pizza."[1] The pitch accent location in this sentence indicates that the focus is on *John,* which is the new information in the discourse and answers the preceding question. In contrast, the appropriate response to "What did John eat?" is "John ate the PIZZA.", with the pitch accent marking the focus on *pizza*. The acquisition of a new focus marking system is more complex than the acquisition of new segments because the acquisition of segments involves primarily the phonetic and phonological systems, while focus marking can involve the interplay between the phonetic, phonological, syntactic, and pragmatic systems.

It is important for non-native speakers to learn how accent placement relates to focus in English. Accenting the wrong word in a sentence can be confusing to the listener because it offers distorted information about which constituents are new or old in the discourse or what the topic of the conversation is. Appropriate focus marking makes a speaker's English easier to understand by providing listeners with more acoustic information about (accented) new items and encouraging listeners to map unaccented items to entities already in the discourse (Terken and Nooteboom 1988). Native listeners can process speech faster when it has context-appropriate accentuation than when it has inappropriate accentuation (Birch and Clifton 1995). Acquiring the ability to understand prosodic focus marking in English allows a listener to take advantage of these processing benefits. Non-native speakers' inappropriate accent placement can also contribute to generally unnatural prosody, which has its own negative consequences.

[1]Capitalization is used throughout this document as a convention to indicate the location of an obligatory pitch accent.

For example, flattened fundamental frequency has been shown to decrease intelligibility (Laures and Weismer 1999). In addition, non-native speakers with unnatural prosody are more likely to be judged by natives as more accented and less comprehensible (Munro and Derwing 1995; Derwing and Munro 1997).

The limited research on second language acquisition of focus marking has led to inadequate language teaching materials on this topic (Celik 2005). In fact, McGory (1997) suggests that some of the unnatural realizations of English pitch accents produced by non-native participants in her experiments could be due to inaccurate advice in their ESL textbooks on how to produce pitch accents. A thorough understanding of what makes acquisition of prosodic focus marking difficult will help in the development of training materials addressing these difficulties. Specifically, knowing which aspects of English focus marking (pitch accent placement, realization, perception, or understanding for particular words, sentence types, and information structures) cause the most difficulties for language learners with a particular L1, or for language learners in general, will allow teachers to focus on the problem areas. Similarly, a better understanding of the relationship between producing and perceiving pitch accents in English will help English teachers determine what types of training (perception and/or production) will be most useful for helping students use pitch accents to communicate effectively.

Other studies have examined particular aspects of non-native acquisition of English focus marking, but none that I know of has examined the complete communicative chain from non-native perception of native production, to non-native production, to native perception of non-native production. For instance, a number of studies have shown that non-natives mark English focus with different prosodic cues than native speakers (Wennerstrom 1994; McGory 1997; Schack 2000; Yeou 2004; Verdugo 2006; Aoyama and Guion 2007) but it is still unclear whether these differences lead to miscommunication with native speakers. A thorough examination of the full chain, involving a single set of participants, will give us a much clearer idea of the causes of non-natives' difficulties in acquiring English focus marking.

1.3 Review of the Literature

1.3.1 Typology of Focus Marking

All languages seem to mark focus in some way; the three methods used cross-linguistically are prosodic, syntactic, and morphological (Büring 2009). English and Korean are examples of languages that mark focus prosodically. English marks focus by placing accents on focused words or words in focused phrases (Gussenhoven 1999; Selkirk 1996; Schwarzschild 1999), and Korean marks focus by placing prosodic boundaries before focused words and removing boundaries after them (Jun and Lee 1998). Hungarian is an example of a language that marks focus syntactically, by placing focused noun phrases in pre-verbal position (Kenesei 2009). Chickasaw is an example of a language that marks focus morphologically, by adding suffixes to focused subjects and objects (Gordon 2008). Using one method to mark focus does not necessarily preclude using another method. In fact, Truckenbrodt's (1995) Prominence Theory of focus realization claims that focus is always marked by prosodic prominence, and that all methods of focus marking have as their goal making the focused segment maximally prominent. For instance, according to this theory, languages in which focus triggers syntactic movement use that movement to place focused constituents in prosodically prominent positions (Büring 2009). Mandarin speakers can use both prosodic and syntactic means to mark focus, marking it syntactically by placing focused words in sentence-final position, and marking it prosodically with larger F0 ranges and longer durations on focused words. However, in contrast to Truckenbrodt's theory, Xu (2004) claims that when focus is marked syntactically in Mandarin it is usually not marked prosodically.

Languages that mark focus prosodically can differ when it comes to marking broad focus on larger constituents, such as verb phrases (VPs) and whole sentences. In these cases, some portion of the focused constituent may be made prominent, or the focus may not be marked at all. Different languages can make different words prominent within broad focused constituents. In English, VP and sentence broad focus are often marked by accenting the last content word in the sentence (Gussenhoven 1999; Selkirk 1996; Schwarzschild 1999). In Korean, VP broad focus is marked prosodically with an expanded pitch range and longer duration on the first word in the VP (Jun, Kim, Lee, and Kim 2006). In Mandarin, broad focus on VPs and sentences is not marked prosodically (Xu 2004). Such differences between languages may lead to difficulties for learners trying to produce and understand focus in a new language.

1.3.2 English Prosody and Focus Marking

1.3.2.1 English Prosody

Within the commonly used autosegmental-metrical prosodic framework (Goldsmith 1976; Liberman and Prince 1977; Pierrehumbert 1980), English intonation is determined by pitch accents, phrase accents and boundary tones. Phrase accents and boundary tones are high or low intonational targets associated with the ends of intermediate phrases and the beginnings and ends of intonation phrases, respectively. An intonation phrase consists of one or more intermediate phrases. Pitch accents are local intonational events associated with particular syllables (Beckman and Pierrehumbert 1986). Every intermediate phrase must contain at least one pitch accent.

In English, pitch accents have particular pitch targets and are associated with stressed syllables, which can be marked by increased duration, higher peak and total amplitude, and spectral changes, making them stress accents (in contrast to non-stress accents, which consist only of pitch movements) (Beckman 1986). The pitch targets associated with pitch accents lead to F0 differences between accented and unaccented words. For example, syllables with H* pitch accents have higher nucleus midpoint and mean F0s, and those with L* pitch accents have lower mean F0s, relative to unaccented syllables (Beckman 1986, Shue et al. 2007). English pitch accents are assigned post-lexically, based on the information structure of the sentence (the relationship between the utterance and the rest of the discourse) and English-specific accent assignment rules (Ladd 1996). There are several different types of English pitch accents, consisting of a high tone (H*[2]), a low tone (L*), or a combination of the two (L+H*, L*+H, H+!H*[3]). English speakers use different pitch accent types to express different meanings (Büring 2003; Steedman 2007; Pierrehumbert and Hirschberg 1990). For instance, Steedman (2007) claims that H* pitch accents are used in phrases that provide new information on some topic, while L+H* pitch accents are used in phrases that refer back to the question or topic being discussed. Pierrehumbert and Hirschberg (1990) claim that different pitch accent types provide information on how the accented word should affect the hearer's beliefs. For instance, an H* pitch accent indicates that the accented item should be treated as new. An L+H* pitch accent evokes a scale of salient items that contrast with the accented item and indicates that a proposition that includes the accented item, rather than one of the other items on the scale, should be believed.

[2] * indicates the tone aligned with the stressed syllable in the accented word.
[3] H! represents a downstepped H tone.

1.3.2.2 English Focus Marking

A number of theories have been proposed on how to relate information structure to pitch accent placement in English (e.g. Büring 2006; Gussenhoven 1983a, 1999; Rooth 1992; Schwarzschild 1999; Selkirk 1996). These theories predict different accentuation patterns for some of the more complex information structures in English. However, they usually agree on the appropriate locations of pitch accents for the simple information structures examined in this study.

In this work, I will assume Gussenhoven's (1983a, 1999) theory of pitch accent placement, because of the empirical support it has received (Birch and Clifton 1995; Gussenhoven 1983b; Welby 2003). In this theory, accent placement is determined by focus domains. A focus domain is a structure that can be marked as entirely [+focus] (i.e. focused) with only one pitch accent (Gussenhoven 1983b). Note that Gussenhoven does not provide a strict definition of [+focus], but says that "[+focus] marks the speaker's declared contribution to the conversation" (Gussenhoven 1983a: 383). Focus domains are created by first dividing a sentence up into three types of semantic constituents: 1) arguments (A), including subjects and objects, 2) predicates (P), including verbs, and 3) modifiers (M), including adverbials (Gussenhoven 1999). The Sentence Accent Assignment Rule (SAAR), given in (1), states how these constituents can combine to form focus domains, and where accents should be placed within a focus domain (Gussenhoven 1983a). In this rule, underlined constituents are [+focus], and constituents without an underline are [-focus]. Square brackets indicate a focus domain, and a star indicates a pitch accent.

(1) SAAR

(a) Domain assignment: P(X)A → [P(X)A]
A(X)P → [A(X)P]
Y → [Y]

(b) Accent assignment: [ ] → [*]. In AP/PA focus domains, accent A.

In brief, the SAAR says that any A, P, or M that is [+focus] forms its own focus domain and receives an obligatory pitch accent. The one exception to this is that As and Ps which are adjacent, or separated only by a [-focus] constituent, can form a single focus domain with an accent on the A. In addition to the obligatory pitch accents assigned in focus domains, the pre-focal pitch accents rule (2) allows optional pitch accents before the nuclear (final) pitch accent in an utterance (Gussenhoven 1999).

(2) Prefocal Pitch Accents: Assign pitch accents to the constituents before the nuclear pitch accent. (Optional)

The examples in the paragraphs below show the relationship between focus and accent placement predicted by Gussenhoven's theory for the three types of information structures used in this study.

Before describing Gussenhoven's predictions for the utterance types used in this experiment, we should first clarify the term 'nuclear pitch accent', which seems to have two overlapping but distinct meanings in the literature. Many researchers use the phrase 'nuclear pitch accent' to describe the last pitch accent in an intermediate prosodic phrase (e.g. Beckman and Pierrehumbert 1986; Pierrehumbert 1980; Ladd 1996; Welby 2003). However, Gussenhoven (1999) seems to use the term to refer to the last pitch accent in a sentence or utterance. He considers this accent to be important because additional optional accents are only allowed before this accent, if the speaker does not want to change the interpretation of the utterance. Ladd (1996: 202) makes a similar point, claiming that the location of the "last accent of a phrase or utterance" determines the types of focus interpretations that are possible for that utterance. Much of the confusion between the two meanings may stem from the fact that if a sentence or utterance contains only one intermediate prosodic phrase (which is quite possible, especially for shorter sentences), then the two definitions pick out the same word. In addition, even if a sentence contains multiple intermediate phrases, the final pitch accent of the utterance will be the nuclear pitch accent for the final phrase. However, sentences with more than one intermediate phrase will have multiple nuclear pitch accents, according to the intermediate-phrase-based definition. Welby (2003) found that listeners treated (intermediate-phrase-based) nuclear pitch accents as more likely to signal focus than pre-nuclear accents. However, her nuclear accent results conflicted with results of a similar experiment by Birch and Clifton (1995), and she concluded that whether an accent is present or absent is more important than whether it is nuclear or pre-nuclear. Due to the uncertainty regarding the role of the (intermediate-phrase-based) nuclear/pre-nuclear distinction in the experimental and theoretical literature, in this study I test the effects of the presence or absence of pitch accents in particular locations. More specifically, I concentrate on the location of the final pitch accent in a sentence, as this plays an important role in Gussenhoven's theory of pitch accent placement. In order to avoid confusion, I will refer to the final pitch accent in an sentence as the "final pitch accent", rather than the "nuclear pitch accent".

The dialogue in (3) can be used to illustrate Gussenhoven's (1983a, 1999) account of narrow focus, in which only one word in a sentence is [+focus] (in this case, the subject). In

(3b), *Sam* is [+focus] because it could be said to contrast will all other possible answers (Rooth 1992), it is new relative the preceding context, and it corresponds to the wh-element in the preceding question. It therefore clearly marks the speaker's contribution to the conversation.

(3) a. Who phoned Maud?
b. [SAM] phoned Maud.

In (3b), *Sam* and *Maud* are both As, and *phoned* is a P. Only the A *Sam* is [+focus], so it forms its own focus domain (indicated by square brackets), which receives an obligatory final pitch accent (indicated by capitalization). There are no words before *Sam* in (3b), so no optional pitch accents are possible in this utterance.

In broad focus, a constituent larger than a word is [+focus]. In verb phrase (VP) broad focus, the VP is [+focus], and the subject is [-focus]. Gussenhoven's (1983a, 1999) focus domain and accenting predictions for this type of sentence are illustrated in (4b).

(4) a. What did Sam do?
b. Sam [phoned MAUD]

The P *phoned* and the A *Maud* are both [+focus] in this sentence. Because this P and A are adjacent to one another, they can be combined to form a single focus domain, with an obligatory final pitch accent on *Maud* (the A). Optional pitch accents on *Sam* and *phoned* are also possible because they come before the final pitch accent. All possible pitch accent patterns for VP broad focus utterances (with or without optional accents) differ from the accent pattern for subject narrow focus utterances. This is because in VP broad focus the final pitch accent is on the object, but in subject narrow focus the final pitch accent is on the subject.

Sentence broad focus arises when an entire sentence is [+focus], as in example (5). Gussenhoven's (1983a, 1999) focus domain and accenting predictions for a sentence broad focus utterance are given in (5b).

(5) a. What happened?
b. [SAM] [phoned MAUD]

In (5b), the As *Sam* and *Maud* and the P *phoned* are all [+focus]. According to an example provided in Gussenhoven (1983a), in this situation, the P and A of the VP form a single focus domain, with an obligatory final pitch accent on the A, and the subject A forms its own focus domain with its own obligatory pitch accent. Therefore, utterances with VP and sentence broad focus can have identical accent patterns, if the VP broad focus utterance has an (optional) pitch accent on the subject. However, utterances with sentence broad focus, like those with VP broad focus, should never have the same accent pattern as utterances with subject narrow focus because of their different final accent locations.

The information structures examined in this study are simpler than those that commonly occur in everyday conversation. The experimental items do not include pronouns, and do not use contexts that put only function words in focus, thereby avoiding the complexity of determining when function words should be accented (German, Pierrehumbert, and Kaufmann 2006; German 2009). The experimental items use question-answer pairs to elicit focus, and constituents in the target sentences are either entirely new or mentioned explicitly in the preceding sentence. This avoids issues like the role of attention in focus, the types of antecedents that affect whether a constituent is focused, and the amount of context that should be considered when determining focus (Chafe 1994; Lambrecht 1994; Baumann and Grice 2006). Finally, the items use only informational focus, in which the focused constituent provides new information, leaving out other sources of accentuation such as contrastiveness (Büring 2003). It is important to study all aspects of focus marking by both first and second language learners. However, by examining the simplest cases, this study seeks to establish a baseline of performance for learners of English. Non-native English speakers' performance with more advanced structures and more realistic discourses should be examined in future work.

1.3.3 Mandarin Prosody and Focus Marking

Mandarin is a tone language, which means that pitch targets (tones) are assigned to words in the lexicon, and are used to distinguish between words with different meanings. For example, the Mandarin word *ma* with a high-level tone means *mother*, while the same string of phonemes with a falling tone means *scold*. Mandarin has four lexical tones: high-level (tone 1), mid-rising (tone 2), falling-rising (tone 3), and falling (tone 4) (Li and Thompson 1989). Some syllables do not have any assigned tone. These syllables are described as having a neutral tone, and their pitch is determined by the tone of the preceding syllable (Li and Thompson 1989). Syllables with an assigned tone are considered stressed, and are longer and have greater intensity than unstressed syllables, which have a neutral tone (McGory 1997). Suffixes and grammatical particles are usually unstressed (Li and Thompson 1989). Mandarin is a topic prominent language, but its word order is difficult to define because Mandarin word order tends to be governed more by meaning than grammatical role (Li and Thompson 1989). That being said, the most common word order is SVO (Li and Thompson 1989).

Focus can be marked in Mandarin both syntactically and prosodically. Words in informational focus can be placed in the most deeply embedded position on the recursive side of the syntactic tree (usually the sentence-final position) (Xu 2004). Words that are in contrastive focus can be surrounded by the focus markers *shi…(de)* (Xu 2004) and *lian…dou/ye* (Shyu 1995). Xu (1999) found that sentences with narrow focus on the final word were prosodically equivalent to sentences with VP broad focus. In this experiment, VP broad focus was described as ‘neutral focus’ and produced with no extra prominence on any word within the VP. This prosodic equivalence makes sense because focus marking on a single word in Mandarin can project up to larger constituents, just as it can in English (Xu 2004). A sentence-final object NP is in focus position, and whether or not it is made prosodically prominent, it licenses several information structure interpretations, including focus on the object alone, focus on the VP, or focus on the entire sentence.

Mandarin, unlike English, does not require that phrases have any prosodically prominent word, and this is generally the case when a sentence is in broad focus or when narrow focus is marked syntactically (Xu 2004). Xu (2004) claims that focus in Mandarin only needs to be marked prosodically when it is not marked syntactically. Focused words (that are marked prosodically) in Mandarin have higher F0 peaks, higher mean F0s, expanded F0 ranges, and longer durations than topics in the same position (Liu and Xu 2005; Wang and Xu 2006; Xu, Xu and Sun 2004). In addition, the words following the contrastively focused word have a lower maximum and mean F0, and a lowered and compressed pitch range compared to those following topics (Liu and Xu 2005; Wang and Xu 2006; Xu, Xu and Sun 2004). As well as producing prosodic markers of narrow focus, native Mandarin listeners are able to determine focus location based on prosody, although they sometimes confused sentences that had narrow focus on the final word with sentences without narrow focus (Liu and Xu 2005).

1.3.4 Korean Prosody and Focus Marking

Jun (2005) proposes that Seoul Korean has four prosodic levels: Intonation Phrases (IPs), made up of Intermediate Phrases (iPs), made up of Accentual Phrases (APs), made up of Words. IPs are prosodically marked with phrase-final lengthening and a boundary tone, and can be followed by a pause. iPs are generally marked with a greater than normal juncture between APs, pitch reset, or a higher than normal AP final boundary tone. APs are prosodically marked with phrasal tones at the beginning and end of each phrase. The AP-initial phrasal tones vary based on the phrase's length and the identity of the first phoneme in the phrase. APs have an LH initial boundary tone, unless the first phoneme is a fortis or aspirated stop, in which case they have an HH initial boundary tone. APs with four or more syllables also have LH final boundary tones. APs with fewer than four syllables exhibit a range of intonational patterns including L/HLH and L/HHH (Jun 1998). Korean has a canonical SOV word order, but the word order is somewhat free because of case-marking particles (Sohn 1999).

Focus in Seoul Korean is marked by a word's position in an AP, with focused information appearing at the beginning of an AP (Jun and Lee 1998). Jun and Lee (1998) found that Korean words in narrow contrastive focus have longer initial consonants, longer first syllables, higher peak F0s, and greater F0 ranges than non-focused words. In the post-focal area, words tend to be shorter, and AP boundaries are often removed, leading to a loss of boundary tones associated with the edges of the APs (Jun and Lee 1998). In addition, the pre-focus sequence is sometimes shorter than its neutral counterpart. Jun, Kim, Lee, and Kim (2006) examined how broad informational VP focus is marked in Korean. They found that, unlike words in narrow contrastive focus, VPs in broad focus had an IP boundary inserted before the focused VP, leading to a boundary tone at the end of the subject (preceding the VP), and did not have dephrasing after the VP-initial word. Acoustically, the first word in focused VPs and often later words in the phrase had an expanded pitch range. Sentences with focused VPs had significantly longer durations for the subject and first word in the VP compared to the same words in a neutral sentence.

These methods of focus marking contrast with the English tendency to accent focused information and deaccent unfocused information. A word is made prominent in Korean by placing it at the beginning of the phrase, as opposed to the English method of placing pitch accents on prominent words somewhere within a phrase. For this reason, Korean is described as an edge prominence language. Interestingly, Jun (2002) still found that Seoul Korean speakers ranked a syntactic constraint over the focus constraint dictating dephrasing words that follow a focused word. They violated the focus constraint more often than the syntactic constraint, but they were more likely to obey the focus constraint when speaking at a fast rate and when the sentence involved a particular syntactic structure. This means that the focus marking rules of Korean are not applied in every situation, and the likelihood that they will be applied depends on competing constraints, speech rate, and the syntactic structures involved.

1.3.5 Do English, Mandarin, and Korean have the Same Focus Marking System?

Sections 1.3.2-1.3.4 have described how the prosody and focus marking systems of English, Korean, and Mandarin differ in a number of important ways. They have also shown that there are certain similarities at the acoustic phonetic level in how the three languages mark narrow focus. It has been claimed (Xu and Xu 2005; Lee and Xu 2010) that these three languages actually have the same focus marking system. Namely, focused words have raised and expanded F0 ranges and longer durations, and words in the post-focal region have lowered and reduced F0 ranges and possibly shorter durations (relative to words in sentences without narrow focus). The Parallel Encoding and Approximation (PENTA) model was designed to capture such commonalities (Xu and Xu 2005). In this model, sentence prosody is determined by the following melodic primatives: pitch target, pitch range, strength, and duration. The values for each of these melodic primatives is determined by some communicative function (e.g. lexical stress, focus, or grouping). Pitch targets are assigned to each syllable, and then modified by the pitch range primitive, which is claimed to be controlled by focus for all three languages. In the English version of the model, pitch targets are determined by lexical stress and sentence type (e.g. questions vs. statement) (Xu and Xu 2005). In Korean, pitch targets may be determined by a syllable's adjacency to a phrase boundary, although this is not explicitly stated. Xu and Xu (2005) support this model for English with data showing that there are small but significant F0 perturbations in the post-focal region, where pitch accents are not predicted, and that the size of post-focal F0 rises do not significantly differ from F0 rises on these words in broad focus (Xu and Xu's 'no narrow focus' condition). Lee and Xu (2010) support this model for Korean by showing that words have higher values for acoustic features like duration, mean F0, max F0 and intensity when they are in narrow focus than when they are in broad focus. They also showed that words have lower values for these features when they follow a word in narrow focus than when they are in broad focus. Finally, the authors point out that the locations of pitch movements on averaged F0 contours do not differ between sentences with narrow focus and sentences with broad focus (Lee and Xu's 'neutral focus' condition), which they suggest provides evidence against Jun and colleagues' dephrasing proposal for focus marking (Jun and Lee 1998, Jun 2002).

The PENTA model differs dramatically from the autosegmental-metrical (AM) models of prosody that have been prominent in recent years. AM models should be able to explain the data described in the preceding paragraph. However, Xu and colleagues' conclusion that these results mean that AM should be abandoned in favor of a model like PENTA may be premature. The general expansion of the pitch range on focused words and reduction of the pitch range on post-focus words is predicted by AM models in which focused words are accented (in English), or adjacent to phrase boundaries (in Korean), while post-focal words are not accented (in English) or less likely to be adjacent to phrase boundaries (in Korean). This is because accents and phrase boundaries are associated with pitch targets which can pull the F0 contour higher or lower. The post-focal pitch peaks found in Xu and Xu (2005) are small and may be explained by micro-prosodic effects related to the segments. In addition, although the rise size of post-focal peaks and peaks on the same words in broad focus did not differ, the peaks in broad focus had significantly higher maximum F0s. Other acoustic cues, such as duration, could also have made the final word in broad focus (which is predicted to be accented) more perceptually prominent than the same word in post-focal position. The notion of prosodic prominence within broad focus is not included in the current PENTA model, and not discussed in Xu and Xu (2005) or Lee and Xu (2010). Therefore this model does not predict a number of findings related to prosodic prominence in English broad focus sentences. For instance, it does not offer an explanation for the extreme F0 peaks on the object found in some of the broad focus productions in Xu and Xu (2005). It also cannot explain native English listeners' difficulty telling apart sentences with narrow focus on the object and sentences with broad focus; participant accuracy was just over 50% (Gussenhoven 1983b). Similarly, it has no way of predicting the higher peak F0s on verbs in unergative broad focus sentences relative to unaccusative broad focus sentences, and the higher peak F0s on subjects in unaccusative broad focus sentences relative to unergative broad focus sentences (Hoskins 1996). All three of these results can be explained by models that predict that accent location within broad focus is rule-governed, like Gussenhoven's (1983a, 1999) model. The reason given for rejecting AM accounts of Korean prosody offered in Lee and Xu (2010) is also open to reinterpretation. The authors claim that "there is no clear evidence of major changes in intonational structure due to focus" (Lee and Xu 2010: 4). However, this observation is based on F0 contours that are averaged across multiple tokens produced by multiple speakers. This averaging method could easily hide categorical differences between utterances with and without dephrasing, a process that does not invariably accompany focus in Korean (Jun and Lee 1998). So dephrasing may be occurring in a subset of the recordings in this study, but its effects are lost in the averaging process.

The arguments above demonstrate that it is far from certain that focus is marked with the same method in English, Mandarin, and Korean. Because of this, and the importance of broad focus in the current, I will continue to assume AM models of English and Korean prosody. However, even if the PENTA model is proved correct, a number of differences between the three languages remain. English and the dialects of Korean spoken by participants in this study are non-tonal languages, while Mandarin is a tonal language. In addition, sentence and VP broad focus do seem to be marked prosodically in English (e.g. Birch and Clifton 1995; Gussenhoven 1983b; Hoskins 1996), and VP broad focus marked prosodically in Korean (e.g. Jun, Kim, Lee, and Kim 2006), but to the best of my knowledge, only prosodic marking of narrow focus has been reported in Mandarin. Xu and colleagues' studies do demonstrate the importance of using comparable methods to investigate features like focus marking cross-linguistically, and the importance of closely examining the assumptions of the models we use.

1.3.6 Discourse-Level Prosody Production in Non-Native English

Learners of English show non-native-like performance in both the placement and realization of pitch accents. Incorrect pitch accent placement can mislead the listener about what information is new and what is old, and can sound unnatural due to misplaced accents. Spanish-speaking learners of English in Verdugo's (2006) study tended to put the nuclear pitch accent on the last word of an utterance in broad focus, while native English speakers put it on the last content word. These Spanish speakers also used broad focus accenting patterns in contexts requiring narrow focus. Nava and Zubizarreta (2010) and Nava (2007) also found that Spanish learners of English tended to place the nuclear accent on the final word of a sentence, even when native English speakers would place it on an earlier constituent (e.g. in sentences with unaccusative verbs). These speakers generally failed to follow the English tendency to deaccent previously-mentioned items. Both the nuclear accent placement and the deaccenting errors are likely due to participants transferring focus marking patterns from their native language to English. In Wang's (2003) analysis of spontaneous conversation, Mandarin-speaking learners of English correctly produced pitch accents on 18 words that carried new or contrastive information, but failed to place pitch accents on six contrastive words, and incorrectly placed nuclear pitch accents on eight words conveying given information. McGory (1997) found that both Mandarin and Korean speakers put pitch accent-like F0 movements both before and after the nuclear pitch accent position in their English sentence productions. This could be because of the extensive use of lexical tone and boundary tones in Mandarin and Korean, respectively. Taken together, these studies demonstrate that language learners can produce both appropriate and inappropriate pitch accent placement in their second language, and that their errors may be the result of transfer from their first language to their second language.

Non-native English speakers also produce pitch accents using different or fewer acoustic cues than native English speakers and may use cues differently. The Spanish speakers in Verdugo's (2006) study failed to produce the highest peak on the nuclear pitch accent, and often used L*+H pitch accents rather than the H* and L+H* pitch accents that the native speakers usually used. One Mandarin speaker of English produced H*+!H pitch accents in *did*-questions rather than the L* pitch accent typically produced by native English speakers in this context (Schack 2000). Similarly, Mandarin and Korean speakers in McGory's (1997) study produced L+H*s on focused words in both statements and questions, while native English speakers only used them in statements. The pitch accent types that these non-native speakers produced depended in part on whether the word was stressed on the first or second syllable. In addition, while speakers of Mandarin (a language with a stress distinction) were able to reduce the durations of unstressed syllables in English, speakers of Korean (a language without a stress distinction) had difficulty reducing unstressed word-initial syllables. Moroccan Arabic-speaking learners of English tended to use increased F0 and vowel lengthening to mark pitch accents to a greater degree than native English speakers, while native speakers used increased intensity more than the Arabic speakers (Yeou 2004). Unlike native English speakers, who marked new and contrastive information with a higher F0 in both read and spontaneous speech, Spanish, Thai, and Japanese participants in another study did not show higher F0s for these words in English read speech, and Spanish and Thai participants did not show higher F0s in spontaneous speech (Wennerstrom 1994). Japanese leaners of English in Aoyama and Guion's (2007) study produced function words, which rarely receive pitch accents, with relatively longer durations than native English speakers. They also produced content words with larger F0 ranges than native speakers. The latter difference could be due to the fact that F0 is the only cue to pitch accents in Japanese. The studies described above show that learners of English differ from native English speakers in the pitch accent types they produce and the acoustic cues they use to mark pitch accents, and that at least some of these differences may be attributed to transfer from their native languages.

1.3.7 Prosody Perception in Non-Native English

Less work has been done on prominence perception by non-native English speakers than on prominence production. Akker and Cutler (2003) found that native English speakers detect words faster when they are accented and when they are in focused position, and these effects interact, so that accenting has a smaller effect when the target word is in focused position. This ability lets native English speakers quickly and efficiently process new and important information. Advanced Dutch learners of English also detected words faster in focused and accented positions; however, these effects did not interact. This lack of interaction makes the Dutch listeners' processing of English prosody less efficient, highlighting the importance of prosody perception acquisition.

Rosenberg, Hirschberg, and Manis (2010) examined how native Mandarin speakers' ability to perceive the location of English pitch accents was affected by features like word order, part of speech, number of syllables, pitch accent type, and boundary tone proximity and type. They found that many of these factors significantly affected participants' performance. The native Mandarin speakers were better at identifying pitch accents on two-syllable words than one-syllable words, better on adverbs and determiners than verbs and nouns, and better on words at the end of an utterance than at the beginning. They were most accurate at perceiving pitch accents in sentences with H* L-L% contours than L* H-H% contours, and were least accurate on H* H-H% contours. Pitch accents that were realized with higher mean and maximum F0s and longer durations were perceived more easily. In addition, pitch accents that acoustically stood out from the rest of the sentence (e.g. with a greater difference between the mean F0 on the accented word the mean F0 for the entire sentence) were perceived more easily, Interestingly, non-native speakers who reported more experience with English were less successful at detecting pitch accents. The authors hypothesized that this is because the experienced participants relied more on the semantic and pragmatic content of the utterance than the acoustic information. Despite this finding, performance improved over the course of the experiment, demonstrating learning. The current study tests whether the effects of number of syllables, word position, and participant language experience/ability found in Rosenberg, Hirschberg, and Manis (2010) can be replicated with a new set of materials and participants.

Zhang, Li, Lo, and Meng (2010) examined native Chinese speakers' ability to understand the meaning communicated by various types of English prosodic structures. Their most relevant experiment for the current study examined participants' knowledge of which word should be accented in narrow focus contexts. The experimental items consisted of question-answer pairs, but the focus type elicited was often contrastive (with the focused word correcting or replacing a word in a yes/no question) rather than informational (with the focused word answering the question). For example, the context "Can doctors give blood tests at this clinic?" was followed by the answer "No. you should go to a hospital for blood test." (*sic*), putting the word *hospital* in contrastive focus. When asked which word should carry emphasis in a paper-and-pencil task, participants identified the correct word 86% of the time, showing relatively good knowledge of the relationship between narrow focus and accent placement. When they heard the question-answer pairs spoken, they selected the emphasized word 98.5% of the time. This result shows that participants are able to perceive at least some pitch accents on narrow focus words, as their performance improved relative to the pencil-and-paper task. However, it is unclear how many pitch accents they would have perceived without the contextual clues to focus location. The current study tests participants' pitch accent perception ability and knowledge of accent placement separately, providing more fine-grained data on pitch accent perception and understanding. It also examines a wider range of information structures, including two types of broad focus in addition to narrow focus.

Studies of lexical stress perception show that language learners often have difficulties perceiving lexical stress like native speakers, possibly due to transfer from their native language. Dupoux et al. (2007) found that native speakers of French, a language without contrastive lexical stress, had difficulty using lexical stress to discriminate between words or to identify words. This was true even for French speakers who had studied Spanish, a language that uses (predictable) lexical stress. The current study expands on this research by examining the influence of L1 prosodic categories on language learners' perception of sentence-level pitch accents in their L2.

1.3.8 Relationship between Perception and Production

Although a full understanding of the relationship between the perception and production of second language prosody is crucial for developing useful training programs, little research has been done in this area (Chun 2002). Coburn (2000) found a relationship between perception and production of pitch accents by native English speakers. In her study, English speakers produced sentences with narrow focus and identified the location of focus in utterances produced by themselves and other speakers. She found a significant correlation between the speakers' accuracy in producing narrow focus and perceiving it. Still, most of her participants were more accurate at perceiving narrow focus than producing it. There was also a greater range in the perception accuracy scores than in the production accuracy scores, and a much larger range in perception scores for the less accurate participants than for the more accurate ones.

A number of studies show that training language learners on perception tasks in their non-native language can lead to improved production. In the prosodic realm, 'tHart and Collier (1975) found that playing a tape illustrating important features of English intonation to native Dutch speakers led to an improvement in their English intonation production. De Bot and Mailfert (1982) also found that perceptual training on English intonation numerically improved native-speaker ratings of the English intonation produced by French speakers, although the study had a small sample size and the results only approached significance.

In the segmental realm, Rochet (1995) found that when Canadian English and Brazilian Portuguese learners of French incorrectly produced the novel sound /y/, their productions (/u/ and /i/, respectively) were determined by their perception of the sound. This study also found that giving Mandarin speakers perceptual training on French voiced vs. voiceless stops significantly improved their productions of some of these stops. Bradlow and colleagues (Bradlow, Pisoni et al. 1997; Bradlow, Akahane-Yamada et al. 1999) found that native Japanese speakers given perception training on the English /ɹ/-/l/ distinction produced more intelligible and native-like /ɹ/ and /l/ tokens even three months after the training. However, they did not find a correlation between the amount of perception and production learning for each participant.

Taken together, these studies demonstrate that there does seem to be a link between language perception and production abilities, both for native speakers and language learners, such that greater skill in one area is often accompanied by greater skill in the other, and training in one can improve performance in the other. However, this relationship is complex: perceptual ability may be greater than production ability, or vice-versa, and it is possible to improve one skill without noticeable improvement in the other. Such imbalances could have a number of causes. It is possible that some language learners have developed appropriate internal representations of sound categories, but have not yet developed the motor skills needed to produce these targets accurately (Bradlow, Pisoni et al. 1997). On the other hand, production could be easier for language learners who have determined the articulatory configurations necessary to produce sound categories well enough to be understood by native speakers, but who have not shifted their perceptual boundaries enough to correctly classify the range of possible native productions. Alternatively, some learners may be focusing on acoustic cues that are adequate for making distinctions in one modality (perception or production), but not in the other (Bradlow, Pisoni et al. 1997). Finally, low perceptual accuracy could result from difficulties with the perception tasks themselves, which are often quite artificial. Such difficulties could have causes that are unrelated to perception ability, such as inadequate short term memory (Coburn 2000).

1.4 Previous Models of Second Language Acquisition

1.4.1 Transfer Models of Phonological and Phonetic Acquisition

Dupoux, Sebastian-Galles et al. (2007) compared three general models of L2 phonological acquisition. The first type of model claims that a language-learner's acoustic perceptual space is warped by their first language experience (e.g. Francis and Nusbaum 2002). This changes the weighting of particular acoustic cues such as F1 or timing. As a result, acoustic cues that do not play an important in their L1 perception also do not play an important role in their L2 perception. The second type of model claims that contrastive phonetic or phonological features such as voicing or nasality that aren't used in the learner's L1 also aren't used in their L2 (e.g. Brown 1998). The third type of model claims that a language learner's L1 categories or prototypes interfere with perception of L2 categories that are similar to those in their L1. Best and colleagues' (Best, McRoberts et al. 2001) Perceptual Assimilation Model (PAM) and Flege's (Flege, Munro et al. 1995) Spoken Language Model (SLM) fall into this third category. PAM predicts how pairs of L2 phonemes will be assimilated into the listener's L1 phonological system. If the two sounds are assimilated into the same L1 category they will be hard for a language learner to distinguish, but if they are assimilated into different categories, if one is a better example of an L1 category than the other, or if they are assimilated as non-speech sounds, they will be easier for the learner to distinguish. SLM classifies L2 categories as identical to, similar to, or different from L1 categories. If an L2 category is similar (but not identical) to an L1 category, the learner will have the most difficulty learning it correctly.

Dupoux, Sebastian-Galles et al. (2007) claim that their data on French speakers' stress deafness support the second type of model for perception of prosodic features, but not the first or third. French does not have lexical stress, and French speakers have difficulty perceiving lexical stress in other languages, in which it is encoded through F0 movement, duration, and energy. French does not have any contrastive prominence features, but does use F0, duration, and energy to mark other prosodic events such as phrase boundaries. Therefore, French speakers have no difficulty perceiving the features themselves, providing evidence against the first type of model.

The lack of similar categories to lexical stress in French provides evidence against the third type of model. However, the second type of model predicts that languages without phonological categories similar to lexical stress will exhibit stress deafness.

Models like PAM and SLM, which seem to work well for predicting patterns of segmental acquisition, may be less applicable in the suprasegmental domain because of the unique nature of suprasegmental categories. For instance, there are fewer suprasegmental categories than segmental categories, leading to a potentially less crowded phonological space. A major difference between the predictions of the second and third types of models is that the third type of model predicts that the existence of a category in a speaker's L1 that is similar, but not identical, to one in their L2 will cause difficulties in acquiring that L2 category. The second type of model does not predict such interference effects. If prosodic categories are in a less crowded phonological space than segmental categories, there may be less chance of confusion if a learner produces a category realization that is similar, but not identical, to the target L2 realization.

Unlike Dupoux et al. (2007), Mennen (1998) found partial support for the SLM model modified for suprasegmentals. In this modified SLM model, when a language learner has similar prosodic categories in their L1 and L2, such as a phonological pitch accent type, they will form a single category, leading to errors in their L2 and even L1 production. She tested this model with a study of native Dutch speaking learners of Greek. Both languages have an L+H* pitch accent, but the pitch accent has different alignment with the syllable in the two languages. She found that most of the non-native Greek speakers did not produce native-Greek-like L+H* alignment in Greek, or monolingual-Dutch-like alignment in Dutch. Instead they collapsed L+H* categories across the two languages. However, one participant did manage to produce alignment that was similar to monolinguals in both languages. These results generally support the SLM model for suprasgemental category production, because having similar categories in their L1 and L2 impaired language learners' production of these categories.

One explanation for these potentially contradictory studies, providing support for different models of prosody acquisition, is the different aspects of prosody examined by Dupoux et al. (2007) and Mennen (1998). Dupoux et al. examined lexical stress, which requires listeners to use a variety of acoustic cues to identify a syllable as prominent, and then match this prominence with a lexical entry. The French listeners may have trouble with the connection between prominence and the lexicon, rather than the connection between the acoustic features and prominence. This hypothesis is supported by the fact that French participants do not exhibit stress deafness in tasks that did not require phonological or lexical representations (i.e. tasks in which the stimuli were short or invariant enough for participants to rely on purely acoustic judgments) (Dupoux et al. 2001). We can conclude that for lexical stress, having a similar category in their L1, which links acoustic prominence to the lexicon, helps language learners. In contrast, Mennen examined fine phonetic details of pitch accent realization. Here, the situation is more similar to the acquisition of segments. The language learners have the L+H* pitch accent category in both their L1 and L2, so they have no trouble producing (and presumably understanding) it in the appropriate context. Instead, the challenge is to use the right acoustic cues for a given language. The differences between Dupoux et al.'s and Mennen's studies highlight a distinction that plays a major role in Mennen's (1999, 2004) theories of second language prosody acquisition: the difference between phonological and phonetic aspects of prosody. Different models may well be needed to explain the acquisition of 1) entirely new prosodic features, such as lexical stress or pitch accents, which connect to the lexical, syntactic, or semantic systems, and 2) phonetic details of prosodic categories that may already exist in a speaker's L1. For this reason, the experiments in this study were designed to distinguish as much as possible between these two levels, which are referred to as 1) pitch accent understanding and placement, and 2) pitch accent perception and realization.

1.4.2 Non-Transfer Models of Morpho-Syntactic Acquisition

While prosody is clearly an aspect of the sound structure of a language, prosodic focus marking also requires knowledge of a language's syntactic and pragmatic systems. Therefore research into the acquisition of these systems may provide useful insights when considering prosodic focus marking acquisition. One such insight is the consideration of factors other than L1 transfer that could influence second language acquisition.

Dulay and Burt (1973) began a very active line of research studying the order of acquisition of function morphemes (e.g. past tense -*ed*, plural -*s*, and articles) by learners of English. The researchers exploring this question found that English learners with a wide variety of L1s seemed to acquire function morphemes in similar orders (e.g. Bailey, Madden, and Krashen 1974; Dulay and Burt 1974; Fathman 1975; Krashen et al. 1976), suggesting that factors other than L1 transfer affect the ease with which learners acquire these morphemes. Many of these studies were attempting to show that L1 is not an important factor in second language acquisition, in a reaction to behaviorist models, which had been prevalent. Recent studies have focused more on specifying precisely what factors determine the ease with which these morphemes are acquired. Goldschneider and DeKeyser (2001) performed a meta-analysis on a subset of these studies, and found that a combination of the morphemes' phonetic salience, semantic complexity, morphophonological regularity, syntactic category, and frequency could explain a large proportion of the variance in their ordering (R^2=.71). Phonetic salience, semantic complexity, and syntactic category seemed to be particularly good predictors of how quickly an English learner would acquire a particular morpheme. Wei (2000) focused on explaining the ordering with a more fully developed model of a single factor: the 4-M Model of morpheme types. Wei examined three morpheme types proposed in the model: 'content morphemes' like nouns and verbs, 'early system morphemes' like prepositions in verbal phrases, and 'late system morphemes' like the 3rd-person singular -*s* marker, which signal grammatical relations. Wei found that, in general, content morphemes were acquired before early system morphemes, which were acquired before late system morphemes.

Contemporary models of phonetics and phonology have not focused on the role that particular features of L2 categories can play in category acquisition. This may be because it is unclear what those features would be for phonemes and allophones. In contrast, prosodic categories lend themselves to this type of analysis because of their relationships with other levels of linguistic structure and the complexity of their usage. The particular factors affecting acquisition may be different for prosodic and morpho-syntactic categories. However, the idea that language acquisition can be affected by features of L2 categories could well be shared. Two features of English prosodic focus marking that could affect focus marking acquisition are discussed in Section 1.5.2, as part of the L2 Challenge Model.

1.5 Two Types of Models of English Prosodic Focus Marking Acquisition

In this dissertation, I aim to design a predictive model of prosodic focus marking acquisition. In order to do this, I must first test the ability of several types of acquisition models to explain aspects of prosodic focus marking by second language learners. As discussed in

Section 1.4, different levels of prosody acquisition (e.g. phonetic and phonological) may behave differently, requiring different types of models to explain them. I describe below the types of models I will be comparing, and the framework I will use to examine these different levels of prosody.

In general, models of second language phonology acquisition, such as those described in Section 1.4, have focused on transfer from the L1 to the L2. If a language learner's native language affects their use of target language prosody, we would expect to see differences between people with different native languages. I will refer to these types of models as 'Transfer Models'. However, I propose that it is also possible for challenging features of L2 prosody to cause similar difficulties for language learners with a range of native language backgrounds. If this is the case, we would expect to see similar patterns of behavior across people with different native languages. I will refer to this type of model as an 'L2 Challenge Model'. In Sections 1.5.1 and 1.5.2 I lay out some predictions that arise from specific types of Transfer Models and L2 Challenge Models of prosodic focus marking acquisition. It is important to note that Transfer Models and L2 Challenge Models do not necessarily make conflicting predictions, and that both types of models could simultaneously capture different aspects of language learner performance.

Models of prosodic focus marking acquisition are more complex than models of phonological category acquisition, like PAM (Best, McRoberts et al. 2001), because focus marking involves not only the phonological system but also the syntactic and pragmatic systems. Therefore, a model of prosodic focus marking acquisition must account for the three levels involved in the prosodic realization of focus. These three levels and the relationships between them are laid out in Figure 1.1.

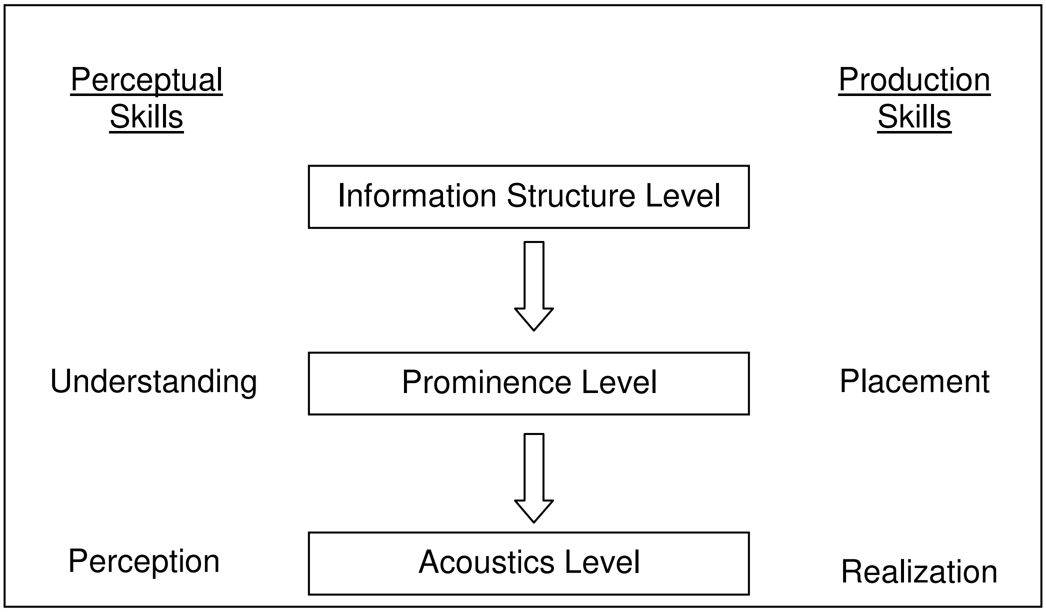

Figure 1.1. Chart demonstrating the three levels of prosodic focus marking and the relationships between them

At the information structure level, the speaker determines the information structure of the utterance (which words are in focus). At this level the prosodic system interacts with the syntactic and pragmatic systems. At the prominence level, the speaker determines how the information structure should be realized within a language's syntactic, morphological, and prosodic structures. In English this involves the selection of a word or words to receive pitch accents given a particular information structure, and the selection of the type of pitch accents that will be used. In Korean it involves choosing the locations for AP and IP boundaries. In Mandarin it involves deciding whether focus will be marked syntactically or prosodically. At the acoustics level, the speaker selects acoustic features to realize the prosodic structures selected at the prominence level. In English this can include F0 movements appropriate to the pitch accent type, increased duration, increased amplitude, and spectral changes. In Korean it can include an increased peak F0 and longer duration on the focused word, and shorter post-focal words. In Mandarin it can include an increased F0 range on the focused word and a decreased F0 range on the following words. A breakdown at any one of these levels could result in an utterance that does not accurately convey the speaker's communicative intent.

1.5.1 Transfer Models

1.5.1.1 Comparisons between L1s and L2 for Transfer Models

A language learner's L1 can interact with their L2 at the information structure, prominence, and acoustics levels. At the level of information structure, languages have the most in common. Languages differ at the prominence and acoustics levels.

At the prominence level, English differs from Mandarin in that Mandarin allows syntactic focus marking, while English requires prosodic focus marking. In addition, English marks broad focus by placing a pitch accent on a word within the focused constituent, while Mandarin generally does not prosodically emphasize any word within broad focus constituents. English differs from Korean at the prominence level in that English uses pitch accents to mark focus while Korean uses prosodic boundary placement. In addition, Korean marks VP broad focus with prosodic prominence at the beginning of the VP while English marks it with prosodic prominence at the end. Finally, Korean does not prosodically mark sentence broad focus, while English generally marks sentence broad focus (for sentences with transitive verbs) with prosodic prominence on the final word.

At the acoustics level, English differs from Mandarin in that Mandarin marks prosodic prominence with an expanded pitch range for the lexically assigned tones, while English marks pitch accents with (among other cues) particular pitch targets, determined by the pitch accent type. English differs from Korean at the acoustics level in that Korean marks prosodic prominence with boundary tones that are associated with a phrase rather than a word, while English aligns pitch accents with the stressed syllable in a particular word. In addition, Korean boundary tone type is influenced by segmental features of the first word in the phrase and the number of syllables in the phrase, which is not true for English pitch accent types. Finally, English has a stress distinction, which can be realized through duration, intensity, and spectral differences between stressed and unstressed syllables, while Korean does not have a stress distinction.

1.5.1.2 Transfer Model Predictions for Pitch Accent Perception and Realization

I will assess a Transfer Model of prosodic prominence acquisition that predicts that language learners will most easily learn how to perceive and realize pitch accents if their L1 also has pitch accents. English learners who do not have pitch accents in their L1 will have an easier time acquiring the pitch accent category if their L1 and L2 share more phonological and phonetic features for marking prosodic prominence. This is the type of model in which shared categories across a speaker's L1 and L2 help with category acquisition, which Dupoux et al. (2007) found support for in their study of lexical stress acquisition. Although Mennen (1998) found evidence in support of an SLM-type model, in which similar L1 and L2 prosodic categories can interfere with each other, her study focused on native-like pitch peak alignment. In the majority of experiments in the current study, learners of English do not have to have entirely native-like pitch accent perception and realization. They simply have to recognize or produce acoustic cues that are similar enough to those used by native English speakers that they can identify pitch accent location and produce pitch accents that can be identified by native speakers. It is entirely possible that the fine phonetic details of their pitch accent productions will be negatively impacted by the acoustic cues used to mark prominence in their native language, as Mennen found.

If the Transfer Model proposed in the preceding paragraph is correct, we can make some specific predictions about the types of problems that native Korean and Mandarin speakers will have when acquiring English prosodic focus marking. For instance, if there is a difference between the two language groups' abilities to accurately perceive and realize pitch accents on words with non-initial stress, the model predicts that Mandarin speakers will be more accurate than Korean speakers. This is partly because Korean does not have a stress contrast, and therefore Korean speakers have difficulty reducing word-initial unstressed syllables (McGory 1997). In addition, Korean speakers put focused words at the beginning of an AP, so they are acoustically marked with phrase-initial boundary tones, which start at the beginning of the word, not at the second syllable. In contrast, Mandarin does have a stress distinction (Li and Thompson 1989).

Excluding difficulties with word-level stress, the model predicts that native Korean speakers will generally be more accurate than native Mandarin speakers at perceiving and realizing English pitch accents. This is because, like English, Seoul Korean and the other Korean dialects spoken by participants in this study use pitch only post-lexically. Korean uses pitch to indicate phrasing, thereby communicating information structure. In contrast, Mandarin uses pitch both post-lexically, to mark narrow focus through cues to prominence like an expanded pitch range, and lexically, to distinguish between words that differ only in their tones. As a result, Korean speakers may more easily tune into pitch as a marker of prominence within a phrase and more easily use pitch movements to mark prominence.

1.5.1.3 Transfer Model Predictions for Pitch Accent Understanding and Placement

1.5.1.3.1 Opposing Transfer Models

There are two possible Transfer Models for the understanding and placement of pitch accents. Both Transfer Models predict that the learner will have the least difficulty appropriately placing and understanding pitch accents when their L1 dictates prominence on the same word for a particular information structure. As a result, both Transfer Models predict that Korean speakers will be better than Mandarin speakers at understanding and appropriately placing pitch accents appropriate to narrow focus. This is because Korean always marks narrow focus prosodically (Jun and Lee 1998), while Mandarin can mark it either prosodically or syntactically (Xu 2004). Therefore, Korean speakers are predicted be able to directly carry over their habit of always marking narrow focus with prosodic cues like expanding the pitch range and lengthening parts of the focused word.

However, the two Transfer Models disagree on the effect of having different prominence locations for a given information structure in the speaker's L1 and L2. In the "Different Prominence Locations Hurt" Transfer Model (described in Section 1.5.1.3.2), having different prominence locations for a particular information structure will lead to errors in the L2, as the prominence location from the L1 is transferred to the L2. In the "Any Prominence Location Helps" Transfer Model (described in Section 1.5.1.3.3), having prosodic prominence predicted at all for a particular information structure will be an advantage in the L2, as the concept of prominence marking for that information structure is transferred to the L2.

1.5.1.3.2 Different Prominence Locations Hurt

Under the assumption that having different prominence locations in their L1 and L2 is detrimental to the language learner, learners should have the least difficulty placing and understanding pitch accents when their L1 dictates prominence on the same word as their L2. They should have the most difficulty when their L1 dictates prominence on a different word. When their L1 predicts no prosodic prominence for a particular information structure, their difficulty level should fall between these two extremes. The difficulty experienced by English learners when their L1 dictates a different pitch accent placement than their L2 is illustrated by the well-documented difficulty that Spanish-speaking learners of English have in appropriately placing pitch accents in English (Nava and Zubizarreta 2010; Nava 2007). Their L1 has pitch accents, like English, but dictates a different pitch accent placement for a number of broad focus sentence types. In contrast, Wang (2003) found that, in general, Mandarin speaking English learners correctly placed their English pitch accents. This provides support for the model described above, as Mandarin does not place prosodic prominence on any words that are not accented in English. The main difference between the two languages is that in English particular words receive a pitch accent in broad focus sentences, while in Mandarin no words are made prosodically prominent in broad focus sentences.

The Different Prominence Locations Hurt model predicts that Mandarin speakers should be better at understanding and appropriately placing pitch accents in VP broad focus sentences than Korean speakers. This is because Korean speakers mark VP broad focus by placing an IP boundary at the beginning of the focused VP, leading to an expanded pitch range and increased duration for the first word in the VP (due to the Korean SOV word order this is often not the verb, but the object of the sentence) (Jun, Kim et al. 2006). These acoustic cues are similar to those of an English pitch accent. However, in English the pitch accent in a broad focus VP is usually placed at the end of the VP – for a transitive verb it goes on the object NP. Such a difference between their L1 and L2 is predicted to block Korean speakers' ability to acquire the English pitch accent placement for VP broad focus sentences. On the other hand, Mandarin does not prosodically mark VP broad focus (Xu 1999). While Mandarin uses many of the same acoustic cues to mark narrow focus as English uses to mark pitch accents, Mandarin speakers do not have to overcome the habit of using these cues to mark VP broad focus on a different constituent.

The model predicts that Korean speakers will more easily understand and appropriately place pitch accents in sentence broad focus sentences than in VP broad focus sentences. This is because Korean does not place extra prosodic prominence on any word in sentence broad focus sentences, in contrast to the extra prosodic prominence on the VP-initial word in VP broad focus sentences.

Finally, the model predicts that Korean and Mandarin speakers should be equally good (or bad) at understanding and appropriately placing pitch accents in sentence broad focus sentences. This is because neither L1 places extra prosodic prominence on any particular word in such sentences.

1.5.1.3.3 Any Prominence Location Helps

Dupoux et al.'s (2007) results support the idea that having the category of prominence for a particular information structure in a speaker's L1 and L2 will help in their placement and understanding of prominence in their L2. They found that native French speakers' lack of the lexical stress category made it very difficult for them to process stress in other languages.

This model predicts that Korean speakers will be better than Mandarin speakers at placing and understanding prominence in VP broad focus contexts. This is because Korean marks VP broad focus prosodically, while Mandarin does not. The model also predicts that Korean speakers might even be better than Mandarin speakers at placing and understanding prominence in sentence broad focus contexts, because Korean speakers mark some kind of broad focus prosodically in their L1. Therefore Korean speakers may have a better mental representation of broad focus than Mandarin speakers.

The model predicts that Korean speakers will be more accurate at understanding and appropriately placing pitch accents in VP broad focus sentences than in sentence broad focus sentences. This is because Korean does not prosodically mark sentence broad focus sentences, but it does prosodically mark VP broad focus sentences.

Finally, the model predicts that Korean speakers could be more accurate than Mandarin speakers at placing and understanding prominence in sentence broad focus contexts. This is because Korean speakers already have the mental concept of marking broad focus prosodically through their treatment of VP broad focus. The concept might extend to broad focus in general, helping them more easily acquire sentence broad focus prosodic marking in English.

1.5.2 L2 Challenge Models

Evidence for an L2 Challenge Model would come from a repeated pattern of behavior across participants with different L1s. Such a repeated pattern has two other potential causes apart from challenges associated with the L2. The first is that the L1s of the participants are similar in the aspect of language being studied, leading to the transfer of similar features from both L1s to the L2. However, this is unlikely to be the cause of repeated patterns if the L1s are known to differ on the relevant dimension. A second potential cause is the additional mental resources required to speak a non-native language. We would expect these to lead to a degraded performance for non-native speakers across all words, sentence types, and contexts. In cases where performance on some words, sentence types, or contexts are degraded to a greater extent than on others, some feature of the problem items must be making them particularly difficult for the learner. Therefore, while it is important to keep all possible explanations for the results in mind, if the same pattern of behavior is seen for language learners with multiple L1s, which are known to vary on the features under examination, and when performance on some items is worse than others, this will be taken as support for an L2 Challenge Model.

1.5.2.1 L2 Challenge Model Predictions for Pitch Accent Perception and Realization

English learners in general may have a particularly hard time perceiving less acoustically prominent pitch accents. There is natural variation in the production of pitch accents by native English speakers, involving differences in duration, pitch range, amplitude, and spectral features. Such variation can arise from the speaker, the situation in which the speech is occurring, or the location of the accent in an utterance. In this study, the speaker and the situation are controlled across items, but accented words are produced in different positions (sentence-initial and -final) and different contexts (subject narrow focus, VP broad focus, and sentence broad focus). Research has shown significant differences in the acoustic characteristics of pitch accents in different sentence positions. Phrase-final H* pitch accents have a significantly earlier and lower F0 peaks than earlier H* pitch accents (Shue et al. 2010). This could make phrase-final H* pitch accents less noticeable. In addition, some researchers have found acoustic differences between pitch accents in broad and narrow focus. They found higher maximum F0s for objects in narrow focus than objects in broad focus (Xu and Xu 2005) and longer durations and higher maximum F0s for indirect objects in narrow focus relative to indirect objects in VP or sentence broad focus (Eady et al. 1986). Because sentences in both broad focus conditions in the current study are predicted to have accents on their final words, these two effects could combine, resulting in broad focus final pitch accents on objects that are less perceptible to language learners than the narrow focus final pitch accents on the subject. If these differences between native speakers' productions of early and late pitch accents and between pitch accents in broad and narrow focus are replicated by non-native speakers, then non-native speakers' pitch accents in broad focus may be less noticeable than those in narrow focus.

1.5.2.2 L2 Challenge Models' Predictions for Pitch Accent Understanding and Placement

Aspects of the target language's focus-marking system may also play a role in non-native acquisition of prosodic focus marking. One potentially important factor is the relationship between focus and prosodic prominence. This will be referred to as the Relationship factor. If the Relationship factor plays a role in non-native performance, English narrow focus will be easier to produce and understand than broad focus. This is because in narrow focus there is a direct correspondence between focus and accent placement, while in broad focus an accent on one word can signal focus on a larger constituent. In broad focus contexts, learners of English may not know which word should be accented, or even that any word should be accented. The Relationship factor also predicts that prosodic marking of VP broad focus should be easier to acquire than sentence broad focus marking. This is because a smaller constituent is being marked in the case of VP broad focus, narrowing the range of possible pitch accent locations within the focused constituent. In addition, in VP broad focus a subset of the sentence is being marked as focused, which may be easier to comprehend than marking an entire sentence as focused.

Another possible target language factor is the frequency with which particular accent patterns are used in the target language. This will be referred to as the Frequency factor. If a certain pattern is used in multiple contexts, non-native speakers may apply it too broadly, believing that it is appropriate in contexts where it is inappropriate. In English, a pitch accent pattern in which the final accent is on the object is appropriate in sentence broad focus, VP broad focus, and object narrow focus contexts. Therefore, language learners may incorrectly use it in other contexts, such as subject narrow focus.

The two factors discussed above (the Relationship factor and the Frequency factor) lead to contradictory predictions. The Relationship factor leads to the prediction that subject narrow focus will be easy for language learners, while the Frequency factor leads to the prediction that subject narrow focus will be hard for them. As a result, the two factors can lead to two different L2 Challenge Models, analogous to the Any Prominence Location Helps, and Different Prominence Locations Hurt Transfer Models. However, I propose that a third possibility also exists: these two factors can be combined into a single model that makes different predictions for different types of tasks. In such a model, non-native participants are predicted to find subject narrow focus marking easy in some tasks and hard in other tasks. The key component of the combined model is the criterion used to distinguish between these two types of tasks. The criterion might be whether a task requires perception or production, or it might be the relative difficulty of the task. I will use the results of the experiments in this study to determine whether the Relationship Factor Model, Frequency Factor Model, or the Hybrid Model is correct. If the

Hybrid Model is correct, the results will be used to determine what distinguishes tasks in which narrow focus is easy from tasks in which narrow focus is hard.

1.5.3 Summary and Comparison of Model Predictions

In section 1.5 I have described the predictions of a Transfer Model and an L2 Challenge Model for pitch accent perception and realization, and described two Transfer Models and three L2 Challenge Models for pitch accent understanding and placement. Here, I summarize these predictions to allow easy comparison across the models. Table 1.1 lists the predictions for pitch accent perception and realization. Table 1.2 lists the predictions for pitch accent understanding and placement.

Transfer Model	**L2 Challenge Model**
- Mandarin>Korean, second syllable stress words - Korean>Mandarin, in general	- Narrow focus>Broad focus

Table 1.1. Model predictions for pitch accent perception and realization, with > signifying greater language learner accuracy

Transfer Model: **Different Prominence Locations Hurt**	**Transfer Model:** **Any Prominence Location Helps**
- Mandarin>Korean, VPBF	- Korean>Mandarin, VPBF
- SBF>VPBF, Korean	- VPBF>SBF, Korean
- Mandarin=Korean, SBF	- Korean>Mandarin, SBF
- Korean>Mandarin, SuNF	- Korean>Mandarin, SuNF

L2 Challenge Model: **Relationship**	**L2 Challenge Model:** **Frequency**	**L2 Challenge Model:** **Hybrid**
- SuNF>VPBF>SBF	- VPBF/SBF>SuNF	- SuNF>VPBF>SBF, some tasks - VPBF/SBF>SuNF, other tasks

Table 1.2. Model predictions for pitch accent understanding and placement, with > signifying greater language learner accuracy and = signifying equal language learner accuracy

1.6 Native Perception of Non-Native Production

It is difficult to make predictions regarding native English listeners' perception of non-native pitch accent production. Based on previous research, it is likely that the non-native English speakers' productions will differ acoustically from native speakers' productions. What is less clear is how native English listeners will interpret these non-native productions. Two possibilities are laid out in Sections 1.6.1 and 1.6.2.

1.6.1 Strict Native Perception Model

Under the Strict Native Perception Model, non-native English speakers who deviate acoustically from the pitch accent productions of native speakers will not be understood by native listeners. This means that only the most native-like language learners will be able to communicate their intended focus to native listeners. The Strict Native Perception Model is based on the idea that native listeners are very dependent on a particular combination of acoustic cues for accented and unaccented words. Such dependence could mean that if any of the cues signaling an intended accentuation is missing, or even weaker than expected, the listeners would judge the word to be unaccented. It could also mean that the general differences between native and non-native speech (such as longer durations in non-native speech) would lead listeners to hear pitch accents on words that were not intended to be accented. A third possible scenario is that even if non-native speakers produce greater distinctions between accented and unaccented words than native speakers, the listeners could interpret the cues that fall outside of the ranges of values they expect to simply be meaningless variation that is part of a non-native accent.

1.6.2 Relaxed Native Perception Model

Under the Relaxed Native Perception Model, non-native speakers can produce sentence prosody that is interpretable by native speakers while still deviating from the productions of native speakers. This means that there is some range of acoustic values that native English speakers will accept to mark a pitch accent which extends beyond the values that they typically hear. As a result, English learners might have prosody that is non-native-like, but still understandable. If this is the case, it is important to determine in what ways non-native speakers can deviate from native prosody production and still be understood.

There are several reasons why non-native-like prosody might be accepted and understood. The first possibility is that native listeners will judge a word to be accented when it has more acoustic cues pointing to it being accented than unaccented. In this case, if a non-native speaker produces a greater acoustic distinction than native speakers between accented and unaccented words (e.g. longer durations or more expanded pitch ranges on accented words), then he will be understood while still differing from the native production norm. Alternatively, a non-native speaker could use a different combination of cues than native speakers and still be understood. For instance, the non-native speaker may produce stronger durational cues and weaker pitch cues relative to native speakers, but if the combination of cues makes it more likely that a word is accented, native listeners would still judge it as accented. A second possible explanation is that native listeners do not perceive subtle acoustic differences between the pitch accent cues produced by native and non-native speakers, even though these differences may show up in acoustic measurements.

If the Relaxed Native Perception Model is supported by the experimental results, the acoustic measurements will be used to determine how the ‘prosodically appropriate’ native and non-native English productions differ. This information will shed light on which of the possibilities outlined in the preceding paragraph is correct. For instance, if the non-native productions have stronger acoustic cues for pitch accents than the native productions, this would support the first explanation.

1.6.3 Summary and Comparison of Model Predictions

In Section 1.6 I have described the predictions of the Strict Native Perception Model and the Relaxed Native Perception Model regarding native perception of non-native production of English prosodic focus marking. In Table 1.3, I summarize these predictions to allow easy comparison across the models.

Strict Native Perception Model	**Relaxed Native Perception Model**
- Any deviation from native production → non-interpretable prosody	- Non-native English speakers can deviate from native-like productions → still interpretable prosody

Table 1.3. Model predictions for native perception of non-native production of English prosodic focus marking

1.7 Relationship between Perception and Production

Two opposing models can be used to make predictions about the relationship between perception and production of English pitch accents by non-native speakers. The Perception/Production Dependence Model assumes a fundamental equivalence between pitch accent use in perception and production. The Perception/Production Independence Model assumes that the skills required to perceive or understand pitch accents can develop separately from those required to realize or place pitch accents.

1.7.1 Perception/Production Dependence Model

For pitch accent perception and realization, this model holds that the ability to accurately realize pitch accents depends in part on accurate perception of pitch accents. An English learner's knowledge of how to perceive or realize English pitch accents can come from several sources. Specifically, it can come from acoustic similarities in prosodic prominence marking between their L1 and L2, or from hearing native English speakers produce pitch accents. Similarities in prosodic prominence between a speaker's L1 and L2 should generally lead to similar performance in perception and production, because Coburn's (2000) results suggest that skill in L1 prominence perception and production are generally correlated (at least for English). Listening to native English speakers can teach the learner both what cues to attend to when listening for pitch accents, and what cues to produce. As the accurate perception of native English speech must precede accurate realizations based on this speech, we would expect this kind of learning to lead to greater accuracy in pitch accent perception than realization. Given these combined effects, the model predicts a correlation between pitch accent perception and realization with greater accuracy for perception than realization.

This model holds that knowledge of pitch accent placement is modality-neutral, so an English learner's accuracy in understanding and producing pitch accent placement should be correlated and roughly equivalent, both overall and for particular information structures. This model is based on the premise that if a language learner knows the relationship between an information structure and a particular pitch accent placement, they should be able to apply it equally well in perception and production. Unlike perception of pitch accent realization, the ability to understand pitch accent placement does not require the learner to adjust the weighting of perceptual cues or change the boundaries between categories. Similarly, in production it should not be physiologically more difficult to place a pitch accent on one word rather than another.

1.7.2 Perception/Production Independence Model

This model assumes that different skills are required for pitch accent perception and understanding on the one hand, and for realization and placement on the other, resulting in no correlation between the perceptual skills and the production skills. As noted in Section 1.3.8, non-native perceptual ability may be greater than production ability, or vice-versa, and it is possible to improve one skill without noticeable improvement in the other.

For perception and realization, some language learners may have developed the internal representations necessary to accurately perceive pitch accents, but not yet developed the motor skills needed to realize them accurately (Bradlow, Pisoni et al. 1997). Alternatively, realization might be easier for some language learners because they have determined the articulatory movements necessary to realize pitch accents well enough to be understood by native speakers, but cannot use acoustic cues to correctly perceive native productions of pitch accents.

For understanding and placement, one of these tasks may require more mental resources than the other, leading to unequal performances across the two modalities. Some language learners may find accent placement more challenging because they have to focus on segment pronunciation in addition to prosody. Other language learners may find accent understanding more challenging because of the unconstrained nature of the task: the speaker may use unfamiliar words or unusual segment pronunciations. If the language learner finds one of these tasks more challenging than the other, her pitch accent use in the more challenging task could degrade.

1.7.3 Summary and Comparison of Model Predictions

In Section 1.7 I have described the predictions of the Perception/Production Dependence Model and the Perception/Production Independence Model regarding the relationship between perception and production in English prosodic focus marking acquisition. In Table 1.4, I summarize these predictions to allow easy comparison across the models.

Perception/Production Dependence Model	**Perception/Production Independence Model**
- Perception and realization correlated, with perception>realization	- No correlation between perception and realization
- Understanding and placement correlated and equal, overall and for individual structures	- No correlation between understanding and placement

Table 1.4. Model predictions for the relationship between perception and production in second language acquisition of prosodic focus marking, with > signifying greater language learner accuracy

Chapter 2

Methods

2.1 Introduction

The goal of this dissertation is to examine English focus marking within a complete communicative chain, from non-native perception of native production, to non-native production, to native perception of non-native production. In order to achieve this goal, a series of experiments were run to test English learners' perception, understanding, placement, and realization of English pitch accents. The same set of participants performed four experimental tasks. These experiments included two perception tasks, a production task, and a computer-based judgment task. Sixty people participated in these experiments: 20 native Korean speakers, 20 native Mandarin speakers, and 20 native English speakers. After this set of experiments was completed, a follow-up experiment was carried out, in which 24 native English speakers listened to the recordings produced by the 60 native and non-native speakers.

The five experiments are discussed in greater detail in the remainder of the chapter. Section 2.2 explains experimental procedures and the order in which the experiments were run. Section 2.3 describes the materials used in the experiments. Section 2.4 describes the 60 native and non-native participants who completed the first four experiments. Sections 2.5 to 2.8 describe the experiments investigating prominence production (Experiment 1), prominence perception (Experiment 2), prominence understanding (Experiment 3), and prominence placement (Experiment 4), respectively. Section 2.9 describes the native perception of non-native production follow-up experiment (Experiment 5). Finally, Section 2.10 discusses the types of statistical tests used to analyze the data.

2.2 General Experiment Structure

Twenty native English speakers, 20 native Mandarin speakers, and 20 native Korean speakers participated in four experiments (Exp. 1-4) during a single one to two hour session. All participants did the experiments in the same order. First, they did a prominence production experiment (Exp. 1), in which they were recorded reading the answers to questions. Second, they did a prominence perception experiment (Exp. 2), in which they were asked to determine which word in a recording was the last prominent word. Third, they did a prominence understanding experiment (Exp. 3), in which they were asked whether the answers in recorded question-answer pairs had context-appropriate prosody. Fourth, they did a prominence placement experiment (Exp. 4), in which they were asked which word they would make prominent if they were producing the answer in a question-answer pair. After all of the experiments were completed, the non-native participants took the Versant English Test. This is an oral test of English learners' ability to understand and produce conversational English. All participants finished the session by filling out a questionnaire on their language background.

The experimental tasks were ordered in this way to try to control the influence of an earlier task on performance in a later task. The prominence production experiment (Exp. 1) was run first so that participants' productions would not be influenced by the native English speaker productions of very similar sentence types that they would be hearing in the perceptual experiments (Exp. 2 and 3). The prominence perception experiment (Exp. 2) was run before the prominence understanding experiment (Exp. 3) because Exp. 2 was expected to be easier than the Exp.3. The understanding experiment (Exp. 3) actually required participants to both perceive and understand pitch accents. One goal of this study was to determine the role that pitch accent perception ability plays in pitch accent understanding. Therefore, it was important to get an accurate measure of participants' perception abilities, free from any practice effects that could occur if the prominence understanding experiment (Exp. 3) came before the perception experiment (Exp. 2). The prominence placement experiment (Exp. 4) was run after the prominence production experiment (Exp. 1), so that participants would not produce unnatural prosody in an effort to replicate the prominence locations they predicted in the placement experiment (Exp. 4). The prominence placement experiment (Exp. 4) was run last to allow the greatest amount of time to pass between the prominence production and placement experiments (Exp. 1 and 4, respectively). This reduced the chance that participants would simply report the prominence locations they produced, which may have been influenced by performance factors related to the difficulty of the speech production task. The prominence production experiment (Exp. 1) was kept free from influence because it involves a more natural task that reflects non-native speakers' actual language performance, and was therefore considered to be more informative than the placement experiment (Exp.4).

2.3 General Materials

These experiments used six sets of subject-verb-object stimulus sentences, containing twelve sentences each. In three of the sets, both target words (subject and object) are monosyllabic. In the other three sets, both target words have three syllables, with stress on the second syllable. There are three sets of sentences for each syllable count to allow a different set to be used for the prominence perception experiment (Exp. 2), the prominence understanding experiment (Exp. 3), and the production and placement experiments (Exp. 1 and 4).[4]

[4] All experimental materials can be found in Appendix A.

In order to make the experiments as comparable as possible, the sets of sentences had a number of features in common. All of the sentences had a subject-verb-object word order. All of the subjects were people's names. In each set (e.g. monosyllabic stimuli for the prominence perception experiment), nine of the objects had the form Determiner Noun, and three of the objects were proper nouns without determiners. In all sentences with determiners, the determiners were one syllable long. Because the phonetic features of some of the determiners are challenging for native Korean and Mandarin speakers (e.g. / ð/ in *the*, /z/ in *his*, /ɚ/ in *her*, /h/ in *his* and *her*, and final consonants in *some, his*, and *her*), the same determiners were used across the six sets of sentences. All sets had five instances of *a*, two instances of *his* or *her*, one instance of *some*, and one instance of *the*. All sentences used only past tense verbs with either one syllable (nine in each set) or two syllables (three in each set). All two-syllable verbs had first-syllable stress. Each verb was used only once.

Non-native speakers' segmental difficulties could cause a number of problems in this study. Trying to produce difficult segments may increase the likelihood of disfluencies, while trying to perceive such segments may keep non-native listeners from concentrating on the prominence perception and understanding tasks. To minimize these problems and avoid having them cause differences between the experiments and across experimental conditions, target words were chosen after considering lists of common segmental problems for native Korean and Chinese speaking learners of English (Swan & Smith, 2001). Because each of the stimulus sentences contained two target words (subject and object), twelve sets of words were matched for phonological features. A number of the problem features were eliminated from all target words, and other problem features were matched, with one exception, across all twelve word sets. Of the segmental problems listed for each language, all problems with English consonants, apart from voicing and phonotactics, were either eliminated or controlled via the stimuli restrictions described in Table 2.1. Segmental problems listed for vowels were not controlled, in order to make it possible to create twelve otherwise comparable lists of twelve words. A distinction was made between problematic vowels and consonants because the continuous nature of the vowel space may make non-native-like productions less likely to cause disfluencies. The segmental features that were controlled across the twelve word sets are listed in Table 2.1.

Language	Segments	Restrictions
Chinese	/n/	- No target word contained more than 1 /n/ - Only 6 target words contained /n/ per set
Chinese, Korean	/ɹ/, /l/	- No target word contained more than 1 /ɹ / or /l/ - Only 5 target words contained /ɹ / or /l/ per set
Chinese	/dʒ/, /tʃ/, /ʃ/	- No target word contained more than 1 /dʒ/, /tʃ/, or /ʃ/ - Only 2 target words contained /dʒ/, /tʃ/, or /ʃ/ per set
Chinese, Korean	/v/, /f/	- No target word contained more than 1 /v/ or /f/ - Only 1 target word contained /v/ or /f/ per set
Chinese	Word-final consonants	- Only 10 target words contained a word-final consonant per 1-syllable set - Only 1 target word contained a word-final consonant per 3-syllable set
Chinese	/h/	- No target word contained /h/
Chinese, Korean	/Θ/, /ð/	- No target word contained /Θ/ or /ð/
Chinese, Korean	/z/	- No target word contained /z/
Korean	/ ʃ / before /i/	- No target word contained / ʃ i/
Chinese	Word-final /l/	- No target word contained word-final /l/
Korean	Word-final / /ʃ/, /tʃ/, /dʒ/, /z/, and /t/	- No target word contained word-final /ʃ/, /tʃ/, /dʒ/, /z/, or /t/
Chinese	Consonant clusters	- No target word contained syllable-initial or -final consonant clusters
Korean	Word-final /s/ and /z/ morphemes	- No target word contained word-final /s/ or /z/ morphemes

Table 2.1. Segmental features that were eliminated or controlled across the twelve target word sets

2.4 Participants

Forty non-native English speakers participated in this experiment. Twenty were native Korean speakers (14 female, 6 male), and twenty were native Mandarin speakers (13 female, 7 male). All of the Korean speakers were from South Korea, and the majority (n=16) were from Seoul. Of the remaining Korean participants, one reported being from Cheongju, one from Incheon, one from Daejeon, and one from a variety of cities in the Jeollabuk-do, Jeollanam-do, and Gyeonggi-do provinces of South Korea. Crucially, the dialects spoken in all of these areas are non-tonal, like the standard dialect spoken in Seoul (Sohn 1999; Lee and Ramsey 2000). The native Korean speakers ranged in age from 19 to 33 (mean=26). The Mandarin speaking participants were more geographically diverse. Nineteen of them were from the Peoples Republic of China (PRC), and one was from Taiwan. Within the PRC, participants were from Beijing, Shanghai, and the provinces of Jiangsu, Liaoning, Henan, Shanxi, Guangxi, Sichuan, Hunan, and Hubei. Although some features of Mandarin vary from one area to another, all varieties of Mandarin have the same basic tonal structure (Li and Thompson 1989). The native Mandarin speakers ranged in age from 20 to 33 (mean=26).

All of the non-native English speakers were living in the U.S. at the time of the experiment. All of them first moved to an English-speaking country when they were 17 or older, and had lived in English-speaking countries for less than six years. Details of the Korean and Mandarin-speaking participants' age-of-arrival to live in an English-speaking country and length of time spent in English-speaking countries are provided in Table 2.2. The non-native participants were recruited from the Northwestern community through word-of-mouth and flyers posted on campus. All non-native participants were paid for their participation. None of the participants reported any speech or hearing impairments.

	Age-of-Arrival Mean (years)	Age-of-Arrival Range (years)	Length of Time Mean (months)	Length of Time Range (months)
Korean	23.75	17-29	22.5	0-66
Mandarin	24.3	18-32	12.65	0-68

Table 2.2. Mean and range of participants' age-of-arrival to live in an English-speaking country (in years), and length of time spent in English-speaking countries (in months), broken down by native language

After completing the experiment, the non-native participants took the Versant English Test (www.ordinate.com/products/english.jsp). This is an oral test of English learners' "ability to understand spoken language and respond intelligibly at a conversational pace on everyday topics." (Bernstein and Cheng 2008: 176). The test assesses non-native English speakers' sentence mastery, vocabulary, fluency, and pronunciation by having them read aloud, repeat sentences, answer questions, rearrange phrases to form sentences, and retell a story. The test takes about 15 minutes, is conducted over a landline telephone, and is automatically scored. Overall scores fall between 20 and 80, with 80 indicating high proficiency. The correlation between the Versant test and the human-scored TSE is 0.88 (n=59) (Bernstein and Cheng 2008). In the current study, there was no significant difference between the Versant scores for the Korean and Mandarin groups (U=230, p=0.42). Figure 2.1 shows the distribution of overall Versant scores for the two groups of non-native participants.

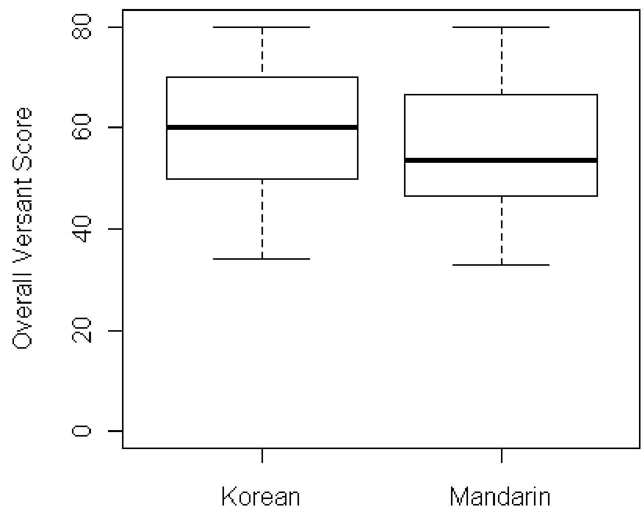

Figure 2.1. Boxplot of overall Versant scores for Korean and Mandarin participants

Twenty native English speaking controls also participated in the experiments (14 female, 6 male). The native English speakers had all lived in the U.S. for their entire lives, and ranged in age from 19 to 22 (mean=20). They were all undergraduate students at Northwestern University, and received course credit for their participation. The majority of these participants reported having studied other languages in addition to English (French, Spanish, German, Latin, and Russian). However, all of them learned these languages when they were 5 years old or older, described themselves as monolingual English speakers, and had been educated entirely in English. None of the participants reported any speech or hearing impairments.

2.5 Experiment 1: Prominence Production Experiment

2.5.1 Experiment 1: Prominence Production Experiment Materials

Twenty-four subject-verb-object sentences were constructed for this experiment using the process described in Section 2.3. Each sentence appeared in three contexts: subject narrow focus (SuNF), verb phrase broad focus (VPBF), and sentence broad focus (SBF). In the SuNF contexts, the sentence was preceded by a question about its subject, e.g. (6).

(6) Who bought a fan?
Kim bought a fan.

In the VPBF contexts, the sentence was preceded by a question about its VP, e.g. (7).

(7) What did Kim do?
Kim bought a fan.

In the SBF contexts, the sentence was preceded by the question "What happened?", e.g. (8).

(8) What happened?
Kim bought a fan.

A female native English speaker was recorded reading each of the context questions. Each question-answer pair was written on a PowerPoint slide, and the appropriate question recording was embedded in the slide.

2.5.2 Experiment 1: Prominence Production Experiment Design

Before the prominence production experiment, the non-native participants were trained on the pronunciations of words that appeared as subjects or objects in the experimental items.

During this training, they saw a series of PowerPoint slides, each with a single word written on it. They could click on a button to hear a female native English speaker producing that word in isolation. They were then asked to repeat the word. They could listen to each word as often as they liked, but were asked to listen to each word at least once.

In the experiment instructions, all participants were told to listen to each question, then to read the answer from the slide. They were instructed to speak as naturally and fluently as possible, as if they were having a real conversation, and to keep the question in mind as they read the answer.[5] Because of the challenges associated with producing context-appropriate prosody in an experimental setting, all of the items for each context type (e.g. SuNF) were grouped together. The order of the contexts was counterbalanced across three conditions.

Participants were recorded reading the answers in a sound-treated booth. They saw the question-answer pairs presented on a computer monitor in the booth and heard the question recordings played over headphones. They were recorded on an AKG C420 microphone, and the recordings were stored on a computer as .wav files.

2.6. Experiment 2: Prominence Perception Experiment

2.6.1 Experiment 2: Prominence Perception Experiment Materials

A second set of twenty-four subject-verb-object sentences was used in this experiment. A female native English speaker was recorded reading these sentences in the three contexts described above: subject narrow focus (SuNF), verb phrase broad focus (VPBF), and sentence broad focus (SBF). For the SuNF sentences, the only pitch accent was placed on the subject, while for the VPBF and SBF sentences, the final pitch accent was placed on the object. Figures

2.2-2.4 show pitch contours for the SuNF, VPBF, and SBF versions of one of the sentences used in the prominence perception experiment. Before being used, each recording was checked by both the experimenter and a second trained linguist (MB-B) to ensure that the prosody was appropriate for its context.

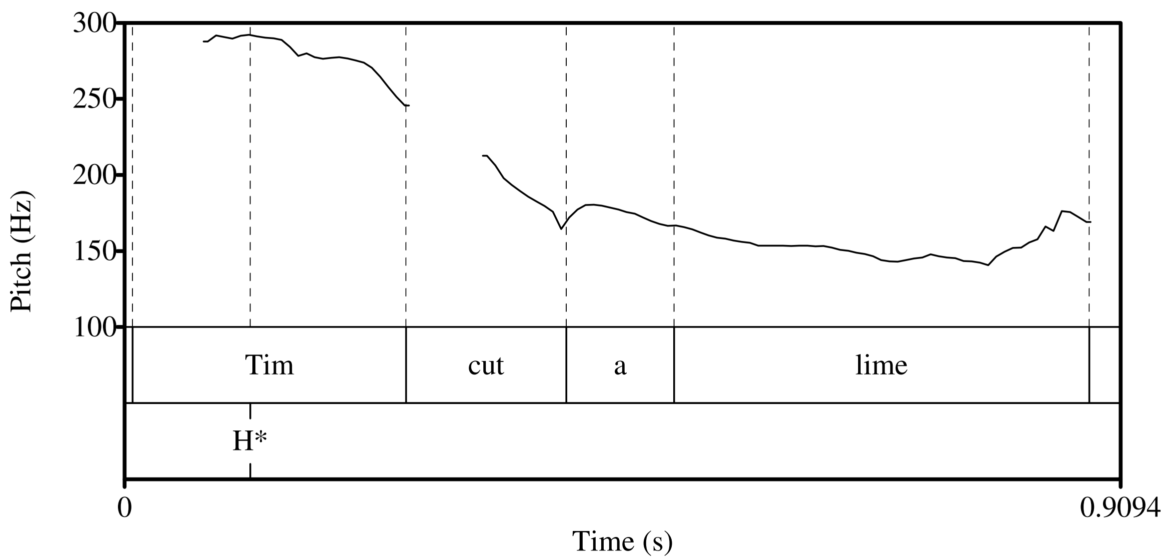

Figure 2.2. Pitch contour labeled with word boundary and pitch accent locations for the SuNF version of a typical stimulus sentence used in the prominence perception experiment (Exp. 2)

[5] The complete instructions for the prominence production experiment (Exp. 1) are provided in Appendix B.

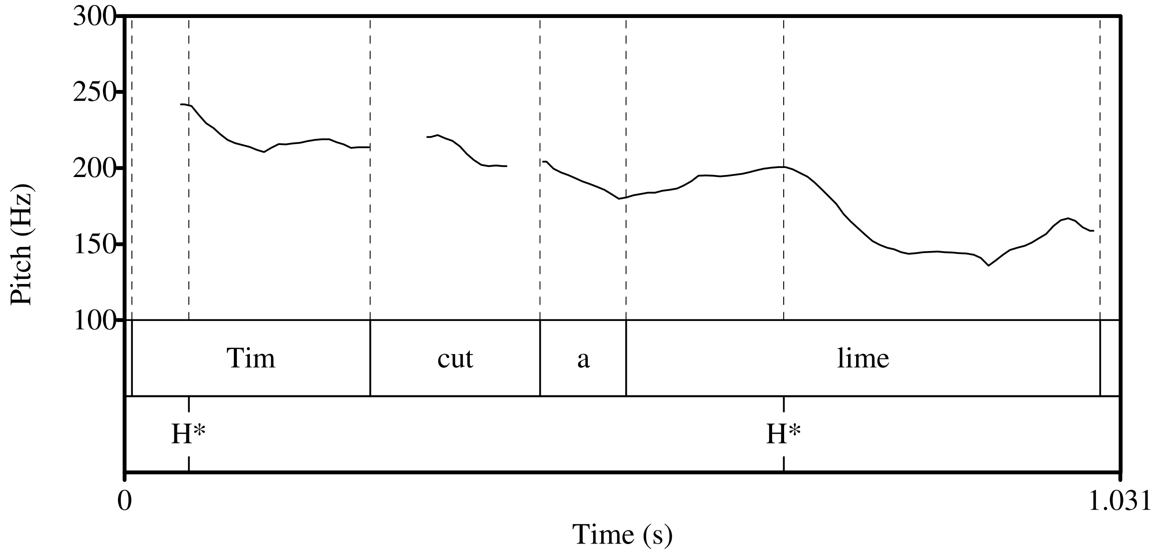

Figure 2.3. Pitch contour labeled with word boundary and pitch accent locations for the VPBF version of a typical stimulus sentence used in the prominence perception experiment (Exp. 2)

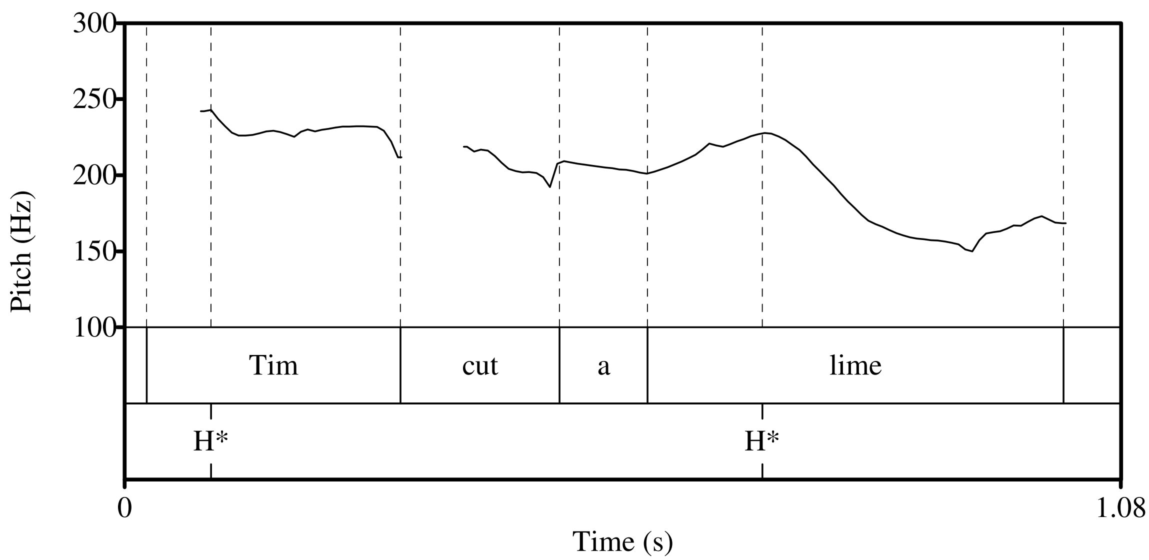

Figure 2.4. Pitch contour labeled with word boundary and pitch accent locations for the SBF version of a typical stimulus sentence used in the prominence perception experiment (Exp. 2)

2.6.2 Experiment 2: Prominence Perception Experiment Design

In the instructions for the prominence perception experiment, participants were trained on the meaning of the word 'prominence' using a description and recorded examples. The description stated that "In English, some words are pronounced in a way that makes them sound more important or prominent than others. A 'prominent' word stands out when you hear it. It may have noticeable intonation or may be especially long. A sentence can have more than one prominent word in it, but today we're interested in the LAST PROMINENT word in a sentence." The participants listened to two recordings of a single sentence. In one recording, the last prominent word was the verb, and in the other, the last prominent word was the object. Participants were told the location of the last prominent word for each recording.[6]

During the experiment, participants saw a sentence written on the screen in standard orthography, with no indication of prominence. They clicked on a button to hear a single recording of the sentence. They then had to answer the question "Is X the last prominent word?", where X was either the subject or object of the sentence, e.g. (9).

(9) Kim bought a fan.
Is Kim the last prominent word?

The last prominent word in the sentence was the subject for half of the items (sentences produced in the SuNF context), and the object for half of the items (sentences produced in the VPBF and SBF contexts). There were twice as many recordings with objects as the last prominent word than recordings with subjects as the last prominent word, because there were two conditions (VPBF and SBF) with prominent objects, but only one condition (SuNF) with prominent subjects. For this reason, the prominent subject recordings were each played twice. For each

[6] The complete instructions for the prominence perception experiment (Exp. 2) are provided in Appendix B.

context type, the question asked about word with the final pitch accent for half of the items (matched items), and the question asked about a different word for the other half of the items (mismatched items). This means that the word in the question was the subject for half the items, and the object for half the items. Participants heard 96 items in all (for each of the 24 sentences, participants heard two recordings with prominence on the subject and two recordings with prominence on the object). The items were presented in pseudo-random order. The experiment was conducted in a sound-treated booth using Max/MSP.[7] Participants saw stimuli presented on a computer monitor, and heard recordings over headphones.

2.7 Experiment 3: Prominence Understanding Experiment

2.7.1 Experiment 3: Prominence Understanding Experiment Materials

A third set of twenty-four subject-verb-object sentences was used in this experiment. The same female native English speaker who recorded the stimuli for the prominence perception experiment (Exp. 2) also recorded these sentences. The sentences were again produced in the three contexts described above: subject narrow focus (SuNF), verb phrase broad focus (VPBF), and sentence broad focus (SBF). For the SuNF sentences, the only pitch accent was on the subject, while for the VPBF and SBF sentences, the final pitch accent was on the object. A different female native English speaker was recorded reading the context questions. Once again, the recordings were checked for prosody appropriateness by the experimenter and a second trained linguist.

[7] Max is a programming environment, and MSP is a set of objects that can be used to present audio files in this programming environment. Together they can be used to design flexible speech perception experiments. For more information, see http://cycling74.com/products/maxmspjitter/.

2.7.2 Experiment 3: Prominence Understanding Experiment Design

In the instructions, participants were trained on the meaning of the word 'prosody' using a description and recorded examples. The description stated that "The word 'prosody' refers to the way that sentences are spoken. This includes things like the intonation and rhythm of words in a sentence. The prosody of a sentence can give information about what the sentence means, and different prosodies are appropriate in different contexts." The participants listened to two versions of an example question-answer pair. The question in this example put narrow focus on the verb (a context type that was not used in the experiment). One of the answer recordings (correctly) had the verb accented, while the other (incorrectly) had the subject accented. Participants were told which recording had appropriate prosody for answering the question.[8]

During the experiment, participants saw question-answer pairs written on the screen in standard orthography, with no indication of prominence. They clicked on a button to hear a recording of the question and the answer. They then had to answer the question "Is the prosody of the answer appropriate given the question?" Participants heard 72 items in all (each of the 24 sentences appeared in the three contexts). The items were presented in pseudo-random order.

For half of the items the question was presented with an answer that had appropriate prosody (matched items), and for the other half of the items the question was presented with an answer that had the final pitch accent on the wrong word (mismatched items). For example, in the SBF mismatched items, the question "What happened?" was followed by a answer that had been recorded in the SuNF context, so it had a final pitch accent on the subject. Both the SBF and VPBF mismatched items used sentences that had been produced in the SuNF context. This is because sentences in SBF and VPBF contexts should have final pitch accents on the object, but sentences produced in SuNF contexts have final pitch accents on the subject. Similarly, the

SuNF mismatched items used sentences that had been produced in SBF contexts. The experiment was conducted in a sound-treated booth using Max/MSP. Participants saw stimuli presented on a computer monitor, and heard recordings over headphones.

2.8 Experiment 4: Prominence Placement Experiment

2.8.1 Experiment 4: Prominence Placement Experiment Materials

The prominence placement experiment used the same 24 sentences as the prominence production experiment. Once again, each of these sentences appeared in SuNF, VPBF, and SBF contexts, resulting in 72 question-answer pairs in total.

2.8.2 Experiment 4: Prominence Placement Experiment Design

In the instructions, participants were reminded that in English "some words are pronounced in a way that makes them sound more important or prominent than others. A 'prominent' word stands out when you hear it. It may have noticeable intonation or may be especially long." [9]

During the experiment, participants saw question-answer pairs on the screen written in standard orthography with no indication of prominence. They did not listen to any recordings of the question or the answer. For each pair, they were asked "If you were answering the question above, which word would you make most prominent?" They could choose the subject, verb, or

[8] The complete instructions for the prominence understanding experiment (Exp. 3) are provided in Appendix B.
[9] The complete instructions for the prominence placement experiment (Exp. 4) are provided in Appendix B.

object of the sentence, or 'NONE'. For example, for the question-answer pair in (10a), they were given the choices in (10b-e).

(10) a. Who drew a line?
Jan drew a line.
b. Jan
c. drew
d. line
e. NONE

They were instructed to answer 'NONE' if they would make all of the words equally prominent. The experiment was conducted in a sound-treated booth using Max/MSP. Participants saw stimuli presented on a computer monitor.

2.9 Experiment 5: Native Perception of Non-Native Production Experiment

2.9.1 Experiment 5: Native Perception of Non-Native Production Participants

A new set of 24 native English speakers (17 female, 7 male) participated in a final perception experiment. These were all undergraduates at Northwestern University, who received course credit for their participation. They ranged in age from 18 to 23 (mean=20). All had lived in the U.S. for their entire lives. The majority of participants reported having studied other languages in addition to English, including Spanish, French, Mandarin, Korean, Jamaican Patois, Hebrew, German, Italian, Dutch, Japanese, Latin, and American Sign Language. However, all the participants learned these languages when they were 5 years old or older, described themselves as monolingual English speakers, and had been educated entirely in English. The only participant who reported any speech or hearing impairment had a minor speech impediment for which they had received speech therapy while in elementary school.

2.9.2 Experiment 5: Native Perception of Non-Native Production Experiment Materials

This experiment used recordings of sentences produced during the prominence production experiment (Exp. 1) described in Section 2.5, above. It included six sentences containing monosyllabic target words produced in each of the three contexts (SuNF, VPBF, SBF) by each speaker, resulting in a possible total of 18 recordings for each speaker. Due to recording errors, one recording for each of three Korean participants was unusable. These unusable recordings were for three different sentences; two were in the SuNF context, and one was in the SBF context. As a result, the experiment included a total of 1077 recordings. The stimulus recordings were paired with the recordings of questions that had been used in the production experiment. The question recordings were all produced by a female native English speaker.

2.9.3 Experiment 5: Native Perception of Non-Native Production Experiment Design

This experiment followed the structure of the prominence understanding experiment (Exp. 3) described in Section 2.7. In the instructions, participants were trained on the meaning of the word ‘prosody’ using a description and recorded examples. The description stated that “The word ‘prosody’ refers to the way that sentences are spoken. This includes things like the intonation and rhythm of words in a sentence. The prosody of a sentence can give information about what the sentence means, and different prosodies are appropriate in different contexts.” The participants listened to two versions of a question-answer pair. The question in this example put narrow focus on the verb (a context type that was not used in the experiment). One of the answer recordings (correctly) had the verb accented, while the other (incorrectly) had the subject accented. Participants were told which recording had appropriate prosody for answering the question. They were instructed that they would be listening to a series of question-answer pairs in which the answers were produced by both native and non-native English speakers. They were also told that sometimes the non-native speakers would mispronounce words, but they should try to overlook these mistakes and focus on the prosody of the answers.[10]

During the experiment, participants saw question-answer pairs written on the screen in standard orthography, with no indication of prominence. They clicked on a button to hear a recording of the question and the answer. They then had to answer the question "Is the prosody of the answer appropriate given the question?". Participants heard 358-360 items in all (six recordings by each of the 60 participants, minus the missing recordings). The items were presented in pseudo-random order. Each recording was rated by eight listeners. For each speaker, each listener heard two sentences from each of the three context conditions (e.g. SBF); these included three matched items and three mismatched items (in each context condition, one item was matched, and one item was mismatched). This means that for each listener, half of the dialogues they heard were matched, and half were mismatched. The experiment was conducted in a sound-treated booth using Max/MSP. Participants saw stimuli presented on a computer monitor, and heard recordings over headphones.

2.10 Statistical Analysis

Experiments 1-4 are all analyzed with mixed-effects logistic regression models, apart from the acoustic analysis of the production data, which is analyzed using mixed-effects linear regression models. Logistic regressions have been shown to avoid spurious effects that can arise when proportion data are analyzed using traditional ANOVAs (Jaeger 2008). The division of

[10] The complete instructions for the native perception of non-native production experiment (Exp. 5) are provided in Appendix B.

items into matched and mismatched items in the prominence perception and understanding experiments (Exp. 2 and 3) may at first suggest that a Signal Detection Theory analysis, such as d', would be more appropriate for these experiments (Stanislaw and Todorov 1999). In this section I will describe the differences between a d' analysis and a regression analysis, and explain why I believe that a regression analysis is more appropriate for these data.

The d' measure allows researchers to separate the effect of sensitivity to the distinction being measured from the participant's bias for one response over another (Stanislaw and Todorov 1999). It is therefore a good analysis to use if there is concern that different groups of participants could have different biases, or that individuals' biases might shift across the experimental conditions. For example, non-native English speakers may be more biased in favor of 'yes' answers than native speakers because of lack of confidence in their English abilities. This would make them seem more accurate on matched items and less accurate on mismatched items. However, it is unclear why participants' biases would be different for utterances produced in different contexts. A d' analysis would result in a separate sensitivity score (combining responses to matched and mismatched items) for each condition. Differences in accuracy across matched and mismatched sets would be captured with separate bias scores for each context. The d' scores for different language groups and contexts could then be compared using an ANOVA, with main effects for language group and context, as well as the interaction between the two.

A regression analysis can determine whether there are significant differences in accuracy between language groups, between contexts, or between matched and mismatched items. When a regression includes match as an independent variable, a significant difference between matched and mismatched items indicates a possible bias effect. A regression can also show interactions, not only between the language group and context variables, but also between these variables and match. It is important to explore the interactions between our variables of interest and match if there is any reason to suspect that different factors may be affecting results for the matched and mismatched items. This is the case for the prominence understanding experiment (Exp. 3), because the relationship between matched and mismatched items differs across the three contexts. In the SuNF context, the matched items differ from the mismatched items not only in the appropriate response (matched: 'yes', mismatched: 'no'), but also in the location of the final pitch accent in the recorded target sentence (matched: subject, mismatched: object) and the location of the final pitch accent relative to focus (matched: inside focus, mismatched: outside focus). However, in the SBF context, the matched items differ from mismatched items in appropriate response (matched: 'yes', mismatched: 'no') and the location of the final pitch accent (matched: object, mismatched: subject), but not in the location of the final pitch accent relative to focus (matched: inside, mismatched: inside). As a result, SBF mismatched items might be more acceptable than SuNF mismatched items, but this prediction does not hold for the SBF and SuNF matched items. This example demonstrates the complex relationship between matched and mismatched items in this experiment, and shows that matched and mismatched item sets might each be interesting in their own right.

A d' analysis is often used for experiments that have a simple relationship between items that should be accepted and items that should be rejected, such as the presence vs. absence of a signal, or two stimuli being either the same or different. This means that different patterns of performance are not expected for the accepted and rejected items, so they can safely be collapsed into a single sensitivity score. As demonstrated above, the relationship between matched and mismatched items in the prominence understanding experiment (Exp. 3) is far from simple. It is possible that different factors govern responses to matched and mismatched items, which could lead to different patterns of performance for the two item types. A d’ analysis would hide these unique patterns by pooling the results for matched and mismatched items within each context. Any interactions between match and context could only be captured by the different bias scores for each context, which would be extremely difficult to interpret.

The preceding paragraph has shown that a d’ analysis has the disadvantage of obscuring potentially interesting interactions between match and context. I will now address whether the regression analysis can handle possible response bias as well as a d’ analysis. Recall that in a regression, a significant difference between matched and mismatched items might signal a response bias. Response bias differences between language groups would be signaled by an interaction between language group and match. If there is an effect of language group, but no effect of match and no interaction between language group and match, then the language group effect can be considered real, and free from the influence of bias. If there is an effect of language group and effect of match, but no interaction between the two, then the difference between the language groups is still unaffected by the bias, which would be comparable across language groups. Only an interaction between language group and match, which would indicate different biases for the two groups, would be a cause for concern. In this case, any language group effect could actually be influenced by the different biases used in the different language groups. However, if both matched and mismatched items show either the same significant language group effect or no significant effect of language group, then this language group result can be considered reliable, regardless of bias. If a significant language group effect only appears for one of the match conditions, than bias could be influencing the overall language group result, and a d’ analysis should be run. This discussion has shown that by including match as an independent variable along with language group and context, we can determine whether response bias is present and whether it could be affecting any language group effects. A d' analysis would only be necessary if 1) there is a significant interaction between match and language group, and 2) different language group effects appear for matched and mismatched items.

This line of argumentation has focused on the prominence understanding experiment (Exp. 3), however, the same arguments apply to the prominence perception experiment (Exp. 2). The relationship between matched and mismatched items in Experiment 2 is not as complex as this relationship in Experiment 3. In Experiment 2, matched and mismatched items differ in appropriate response (matched: 'yes', mismatched: 'no'), and the location of the word asked about in the question (SuNF matched: subject, SuNF mismatched: object; SBF matched: object, SBF mismatched: subject). In Experiment 2, participants were asked whether word X is the last prominent word in a sentence. Participants may be more accurate at answering questions about early words or late words, leading to an advantage for SuNF items for one match set and an advantage for VPBF and SBF items in the other set. Like the potential interaction between language group and match in Experiment 3, such an effect would be hidden in a d' analysis because the d' analysis pools together results from matched and mismatched items. In contrast, the effect would be apparent in a regression, along with any simpler language group and context effects. In addition, a regression on Experiment 2 data could model any bias effects, just as in Experiment 3.

This section has shown that a regression analysis is more appropriate than a d' analysis for the prominence perception and understanding experiments (Exp. 2 and 3), because a regression allows for separate analyses of matched and mismatched items, which can be different in several ways. Like the d' analysis, it tests whether response bias is playing a role in the responses of participants in each language group. As a result, a d' analysis would only be necessary for comparing the sensitivities of different language groups if there is an interaction between language group and match and different language group effects are found for matched and mismatched items.

Chapter 3

Non-Native Perception and Understanding of English Prosodic Prominence - Experiments 2 and 3

3.1 Introduction

The two perceptual experiments were designed to tease apart the English learners' abilities to (1) perceive the location of English final pitch accents based on acoustic cues, and (2) understand the meaning of those pitch accent locations. The first skill was tested in the prominence perception experiment (Exp. 2), and the second skill was tested in the prominence understanding experiment (Exp. 3). Due to the nature of auditory speech perception, the prominence understanding experiment (Exp. 3) requires participants to use both their perception and understanding abilities. However, if they are more successful at the perception experiment than the understanding experiment, we can conclude that at least some of their difficulties in the latter experiment are due to problems interpreting prosodic meaning, rather than problems interpreting the acoustic cues for pitch accents.

The remainder of Chapter 3 provides details on the analyses and results for these two experiments, and discusses the implications of the results. Section 3.2 discusses the analysis and results of the prominence perception experiment (Exp. 2). Section 3.3 discusses the analysis and results of the prominence understanding experiment (Exp. 3). Section 3.4 provides a general discussion of the two perceptual experiments, and how their results fit together.

3.2 Experiment 2: Prominence Perception Experiment

3.2.1 Experiment 2: Prominence Perception Experiment Analyses

The participants' responses were analyzed using mixed-effects logistic regression models (Jaeger 2008), with participants and items as random factors. The regressions were run using the lmer function on R, version 2.9.1. The dependent variable in all regression models was item accuracy (correct or incorrect). Categorical variables were 'dummy' coded, so that they had a baseline level, to which the other level or levels in the variable were compared. The results of logistic regressions include estimates of the coefficients associated with each effect, in log odds. For categorical variables, a positive estimate means that participants were more likely to respond correctly to items in the target level than the baseline level. For continuous variables, a positive estimate means that higher values of the variable are associated with more correct answers.

It is important to note that if a regression includes an interaction, then the estimates for the individual effects associated with the interaction actually describe conditional effects, rather than main effects. For example, imagine that two categorical variables (e.g. gender and native language) with two levels each (male, female; native English speaker, non-native English speaker) are included in an interaction. They each have a baseline level (male; native), so the coefficients for the simple effects when an interaction is not included describe how being female (vs. male) relates to the dependent variable and how being non-native (vs. native) relates to the dependent variable. However, when the interaction is included, each of these individual effects is calculated as if the other variable was set to its baseline. So the coefficient for the gender distinction is calculated only for English natives, and the coefficient for native language distinction is calculated only for males (Aiken and West 1996). This fact helps with the follow-up analyses for significant interactions. We can determine the significance of each individual effect that is involved in an interaction by simply changing the baseline for the other variable in the interaction (Aiken and West 1996). So if there is a significant interaction between gender and native language, and we want to know whether there is a difference between males and females for both native and non-native English speakers, we can run the regression again with the non-natives as the native language baseline. Continuous variables involved in interactions were centered by subtracting the mean for all scores from each score. This reduces collinearity in the model, and makes the conditional effects more interpretable (Aiken and West 1996).

The items used in the prominence perception experiment (Exp. 2) can be described with three related variables: match (matched, mismatched), discourse context (SuNF, VPBF, SBF), and question word (subject, object). Because of the relationship among these three variables (if you know the match and context status for an item, you also know the question word), they cannot all be examined in the same statistical analysis. Therefore, context and question word were examined in two separate statistical analyses. The context analysis had context, match, language group, and number of syllables in the target words as independent variables. The question word analysis had question word, match, language group, and number of syllables as independent variables.

Regressions were run on both the full set and a subset of the prominence perception experiment (Exp.2) data. The native/non-native regressions were run on the full dataset, so they included all three language groups (English, Korean, Mandarin), while the non-native regressions were run on a subset of the data, which included only the Korean and Mandarin language groups. The native/non-native regressions compared pitch accent detection by native English speakers to detection by non-native English speakers. The non-native regressions let us look for differences between the two groups of non-native speakers and control for proficiency using the Versant English Test scores.

3.2.2 Experiment 2: Prominence Perception Native/Non-Native Regressions

The fixed variables in the final prominence perception native/non-native context regression were selected by first building a model that included all possible variables as main effects (with no interactions). All variables that were significant in this model were retained, and insignificant variables were eliminated. Interactions between the remaining variables were tested by building models that each included one of the possible two-way interactions. These new models were compared to the model with only the significant main effects, using likelihood ratio tests to determine whether adding the interaction significantly improved the fit of the model. The final model included all effects that were significant in the original model, plus all interactions that significantly improved the fit of the model. The variables tested were: language group (English, Korean, Mandarin), match (matched, mismatched), context (Subject Narrow Focus, VP Broad Focus, Sentence Broad Focus), and number of syllables in the target words (one syllable, three syllables). In the first context regression, with all four variables as main effects, only language group had a significant effect on the likelihood of participants producing a correct answer, so only this variable was retained.

The fixed variables in the final prominence perception native/non-native question word regression were selected using the same process as the context regression. The variables tested were: language group (English, Korean, Mandarin), match (matched, mismatched), question word (subject, object), and number of syllables in the target words (one syllable, three syllables).

In the first question word regression, with all four variables as main effects, only language group had a significant effect on the likelihood of participants producing a correct answer.

As a result of the comparisons described above, only one final prominence perception native/non-native regression model was built, because in both the context and the question word regressions only language group was significant. The final native/non-native model had only language group (English, Mandarin, Korean) as a fixed variable. Because there was only one fixed variable, no interactions were tested. For the language group variable, the English group served as a baseline, and was compared to the Korean group and the Mandarin group. As Table 3.1 shows, the only significant effect was the comparison between English and Mandarin, although the comparison between English and Korean approached significance. Native English speakers were significantly better at detecting English final pitch accent location than native Mandarin speakers. These results are illustrated in Figure 3.1.

	Estimate	Std. Error	z-value	p
Intercept	3.3037	0.2552	12.947	<0.001
Korean	-0.5894	0.3302	-1.785	0.0743
Mandarin	-1.5208	0.3246	-4.686	<0.001

Table 3.1. Parameter values for fixed variables in the prominence perception native/non-native regression, with English as the language group baseline

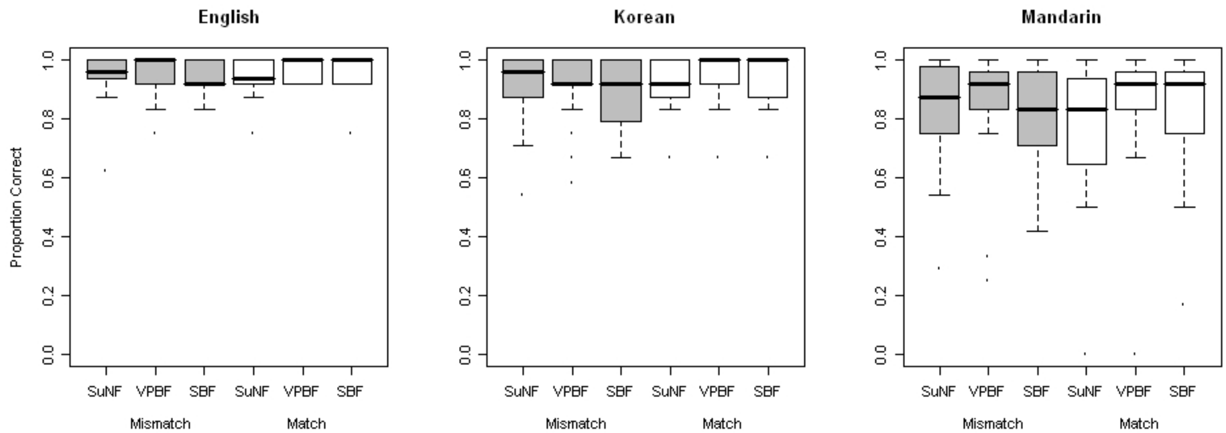

Figure 3.1. Boxplots showing proportions of correctly identified accent locations by language group, match, and context

3.2.3 Experiment 2: Prominence Perception Non-Native Regressions

The variables considered for the context and question word non-native regressions were the same as those considered for the native/non-native regressions with two exceptions. First, participants' Versant scores were included as a control variable. Second, the variable language group had only two levels (Korean and Mandarin), which were contrasted with each other.

Once again, the only significant effect in the context model that included all main effects and the question word model that included all main effects was language group. As a result, only one final prominence perception non-native regression model was built. This model included only language group as a fixed variable, and included no interactions. As Table 3.2 shows, the Korean language group was significantly more accurate than the Mandarin group.

	Estimate	Std. Error	z-value	p
Intercept	2.7243	0.2479	10.989	<0.001
Mandarin	-0.9337	0.3201	-2.916	<0.005

Table 3.2. Parameter values for fixed variables in the prominence perception non-native regression, with Korean as the language group baseline

3.2.4 Experiment 2: Prominence Perception Discussion

These results have several implications. First, native Mandarin participants were significantly less accurate at perceiving English pitch accent location than native Korean and English participants, but there was no significant difference between the Korean and English groups. English proficiency (based on Versant score) was not a significant variable in the non-native regressions, indicating that the difference between the Korean and Mandarin groups was not due to any proficiency differences between the groups. This difference between the two non-native language groups may be due to transfer from the English learners' native languages, as predicted by the Transfer Model. Seoul Korean uses pitch post-lexically only, while Mandarin, as a tone language, uses it primarily lexically. It is possible that native Korean speakers can more easily use intonational cues to sentence-level prominence than native Mandarin speakers, as they have more experience doing this in their native language.

These results also show that features of the stimuli themselves (e.g. the location of the accent in the sentence, the information structure of the sentence, and the number of syllables in the target words) did not significantly affect participants' accuracy. This means that non-native English speakers can apply their knowledge of the acoustic cues marking pitch accents across a variety of word types and contexts. It also means that any differences that arise between different types of items in the pitch accent understanding experiment are unlikely to be due to differences in participants' ability to perceive the location of a pitch accent across contexts. Rather, pitch accent understanding differences are more likely to be due to differences in their ability to map between pitch accent location and the information structure of the sentence.

Finally, it is interesting to note that although there were significant differences between the native Mandarin group and the native English group, the non-native participants were quite successful at this task. The median proportion of correct answers for the non-native participants was .91. This compares to the median proportion of correct answers for native participants of .96. This relative success may be due to the similar acoustic cues used to prosodically mark (narrow) focus across the three languages: higher F0s (if we consider only H* pitch accents in English) and increased duration.

3.3 Experiment 3: Prominence Understanding Experiment

3.3.1 Experiment 3: Prominence Understanding Experiment Analyses

The items used in the prominence understanding experiment (Exp. 3) can be described with three variables: match (matched, mismatched), discourse context (SuNF, VPBF, SBF), and accent location (subject, object). Because of the relationship among these three variables (if you know the match and context status for an item, you also know its accent location), they cannot all be examined in the same statistical analysis. Therefore, context and accent location were examined in two separate statistical analyses. The analyses in Section 3.3.2 examine the effect of context (among other factors) on native and non-native listeners' interpretations of utterances. In these analyses, items were grouped by the type of context question (SuNF, VPBF, SBF). The analyses in Section 3.3.3 examine the effect of accent location and other factors on sentence interpretation. In these analyses, items were grouped by the location of the final pitch accent in the answer (subject, object).

3.3.2 Experiment 3: Prominence Understanding Context Regressions

3.3.2.1 Experiment 3: Prominence Understanding Native/Non-Native Context Regressions

As in the prominence perception analysis, the analysis of prominence understanding began by building a model with all possible main effects. The variables tested were: language group (English, Korean, Mandarin), match (matched, mismatched), context (SuNF, VPBF, SBF), and number of target word syllables (one syllable, three syllables). All of these variables were significant except for number of syllables, therefore the syllables variable was removed from later models. Interactions between the remaining fixed variables were tested by adding each interaction to a new model along with all the significant fixed variables from the first model, then using a likelihood ratio test to compare the new model to the original model, with only the language group, match, and context main effects. Only the interaction between match and context significantly improved the fit of the model ($X^2(2)=68.427$; $p<0.001$).

As a result of the preceding comparisons, the final model included language group, match, context, and the interaction between match and context as fixed variables. For the language group variable, the English group served as a baseline, and was compared to the Korean group and the Mandarin group. For the match variable, the matched items served as a baseline. Two native/non-native regressions were run, one with SBF as the context baseline, and the other with VPBF as the context baseline.

Both native/non-native regressions showed that the English group was significantly more accurate than the Korean and Mandarin groups. These regressions also included significant interactions between match and all pairs of contexts. Because of these interactions, the match and context individual effects are actually conditional effects, and cannot be interpreted as representing the effects of context or match over the entire dataset. However, as the main purpose of the native/non-native regression is to determine whether the two non-native language groups differed from the English group, a complete exploration of the interactions between match and context is left until the non-native regression. The parameter values for the native/non-native regressions are listed in Table 3.3. These results are illustrated in Figure 3.2.

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	4.57366	0.32746	13.967	<0.001
Mismatched (for SBF condition)	-1.40976	0.18526	-7.610	<0.001
Korean	-2.18581	0.37986	-5.754	<0.001
Mandarin	-2.27262	0.38000	-5.981	<0.001
SuNF (for matched items)	1.73333	0.32528	5.329	<0.001
VPBF (for matched items)	-0.04904	0.20494	-0.239	0.81086
Mismatched: SuNF	-2.00343	0.36466	-5.494	<0.001
Mismatched: VPBF	0.85511	0.27285	3.134	<0.005
VPBF Baseline				
Intercept	4.52464	0.32616	13.873	<0.001
Mismatched (for VPBF condition)	-0.55475	0.19329	-2.870	<0.005
Korean	-2.18582	0.37986	-5.754	<0.001
Mandarin	-2.27263	0.38001	-5.981	<0.001
SBF (for matched items)	0.04905	0.20494	0.239	0.81085
SuNF (for matched items)	1.78235	0.32386	5.504	<0.001
Mismatched: SuNF	-2.85833	0.36947	-7.736	<0.001
Mismatched: SBF	-0.85502	0.27285	-3.134	<0.005

Table 3.3. Parameter values for fixed variables in the prominence understanding native/non-native context regressions, with English as the language group baseline and matched as the match baseline

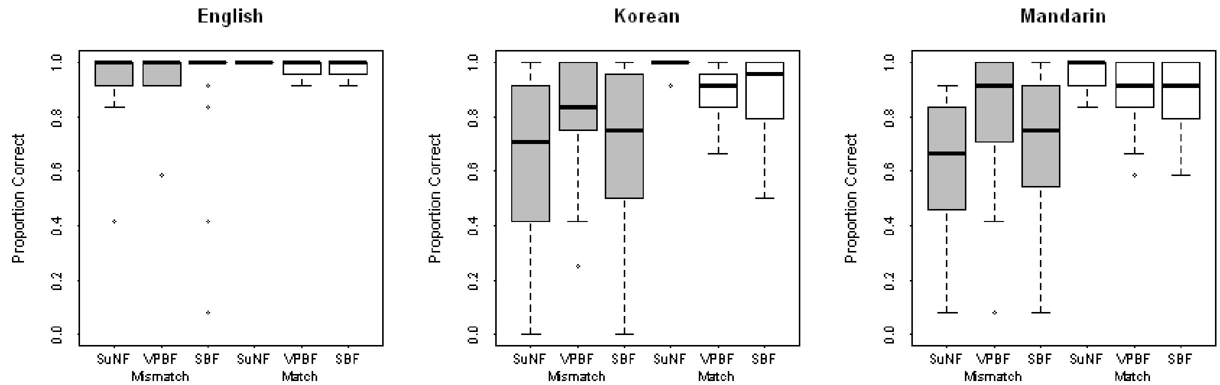

Figure 3.2. Boxplots showing proportions of correctly interpreted accents by language group, match, and context

3.3.2.2 Experiment 3: Prominence Understanding Non-Native Context Regressions

The variables considered for the non-native regression were the same as those considered for the native/non-native regression with two exceptions. First, participants' Versant scores were tested as a control factor. Second, the variable language group had only two levels (Korean and Mandarin), which were contrasted with each other. The regression containing all possible main effects showed that only match, context, and Versant score were significant, so language group and syllables were not retained. Tests of individual interactions revealed that only the interaction between match and context significantly improved the fit of the model over a model with the significant main effects alone ($X^2(2)= 67.403; p<0.001$).

As a result, the final regression included match, context, Versant score, and the interaction between match and context as fixed variables. The baseline values were the same as those in the native/non-native regressions. Versant score did not have a baseline value because it is a continuous variable.

These regressions show that Versant score was a significant predictor of accuracy;

participants with higher Versant scores tended to be more accurate. There were also significant interactions between match and all pairs of context types. These regressions provide information on the context effects for matched items, but the significant interaction shows that it is also important to determine the context effects for mismatched items. To further explore the interactions between match and context, a second pair of non-native regressions was run, this time with mismatched items as the match baseline. For the matched items, non-natives were significantly more accurate on SuNF items than VPBF and SBF items. For the mismatched items, non-natives were significantly more accurate on VPBF items than SuNF and SBF items, and more accurate on SBF than SuNF items. Finally, participants were significantly more accurate on matched items than mismatched items in the SBF and VPBF conditions. The parameter values for the regression on matched items are listed in Table 3.4, and those for the regression on mismatched items are listed in Table 3.5.

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	0.35269	0.63785	0.553	0.58031
Mismatched (for SBF condition)	-1.34190	0.19427	-6.908	<0.001
Versant	0.03453	0.01070	3.228	<0.005
SuNF (for matched items)	1.66639	0.32844	5.074	<0.001
VPBF (for matched items)	-0.04613	0.21370	-0.216	0.82911
Mismatched:SuNF	-2.07301	0.37126	-5.584	<0.001
Mismatched:VPBF	0.83541	0.28524	2.929	<0.005
VPBF Baseline				
Intercept	0.30657	0.63725	0.481	0.63046
Mismatched (for VPBF condition)	-0.50657	0.20241	-2.503	<0.05
Versant	0.03453	0.01070	3.228	<0.005
SBF (for matched items)	0.04612	0.21370	0.216	0.82912
SuNF (for matched items)	1.71252	0.32687	5.239	<0.001
Mismatched:SBF	-0.83532	0.28524	-2.928	<0.005
Mismatched:SuNF	-2.90824	0.37640	-7.727	<0.001

Table 3.4. Parameter values for fixed variables in the prominence understanding non-native context regressions, with matched as the match baseline

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	-0.98921	0.63297	-1.563	0.11810
Matched (for SBF condition)	1.34189	0.19427	6.907	<0.001
Versant	0.03453	0.01070	3.228	<0.005
SuNF (for mismatched items)	-0.40662	0.15648	-2.599	<0.01
VPBF (for mismatched items)	0.78920	0.17402	4.535	<0.001
Matched:SuNF	2.07301	0.37126	5.584	<0.001
Matched:VPBF	-0.83521	0.28524	-2.928	<0.005
VPBF Baseline				
Intercept	-0.20000	0.63422	-0.315	0.75249
Matched (for VPBF condition)	0.50657	0.20241	2.503	<0.05
Versant	0.03453	0.01070	3.228	<0.005
SBF (for mismatched items)	-0.78920	0.17402	-4.535	<0.001
SuNF (for mismatched items)	-1.19581	0.17115	-6.987	<0.001
Matched:SBF	0.83532	0.28524	2.928	<0.005
Matched:SuNF	2.90851	0.37641	7.727	<0.001

Table 3.5. Parameter values for fixed variables in the prominence understanding non-native context regressions, with mismatched as the match baseline

3.3.3 Experiment 3: Prominence Understanding Accent Location Regressions

3.3.3.1 Experiment 3: Prominence Understanding Native/Non-Native Accent Location Regressions

The location of the final pitch accent in the answer was examined because it could influence listeners' perceptions of items, independently of context. For example, listeners may find it easier to understand the meaning of a sentence's prosody when the final pitch accent is on the subject. In the following analyses, the context factor is replaced by the accent location factor, in order to determine whether final pitch accent location plays a role in the accuracy of participants' responses.

The prominence understanding native/non-native accent location analysis began by building a model with all possible main effects. The variables tested were: language group (English, Korean, Mandarin), match (matched, mismatched), final pitch accent location (subject, object), and number of target word syllables (one syllable, three syllables). All of these variables were significant except for syllables, therefore the syllables variable was removed from later models. Interactions between the remaining fixed variables were tested by adding each one to a new model, then comparing the new model to the original model, with only the language group, match, and accent location main effects. Only the interaction between match and accent location significantly improved the fit of the model ($X^2(1)= 14.36$; $p<0.001$).

As a result of the preceding comparisons, the final model included language group, match, accent location, and the interaction between match and accent location as fixed variables. For the language group variable, the English group served as a baseline, and was compared to the Korean group and the Mandarin group. For the match variable, the matched items served as a baseline. For the accent location variable, the object position served as a baseline.

The prominence understanding native/non-native accent location regression confirmed that English participants were significantly more accurate than Korean and Mandarin participants. It also included a significant interaction between the match and accent location factors. As a result, we cannot interpret the match and accent location variables as main effects. However, an exploration of the match and accent location interaction is put off until the non-native analysis, because the main purpose of the native/non-native regression was to determine whether the English language group differed from the Korean and Mandarin groups. The parameter values for these regressions are listed in Table 3.6. The results are illustrated in Figure 3.3.

	Estimate	Std. Error	z-value	p
Intercept	4.5513	0.3115	14.609	<0.001
Mismatched (for object condition)	-1.6594	0.1426	-11.636	<0.001
Korean	-2.1750	0.3779	-5.756	<0.001
Mandarin	-2.2610	0.3780	-5.981	<0.001
Accent Subject (for matched items)	1.7516	0.3104	5.643	<0.001
Mismatched: Accent Subject	-1.1248	0.3280	-3.429	<0.001

Table 3.6. Parameter values for fixed variables in the prominence understanding native/non-native accent location regression, with matched as the match baseline, English as the language group baseline, and object as the accent location baseline

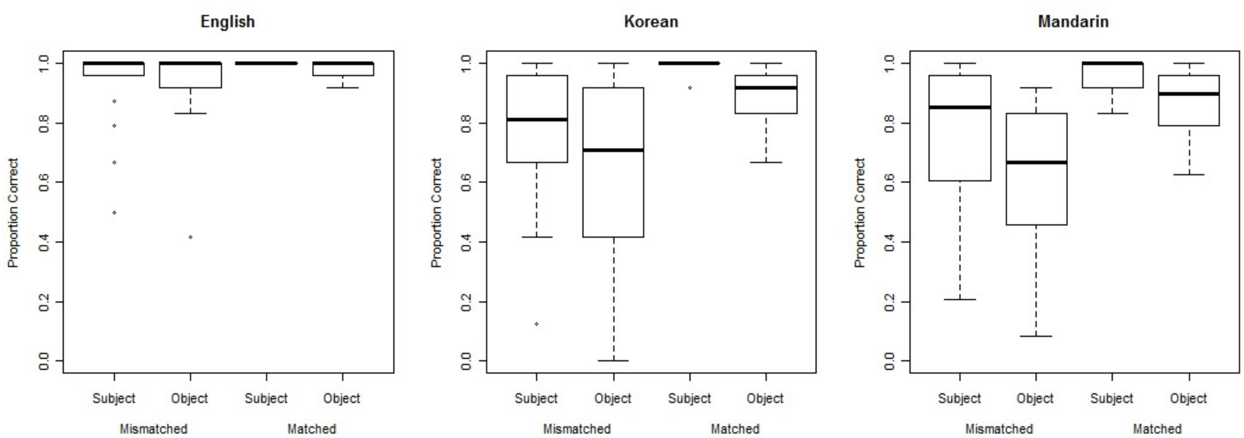

Figure 3.3. Boxplots showing proportions of correctly interpreted items by language group, match, and context

3.3.3.2 Experiment 3: Prominence Understanding Non-Native Accent Location Regressions

The variables considered for the prominence understanding non-native accent location regression were the same as those considered for the native/non-native regression with two exceptions. First, participants' Versant scores were tested as a control factor. Second, the variable language group had only two levels (Korean and Mandarin), which were contrasted with each other. The non-native regression containing all possible main effects showed that only match, accent location, and Versant score were significant, so language group and syllables were eliminated. Tests of individual interactions revealed that only the interactions between match and accent location ($X^2(1)= 8.9914$; $p<0.005$) and between Versant score and accent location ($X^2(1)= 7.6189$; $p<0.01$) significantly improved the fit of the model over a model with the significant main effects alone.

As a result, the final regression included match, accent location, Versant score, and the interactions between match and accent location and between Versant score and accent location as fixed variables. The baseline values were the same as those in the native/non-native regression. Versant score did not have a baseline value because it is a continuous variable.

This regression confirms the significant interaction between accent location and match found in the native/non-native regression, and shows that there is also a significant interaction between accent location and Versant score. In order to fully explore the interaction between match and accent location, a second regression was run, which was identical to the first except with mismatched items as the match baseline rather than matched items. Taken together, these regressions reveal that participants were significantly more accurate on subject items than object items regardless of whether they were matched or mismatched. The interaction seems to be a matter of degree rather than the direction: there is a greater difference between subjects and objects for matched items than mismatched items. The parameter values for the matched baseline and mismatched baseline regressions are listed in Tables 3.7 and 3.8, respectively. In order to explore the interaction between accent location and Versant score, a third regression was run, this one identical to the original regression (Table 3.7), but with subject as the accent location baseline. The parameter values for the matched/subject baseline regression are listed in Table 3.9. The regressions in Tables 3.7 and 3.9 show that general English proficiency (as measured by Versant score) plays a significant role in participant accuracy at understanding pitch accents for both object and subject items. More proficient participants were better at understanding pitch accents than less proficient ones. Once again, the interaction seems to be a matter of degree, with a stronger effect of proficiency for subject than for object items. The relationship between Versant score and accuracy at understanding pitch accents for subject and object items can be seen in Figure 3.4. The regressions in Tables 3.7 and 3.9 also show that participants were more accurate on matched items than mismatched items (across both subject and item conditions).

	Estimate	Std. Error	z-value	p
Intercept	2.274726	0.177104	12.844	<0.001
Mismatched (for object condition)	-1.695851	0.149582	-11.337	<0.001
Versant (for object condition)	0.023224	0.011184	2.076	<0.05
Accent Subject (for matched items)	1.824956	0.317675	5.745	<0.001
Mismatched: Accent Subject	-1.017153	0.335526	-3.032	<0.005
Versant: Accent Subject	0.024921	0.008116	3.070	<0.005

Table 3.7. Parameter values for fixed variables in the prominence understanding non-native accent location regression, with matched as the match baseline and object as the accent location baseline

	Estimate	Std. Error	z-value	p
Intercept	0.578869	0.176592	3.278	<0.005
Matched (for object condition)	1.695853	0.149582	11.337	<0.001
Versant (for object condition)	0.023224	0.011184	2.076	<0.05
Accent Subject (for mismatched items)	0.807808	0.144866	5.576	<0.001
Matched: Accent Subject	1.017157	0.335527	3.032	<0.005
Versant: Accent Subject	0.024920	0.008116	3.070	<0.005

Table 3.8. Parameter values for fixed variables in the prominence understanding non-native accent location regression, with mismatched as the match baseline and object as the accent location baseline

	Estimate	Std. Error	z-value	p
Intercept	4.099684	0.329648	12.437	<0.001
Mismatched (for subject condition)	-2.713005	0.305317	-8.886	<0.001
Versant (for subject condition)	0.048140	0.011632	4.138	<0.001
Accent Object (for matched items)	-1.824957	0.317678	-5.745	<0.001
Mismatched: Accent Object	1.017154	0.335530	3.031	<0.005
Versant: Accent Object	-0.024910	0.008116	-3.069	<0.005

Table 3.9. Parameter values for fixed variables in the prominence understanding non-native accent location regression, with matched as the match baseline and subject as the accent location baseline

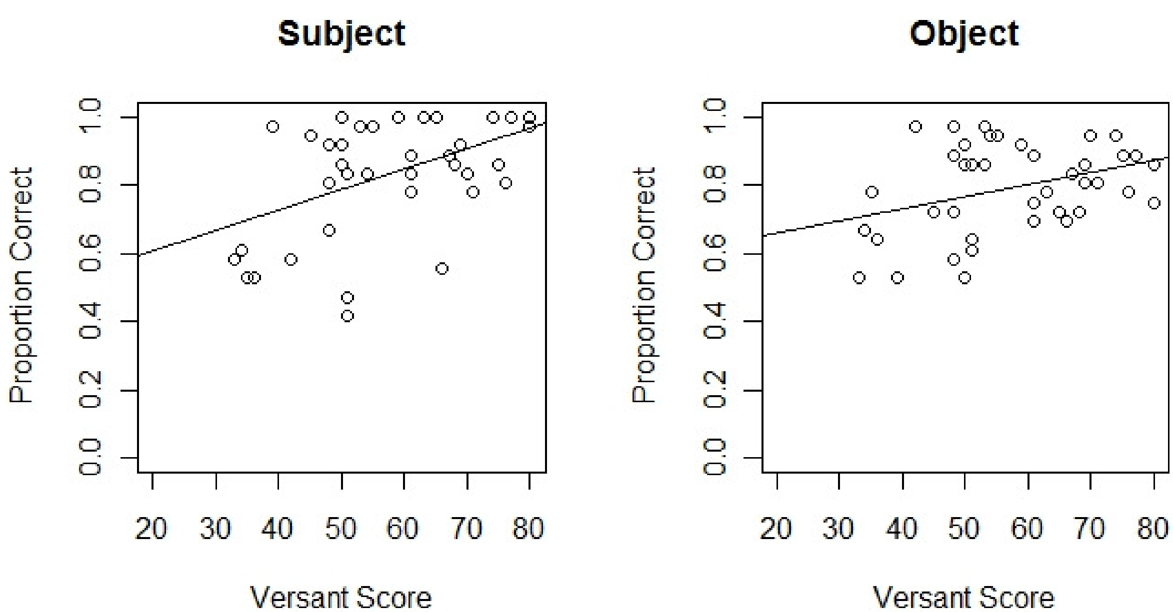

Figure 3.4. Scatterplots showing relationship between accuracy and Versant score for prominence understanding non-native subject and object items

3.3.4 Experiment 3: Prominence Understanding Discussion

These results show that non-native English speakers struggle to match pitch accent location with intended focus when listening to native English speech. A non-native English speaker's ability to correctly interpret English pitch accent placement varies with their proficiency. It is also influenced by the two factors proposed in the L2 Challenge Models: the Frequency factor (the number of contexts in which an accent placement is used), and the Relationship factor (the size of the focused constituent). Therefore, these results support the Hybrid L2 Challenge Model. The results do not support either Transfer Model, because there were no significant differences between the Korean and Mandarin participants in this experiment.

When asked whether the prosody of a sentence was appropriate, non-native participants were overly lenient. They were more likely to correctly accept prosody as appropriate for matched items than to correctly reject it for mismatched items. However, non-natives' performance on this task did improve with greater English proficiency. Participants with higher Versant scores were more accurate than those with lower scores.

For mismatched items, non-native participants were significantly worse at identifying the prosody of a sentence as inappropriate in the SuNF context, relative to the VPBF and SBF contexts. In the SuNF context, the question asks about the subject of the sentence, but for the mismatched items, the final pitch accent is on the object instead of the subject (11).

(11) a. Who bought a fan?
b. [Kim] bought a FAN.

As predicted by the L2 Challenge Models' Frequency factor, non-natives may incorrectly accept this accent placement because having a final pitch accent on the object is common in English (it is used in SBF, VPBF, and object narrow focus contexts). They could be over-extending this pattern to contexts where it is not appropriate (e.g. the SuNF context).

Non-natives were also significantly worse at correctly rejecting inappropriate prosody for mismatched items in the SBF context than in the VPBF context. In SBF items, the question "What happened?" puts the whole sentence in focus, but for the mismatched items, the final pitch accent is on the subject instead of the object (12).

(12) a. What happened?
b. [KIM bought a fan].

The L2 Challenge Models' Relationship factor correctly predicts this result. It proposes that SBF focus will be more difficult to acquire because it focuses a larger constituent than VPBF focus, thereby providing more possible (incorrect) final pitch accent locations within the focused area. In this case, the word with the final pitch accent is within the focused constituent (the whole sentence). Non-natives may more easily accept final pitch accents when they are within the focused constituent, even if they are not placed in the standard location.

The non-natives were most successful at rejecting inappropriate prosody in the VPBF context. In this context, the question asks about what somebody did, but for the mismatched items, the final pitch accent is on the subject instead of the object (13).

(13) a. What did Kim do?
b. KIM [bought a fan].

For these items, the final pitch accent is not in a common location or within the focused constituent, so they may have seemed more clearly incorrect.

For the matched items, non-native participants were especially good at identifying appropriate prosody in the SuNF context, as predicted by the L2 Challenge Model's Relationship factor. In these items, the question asks about the subject, and the final pitch accent is on the subject (14).

(14) a. Who bought a fan?
b. [KIM] bought a fan.

Narrow focus items like (14) have the most straightforward relationship between focus and accent location: only one word is in focus, and it receives the only pitch accent. This makes such items easier to interpret. In contrast, broad focus items like (15) and (16) have an indirect relationship between final pitch accent location and focus.

(15) a. What did Kim do?
b. Kim [bought a FAN].

(16) a. What happened?
b. [Kim bought a FAN].

The analyses of accent location indicate that understanding sentences with a final pitch accent on the subject was easier than sentences with a final pitch accent on the object. This held true across matched and mismatched items. Two possible causes may explain this result. One possibility is suggested by the L2 Challenge Model of pitch accent perception. It proposes that a narrow-focus final pitch accent on the subject (as found in the SuNF stimuli) will be easier to perceive than a broad-focus final pitch accent on the object (as found in the VPBF and SBF stimuli). This is because the subject pitch accent may be acoustically more pronounced than the object pitch accent, either because of its location at the beginning of the phrase or because it is in narrow focus. If this is true, it would suggest that listeners have an easier time understanding the meaning of utterances with clear final pitch accents, which would not be surprising. When the final pitch accent is clear, the listener is less likely to mistakenly believe that a different word has the final pitch accent. Even if the final pitch accent is still identifiable, less obvious prominence could require listeners to use more cognitive resources to identify its location, leaving them less able to interpret the meaning of the prosody. Surprisingly, the pitch accent perception L2 Challenge Model's prediction was not supported in the prominence perception experiment. In that experiment, there were no significant differences between items with final pitch accent on the subject and the object. Perhaps the differences in perceptibility only play a role in performance during tasks that require more mental resources, such as the prominence understanding experiment.

A second possible explanation for the accented subject advantage is that there is something about the location of the subject in the sentence that makes it easier to interpret sentences with a final pitch accent on the subject. For instance, because the most common final pitch accent location is on the object, listeners may be more attentive to final pitch accents on the subject, which is a more unusual location. Even if the accents on subjects and objects are acoustically equivalent, the novelty of a pitch accent on a subject might attract listeners' attention, making them more accurate at interpreting the meaning of the accent placement.

More work is needed to determine the true cause of the subject advantage. Acoustic analyses of the stimuli recordings could indicate whether the accented subjects in this study have stronger acoustic markers of prominence (e.g. greater pitch range or duration) than the accented objects. If such a difference were found, a follow-up study could be conducted in which final pitch accent location were manipulated artificially by re-synthesizing recordings with varying pitch ranges and durations on the target words. This would ensure that any differences found between subject and object positions were not due to acoustic differences between pitch accents.

Another important follow-up study would compare accuracy on items with narrow focus on the subject and the object. The current study confounds accent location and focus type (broad vs. narrow): only the narrow focus items have final pitch accents on the subject and only the broad focus items have final pitch accents on the object. By examining narrow focus in a variety of positions, researchers could more accurately evaluate the role of sentence position on interpretation, independent of focus type.

It is encouraging that the more proficient English learners in this study were better at understanding the meaning of final pitch accent location than the less proficient language learners. This suggests that as language learners gain experience with a non-native language, they become better able to understand its prosody. Interestingly, the more proficient non-native English speakers had a greater advantage for items with a final pitch accent on the subject than those with a final pitch accent on the object. It is possible that over time, language learners become aware of the relative rarity of sentences with final pitch accents on their subjects, and as a result, attend to these sentences more.

Taken together, these results do not provide evidence for the Transfer Models of non-native prosodic focus-marking understanding, as there were no significant differences between the Korean and Mandarin language groups. However, the non-native participants were most accurate at identifying appropriate prosody in the SuNF context, which is the context that is most similar across the three languages. It may be that Korean and Mandarin are too similar to cause differences in behavior in this type of task. The next step is to examine this issue with a wider variety of native and non-native languages, in order to find more conclusive evidence for or against a Transfer Model.

3.4 Experiments 2 and 3: Prominence Perception and Understanding General Discussion

The results discussed in this chapter show that different processes are at work in the acquisition of the ability to perceive prosodic prominence in an L2, and the ability to understand prosodic prominence in an L2. One prediction of the Transfer Model of prominence perception was realized in the data. Native speakers of Korean, a language that is non-tonal, like English, had a general advantage in perceiving prosodic prominence in English. However, for prominence understanding acquisition, the Hybrid L2 Challenge Model made more accurate predictions. As predicted by the Relationship factor, native speakers of both Korean and Mandarin more easily understood narrow focus than broad focus (in the matched condition), and more easily understood VPBF utterances than SBF utterances (in the mismatched condition). As predicted by the Frequency factor, native speakers of Korean and Mandarin seemed to think that the most commonly used pattern of prominence (with a final pitch accent on the object) could apply in any of the contexts examined. These results highlight the importance of separating phonological and phonetic effects in studies of second language prosody acquisition, and the importance of considering challenging features of the L2 in addition to possible transfer effects.

Chapter 4

Non-Native Placement and Realization of English Prosodic Prominence - Experiments 1, 4, and 5

4.1 Introduction

In this study, English learners' production skills were analyzed in several ways. Their knowledge of which word should receive a final pitch accent in a given context was tested with the computer-based prominence placement experiment (Exp. 4). The prominence placement experiment did not require participants to actually speak. This was to avoid mistakes that may arise from the challenge of producing English speech, which requires speakers to consider how to produce segmental features as well as suprasegmental features. Like the prominence understanding experiment (Exp. 3), discussed in Chapter 3, the prominence production experiment (Exp. 1) required participants to use multiple skills simultaneously. In this case, they had to use their prominence placement skills as well as their realization skills. A subset of the recordings made during the production experiment was used as stimuli for the native perception of non-native production experiment (Exp. 5). The native perception of non-native production experiment provided a listener-oriented perspective on the appropriateness of the non-natives' prominence production.

The remainder of Chapter 4 describes the results of these three experiments. Section 4.2 discusses the analysis and results of the prominence placement experiment (Exp. 4). Section 4.3 discusses the analysis and results of the native perception of non-native production experiment (Exp. 5), which uses recordings from the prominence production experiment (Exp. 1).

4.2 Experiment 4: Prominence Placement Experiment

4.2.1 Experiment 4: Prominence Placement Analysis

The prominence placement experiment (Exp. 4) tested which word participants believed should be most prominent when producing a sentence in a particular context. It aimed to study participants' knowledge of English prosody, which they may not be able to put to use in perception and production tasks because of the multiple concurrent demands on their attention during these tasks. Recall that participants could select the target sentence's subject, verb, or object, or could select 'NONE' if they thought that no word should be prominent. A participant's response was considered correct if they reported that they would accent the word predicted to be have the obligatory final pitch accent for that context. For SBF and VPBF contexts, the sentence object was considered correct. For SuNF contexts, the sentence subject was considered correct. Selections of the verb or 'NONE' were never considered correct. The responses were first analyzed to determine whether the language groups differed in their accuracy on this task. Two sets of regressions were run: native/non-native, comparing the native English language group to the Korean and Mandarin groups; and non-native, comparing the Korean and Mandarin groups to each other.

4.2.2 Experiment 4: Prominence Placement Native/Non-Native Regressions

The potential fixed variables in this regression were: language group (English, Korean, Mandarin), context (SuNF, VPBF, SBF), and number of target word syllables (one syllable, three syllables). The fixed variables were selected using the same procedure used for the prominence perception and understanding experiments. A model with all main effects was built, and variables that were significant in this model were retained. In this model, only language group and context were significant, so they were retained. Next, a model with only these two main effects was compared to a model with these effects plus the interaction between them. A likelihood ratio test showed that including the interaction significantly improved the fit of the model ($X^2(4)=28.436$, $p<0.001$).

The final regressions included language group and context and the interaction between them as fixed variables. The baseline for the language group variable was English; this was compared to Korean and Mandarin. Because of the inclusion of the interaction between language group and context, these regressions do not provide general information about the difference between the English group and the two non-native groups across all items. Instead, each regression only provides information about the difference between the English group and the non-native groups for the context baseline for that regression. As a result, three regressions were run to allow for comparisons between the English group and the Korean and Mandarin groups in all three contexts. In the first regression SBF was the baseline context, in the second regression VPBF was the baseline context, and in the third regression SuNF was the baseline context.

The parameter values for the fixed variables in these native/non-native regressions are listed in Table 4.1. There were significant interactions between some context comparisons (SBF/VPBF, SuNF/VPBF) and the comparison between the English and Mandarin groups. The English group was significantly more accurate than the Mandarin group in the SBF and VPBF conditions, but not the SuNF condition. The English language group was also significantly more accurate in the SuNF condition than the SBF or VPBF conditions, and more accurate in the VPBF condition than the SBF condition. The three language groups' accuracies for each context can be seen in Figure 4.1.

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	-1.3047	0.4286	-3.044	<0.005
Korean (for SBF condition)	0.2538	0.5926	0.428	0.668473
Mandarin (for SBF condition)	-1.2468	0.6091	-2.047	<0.05
SuNF (for English group)	7.3553	0.5613	13.104	<0.001
VPBF (for English group)	2.3011	0.1897	12.131	<0.001
SuNF:Korean	-1.2890	0.7231	-1.783	0.074668
SuNF:Mandarin	1.1724	0.8443	1.389	0.164974
VPBF:Korean	-0.3542	0.2506	-1.414	0.157480
VPBF:Mandarin	-0.9431	0.2743	-3.439	<0.001
VPBF Baseline				
Intercept	0.9965	0.4242	2.349	<0.05
Korean (for VPBF condition)	-0.1004	0.5878	-0.171	0.864378
Mandarin (for VPBF condition)	-2.1897	0.5956	-3.676	<0.001
SBF (for English group)	-2.3011	0.1897	-12.131	<0.001
SuNF (for English group)	5.0542	0.5427	9.313	<0.001
SBF:Korean	0.3542	0.2506	1.414	0.157484
SBF:Mandarin	0.9429	0.2743	3.438	<0.001
SuNF:Korean	-0.9348	0.7030	-1.330	0.183620
SuNF:Mandarin	2.1148	0.8207	2.577	<0.01
SuNF Baseline				
Intercept	6.05066	0.66434	9.108	<0.001
Korean (for SuNF condition)	-1.03523	0.88613	-1.168	0.24270
Mandarin (for SuNF condition)	-0.07443	0.95591	-0.078	0.93794
SBF (for English group)	-7.35531	0.56129	-13.104	<0.001
VPBF (for English group)	-5.05418	0.54270	-9.313	<0.001
SBF:Korean	1.28901	0.72314	1.783	0.07467
SBF:Mandarin	-1.17238	0.84433	-1.389	0.16497
VPBF:Korean	0.93483	0.70303	1.330	0.18362
VPBF:Mandarin	-2.11545	0.82079	-2.577	<0.01

Table 4.1. Parameter values for fixed variables in the prominence placement native/non-native regressions, with English as the language group baseline

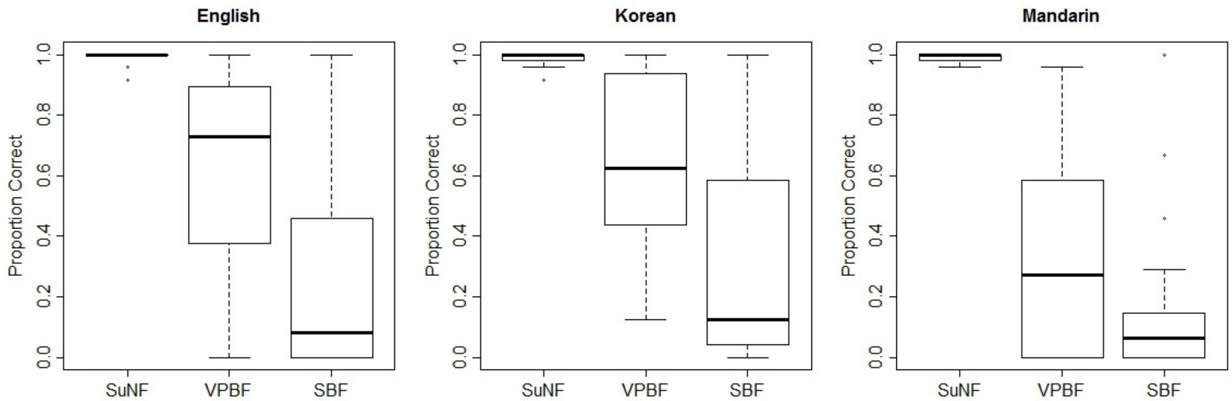

Figure 4.1. Boxplots showing proportions of correct predictions of accent placement by language group and context

4.2.3 Experiment 4: Prominence Placement Non-Native Regressions

The non-native regressions compared the Korean and Mandarin groups to each other, and had language, context, number of target word syllables, and participant Versant score as potential fixed variables. The final regression variables were determined using the procedure described above for the native/non-native regressions. In the regression with all possible main effects, only language group and context were significant, so these variables were retained. A likelihood ratio test showed that the model that included language group and context as well as the interaction between them was a significantly better fit for the data than a model that included only the two main effects ($X^2(2)= 19.989$, $p<0.001$).

The final regressions included language group and context and the interaction between them as fixed variables. Korean was the baseline for the language group variable. As in the native/non-native regression, the inclusion of the interaction between language and context means that these regressions do not provide information on the effects of either language group or context over all items. Therefore, three versions of the non-native regression were run, one with SBF as the context baseline, one with VPBF as the context baseline, and one with SuNF as the context baseline.

The parameter values for the fixed variables in these native/non-native regressions are listed in Table 4.2. There were significant interactions between all pairs of contexts and the contrast between the Korean and Mandarin groups. The Korean group was significantly more accurate than the Mandarin group for the SBF and VPBF items, but not the SuNF items. Because of the interaction between language and context, the regressions in Table 4.2 only provide information on the context effect for the Korean group (the language group baseline). To determine the context effect for the Mandarin group, another two regressions were run with Mandarin as the language group baseline, one with SBF as the context baseline, and one with VPBF as the context baseline. These two regressions provide information about the Mandarin participants' relative accuracy on all three pairs of contexts. The parameter values for these two regressions are listed in Table 4.3. The regressions in Tables 4.2 and 4.3 show that both groups of non-native participants were significantly more accurate on SuNF items than SBF or VPBF items, and more accurate on VPBF items than SBF items.

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	-1.0657	0.3901	-2.732	<0.01
Mandarin (for SBF condition)	-1.4789	0.5543	-2.668	<0.01
SuNF (for Korean group)	6.1022	0.4591	13.291	<0.001
VPBF (for Korean group)	1.9636	0.1648	11.915	<0.001
SuNF:Mandarin	2.3698	0.7686	3.083	<0.005
VPBF:Mandarin	-0.6003	0.2575	-2.331	<0.05
VPBF Baseline				
Intercept	0.8980	0.3875	2.317	<0.05
Mandarin (for VPBF condition)	-2.0791	0.5412	-3.842	<0.001
SBF (for Korean group)	-1.9636	0.1648	-11.915	<0.001
SuNF (for Korean group)	4.1386	0.4493	9.211	<0.001
SBF:Mandarin	0.6002	0.2575	2.331	<0.05
SuNF:Mandarin	2.9696	0.7507	3.956	<0.001
SuNF Baseline				
Intercept	5.0366	0.5751	8.758	<0.001
Mandarin (for SuNF condition)	0.8908	0.8728	1.021	0.30745
SBF (for Korean group)	-6.1022	0.4591	-13.291	<0.001
VPBF (for Korean group)	-4.1386	0.4493	-9.211	<0.001
SBF:Mandarin	-2.3698	0.7686	-3.083	<0.005
VPBF:Mandarin	-2.9700	0.7508	-3.956	<0.001

Table 4.2. Parameter values for fixed variables in the prominence placement non-native regressions, with Korean as the language group baseline

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	-2.5446	0.4134	-6.156	<0.001
Korean (for SBF condition)	1.4789	0.5543	2.668	<0.01
SuNF (for Mandarin group)	8.4720	0.6179	13.712	<0.001
VPBF (for Mandarin group)	1.3634	0.1985	6.869	<0.001
SuNF:Korean	-2.3698	0.7686	-3.083	<0.005
VPBF:Korean	0.6001	0.2575	2.330	<0.05
VPBF Baseline				
Intercept	-1.1812	0.3978	-2.969	<0.005
Korean (for VPBF condition)	2.0791	0.5412	3.842	<0.001
SBF (for Mandarin group)	-1.3634	0.1985	-6.869	<0.001
SuNF (for Mandarin group)	7.1086	0.6020	11.807	<0.001
SBF:Korean	-0.6002	0.2575	-2.331	<0.05
SuNF:Korean	-2.9703	0.7507	-3.956	<0.001

Table 4.3. Parameter values for fixed variables in the prominence placement non-native regressions, with Mandarin as the language group baseline

4.2.4 Experiment 4: Prominence Placement Error Analysis

In addition to looking at the participants' accuracy in placing accents in various contexts, it is also informative to consider the kinds of mistakes they make. Table 4.4 shows the percentage of times participants in each language group gave each answer, for the three contexts.

Language	Structure	Subject	Verb	Object	'NONE'
English	SuNF	99.0%	0%	0.8%	0.2%
	VPBF	1.9%	20.6%	63.5%	14.0%
	SBF	1.3%	10.8%	29.0%	59.0%
Korean	SuNF	98.8%	0.4%	0.6%	0.2%
	VPBF	0.6%	29.6%	65.0%	4.8%
	SBF	6.3%	14.8%	29.4%	49.6%
Mandarin	SuNF	99.0%	0.6%	0.2%	0.2%
	VPBF	0%	56.5%	31.5%	12.1%
	SBF	0.6%	10.8%	15.4%	73.1%

Table 4.4. Mean percentages of responses by language group and context. The shaded squares indicate the 'correct' answer for each context.

In four cases, participants within a language group selected one of the incorrect answers more often than one of the correct answers. This was true for the SBF context in all three language groups: participants selected 'NONE' more often than the object when asked which word they would make most prominent in this context (English: 'NONE' 59%, object 29%; Korean: 'NONE' 50%, object 29%; Mandarin: 'NONE' 73%, object 15%). In addition, Mandarin participants selected the verb more often than the object in the VPBF context (verb 57%, object 32%).

Beyond the cases in which participants preferred an incorrect answer over a correct answer, there were a number of cases in which participants chose an incorrect answer for 10% of items or more. For the VPBF context, the verb was a popular response across all three language groups (English: verb 21%; Korean: verb 30%; Mandarin: verb 57%). Also for the VPBF context, the English and Mandarin groups often chose 'NONE', although this was a less popular answer than the verb (English: 'NONE' 14%; Mandarin: 'NONE' 12%). For the SBF context, as discussed above, the most popular response was 'NONE' for all three language groups (English: 'NONE' 59%; Korean: 'NONE' 50%; Mandarin: 'NONE' 73%). The verb was selected less often for the SBF context, but still formed a sizable minority of answers in all three language groups (English: verb 11%; Korean: verb 15%; Mandarin: verb 11%).

4.2.5 Experiment 4: Prominence Placement Discussion

One of the most striking features of the results for the prominence placement experiment (Exp. 4) is that the native English speakers gave unexpected answers so often. The mean accuracy score for English participants was 64%. This suggests that the task may tap into meta-linguistic knowledge that even native English speakers do not necessarily possess. However, the

English participants, like participants in the other two language groups, greatly lowered their accuracy scores through their preference for the 'NONE' answer in the SBF context. In the SBF context, all of an utterance's constituents are [+focus], so there is no contrast between [+focus] and [-focus] constituents. Participants' answers for the SBF context may reflect this fact, rather than a complete lack of knowledge of its information structure. Despite these unexpected results and complications, the prominence placement experiment (Exp. 4) can still provide insights into native and non-native English speakers' knowledge and beliefs about accent placement in various contexts.

All three language groups were very successful at determining the correct accent location for the SuNF context, with 99% accuracy for each group. Participants in all three language groups were significantly more accurate on SuNF items than on items in the broad focus contexts. This result could support either one of the Transfer Models, or the Relationship or Hybrid L2 Challenge Models of prosody acquisition. Both Transfer Models predict that the learner will most accurately place pitch accents when their L1 dictates prominence on the same word for a particular information structure. Both Korean and Mandarin provide the option of marking narrow focus on the subject by making the subject prosodically prominent. The Relationship factor of the L2 Challenge Model predicts that the learner will find it easier to correctly place pitch accents in English narrow focus sentences than in broad focus sentences. In narrow focus, there is a one-to-one correspondence between focus and accent: only one word is focused and it is accented. As a final note, we must be cautious of over-interpreting the results of this experiment, because native English speakers also struggled with the VPBF and SBF contexts in this task. Still, it is clear that the non-native participants have learned that the subject should be prominent when it is in narrow focus.

While participants in all three language groups struggled with the broad focus items, they were all more successful at placing prominence for the VPBF items than the SBF items. This result is also predicted by the Relationship factor of the L2 Challenge Model. This can be attributed to the common belief (at least among these participants) that items in the SBF context do not have any prominent word. This was the most common response for SBF items, and it was much more common for SBF items than for VPBF items. The Relationship factor predicts that VPBF items should be easier than SBF items because in VPBF contexts participants mark focus on a phrase that provides new information (in contrast to the old information in the rest of the sentence). In SBF contexts, the whole sentence provides new information, so it does not contrast with any given constituent, making its focused status less clear.

There were no significant differences between the English and Korean groups on this task, and their pattern of responses were surprisingly similar. However, the Mandarin group was significantly less accurate than both the English group and the Korean group for the VPBF and SBF contexts. Versant score was not a significant variable in this analysis, so the difference is unlikely to be due to lower proficiency. The difference between the Korean and Mandarin groups supports the Any Prominence Location Helps Transfer Model. This model posits that if a speaker's L1 and L2 both prosodically mark focus for a particular constituent, this will help the speaker to place and understand pitch accents marking focus on that constituent in their L2. The model predicts that Korean speakers will be better than Mandarin speakers at placing prominence in VPBF contexts. This is because Korean marks VP broad focus prosodically, while Mandarin does not. The model also predicts that Korean speakers could be better than Mandarin speakers at placing prominence in SBF contexts, because Korean speakers mark some kind of broad focus prosodically in their L1. Therefore Korean speakers may have a better mental representation of broad focus than Mandarin speakers.

Interestingly, general English proficiency did not seem to play a role in a participant's success at this task. This can be seen both in the poor performance by the native English speakers, and in the fact that Versant score was not a significant variable in the non-native regression model. As Figure 4.2 shows, similar numbers of high proficiency and low proficiency non-natives struggled with this task. This does not seem to be something that language learners pick up as they gain experience with a language, nor does it seem to be (successfully) taught in language classes. In future work, it would be interesting to investigate how explicit instruction on prominence location in different contexts affects non-natives' performance on more naturalistic perception and production tasks.

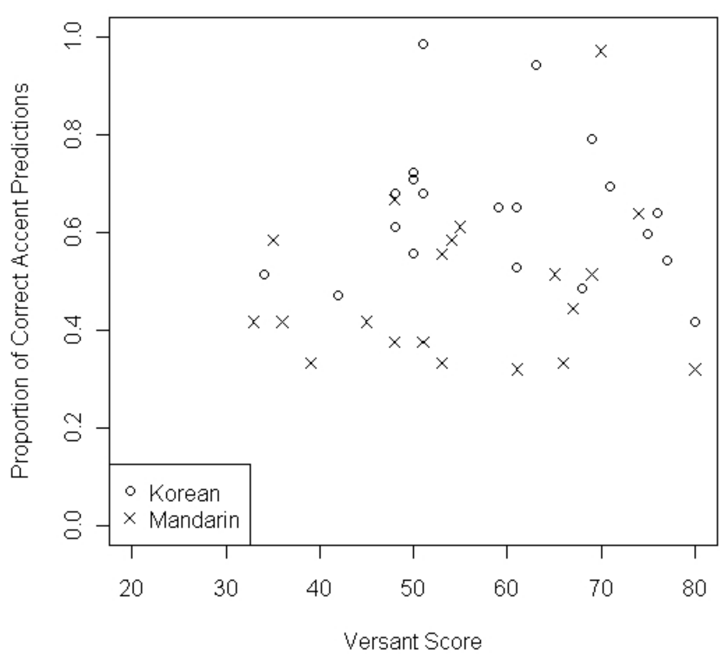

Figure 4.2. Scatterplot of participants' proportions of correct accent placement by their Versant scores

4.3 Experiments 1 and 5: Native Perception of Non-Native Production Experiment

4.3.1 Experiments 1 and 5: Native Perception of Non-Native Production Analysis

In the native perception of non-native production experiment (Exp. 5), a new set of native English speakers listened to sentences produced by the native English, Korean, and Mandarin speakers in the prominence production experiment (Exp. 1). The listeners judged whether the prosody of each sentence production was appropriate for a given context. The listeners' judgments of the productions were analyzed using mixed-effects logistic regression models. The dependent variable in all models was production accuracy (whether a production was judged to be correct or incorrect by a particular listener). The random effects were sentence and listener.

Two types of regressions were run on the data: native/non-native and non-native. The native/non-native regressions included all three language groups, while the non-native regressions included only the Mandarin and Korean groups. The native/non-native regressions compared native English speakers to non-native speakers. The non-native regressions compared the two groups of non-native speakers and controlled for proficiency with the Versant scores. The fixed variables for each regression were chosen using the model-comparison method used in the analyses of the prominence perception, understanding, and placement experiments.

4.3.2 Experiments 1 and 5: Native Perception of Non-Native Production Native/Non-Native Regressions

In the native/non-native regression, the language group (English, Korean, Mandarin) and context (SuNF, VPBF, SBF) variables were considered for inclusion as fixed variables. Number of syllables in the target word was not considered because this experiment only used sentences with monosyllabic subjects and objects. Match was also excluded from the analysis because this only affected listener behavior, not speaker behavior, which was the factor under investigation. The fixed variables were selected by first building a model with both main effects, and retaining the variables that were significant in this model. In this model, both language group and context were significant, so they were both retained. Next, a model with only these two main effects was compared to a model with these variables plus the interaction between them. A likelihood ratio test showed that including the interaction significantly improved the fit of the model ($X^2(4)= 217.6$, $p<0.001$).

The final regressions included language group, context, and the interaction between them as fixed variables. For the language group variable, the English group served as a baseline, and was compared to the Korean group and the Mandarin group. The interaction between context and language group means that the differences between the English group and the non-native groups over all items cannot be determined in these regressions. As a result, three native/non-native regressions were run, one with SBF as the context baseline, one with VPBF as the context baseline, and one with SuNF as the context baseline.

The parameter values for the fixed variables in these regressions are listed in Table 4.5. There were significant interactions between the SuNF/VPBF and SuNF/SBF context pairs and the contrasts between both non-native groups and English. In the SuNF context, productions by native English speakers were judged to have appropriate prosody by native listeners significantly more often than those by Korean and Mandarin speakers. However, in the VPBF and SBF contexts, Mandarin productions were judged to be appropriate significantly more often than English productions. The English speakers' SuNF productions were judged to be appropriate significantly more often than their VPBF productions. These effects can be seen in Figure 4.3.

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	1.0502	0.1040	10.096	<0.001
Korean (for SBF condition)	-0.1668	0.1031	-1.617	0.1058
Mandarin (for SBF condition)	0.2331	0.1078	2.162	<0.05
SuNF (for English group)	0.1360	0.1064	1.278	0.2013
VPBF (for English group)	-0.1262	0.1033	-1.222	0.2218
SuNF:Korean	-1.0469	0.1444	-7.251	<0.001
SuNF:Mandarin	-1.7236	0.1479	-11.654	<0.001
VPBF:Korean	0.2121	0.1453	1.460	0.1444
VPBF:Mandarin	0.1622	0.1519	1.067	0.2858
VPBF Baseline				
Intercept	0.92405	0.10259	9.007	<0.001
Korean (for VPBF condition)	0.04527	0.10235	0.442	0.658275
Mandarin (for VPBF condition)	0.39541	0.10704	3.694	<0.001
SBF (for English group)	0.12614	0.10328	1.221	0.221936
SuNF (for English group)	0.26215	0.10504	2.496	<0.05
SBF:Korean	-0.21202	0.14529	-1.459	0.144478
SBF:Mandarin	-0.16233	0.15192	-1.069	0.285278
SuNF:Korean	-1.25893	0.14385	-8.751	<0.001
SuNF:Mandarin	-1.88606	0.14735	-12.800	<0.001
SuNF Baseline				
Intercept	1.1862	0.1058	11.213	<0.001
Korean (for SuNF condition)	-1.2137	0.1011	-12.008	<0.001
Mandarin (for SuNF condition)	-1.4905	0.1012	-14.728	<0.001
SBF (for English group)	-0.1361	0.1064	-1.278	0.2011
VPBF (for English group)	-0.2622	0.1050	-2.496	<0.05
SBF:Korean	1.0470	0.1444	7.251	<0.001
SBF:Mandarin	1.7236	0.1479	11.654	<0.001
VPBF:Korean	1.2590	0.1439	8.752	<0.001
VPBF:Mandarin	1.8857	0.1473	12.798	<0.001

Table 4.5. Parameter values for fixed variables in the native perception of non-native production native/non-native regressions, with English as the language group baseline

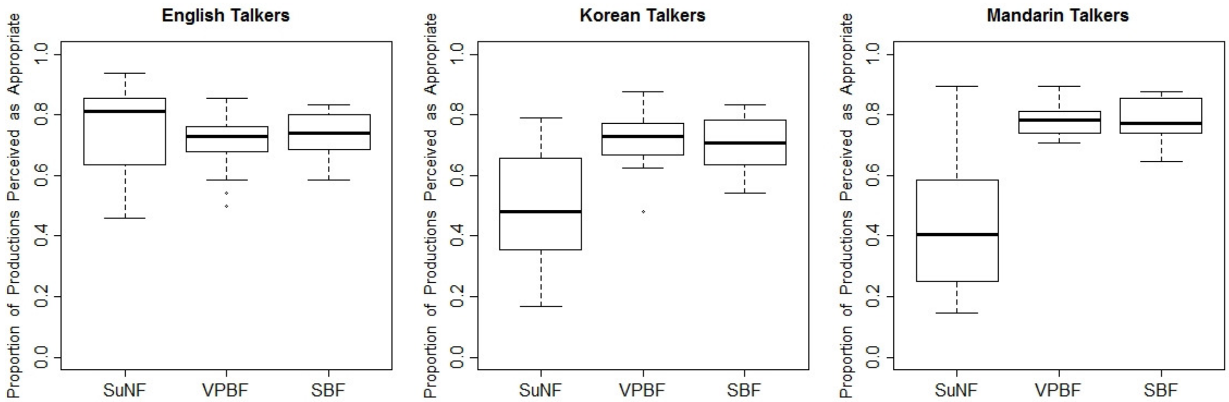

Figure 4.3. Boxplots of proportion of productions perceived to have context-appropriate prosody, by talker native language and context

4.3.3 Experiments 1 and 5: Native Perception of Non-Native Production Non-Native Regressions

The fixed variables considered for the non-native regressions were: language group (English, Korean, Mandarin), context (SuNF, VPBF, SBF), and Versant score. A model with all three main effects was built, and variables that were significant in this model were retained. In this model, all three variables were significant, so they were all retained. Next, a series of models with the three variables plus an interaction between two of them was compared to a model with only the three variables. Only the interaction between language group and context significantly improved the fit of the model ($X^2(2)= 29.599$, $p<0.001$).

The final regressions included language group, context, Versant score, and the interaction between language group and context as fixed variables. For the language group variable, the Korean group served as a baseline, and was compared to the Mandarin group. Two non-native regressions were run, one with SBF as the context baseline, and the other with VPBF as the context baseline.

The parameter values for the fixed variables in these regressions are listed in Table 4.6. More proficient non-native speakers (as measured by Versant score) produced more appropriate prosody than less proficient speakers. There were also significant interactions between the SuNF/VPBF and SuNF/SBF context pairs and the contrast between the Mandarin and Korean groups. VPBF and SBF productions by Mandarin speakers were judged to have appropriate prosody significantly more often than those by Korean speakers. However, SuNF productions by Korean speakers were judged to have appropriate prosody significantly more often than those by Mandarin speakers. Because of the interaction between language and context, the regressions in Table 4.6 only provide information on the context effects for the Korean group (the language group baseline). To determine the context effects for the Mandarin group, another two regressions were run with Mandarin as the language group baseline, one with SBF as the context baseline, and one with VPBF as the context baseline. The parameter values for these two regressions are listed in Table 4.7. The regressions in Tables 4.6 and 4.7 show that both Korean and Mandarin participants were more successful at producing contextually appropriate prosody in the VPBF and SBF contexts than in the SuNF context.

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	0.460585	0.164794	2.795	<0.01
Mandarin (for SBF condition)	0.428457	0.106591	4.020	<0.001
SuNF (for Korean group)	-0.907160	0.097373	-9.316	<0.001
VPBF (for Korean group)	0.086107	0.102005	0.844	0.39858
Versant	0.007103	0.002262	3.141	<0.005
SuNF:Mandarin	-0.675857	0.141271	-4.784	<0.001
VPBF:Mandarin	-0.049954	0.150940	-0.331	0.74068
VPBF Baseline				
Intercept	0.546675	0.164971	3.314	<0.001
Mandarin (for VPBF condition)	0.378492	0.107795	3.511	<0.001
SBF (for Korean group)	-0.086103	0.102005	-0.844	0.398611
SuNF (for Korean group)	-0.993247	0.098087	-10.126	<0.001
Versant	0.007103	0.002262	3.141	<0.005
SBF:Mandarin	0.049950	0.150940	0.331	0.740699
SuNF:Mandarin	-0.626007	0.142197	-4.402	<0.001
SuNF Baseline				
Intercept	-0.446571	0.162815	-2.743	<0.01
Mandarin (for SuNF condition)	-0.247428	0.093630	-2.643	<0.01
SBF (for Korean group)	0.907145	0.097373	9.316	<0.001
VPBF (for Korean group)	0.993248	0.098087	10.126	<0.001
Versant	0.007103	0.002262	3.141	<0.005
SBF:Mandarin	0.675869	0.141271	4.784	<0.001
VPBF:Mandarin	0.625816	0.142195	4.401	<0.001

Table 4.6. Parameter values for fixed variables in the native perception of non-native production non-native regressions, with Korean as the language group baseline

	Estimate	Std. Error	z-value	p
SBF Baseline				
Intercept	0.889013	0.159966	5.557	<0.001
Korean (for SBF condition)	-0.428441	0.106591	-4.019	<0.001
SuNF (for Mandarin group)	-1.583013	0.102542	-15.438	<0.001
VPBF (for Mandarin group)	0.036153	0.111256	0.325	0.74522
Versant	0.007103	0.002262	3.141	<0.005
SuNF:Korean	0.675868	0.141271	4.784	<0.001
VPBF:Korean	0.049851	0.150939	0.330	0.74119
VPBF Baseline				
Intercept	0.925162	0.160341	5.770	<0.001
Korean (for VPBF condition)	-0.378490	0.107795	-3.511	<0.001
SBF (for Mandarin group)	-0.036151	0.111256	-0.325	0.745229
SuNF (for Mandarin group)	-1.619165	0.103167	-15.695	<0.001
Versant	0.007103	0.002262	3.141	<0.005
SBF:Korean	-0.049950	0.150940	-0.331	0.740698
SuNF:Korean	0.625831	0.142197	4.401	<0.001

Table 4.7. Parameter values for fixed variables in the native perception of non-native production non-native regressions, with Mandarin as the language group baseline

4.3.4 Experiments 1 and 5: Native Perception of Non-Native Production Discussion

Both groups of non-native English speakers were more successful at producing prosody that was perceived to be appropriate by native listeners in the broad focus contexts than in the SuNF context. This pattern did not appear for the native English speakers. In fact, the native English speakers' SuNF productions were judged to be appropriate more often than their VPBF productions. SuNF productions by native English speakers were perceived to have significantly more appropriate prosody than SuNF productions by both Korean and Mandarin speakers. The difficulty that non-native speakers had in producing prosody that was perceived to be appropriate for the SuNF context may be due to their over-application of broad focus prosody to the narrow focus context. This supports the L2 Challenge Models' Frequency factor prediction that a common pitch accent pattern will be over-applied by language learners.

A non-native English speaker's native language also seems to play a role in the appropriateness of the prosody that they produce. For SuNF items, Korean speakers' productions were perceived as more appropriate than Mandarin speakers' productions. However, for the VPBF and SBF items, Mandarin speakers' productions were perceived as more appropriate than Korean speakers' productions. Two of these three results can be explained by the Different Locations Hurt Transfer Model. This model predicts that Korean speakers will be better than Mandarin speakers at appropriately placing final pitch accents in narrow focus. This is because Korean always marks narrow focus prosodically (Jun and Lee 1998), while Mandarin can mark it either prosodically or syntactically (Xu 2004). Therefore, Korean speakers should transfer their habit of always marking narrow focus with prosodic cues like expanding the pitch range and lengthening parts of the focused word. This model also predicts that Mandarin speakers will be better at placing final pitch accents in VPBF contexts because, unlike Korean speakers, they do not have to overcome an L1 tendency to make the first word in a focused VP prosodically prominent. However, this model, like all models set out in Chapter 1, fails to predict that Mandarin speakers will be better than Korean speakers at placing pitch accents in SBF contexts.

One possible explanation for the superior Mandarin performance in both the SBF and the VPBF conditions can be found in the poorer performance of Mandarin speakers in the SuNF condition. The Mandarin speakers' inappropriate SuNF productions may have pitch accent placement that is more appropriate for VPBF and SBF contexts, as predicted by the L2 Challenge Models' Frequency factor. If so, the Mandarin speakers would be producing more sentences with this pattern of pitch accents than Korean speakers. The extra repetitions could reinforce VPBF/SBF accent pattern, leading the Mandarin participants to produce it more consistently or realize it with stronger acoustic cues than the Korean speakers, who were switching to SuNF-appropriate accent patterns more successfully. This hypothesis is supported by the fact that the Mandarin participants even outperformed the native English participants in the SBF and VPBF conditions. The English language group's relative success in the SuNF condition shows that they were switching between SuNF-appropriate and VPBF/SBF-appropriate accent patterns.

The significance of the Versant variable in the non-native regressions indicates that a language learner's proficiency influences their ability to produce context-appropriate prosody. Specifically, native English listeners were more likely to judge utterances produced by more proficient speakers (as determined by the Versant test) to have appropriate prosody than those produced by less proficient speakers. This suggests that context-appropriate pitch accent placement and realization can improve with increased experience with a non-native language.

4.4 Prominence Placement and Realization General Discussion

In this chapter, non-native prosodic prominence production was examined in several ways. The participants' ability to determine the correct placement of a pitch accent for a given context was tested in the prominence placement experiment (Exp. 4). Their ability to place and realize pitch accents in actual speech production was tested in the prominence production experiment (Exp. 1). Recordings of sentences from the prominence production experiment were played for native English listeners, who judged whether their prosody was context-appropriate in the native perception of non-native production experiment (Exp. 5).

The results of the production and listening experiments (Exp. 1 and 5) provide support for one prediction made by both Transfer Models. These models correctly predicted that Korean participants would be better at producing context-appropriate prosody in the SuNF context than Mandarin participants. The models predict that this should be the case because narrow focus is always marked prosodically in Korean, but is only sometimes marked prosodically in Mandarin.

Problematically, the prominence placement experiment (Exp. 4) and the production/listening experiments (Exp. 1 and 5) seem to support conflicting predictions made by the two Transfer Models. The Any Prominence Location Helps Transfer Model correctly predicted that Korean speakers would be better than Mandarin speakers at placing prominence in the VPBF context in Exp. 4, because VPBF is marked prosodically in Korean, but not in Mandarin. However, the Different Prominence Locations Hurt Transfer Model correctly predicted that Mandarin speakers would be better than Korean speakers at producing context-appropriate prosody in the VPBF context in Exp. 1, because Koreans have to overcome their L1 tendency to make the first word in a focused VP prosodically prominent. Fortunately, there is another possible explanation for the Mandarin participants' superior prosody production in VPBF contexts. This alternative explanation is suggested by the fact that Mandarin participants were better than Koreans at producing context-appropriate VPBF and SBF prosody, but were worse than Koreans at producing context-appropriate SuNF prosody. Therefore, the Mandarin speakers may have produced more sentences with broad focus-appropriate patterns of pitch accents than Korean speakers. The extra repetitions could have reinforced the VPBF/SBF accent patterns, leading the Mandarin participants to produce them more consistently or realize them with stronger acoustic cues than the Korean participants, who were switching to SuNF-appropriate accent patterns more successfully. When this explanation is taken into account, the production results support the Any Prominence Location Helps Transfer Model. The hypothesis that inappropriate SuNF productions have prosody that is appropriate for VPBF and SBF contexts will be tested in the acoustic analysis in Chapter 5.

The results of the production/listening and placement experiments (Exp. 1, 4, and 5) also support the Hybrid L2 Challenge Model. Participants were most successful at correctly placing pitch accents in the SuNF context of the prominence placement experiment (Exp. 4). However, they were most successful at correctly producing context-appropriate prosody in the VPBF and SBF contexts of the prominence production experiment (Exp. 1 and 5). It seems that language learners are both especially bad and especially good at pitch accent placement in SuNF contexts. The Hybrid L2 Challenge Model predicts such contradictory results, as it posits two opposing forces that can influence language learner behavior. The first is the relationship between focus and prosodic prominence (the Relationship factor). This relationship is the most direct in narrow focus contexts, making it easier for language learners to acquire appropriate focus marking in such contexts. The second is the frequency with which a pattern of prominence is used (the Frequency factor). In SuNF contexts an unusual pattern of prominence is used, so language learners may incorrectly produce a more common pattern of prominence in this context. In order to be predictive, the Hybrid L2 Challenge Model still needs a criterion for determining when the Relationship factor is more important and when the Frequency factor is more important. This criterion is discussed in Chapter 7.

Proficiency played a role in non-native participants' ability to produce context-appropriate prosody (Exp. 1 and 5). Non-native speakers with higher general English proficiency were more likely to have their utterances judged to have appropriate prosody. In contrast, proficiency did not play a role in a participant's ability to say which word should be accented for a given context in the prominence placement experiment (Exp. 4). Even native

English speakers often provided unexpected answers in this task. Clearly, speakers can produce context-appropriate English prosody without consciously knowing which words they are accenting. What is not yet known is whether such knowledge can help non-native speakers produce more appropriate prosody.

Chapter 5

Acoustic Analysis of Production Recordings - Experiment 1

5.1 Purpose of Acoustic Analysis

The acoustic analysis of the recordings made in the prominence production experiment (Exp. 1) had two goals. First, it was used to test the hypothesis that the non-native SuNF productions that were judged to have inappropriate prosody had pitch accent placement patterns that were appropriate for VPBF or SBF contexts rather than SuNF contexts. To test this hypothesis, the non-native productions of sentences in the SuNF context that had been judged to have inappropriate prosody were compared to non-native productions in all three contexts that were judged to have appropriate prosody. Second, the acoustic analysis was used to determine whether there is more than one way to produce prosody that will be judged appropriate by native listeners. To answer this question, acoustic features of native productions that had been judged to have appropriate prosody were compared to non-native productions with appropriate prosody, within each context. The acoustic analysis involved both labeling selected recordings for accent location, and making acoustic measurements of them.

The remainder of Chapter 5 explores the methods used in the acoustic analysis and its results. Section 5.2 describes how recordings were selected for acoustic analysis. Section 5.3 describes the procedure used for labeling and measurement. Section 5.4 describes the results of the analysis of non-native SuNF productions that were judged to be inappropriate by native listeners. Section 5.5 presents the results of the comparison between native productions and non-native productions judged to have appropriate prosody. Section 5.6 summarizes the findings in this chapter and discusses their implications.

5.2 Selection of Recordings for Acoustic Analysis

Recordings were only acoustically analyzed if at least 7 out of the 8 native listeners judged them to be either appropriate or inappropriate. This criterion was used to ensure that the recordings analyzed had prosody that was unambiguously appropriate or inappropriate. Table 5.1 shows the breakdown of the 356 recordings that were judged to have unambiguously appropriate prosody. Table 5.2 shows the breakdown of the 42 recordings that were judged to have unambiguously inappropriate prosody.

	SuNF	VPBF	SBF	Total
English	56	42	42	140
Korean	18	34	38	90
Mandarin	12	58	56	126
Total	86	134	136	356

Table 5.1. Number of productions classified as prosodically appropriate, by language group and context

	SuNF	VPBF	SBF	Total
English	0	1	1	2
Korean	13	0	0	13
Mandarin	27	0	0	27
Total	40	1	1	42

Table 5.2. Number of productions classified as prosodically inappropriate, by language group and context

The recordings judged to have unambiguously inappropriate prosody included 40 sentences produced by non-native English speakers in the SuNF context, and two sentences produced by native English speakers, one in the SBF and on in the VPBF context. Because of the very small number of recordings by native speakers that were judged to be inappropriate, the analyses of inappropriate prosody focused exclusively on the non-native SuNF recordings.

5.3 Acoustic Labeling and Measurements

The acoustic analysis had two components, phonological and phonetic. In the phonological component, the location of the accented word(s) was determined by the experimenter (a native English speaker and trained ToBI labeler), who was blind to the recordings' contexts and native listener judgments. Words were labeled as accented or unaccented based on the ToBI labeling guidelines (Beckman and Ayers Elam 1997), using auditory perception and examination of the pitch contour, displayed in Praat (Boersma and Weenink 2009). In the phonetic component, several acoustic features were extracted for the subject, verb, and object of each sentence. Acoustic features of each sentence as a whole were also extracted to allow for normalization of the individual words' acoustic values. The acoustic features extracted were the duration, RMS amplitude, and F0 maximum and range for each word. These features were chosen because of the important roles they play in prosodic focus marking in English, Korean, and Mandarin. (Narrow) focused words tend to have longer durations in all three languages (Beckman 1986; Jun and Lee 1998; Liu and Xu 2005). They also tend to have greater amplitudes in English and Korean (Beckman 1986; Lee and Xu 2010). They have higher F0 peaks and expanded F0 ranges in Korean and Mandarin (Jun and Lee 1998; Liu and Xu 2005). The high and low pitch targets in English pitch accents can also lead to more extreme F0 values on accented words than unaccented words (Beckman 1986; Shue et al. 2007).

The acoustic feature extraction involved several steps. First, the selected recordings were automatically aligned with their transcriptions using the NU-Aligner program. NU-Aligner (http://groups.linguistics.northwestern.edu/documentation/nualigner_home.html) is a program developed by Chun Chan as an add-on to the SONIC speech recognition system (http://cslr.colorado.edu/beginweb/speech_recognition/sonic.html). It takes as input a sound file and a transcription, and outputs a Praat textgrid in which the transcription is aligned with the sound file. Next, these textgrids were hand-corrected by the experimenter, based on features of the waveform and spectrogram. Finally, the acoustic values for the subject, verb, object, and entire sentence in each recording were automatically extracted using a Praat script. The Praat script was run separately on recordings of male and female speakers because of the differences in their typical F0 ranges. For male participants the pitch range was set to 75-350 Hz, and for female participants the pitch range was set to 100-400 Hz. These ranges were chosen by testing the ranges recommended in the Praat manual on selected recordings, and adjusting them to produce the most accurate pitch contours. The default Praat values were used for all other variables.

5.4 Non-Native SuNF Sentences with Inappropriate Prosody

5.4.1 Phonological Features

The goal of this analysis was to determine whether the inappropriate non-native SuNF productions were prosodically similar to appropriate non-native broad focus productions. Forty SuNF productions by non-native English speakers with unambiguously inappropriate prosody were compared to the 30 non-native SuNF productions, 92 non-native VPBF productions, and 94 non-native SBF productions with unambiguously appropriate prosody. As discussed in Section 1.3.2.2, the predicted accent pattern for the SuNF context is an accented subject and unaccented verb and object. Sentences in the VPBF and SBF contexts can have several pitch accent patterns, as long as they have a pitch accent on the object[11], although in Gussenhoven's model SBF productions are also required to have pitch accents on the subject.

Non-native SuNF productions that were judged to have inappropriate prosody had patterns of accent placement that are predicted for sentences in VPBF contexts. Figure 5.1 shows how the inappropriate non-native SuNF accent patterns differ from the appropriate non-native SuNF patterns, and overlap with the appropriate non-native VPBF and SBF patterns. For ease of reference, pitch accent patterns will be labeled based on whether each content word was accented. The labels all have three slots, for subject, verb, and object, and these slots contain a 'Y' (yes) if the word is accented, and an 'n' (no) if the word is unaccented. Under this system, a sentence with an accented subject and unaccented verb and object will be labeled 'Ynn'. Out of the non-native SuNF productions that were judged to have appropriate prosody, 100% had the predicted Ynn pattern of pitch accents. However, none of the non-native SuNF productions that were judged inappropriate had the Ynn pattern. Instead, the inappropriate productions had YnY (53%), YYY (43%), and nYY (5%) accent patterns. All of these accent patterns are predicted to be appropriate for sentences produced in VPBF contexts, and the YnY and YYY patterns are predicted to be appropriate for sentences in SBF contexts. Section 5.4.2 will explore the acoustic associated with these accent patterns.

[11] This prediction applies to the sentences with transitive verbs, of the type used in this study.

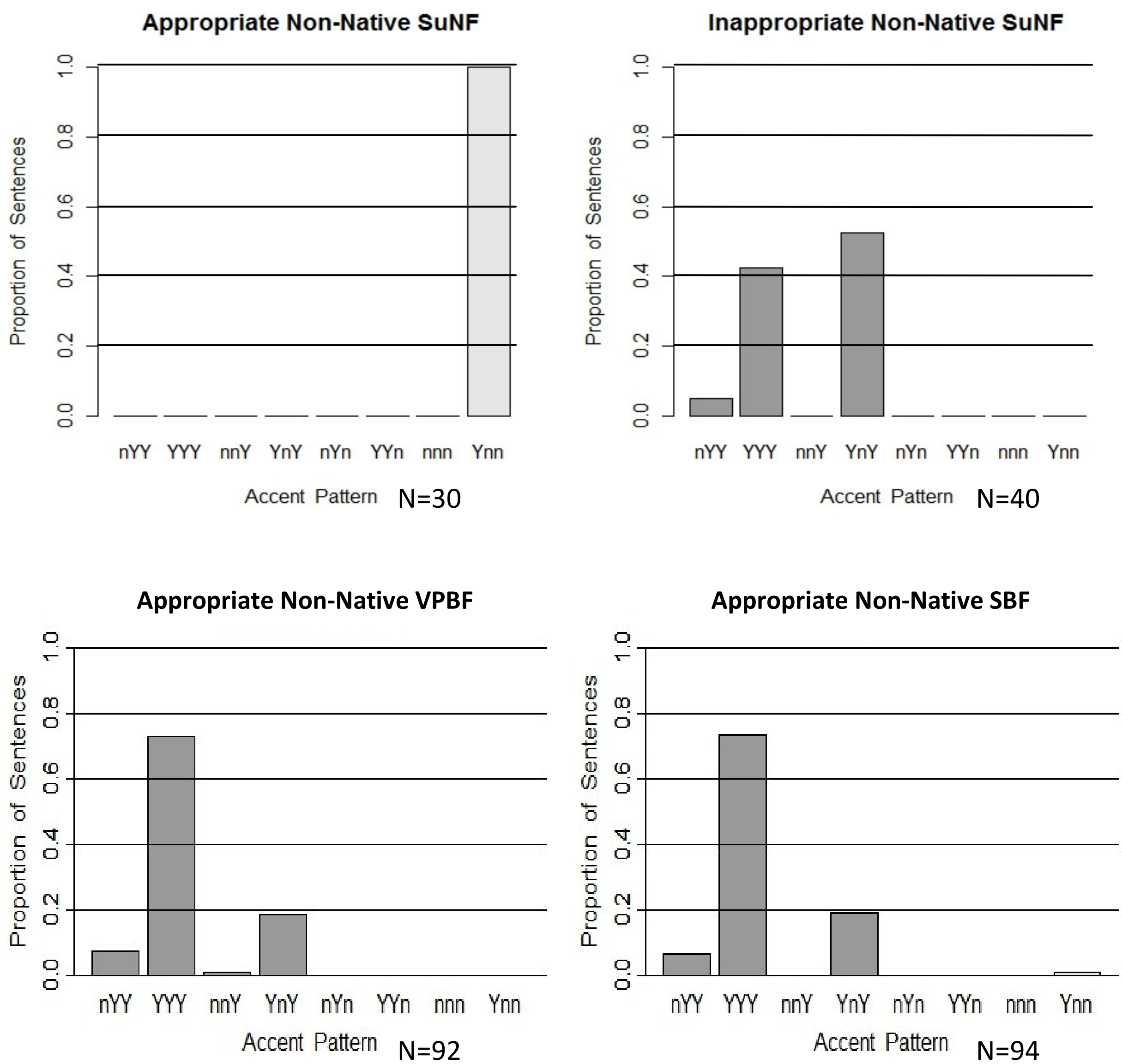

Figure 5.1. Barplots showing proportions of non-native SuNF sentences perceived to have context-inappropriate prosody and non-native SuNF, VPBF, and SBF sentences perceived to have context-appropriate prosody. Bars in dark gray represent accent patterns that are predicted to be acceptable for VPBF contexts, bars in light gray represent accent patterns that are predicted to be acceptable for SuNF contexts.

5.4.2 Phonetic Features

5.4.2.1 Statistical Analysis

The goal of the phonetic comparisons was to determine the acoustic differences (if any) between the inappropriate non-native SuNF productions and appropriate non-native SuNF, VPBF, and SBF productions. The productions were analyzed with a series of mixed-effect linear regressions. These regressions all had speaker and sentence as random effects. The dependent variable was always an acoustic feature (e.g. duration) for a particular word position (subject, verb, or object). The fixed variable of interest was an appropriateness/context variable, with inappropriate SuNF, appropriate SuNF, appropriate VPBF, and appropriate SBF as possible values. Inappropriate SuNF productions served as the baseline, so that they could be compared to the three types of appropriate productions.

One challenge that arose from only using recordings that were unambiguously judged to have appropriate or inappropriate prosody was that the same speakers and the same sentences were not necessarily included in the sets of appropriate and inappropriate productions. Theoretically, this could lead to inaccurate results. For example, if all the appropriate SuNF recordings happened to be produced by speakers with unusually large F0 ranges, it would give the impression that words in appropriate SuNF productions generally have large F0 ranges. Fortunately, the regressions allow us to statistically control for features particular to each speaker or the words in each sentence. The speakers' values for various acoustic features (e.g. F0 range) over the course of a particular production were controlled by including as a fixed variable the sentence-level values (either over the entire sentence or averaging the values for each word in the sentence) for the acoustic feature under investigation. For example, in a regression with object F0 range as the dependent variable, the mean of the F0 ranges for the subject, verb, and object in that sentence production was included as a fixed variable. Features of individual words were controlled by including as a fixed variable in the regression the mean value of the feature under investigation for each word in the position being investigated. For example, in a regression with object duration as the dependent variable, the mean durations for each of the six objects were included as a fixed variable. The value of this variable for items with *Maine* as the object, for instance, would be the mean duration of all productions of *Maine* that were being analyzed in the regression. Because there was only one fixed variable of interest, all of the regressions in this section included the appropriateness/context variable, and the sentence and mean word control variables as fixed variables. The dependent and control variables for each acoustic feature regression are discussed in more detail in the results section for that feature.

5.4.2.2 Analysis of Duration

In the duration regressions, the dependent variable was word duration in seconds. The sentence control variable was speech rate (syllables/second) for the sentence. The word control variable was the mean duration of the word in the target position (subject, verb, or object) over all productions of that sentence in the analysis. For example, in the subject duration regression, the word control variable had six possible values: the mean durations for each of the six sentences' subjects. The parameter values for the fixed variables in the subject, verb, and object duration regressions are provided in Tables 5.3, 5.4, and 5.5, respectively. Speech rate was not a significant variable in the verb duration regression, so it was removed from the final regression.

	Estimate	95% CI – lower	95% CI - upper	p
Intercept	0.0739	0.0132	0.1757	<0.05
Appropriate SuNF	0.0105	-0.0285	0.0375	0.8066
Appropriate VPBF	-0.0281	-0.0536	-0.0048	<0.05
Appropriate SBF	-0.0085	-0.0351	0.0137	0.4078
Speech Rate	-0.0119	-0.0260	-0.0054	<0.005
Mean Subject Dur.	0.9126	0.6879	1.0965	<0.001

Table 5.3. Parameter values for fixed variables in the subject duration regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

	Estimate	95% CI – lower	95% CI - upper	p
Intercept	-0.0389	-0.1091	0.0345	0.2616
Appropriate SuNF	0.0040	-0.0498	0.0345	0.7436
Appropriate VPBF	0.0591	0.0251	0.0858	<0.001
Appropriate SBF	0.0463	0.0163	0.0770	<0.005
Mean Verb Dur.	1.0069	0.8514	1.1745	<0.001

Table 5.4. Parameter values for fixed variables in the verb duration regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

	Estimate	95% CI – lower	95% CI - upper	P
Intercept	0.1558	0.0476	0.3195	<0.01
Appropriate SuNF	-0.0705	-0.1239	-0.0420	<0.001
Appropriate VPBF	0.0302	-0.0070	0.0529	0.1208
Appropriate SBF	0.0037	-0.0333	0.0261	0.8380
Speech Rate	-0.0234	-0.0415	-0.0165	<0.001
Mean Object Dur.	0.8164	0.5597	1.0247	<0.005

Table 5.5. Parameter values for fixed variables in the object duration regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

The analyses of durational differences between appropriate and inappropriate non-native productions show that inappropriate SuNF productions had significantly longer object durations than appropriate SuNF productions. The longer object durations for inappropriate SuNF productions signal pitch accents on the object. The phonological prosodic analysis showed that none of the appropriate SuNF productions had accented objects, but all of the inappropriate ones did. Because sentences in SuNF contexts are predicted to have final pitch accents on the subject, a pitch accent on the object would lead to inappropriate prosody for the context. This difference is illustrated in Figure 5.2.

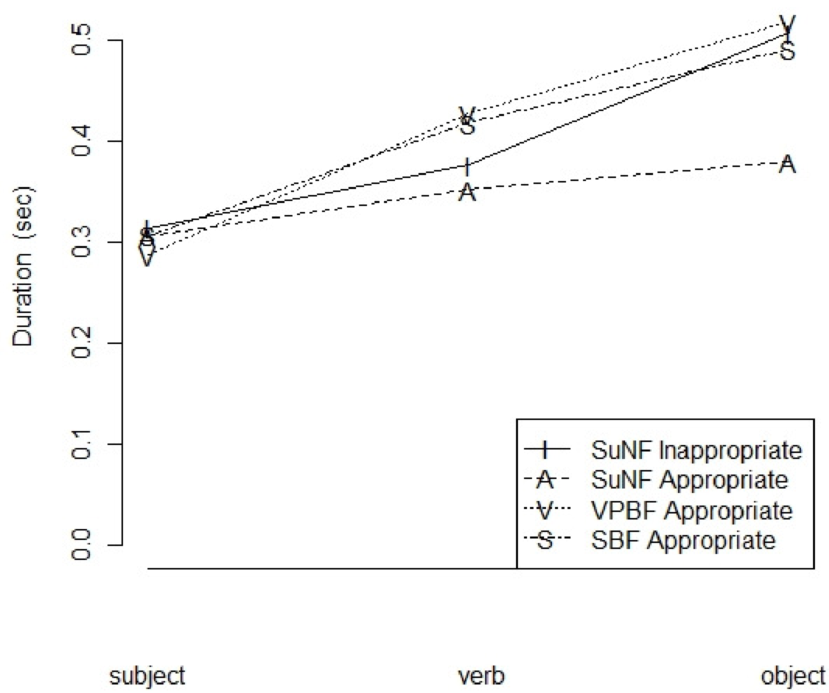

Figure 5.2. Line graph showing the mean durations in seconds for the subject, verb, and object in inappropriate SuNF productions and appropriate SuNF, VPBF, and SBF productions by non-native speakers

However, the inappropriate SuNF productions also differed durationally from the appropriate VPBF and SBF productions. The inappropriate SuNF productions had significantly longer subject durations than appropriate VPBF productions, shorter verb durations than appropriate VPBF and SBF productions. This partially reflects differences in accent patterns revealed in the phonological analysis. Non-native speakers were more likely to accent the verb in the appropriate VPBF and SBF productions than in the inappropriate SuNF productions, leading to longer verb durations in these appropriate broad focus productions. Accenting the subject was quite common in both the inappropriate SuNF productions (95%) and the appropriate VPBF productions (91%). The longer subject durations in the inappropriate SuNF productions than in the appropriate VPBF productions might reflect this small difference in the percentage of accented subjects.

5.4.2.3 Analysis of RMS Amplitude

In the RMS amplitude regressions, the dependent variable was RMS amplitude of the word. Praat returns RMS amplitude in Pascals, which were converted to decibels, with the RMS amplitude for the entire sentence (in Pascals) as the reference value. This method of conversion controlled for differences in overall amplitudes across recordings, so no separate sentence control variable was needed in the RMS amplitude regressions. The word control variable was the mean RMS amplitude of the word in the target position, after conversion to dB (averaging over all productions of that sentence). The parameter values for the fixed variables in the subject, verb, and object RMS amplitude regressions are provided in Tables 5.6, 5.7, and 5.8, respectively.

	Estimate	95% CI - lower	95% CI - upper	p
Intercept	-0.3597	-1.1513	0.3633	0.3312
Appropriate SuNF	1.6742	0.9656	2.3635	<0.001
Appropriate VPBF	0.1426	-0.4080	0.6494	0.6616
Appropriate SBF	0.1441	-0.4101	0.6506	0.6210
Mean Subject RMS	1.0350	0.7883	1.2966	<0.001

Table 5.6. Parameter values for fixed variables in the subject RMS amplitude regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

	Estimate	95% CI - lower	95% CI - upper	p
Intercept	-0.2541	-0.8193	0.2524	0.2864
Appropriate SuNF	-1.1706	-1.8546	-0.4237	<0.005
Appropriate VPBF	0.6341	0.1137	1.2080	<0.05
Appropriate SBF	0.5229	0.0229	1.1018	<0.05
Mean Verb RMS	0.9847	0.5973	1.3305	<0.005

Table 5.7. Parameter values for fixed variables in the verb RMS amplitude regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

	Estimate	95% CI - lower	95% CI - upper	p
Intercept	0.5861	-0.2470	1.3359	0.1290
Appropriate SuNF	-3.1342	-4.0793	-2.3517	<0.001
Appropriate VPBF	-0.3493	-0.9420	0.3377	0.3502
Appropriate SBF	0.0295	-0.5783	0.6876	0.8930
Mean Object RMS	1.0834	0.7985	1.3547	<0.001

Table 5.8. Parameter values for fixed variables in the object RMS amplitude regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

The analyses of RMS amplitude revealed that non-native speakers' inappropriate SuNF productions had significantly lower subject RMS amplitudes and higher verb and object RMS amplitudes than appropriate SuNF productions. A higher RMS amplitude on the subject can signal a pitch accent on the subject, which is predicted to be required in the SuNF context. However, the phonological analysis showed that 96% of inappropriate SuNF productions had accented subjects (compared with 100% of appropriate productions). The higher RMS amplitude for subjects in appropriate SuNF productions could reflect this small difference in accentuation. Alternatively, the non-native speakers may be producing final pitch accents with higher RMS amplitudes than pre-final pitch accents. A higher RMS amplitude on verbs and objects can signal pitch accents on these words, which is predicted to be inappropriate for sentences in the SuNF context. The phonological analysis showed that 48% of inappropriate SuNF productions had accented verbs all had accented objects, but none of the appropriate SuNF productions had accented verbs or objects. The RMS effects can be seen in Figure 5.3.

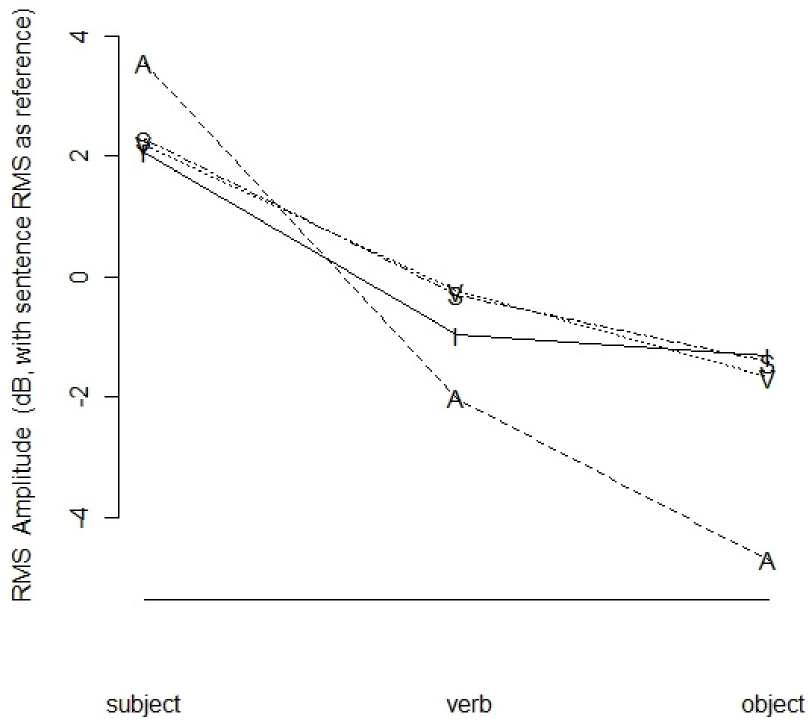

Figure 5.3. Line graph showing the mean RMS amplitudes in dB for the subject, verb, and object in inappropriate SuNF productions and appropriate SuNF, VPBF, and SBF productions by non-native speakers

Non-native speakers' inappropriate SuNF productions also had lower verb RMS amplitudes than appropriate VPBF and SBF productions. Like the longer durations for verbs in the appropriate VPBF and SBF productions, this reflects the greater number of pitch accents on the verb in the appropriate VPBF and SBF productions.

5.4.2.4 Analysis of F0 Maximum

In the F0 maximum regressions, the dependent variable was the maximum F0 in Hz for the word. The sentence control variable was the mean of the F0 maxima for the subject, verb, and object in the sentence. The word control variable was the mean F0 maximum for the word in the target position (averaging over all productions of that sentence). The parameter values for the fixed variables in the subject, verb, and object F0 maximum regressions are provided in Tables 5.9, 5.10, and 5.11, respectively. The mean target word maximum F0 variables (e.g. mean subject F0 maximum) were not significant in any of the three F0 maximum regressions, so they were excluded from the final subject, verb, and object regressions.

	Estimate	95% CI - lower	95% CI – upper	p
Intercept	-0.8979	-19.778	12.367	0.7146
Appropriate SuNF	27.0624	16.226	38.283	<0.001
Appropriate VPBF	-9.3053	-16.910	-0.706	<0.05
Appropriate SBF	-6.6820	-15.186	1.196	0.0946
Mean Sentence F0 Max	1.0867	1.032	1.163	<0.001

Table 5.9. Parameter values for fixed variables in the subject F0 maximum regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

	Estimate	95% CI - lower	95% CI – upper	p
Intercept	7.3848	-9.2799	25.231	0.3586
Appropriate SuNF	-11.6443	-22.9930	0.356	0.0522
Appropriate VPBF	4.0071	-4.3295	13.564	0.3774
Appropriate SBF	-2.5204	-11.9830	6.392	0.6064
Mean Sentence F0 Max	0.9957	0.9321	1.051	<0.001

Table 5.10. Parameter values for fixed variables in the verb F0 maximum regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

	Estimate	95% CI - lower	95% CI – upper	p
Intercept	-0.2858	-20.9588	18.4894	0.8914
Appropriate SuNF	-16.7465	-29.7365	-2.6877	<0.05
Appropriate VPBF	5.3019	-5.1422	15.1019	0.3278
Appropriate SBF	9.5387	-0.7340	19.8365	0.0660
Mean Sentence F0 Max	0.8874	0.8133	0.9661	<0.001

Table 5.11. Parameter values for fixed variables in the object F0 maximum regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

The analyses of F0 maxima revealed that non-native SuNF productions with inappropriate prosody had significantly lower F0 maxima on subjects and higher F0 maxima on objects than SuNF productions with appropriate prosody. As discussed for the RMS amplitude regression, the phonological analysis showed that 96% of inappropriate SuNF productions had accented subjects (compared with 100% of appropriate productions). This suggests three possibilities. First, the small difference in accenting may be adequate to explain the phonetic differences. Second, the pre-final subject pitch accents in the inappropriate SuNF productions may be reduced acoustically relative to the final subject pitch accents in the appropriate SuNF productions. Third, the speakers could be using more L* pitch accents, which do not contain an H tone. All of the inappropriate SuNF productions had accented objects, but none of the appropriate SuNF productions had accented objects. This reflects the fact that SuNF sentences are predicted not to have pitch accents on verbs and objects. The higher F0 maxima on inappropriate SuNF objects are likely to be cues to pitch accents containing an H tone (H* or L+H*), inappropriately placed on these words.

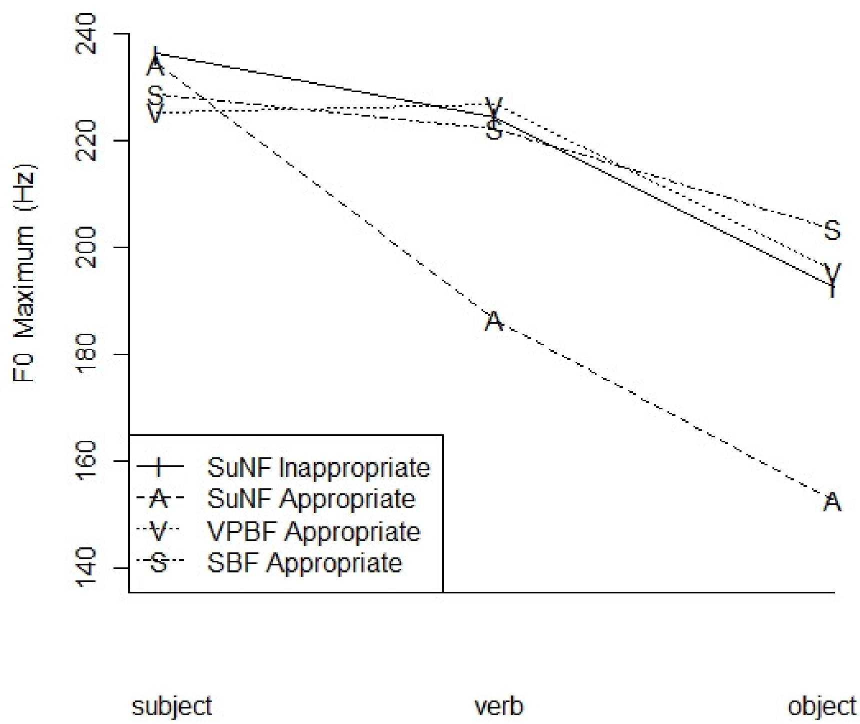

Figure 5.4. Line graph showing the mean F0 maxima in Hz for the subject, verb, and object in inappropriate SuNF productions and appropriate SuNF, VPBF, and SBF productions by non-native speakers

Non-native inappropriate SuNF productions also had significantly higher F0 maxima on subjects than appropriate VPBF productions. This result is parallel to the longer subject durations in inappropriate SuNF productions than appropriate VPBF productions. The subject was accented in 95% of inappropriate SuNF productions and 91% of appropriate VPBF productions. The longer subject durations in the inappropriate SuNF productions than in the appropriate VPBF productions might reflect this small difference in accenting rate.

5.4.2.5 Analysis of F0 Range

In the F0 range regressions, the dependent variable was the range in F0 values in Hz for each target word. The sentence control variable was the mean of the F0 range values for the subject, verb, and object in each sentence. The word control variable was the mean F0 range for the word in the target position (averaging over all productions of that sentence). The parameter values for the fixed variables in the subject, verb, and object F0 range regressions are provided in Tables 5.12, 5.13, and 5.14, respectively. The mean object F0 range control variable was not significant in the object F0 range regression, so it was not included in the final version of this regression.

	Estimate	95% CI - lower	95% CI - upper	p
Intercept	-33.8711	-59.2024	-7.154	<0.05
Appropriate SuNF	21.7669	8.6847	33.894	<0.001
Appropriate VPBF	-8.0689	-17.1007	2.058	0.1230
Appropriate SBF	-6.5834	-16.0340	2.915	0.1728
Mean Sentence F0 Range	0.9233	0.7953	1.059	<0.001
Mean Subject F0 Range	0.6946	0.2040	1.154	<0.05

Table 5.12. Parameter values for fixed variables in the subject F0 range regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

	Estimate	95% CI - lower	95% CI - upper	p
Intercept	-33.7486	-64.4882	-1.449	<0.05
Appropriate SuNF	-2.3515	15.9627	10.160	0.6938
Appropriate VPBF	-0.7364	-10.9964	9.079	0.8982
Appropriate SBF	-7.4660	-17.6464	2.560	0.1454
Mean Sentence F0 Range	1.0579	0.9325	1.208	<0.001
Mean Verb F0 Range	0.6110	0.1012	1.124	<0.05

Table 5.13. Parameter values for fixed variables in the verb F0 range regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

	Estimate	95% CI - lower	95% CI - upper	p
Intercept	1.9784	-16.1338	16.920	0.9388
Appropriate SuNF	-18.7625	-34.0800	-4.485	<0.05
Appropriate VPBF	7.8254	-3.7750	19.615	0.2016
Appropriate SBF	13.0438	2.1741	25.298	<0.05
Mean Sentence F0 Range	0.9345	0.7972	1.124	<0.001

Table 5.14. Parameter values for fixed variables in the object F0 range regression for appropriate and inappropriate productions by non-native speakers, with inappropriate SuNF as the appropriateness/context baseline

The analyses of F0 range revealed that non-native SuNF sentences with inappropriate prosody had significantly smaller F0 ranges on subjects and larger F0 ranges on objects than those with appropriate prosody. The smaller F0 ranges on subjects in inappropriate SuNF productions are closely linked to the lower F0 maxima on subjects in inappropriate SuNF productions. If the F0 maximum is lowered, this reduces the F0 range (unless the F0 minimum is similarly lowered). In the F0 maximum discussion, it was suggested that the lower F0 maxima on inappropriate SuNF subjects may be due to these non-final pitch accents being reduced acoustically relative to the final subject pitch accents in the appropriate SuNF productions, or due to the speakers using more L* pitch accents, which do not contain an H tone. The fact that the F0 range was also reduced on these subjects supports that former hypothesis, because a low pitch target associated with an L* pitch accent could drag down the F0 minimum, thereby increasing the F0 range. The larger F0 ranges on objects in SuNF productions with inappropriate prosody are likely to represent the pitch movements associated with pitch accents. Recall that all of the inappropriate productions had accented objects, but none of the appropriate productions did. Sentences in SuNF contexts are not predicted to have pitch accents on objects. These differences are illustrated in Figure 5.5.

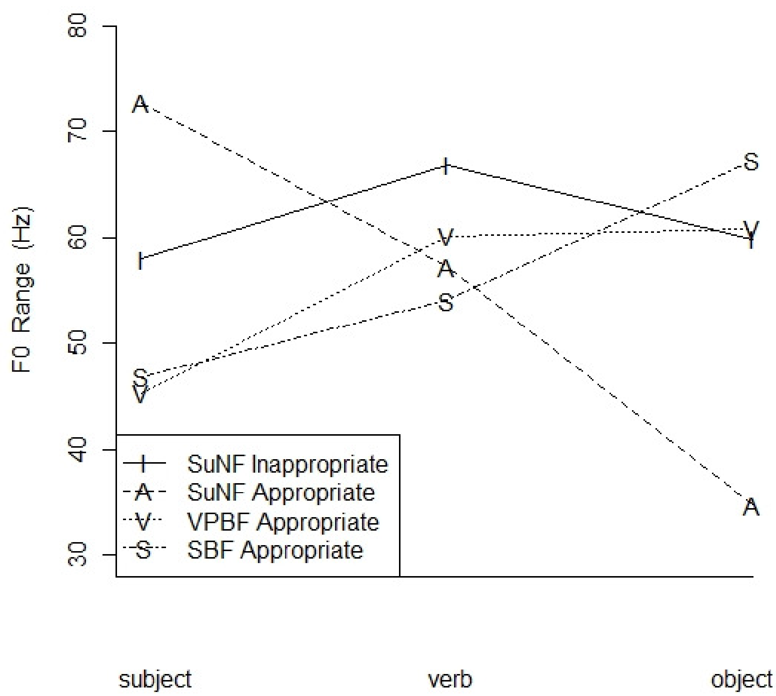

Figure 5.5. Line graph showing the mean F0 ranges in Hz for the subject, verb, and object in inappropriate SuNF productions and appropriate SuNF, VPBF, and SBF productions by non-native speakers

The inappropriate SuNF productions also had significantly smaller object F0 ranges than appropriate SBF productions. The object was accented in 100% of inappropriate SuNF productions, and in 99% of appropriate SBF productions. Therefore, it seems that the unpredicted final object pitch accent in the inappropriate SuNF productions is acoustically reduced (with a reduced F0 range) relative to the predicted final object pitch accent in the appropriate SBF productions.

5.5 Acoustic Differences between Native and Non-Native Productions with Appropriate Prosody

5.5.1 Phonological Features

Prosodically appropriate sentence productions by native and non-native speakers consistently had the pitch accent patterns predicted for their context. Recall that the predicted accent pattern for the SuNF context is an accented subject and unaccented verb and object. SuNF recordings judged to have appropriate prosody overwhelming had the Ynn accent pattern. An unpredicted pitch accent pattern (YYn) was used in only one out of 86 SuNF recordings with appropriate prosody.

As discussed in Section 1.3.2.2, several patterns of pitch accent placements are acceptable in VPBF and SBF contexts, as long as the object is accented. Each of the four acceptable patterns for VPBF productions (nYY, YYY, nnY, YnY) was produced by some speaker in the VPBF context. Gussenhoven predicts that only YYY and YnY should be acceptable in the SBF context. However, all four patterns with an accented subject were produced in acceptable SBF productions. This is true even if non-native speakers are excluded. Still, accent patterns with unaccented subjects were much less common than patterns with accented subjects in both VPBF and SBF productions. In both of the broad focus contexts, the most popular pattern was YYY, followed by YnY. An unpredicted pitch accent pattern was used in only one out of 134 VPBF recordings (YYn), and only one out of 136 SBF recordings (Ynn). The proportions of recordings with each accent pattern produced in each context by native and non-native English speakers can be seen in Figure 5.6.

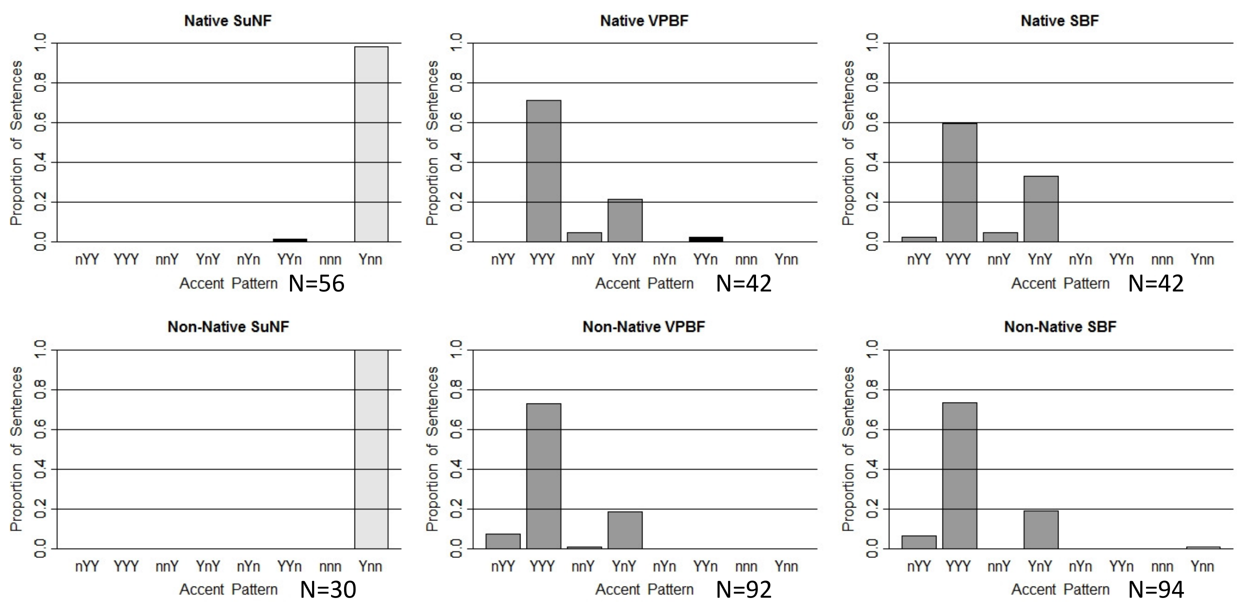

Figure 5.6. Barplots showing proportions of prosodically appropriate sentences in each context produced with each accent pattern by native and non-native English speakers. Bars in dark gray represent accent patterns that are predicted to be acceptable for VPBF contexts, bars in light gray represent accent patterns that are predicted to be acceptable for SuNF contexts, bars in black represent patterns that are not predicted to be acceptable in any of these three contexts.

The distribution of pitch accent patterns across the three contexts was remarkably similar for native and non-native English speakers. In the SuNF context, the Ynn pattern was used in 99% of native recordings, and 100% of non-native recordings. Despite the greater number of pitch accent pattern options available in the VPBF and SBF contexts, the patterns were again consistent across the native and non-native groups. In the VPBF context, the most common pattern (YYY) was used in 71% of native recordings and 73% of non-native recordings; the second most common pattern (YnY) was used in 21% of native recordings and 18% of non-native recordings. In the SBF context, the most common pattern (YYY) was used in 60% of native recordings and 73% of non-native recordings; the second most common pattern (YnY) was used in 33% of native recordings and 19% of non-native recordings. The main difference between the native and non-native groups is the non-native speakers' slightly greater preference for accenting all of the content words in a VPBF or SBF sentence, relative to native speakers.

5.5.2 Phonetic Features

5.5.2.1 Statistical Analysis

The phonetic features of the recordings judged to have appropriate prosody were analyzed with a series of mixed-effect linear regressions, one for each acoustic feature on each word. These regressions all had speaker and sentence as random effects. The dependent variable was always an acoustic feature (e.g. duration) for a particular word (subject, verb, or object).

The fixed variables of interest were whether the speaker was a native English speaker (native, non-native) and the sentence context (SuNF, VPBF, SBF). In order to examine all three pairs of contexts, two regressions were run, one with SuNF as the context baseline, and one with VPBF as the context baseline. The native English speaking group served as the nativeness baseline. When the interaction between nativeness and context is included, it allows us to examine whether there is any difference between native and non-native speakers in the degree to which a particular acoustic feature is used to signal the difference between contexts in sentence productions with acceptable prosody.

Just like in the comparison of appropriate and inappropriate non-native productions, these regressions controlled for features of individual speakers and words by including control variables. The speakers' values for the acoustic feature under investigation (e.g. F0 range) over the course of a particular production were controlled by including sentence-level values for that acoustic feature as fixed variable. Features of individual words were controlled by including as a fixed variable the mean value of the feature under investigation for the word in the target position (averaging over all productions of that sentence). Details of the control and dependent variables for each regression are included in the analysis results sections.

The fixed variables for the regressions were selected using essentially the same process used in the other experiments. First, a no-interaction regression was run on the two fixed variables of interest and the two control variables. All variables that were significant in this regression were retained. If both variables of interest were retained, a second regression was built that included the original significant variables plus the interaction between the two variables of interest. This regression was compared to an identical regression without the interaction, using a likelihood ratio test. If the interaction significantly improved the fit of the model, then it was retrained in the final regression.

5.5.2.2 Analysis of Duration

In the duration regressions, the dependent variable was word duration in seconds. The sentence control variable was speech rate (syllables/second) for the sentence. The word control variable was the mean duration for the word in the target position (averaging over all productions of that sentence).

In the first subject duration regression model, the nativeness variable was not significant, so it was not retained. No interactions were tested because only one variable of interest remained. The final model included context, speech rate, and mean subject duration as fixed variables. The parameter values for the SuNF and VPBF baseline subject duration regressions are listed in Table 5.15.

	Estimate	95% CI – lower	95% CI - upper	p
SuNF Baseline				
Intercept	0.2520	0.1260	0.4247	<0.005
SBF	-0.0170	-0.0342	-0.0015	<0.05
VPBF	-0.0276	-0.0447	-0.0119	<0.005
Speech Rate	-0.0488	-0.0642	-0.0394	<0.001
Mean Subject Dur.	0.7236	0.2555	1.1707	<0.05
VPBF Baseline				
Intercept	0.2244	0.0929	0.3853	<0.01
SBF	0.0106	-0.0039	0.0230	<0.05
SuNF	0.0276	0.0118	0.0441	<0.001
Speech Rate	-0.0488	-0.0646	-0.0401	<0.001
Mean Subject Dur.	0.7236	0.2711	1.1688	<0.01

Table 5.15. Parameter values for fixed variables in the subject duration regressions on prosodically appropriate productions

In the first verb duration regression model, all four variables were significant, so they were all retained. Because both variables of interest were retained, a model containing all four original variables was compared to a model with these variables plus the interaction between nativeness and context with a likelihood ratio test. However, adding the interaction did not significantly improve the fit of the model ($X^2(2)=2.282$, p=0.32). The final model included context, nativeness, speech rate, and mean verb duration as fixed variables. The parameter values for these regressions are listed in Table 5.16.

	Estimate	95% CI – lower	95% CI - upper	p
SuNF Baseline				
Intercept	0.1260	-0.0788	0.3366	0.1524
Non-Native	0.0288	0.0030	0.0484	<0.05
SBF	0.0145	-0.0017	0.0371	0.0722
VPBF	0.0276	0.0105	0.0485	<0.005
Speech Rate	-0.0516	-0.0713	-0.0417	<0.001
Mean Verb Dur.	1.0113	0.4985	1.4905	<0.005
VPBF Baseline				
Intercept	0.1535	-0.0484	0.3777	0.1000
Non-Native	0.0288	0.0037	0.0490	<0.05
SBF	-0.0131	-0.0269	0.0041	0.1510
SuNF	-0.0276	-0.0481	-0.0102	<0.005
Speech Rate	-0.0516	-0.0718	-0.0420	<0.001
Mean Verb Dur.	1.0113	0.5140	1.5537	<0.005

Table 5.16. Parameter values for fixed variables in the verb duration regressions on prosodically appropriate productions, with native as the nativeness baseline

In the first object duration regression model, the mean object duration variable was not significant, so it was not retained. Including the interaction between the two variables of interest significantly improved the fit of the model ($X^2(2)=9.6731$, $p<0.01$). The final model included context, nativeness, and speech rate, plus the interaction between context and nativeness, as fixed variables. In order to determine the nature of the significant interaction between the SBF/VPBF and SuNF/VPBF context contrasts and the native/non-native contrast, a third regression was run with SBF as the baseline. The parameter values for these three regressions are listed in Table 5.17.

	Estimate	95% CI – lower	95% CI – upper	p
SuNF Baseline				
Intercept	0.6078	0.5651	0.6962	<0.001
Non-Native (for SuNF condition)	0.0250	-0.0162	0.0526	0.3172
SBF (for native group)	0.0445	0.0170	0.0709	<0.001
VPBF (for native group)	0.0351	0.0074	0.0601	<0.05
Speech Rate	-0.0686	-0.0892	-0.0606	<0.001
Non-Native:SBF	0.0149	-0.0223	0.0562	0.3660
Non-Native:VPBF	0.0535	0.0195	0.0967	<0.005
VPBF Baseline				
Intercept	0.6429	0.6027	0.7282	<0.001
Non-Native (for VPBF condition)	0.0786	0.0458	0.1045	<0.001
SBF (for native group)	0.0094	-0.0173	0.0391	0.4570
SuNF (for native group)	-0.0351	-0.0608	-0.0085	<0.05
Speech Rate	-0.0686	-0.0887	-0.0609	<0.001
Non-Native:SBF	-0.0387	-0.0735	-0.0058	<0.05
Non-Native:SuNF	-0.0535	-0.0990	-0.0209	<0.005
SBF Baseline				
Intercept	0.6523	0.6113	0.7416	<0.001
Non-Native (for SBF condition)	0.0399	0.0059	0.0652	<0.05
SuNF (for native group)	-0.0445	-0.0722	-0.0185	<0.001
VPBF (for native group)	-0.0094	-0.0378	0.0180	0.4414
Speech Rate	-0.0686	-0.0887	-0.0597	<0.001
Non-Native:SuNF	-0.0149	-0.0568	0.0215	0.3718
Non-Native:VPBF	0.0387	0.0055	0.0743	<0.05

Table 5.17. Parameter values for fixed variables in the object duration regressions on prosodically appropriate productions, with native as the nativeness baseline

The validity of the duration measure is confirmed by the findings that verb durations were significantly longer in VPBF productions than SuNF productions, and object durations were significantly longer in native VPBF and SBF productions than in native SuNF productions. The longer verb durations in the VPBF context and longer object durations in VPBF and SBF contexts are due to the fact that the object should and the verb can receive pitch accents in the VPBF and SBF conditions, but the object and verb should not receive pitch accents in the SuNF condition. The phonological analysis showed that 78% of the VPBF verbs were accented, while only 1% of the SuNF verbs were accented. Similarly, almost 100% of native VPBF and SBF objects were accented, while none of the native SuNF objects were accented. Accented words tend to have longer durations than unaccented words.

Interestingly, although subjects can be accented in all three contexts, the subject durations were longest in the SuNF context, followed by the SBF context, and were shortest in the VPBF context. The longer subject durations for the SuNF context relative to the broad focus contexts might reflect the higher percentage of accented subjects in the SuNF context (100%) than in the broad focus contexts (both 93%). The longer subject durations in the SBF context than the VPBF context are worthy of note because Gussenhoven's (1983a, 1999) model predicts that subjects should be obligatorily accented in the SBF context, but not the VPBF context. This prediction does not seem to be supported by the phonological analysis, although at the phonetic level this distinction may be leading to the longer subject durations in SBF productions.

Non-native speakers did produce longer verb durations than native speakers, even after speech rate was statistically controlled. This difference may have led to the perception of a non-native accent, but it did not affect the context-appropriateness of the non-native prosody.

Most importantly for this study, there were significant interactions between nativeness and the VPBF/SuNF and the VPBF/SBF context contrasts for object durations. These regressions reveal that object durations are significantly longer for non-native productions than native productions in the VPBF and SBF conditions, but there is no significant difference between native and non-native productions in the SuNF condition. Recall that in the VPBF and SBF conditions the object receives an obligatory final pitch accent. In the phonological analysis of VPBF sentences, 98% of native VPBF productions and 100% of non-native VPBF productions native had accented objects. Similarly, 100% of native SBF productions and 99% of non-native SBF productions had accented objects. Given the high degree of accentuation for

VPBF and SBF objects, the longer object durations in these contexts for non-native speakers relative to native speakers suggests that the non-natives may be using duration as a more important cue to accentuation than the natives, or may be providing stronger cues to accentuation in general. The differences between native and non-native object durations across the three contexts can be seen in Figure 5.7.

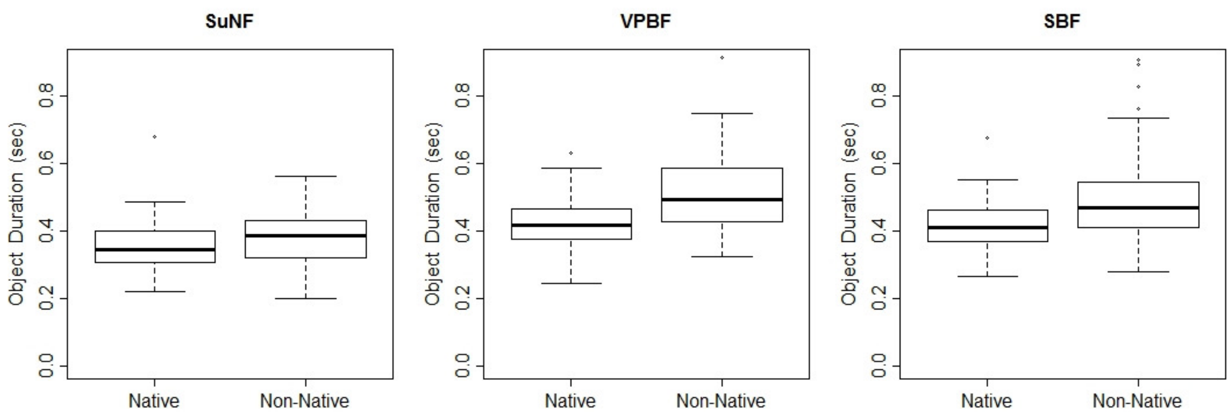

Figure 5.7. Boxplots showing the object durations in seconds in appropriate SuNF, VPBF and SBF productions by native and non-native speakers

Whatever their source, these durational differences between native and non-native speakers did not lead native listeners to judge the non-native prosody to be inappropriate. This is likely to be because duration is a cue to pitch accent location, and the non-native speakers lengthened the word containing the obligatory final pitch accent in VPBF and SBF productions. Therefore, native listeners would be more likely to perceive the obligatory object pitch accents produced by non-native speakers. This would lead the listeners to judge the non-native productions to have appropriate prosody.

5.5.2.3 Analysis of RMS Amplitude

In the RMS amplitude regressions, the dependent variable was RMS amplitude for each target word. Praat returns RMS amplitude in Pascals, which were converted to decibels, with the RMS amplitude for the entire sentence (in Pascals) as the reference value. This method of conversion controlled for differences in overall RMS amplitudes across recordings, so no separate sentence control variable was needed in the RMS amplitude regressions. The word control variable was the mean RMS amplitude for the word in the target position, after conversion to dB (averaging over all productions of that sentence).

In the first subject RMS amplitude regression model, nativeness was not significant, so it was not retained. Because of this, the interaction between nativeness and context was not explored. The final model included context and mean subject RMS as fixed variables. The parameter values for the final subject RMS regressions with SuNF and VPBF baselines are listed in Table 5.18.

	Estimate	95% CI - lower	95% CI – upper	p
SuNF Baseline				
Intercept	1.277	0.6702	1.837	<0.01
SBF	-1.715	-2.0810	-1.389	<0.001
VPBF	-1.815	-2.1876	-1.498	<0.001
Mean Subject RMS	1.034	0.8396	1.227	<0.001
VPBF Baseline				
Intercept	-0.5382	-1.1509	-0.0213	<0.05
SBF	0.1004	-0.1619	0.4242	0.4362
SuNF	1.8153	1.4928	2.1810	<0.001
Mean Subject RMS	1.0344	0.8435	1.2329	<0.001

Table 5.18. Parameter values for fixed variables in the subject RMS amplitude regressions on prosodically appropriate productions

In the first verb RMS regression model, nativeness was not significant, so it was not retained. Because of this, the interaction between nativeness and context was not explored. The final model included context and mean verb RMS as fixed variables. The parameter values for the final verb RMS regressions with SuNF and VPBF baselines are listed in Table 5.19.

	Estimate	95% CI - lower	95% CI – upper	p
SuNF Baseline				
Intercept	-1.5630	-2.1740	-1.030	<0.001
SBF	1.9780	1.5773	2.449	<0.001
VPBF	1.9738	1.5649	2.434	<0.001
Mean Verb RMS	0.9065	0.3567	1.504	<0.05
VPBF Baseline				
Intercept	0.4108	-0.0966	0.9284	0.1112
SBF	0.0042	-0.3650	0.3954	0.9424
SuNF	-1.9738	-2.4347	-1.5534	<0.001
Mean Verb RMS	0.9065	0.3818	1.4721	<0.05

Table 5.19. Parameter values for fixed variables in the verb RMS amplitude regressions on prosodically appropriate productions

In the first object RMS regression model, all three variables were significant, so they were all retained. A model containing these three significant variables was compared to a model that contained the three variables plus the interaction between context and nativeness with a likelihood ratio test. The addition of the interaction did not significantly improve the fit of the model ($X^2(2)= 2.032$, $p= 0.3620$). Therefore, the final model included context, nativeness, and mean object RMS as fixed variables. The parameter values for the final object RMS regressions with SuNF and VPBF baselines are listed in Table 5.20.

	Estimate	95% CI - lower	95% CI – upper	p
SuNF Baseline				
Intercept	-3.323	-4.2616	-2.378	<0.001
Non-Native	1.331	0.7201	1.903	<0.001
SBF	3.478	2.9540	4.102	<0.001
VPBF	3.252	2.7289	3.897	<0.001
Mean Object RMS	1.025	0.7836	1.298	<0.001
VPBF Baseline				
Intercept	-0.0710	-0.9382	0.8925	0.9874
Non-Native	1.3305	0.7694	1.9562	<0.001
SBF	0.2263	-0.2559	0.6923	0.3726
SuNF	-3.2522	-3.8773	-2.7274	<0.001
Mean Object RMS	1.0249	0.7796	1.2697	<0.001

Table 5.20. Parameter values for fixed variables in the object RMS amplitude regressions on prosodically appropriate productions, with native as the nativeness baseline

The RMS amplitude analysis revealed that amplitude served as a cue to accent location in these productions. RMS amplitude was higher for subjects in the SuNF context than in the VPBF and SBF contexts. Gussenhoven predicts that the subject should have a final pitch accent in the SuNF context, while the VPBF and SBF contexts are predicted to have final pitch accents on the object. As a result, pre-final subject pitch accents are optional, rather than required, for the VPBF context, although pre-final subject pitch accents are predicted to be required in SBF contexts. The phonological analysis shows that subjects received pitch accents in 100% of SuNF productions and 93% of VPBF and SBF productions. The higher RMS values for subjects in SuNF contexts have two possible causes. The first is that subjects were accented more frequently in SuNF contexts, and these accents were signaled by higher RMS values. The second is that subject accents in SuNF contexts are final, while subject accents in VPBF and SBF contexts are pre-final, and final accents could tend to have higher RMS values than pre-final accents. Of course, these possibilities are not mutually exclusive, and both may be playing a role in these results.

RMS amplitude was higher for verbs and objects in the SBF and VPBF contexts than in the SuNF context. Sentences in SuNF contexts are predicted to have final pitch accents on the subject, so the verb and object are not predicted to be accented in this context. In the VPBF and SBF contexts, pre-final pitch accents are allowed on the verb and final pitch accents are required on the object. In the phonological analysis, the verb was accented in only 1% of SuNF productions, but it was accented in 78% of VPBF and 74% of SBF productions. The object was accented in 0% of SuNF productions, but in 99% of VPBF and SBF productions. The higher RMS amplitudes for verbs in the VPBF and SBF contexts are likely to be due to higher RMS amplitudes on accented verbs and objects, which were vastly more common in these contexts than in the SuNF context.

There was a difference between native and non-native RMS values in the object regression. Non-native English speakers had higher RMS values on the object than native speakers, even after controlling for RMS over the entire sentence by using it as a reference when calculating dBs. Despite this difference between native and non-native productions, native English listeners judged these non-native productions to have appropriate prosody. These listeners may not have noticed the difference in RMS amplitude, or may not have judged it to be an adequate cue for object accentuation in the SuNF context.

5.5.2.4 Analysis of F0 Maximum

In the F0 maximum regressions, the dependent variable was maximum F0 in Hz for each target word. The sentence control variable was the mean of the F0 maxima for the subject, verb, and object in the sentence. The word control variable was the mean F0 maximum for the word in the target position (averaging over all productions of that sentence).

In the first subject F0 maximum regression model, nativeness and mean subject F0 maximum were not significant, so they were not retained. Because of this, the interaction between nativeness and context was not explored. The final model included context and mean sentence F0 maximum as fixed variables. The parameter values for the final subject F0 maximum regressions with SuNF and VPBF baselines are listed in Table 5.21.

	Estimate	95% CI - lower	95% CI – upper	p
SuNF Baseline				
Intercept	26.591	13.4952	39.042	<0.001
SBF	-24.266	-30.6055	-17.293	<0.001
VPBF	-26.767	-33.0500	-19.685	<0.001
Mean Sentence F0 Maximum	1.039	0.9854	1.094	<0.001
VPBF Baseline				
Intercept	-0.1756	-14.1174	11.673	0.8944
SBF	2.5008	-3.2743	7.996	0.3996
SuNF	26.7671	19.9248	33.306	<0.001
Mean Sentence F0 Maximum	1.0386	0.9889	1.098	<0.001

Table 5.21. Parameter values for fixed variables in the subject F0 maximum regressions on prosodically appropriate productions

In the first verb F0 maximum regression model, all four variables were significant, so they were all retained. A model containing these four significant variables was compared to a model that contained the four variables plus the interaction between context and nativeness, with a likelihood ratio test. The addition of the interaction did not significantly improve the fit of the model ($X^2(2)=0.4119$, $p=0.8139$). The final model included nativeness, context, mean sentence F0 maximum, and mean verb F0 maximum as fixed variables. The parameter values for the final verb F0 maximum regressions with SuNF and VPBF baselines are listed in Table 5.22.

	Estimate	95% CI - lower	95% CI – upper	p
SuNF Baseline				
Intercept	-151.2671	-286.3381	-25.347	<0.05
Non-Native	7.1390	0.9787	13.292	<0.05
SBF	10.1887	2.7984	17.595	<0.01
VPBF	15.1626	7.6751	22.743	<0.001
Mean Sentence F0 Maximum	0.9993	0.9409	1.048	<0.001
Mean Verb F0 Maximum	0.6478	0.0465	1.261	<0.05
VPBF Baseline				
Intercept	-136.1038	-271.5696	-8.8910	<0.05
Non-Native	7.1391	1.0966	13.3930	<0.05
SBF	-4.9740	-11.3880	0.7068	0.0834
SuNF	-15.1626	-22.4963	-8.0204	<0.001
Mean Sentence F0 Maximum	0.9993	0.9408	1.0457	<0.001
Mean Verb F0 Maximum	0.6478	0.0260	1.2526	<0.05

Table 5.22. Parameter values for fixed variables in the verb F0 maximum regressions on prosodically appropriate productions, with native as the nativeness baseline

In the first object F0 maximum regression model, mean object F0 maximum was not significant, so it was not retained. Because both nativeness and context were retained, a model containing only the three significant variables was compared to a model with these variables plus the interaction between nativeness and context. However, a likelihood ratio test showed that adding the interaction did not significantly improve the fit of the model ($X^2(2)= 3.5577$, p=0.1688), so the interaction was not retained in the final model. The final model included nativeness, context, and mean sentence F0 maximum as fixed variables. The parameter values for the final object F0 maximum regressions with SuNF and VPBF baselines are listed in Table 5.23.

	Estimate	95% CI – lower	95% CI – upper	p
SuNF Baseline				
Intercept	-8.597	-26.9495	5.096	0.1982
Non-Native	-11.964	-19.8339	-4.128	<0.005
SBF	16.387	8.5300	24.905	<0.001
VPBF	13.992	5.5950	21.918	<0.005
Mean Sentence F0 Maximum	0.940	0.8811	1.014	<0.001
VPBF Baseline				
Intercept	5.395	-13.0403	20.314	0.6886
Non-Native	-11.964	-19.7837	-3.861	<0.005
SBF	2.395	-4.0410	9.605	0.4340
SuNF	-13.992	-21.7044	-5.377	<0.005
Mean Sentence F0 Maximum	0.940	0.8829	1.014	<0.001

Table 5.23. Parameter values for fixed variables in the object F0 maximum regressions on prosodically appropriate productions, with native as the nativeness baseline

The F0 maximum analysis showed that the maximum F0 was significantly lower for subjects and higher for verbs and objects in the SBF and VPBF contexts relative to the SuNF context. Like the higher RMS values, the higher F0 maxima for verbs and objects in the broad focus contexts are likely due to the much higher percentage of pitch accents on verbs and objects in SBF and VPBF conditions. If these pitch accents contain an H (high) target (e.g. H* or L+H*), this high pitch target would raise the maximum F0. The high accentuation rate for subjects in VPBF and SBF contexts means that several explanations for the difference in F0 maxima are possible. The first possible explanation is that the slightly higher accentuation rate for subjects in the SuNF contexts led to more H pitch targets and therefore generally higher F0 maxima. The second is that the final pitch accents for subjects in the SuNF context were realized with stronger acoustic cues, including higher F0 maxima for H pitch accent targets. The third is that the subject pitch accents in SuNF contexts contained an H target more often than the subject pitch accents in the VPBF and SBF contexts. Further research is needed to tease apart these possibilities.

The verb and object F0 maximum regressions showed that non-native speakers had significantly higher F0 maxima on verbs and lower F0 maxima on objects than native speakers. This could reflect a sharper F0 declination for non-native speakers than native speakers. Despite these differences between native and non-native productions, native English listeners judged these non-native productions to have appropriate prosody. It is possible that these listeners did not notice the difference in F0 maxima, or that when combined with other pitch accent cues, it did not lead to misinterpretation.

5.5.2.5 Analysis of F0 Range

In the F0 range regressions, the dependent variable was the range in F0 values in Hz for each target word. The sentence control variable was the mean of the F0 range values for the subject, verb, and object in the sentence. The word control variable was the mean F0 range for the word in the target position (averaging over all productions of that sentence).

In the first subject F0 range regression model, the mean subject F0 range and both variables of interest were not significant, so a new model was built that did not include mean subject F0 range. The two variables of interest were also not significant in this model. The final model included context, nativeness, and sentence F0 range as fixed variables. The parameter values for the final subject F0 range regressions with SuNF and VPBF baselines are listed in Table 5.24.

	Estimate	95% CI – lower	95% CI - upper	p
SuNF Baseline				
Intercept	1.4217	-7.9091	15.6952	0.5634
Non-Native	4.2718	-3.3363	11.9100	0.2812
SBF	-5.7473	-14.7852	2.8461	0.1718
VPBF	-6.3085	-15.8938	1.8124	0.1396
Mean Sentence F0 Range	0.8831	0.7381	0.9866	<0.001
VPBF Baseline				
Intercept	-4.8881	-14.6750	8.6535	0.5738
Non-Native	4.2719	-3.2457	11.7642	0.2656
SBF	0.5612	-6.6628	8.2514	0.8966
SuNF	6.3085	-2.6106	15.3420	0.1396
Mean Sentence F0 Range	0.8831	0.7276	0.9747	<0.001

Table 5.24. Parameter values for fixed variables in the subject F0 range regressions on prosodically appropriate productions, with native as the nativeness baseline

In the first verb F0 range regression model, all four variables were significant, so they were all retained. Because both nativeness and context were both retained, a model containing only the four original variables was compared to a model with these variables plus the interaction between nativeness and context. However, a likelihood ratio test showed that adding the interaction did not significantly improve the fit of the model ($X^2(2)=4.44$, p=0.1086), so the interaction was not retained in the final model. The final model included context, nativeness, mean sentence F0 range, and mean verb F0 range as fixed variables. The parameter values for the final verb F0 range regressions with SuNF and VPBF baselines are listed in Table 5.25.

	Estimate	95% CI – lower	95% CI - upper	p
SuNF Baseline				
Intercept	-29.1108	-53.3360	-6.625	<0.05
Non-Native	8.3998	2.1632	14.851	<0.01
SBF	-9.3939	-17.3518	-1.617	<0.05
VPBF	-5.8029	-14.1312	1.626	0.1406
Mean Sentence F0 Range	0.9047	0.7979	1.010	<0.001
Mean Verb F0 Range	0.6015	0.2324	1.011	<0.05
VPBF Baseline				
Intercept	-34.9137	-58.5081	-13.338	<0.01
Non-Native	8.3998	2.2671	14.913	<0.01
SBF	-3.5910	-10.2453	3.012	0.2914
SuNF	5.8029	-1.7813	14.052	0.1432
Mean Sentence F0 Range	0.9047	0.8031	1.015	<0.001
Mean Verb F0 Range	0.6015	0.2513	1.008	<0.01

Table 5.25. Parameter values for fixed variables in the verb F0 range regressions on prosodically appropriate productions, with native as the nativeness baseline

In the first object F0 range regression model, only mean object F0 range was not significant, so it was not retained. Because both nativeness and context were retained, a model containing only the three significant variables was compared to a model with these variables plus the interaction between nativeness and context. A likelihood ratio test showed that adding the interaction significantly improved the fit of the model ($X^2(2)=8.3606$ $p<0.05$), so the interaction was retained in the final model. The final model included nativeness, context, mean sentence F0 range, and the interaction between nativeness and context as fixed variables. In order to determine the nature of the significant interaction between the SBF/SuNF context contrast and the native/non-native contrast, a third regression was run with SBF as the baseline. The parameter values for the three final object F0 range regressions are listed in Table 5.26.

	Estimate	95% CI – lower	95% CI - upper	p
SuNF Baseline				
Intercept	0.4454	-15.439	13.185	0.8866
Non-Native (for SuNF condition)	-31.4351	-46.209	-16.659	<0.001
SBF (for native group)	4.1600	-9.493	16.839	0.5462
VPBF (for native group)	5.9567	-7.018	19.254	0.3976
Mean Sentence F0 Range	1.1945	1.089	1.350	<0.001
Non-Native:SBF	27.3401	9.031	46.678	<0.005
Non-Native:VPBF	20.2910	2.308	40.088	<0.05
VPBF Baseline				
Intercept	6.402	-10.220	19.682	0.5272
Non-Native (for VPBF condition)	-11.144	-11.144	1.128	0.0818
SBF (for native group)	-1.797	-15.685	12.016	0.7986
SuNF (for native group)	-5.957	-18.796	7.456	0.4044
Mean Sentence F0 Range	1.194	1.097	1.351	<0.001
Non-Native:SBF	7.049	-10.135	24.063	0.3806
Non-Native:SuNF	-20.291	-39.211	-1.261	<0.05
SBF Baseline				
Intercept	4.605	-11.115	17.574	0.6940
Non-Native (for SBF condition)	-4.095	-15.483	8.950	0.5584
SuNF (for native group)	-4.160	-17.488	8.789	0.5512
VPBF (for native group)	1.797	-12.432	15.490	0.7972
Sentence F0 range	1.194	1.088	1.349	<0.001
Non-Native:SuNF	-27.340	-46.407	-8.485	<0.01
Non-Native:VPBF	-7.049	-23.822	9.854	0.3800

Table 5.26. Parameter values for fixed variables in the object F0 range regressions on prosodically appropriate productions, with native as the nativeness baseline

Surprisingly, SuNF productions had larger F0 ranges on verbs than SBF productions. The verb F0 maximum regression showed that SBF productions had higher maximum F0s than SuNF productions, so the larger ranges for SuNF verbs must be due to lower F0 minima. Verbs were almost never accented in SuNF productions, so one possible explanation is that the end of an H tone from the subject accent continued into the verb. This would result in F0s on the verb that ranged from mid values (due to a pitch falling after the preceding accent) to low values from the L- intermediate phrase accent that is common in declarative sentences. In contrast, in SBF productions, 93% of accented verbs were surrounded by accented subjects and objects, and the remaining accented verbs were adjacent to accented objects. If all the pitch accents in a sentence were of the same type (e.g. H*), and there were no sentence-internal phrase boundaries, then there would not need for much pitch movement between the accents. This would result in a small F0 range on these SBF verbs.

Non-native speakers produced larger F0 ranges on verbs than native speakers. This could be related to the higher F0 maxima found for non-native verbs in the previous analysis. That analysis also revealed lower F0 maxima for non-native objects, which pointed to the possibility that non-native speakers had steeper F0 declinations than native speakers. This explanation is also compatible with the larger F0 ranges on non-native verbs. If the F0 in a non-native utterance is moving from a higher point during the verb to a lower point during the object (relative to a native utterance), we would expect a larger F0 range on the verb, resulting from this F0 movement.

There was a significant interaction between nativeness and the SuNF/SBF and SuNF/VPBF contrasts in the object regression. A comparison of the three regressions shows that in the SuNF condition, objects in native productions had significantly larger F0 ranges than those in non-native productions, but there were no significant differences between native and non-native object F0 ranges in the VPBF or SBF conditions. The non-natives' tendencies to produce smaller F0 ranges on SuNF objects can help listeners determine that they are unaccented. All of the SuNF productions had unaccented objects. By decreasing the F0 ranges on these SuNF objects, non-native speakers are making the distinction between accented and unaccented words more acoustically apparent. This difference between native and non-native productions is illustrated in Figure 5.8.

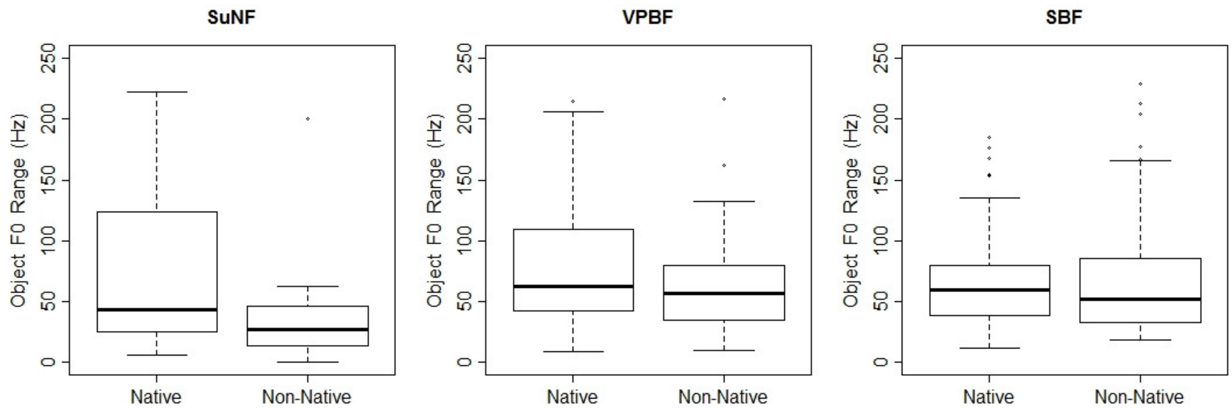

Figure 5.8. Boxplots showing the object F0 range values in Hz in appropriate SuNF, VPBF and SBF productions by native and non-native speakers

5.6 Acoustic Analysis of Production Discussion

The acoustic analyses of the production recordings had two goals. The first goal was to test the hypothesis that the non-native SuNF productions which were judged to have inappropriate prosody had pitch accent placement patterns that were appropriate for VPBF or SBF rather than SuNF contexts. The second goal was to determine whether there is more than one way to produce prosody that will be considered appropriate by native listeners.

5.6.1 Discussion of Inappropriate Non-Native SuNF Productions

The phonological prosodic analysis of the appropriate and inappropriate productions by non-native speakers revealed that the SuNF productions that had been judged inappropriate had very different patterns of accent placement than the SuNF productions that had been judged appropriate. The inappropriate SuNF productions all had final pitch accents on the object, which is predicted to be appropriate for VPBF and SBF productions. In contrast, the appropriate SuNF productions all had final pitch accents on the subject, a pattern that is predicted for sentences in

SuNF contexts. These differences in perceived accent placement were tied to a number of prosodic differences at the phonetic level. Inappropriate SuNF productions had subjects with lower RMS amplitudes, lower F0 maxima, and smaller F0 ranges, verbs with higher RMS amplitudes, and objects with longer durations, higher RMS amplitudes, higher F0 maxima, and larger F0 ranges than appropriate SuNF productions. These analyses show that the non-native SuNF productions that were judged to have inappropriate prosody differed from those with appropriate prosody at the phonological and phonetic levels.

The analysis of non-native productions also compared the inappropriate SuNF productions to appropriate VPBF and SBF productions, to see if they were indeed identical. The phonological analysis revealed some differences. The most common accent pattern for the inappropriate SuNF productions was YnY (53%), followed by YYY (43%). These preferences were reversed for the appropriate VPBF and SBF productions: YYY was the most common pattern (VPBF: 73%, SBF: 73%), followed by YnY (VPBF: 18%, SBF: 19%). This difference is logical if we consider that YnY is closer than YYY to the appropriate SuNF pattern of Ynn. Therefore, even speakers who did not mark subject narrow focus correctly may be moving in the right direction. The phonetic analysis also revealed acoustic differences between inappropriate SuNF productions and appropriate VPBF and SBF productions. The inappropriate SuNF productions had subjects with longer durations and higher F0 maxima, and verbs with shorter durations and lower RMS amplitudes than appropriate VPBF productions. The inappropriate SuNF productions also had verbs with shorter durations and lower RMS amplitudes, and objects with smaller F0 ranges than appropriate SBF productions.

All of these acoustic analysis results show that inappropriate SuNF productions fall somewhere between appropriate SuNF productions and appropriate VPBF and SBF productions, although they are closer to the broad focus productions. At the phonological level, the inappropriate SuNF accent patterns are predicted to be appropriate for broad focus, but not subject narrow focus. At the phonetic level, there were more significant acoustic feature differences between the inappropriate and appropriate SuNF productions than between the inappropriate SuNF productions and the two types of broad focus productions combined.

5.6.2 Discussion of Appropriate Native and Non-Native Productions

The phonological prosodic analysis of the appropriate productions by native and non-native speakers revealed striking similarities in the patterns of pitch accent placement that the two groups used across the three contexts. However, the phonetic analysis showed that there were still some significant prosodic differences between productions by native and non-native speakers. Thus, the data support the Relaxed Native Perception Model described in Chapter 1.

The analysis of productions judged to have context-appropriate prosody showed that non-native prosody did differ from native prosody at the phonetic level, in ways that did not always interact with context. Non-native productions had longer durations, higher maximum F0s and larger F0 ranges on the verb, as well as higher RMS amplitudes and lower F0 maxima on the object, relative to native productions. These differences were either not noticed, or not deemed to affect the context-appropriateness of the productions. It would be interesting to run a follow-up experiment in which native listeners gave likert-scale ratings of the appropriateness of the prosody, rather than making a binary appropriate/inappropriate decision. In such an experiment, these types of general differences between native and non-native productions may lower the ratings for non-native prosody.

Interestingly, in both of the context-specific phonetic differences found between native and non-native productions, the non-native speakers enhanced the acoustic cues signaling the presence or absence of a pitch accent, relative to native speakers. SuNF productions should not (and did not) have pitch accents on objects, and non-native speakers produced SuNF objects with smaller F0 ranges than native speakers. VPBF and SBF productions have obligatory final pitch accents on objects, which the non-native speakers produced with longer durations that native speakers. As long as pitch accents are placed on the correct words, such an increase in the acoustic differences between accented and unaccented words should make the prosody of a production easier to interpret, and therefore more likely to be judged appropriate. These results demonstrate that non-native speakers can use prosodic cues differently from native speakers and still be understood, as long as the non-native speakers' cues still allow listeners to distinguish between accented and unaccented words and the accents are placed on the correct word or words for the context. These results raise the question of whether non-native speakers have to produce stronger cues to accentuation in order to have a production judged to be as appropriate as a native speaker's production. This possibility should be explored in future work.

5.6.3 Acoustic Analysis of Production Summary

Taken together, the two acoustic analyses showed that incorrect accent placement was the main problem with non-native productions that were judged to have inappropriate prosody. Some non-native speakers produced utterances in the SuNF context that had inappropriate accent patterns for that context, but appropriate accent patterns for the broad focus contexts. However, when non-native speakers did use context-appropriate accent patterns, they seemed to have a good mastery of the acoustic cues used to signal pitch accents. When non-native productions with appropriate accent patterns differed acoustically from native productions in context-specific ways, the non-native speakers actually increased the acoustic distinction between accented words and unaccented words, relative to native speakers.

Chapter 6

Relationship between Perception and Production in the Prosodic Prominence of Non-Native English Speakers

6.1 Introduction

One goal of this dissertation was to determine whether an English learner's ability to place a final pitch accent on the appropriate word for a particular information structure and to produce this accent in a native-like manner depends on their ability to accurately perceive and understand such accents. This is an under-studied research area (Chun 2002), although a better understanding of the relationship between the perception and production of second language prosody is crucial for developing useful training programs. For example, information on this topic would help English teachers determine whether perception or production training would be most useful for helping students use pitch accents to communicate effectively.

This chapter explores the relationship between participants' perceptual and production skills, explored through a series of correlations. Section 6.2 discusses the correlations between participants' accuracy in the prominence placement experiment (Exp. 4) and the prominence understanding experiment (Exp. 3). Section 6.3 discusses the correlations between accuracy in the prominence production experiment (Exp. 1, as judged by native listeners in Exp. 5), and the prominence understanding experiment (Exp. 3). Section 6.4 discusses the correlation between accuracy in the prominence production experiment (Exp. 1 and 5) and the prominence perception experiment (Exp. 2). Finally, Section 6.5 offers a general discussion of all of these correlations and provides suggestions for future experiments that could be used to explore this topic.

6.2 Correlations between Prominence Placement (Exp. 4) and Understanding (Exp. 3)

Spearman correlations were used to compare non-native participants' accuracy on the computer-based prominence placement task (Exp. 4) to their accuracy on the prominence understanding task (Exp. 3). Non-parametric Spearman correlations were used because of the non-normal distributions of the data. Participants' overall performance on these tasks was compared, as well as their performance for individual contexts (SuNF, VPBF, SBF). There were no significant correlations between an English learner's ability to accurately place prominence and their ability to understand the meaning of prominence placement (Overall: ρ=0.2716, S = 7765, p = 0.09; SuNF: ρ=0.0513, S = 10113, p = 0.7534; VPBF: ρ= 0.1807, S = 8733, p = 0.2644; SBF: ρ=0.0156, S = 10494, p = 0.924). These results can be seen in Figure 6.1.

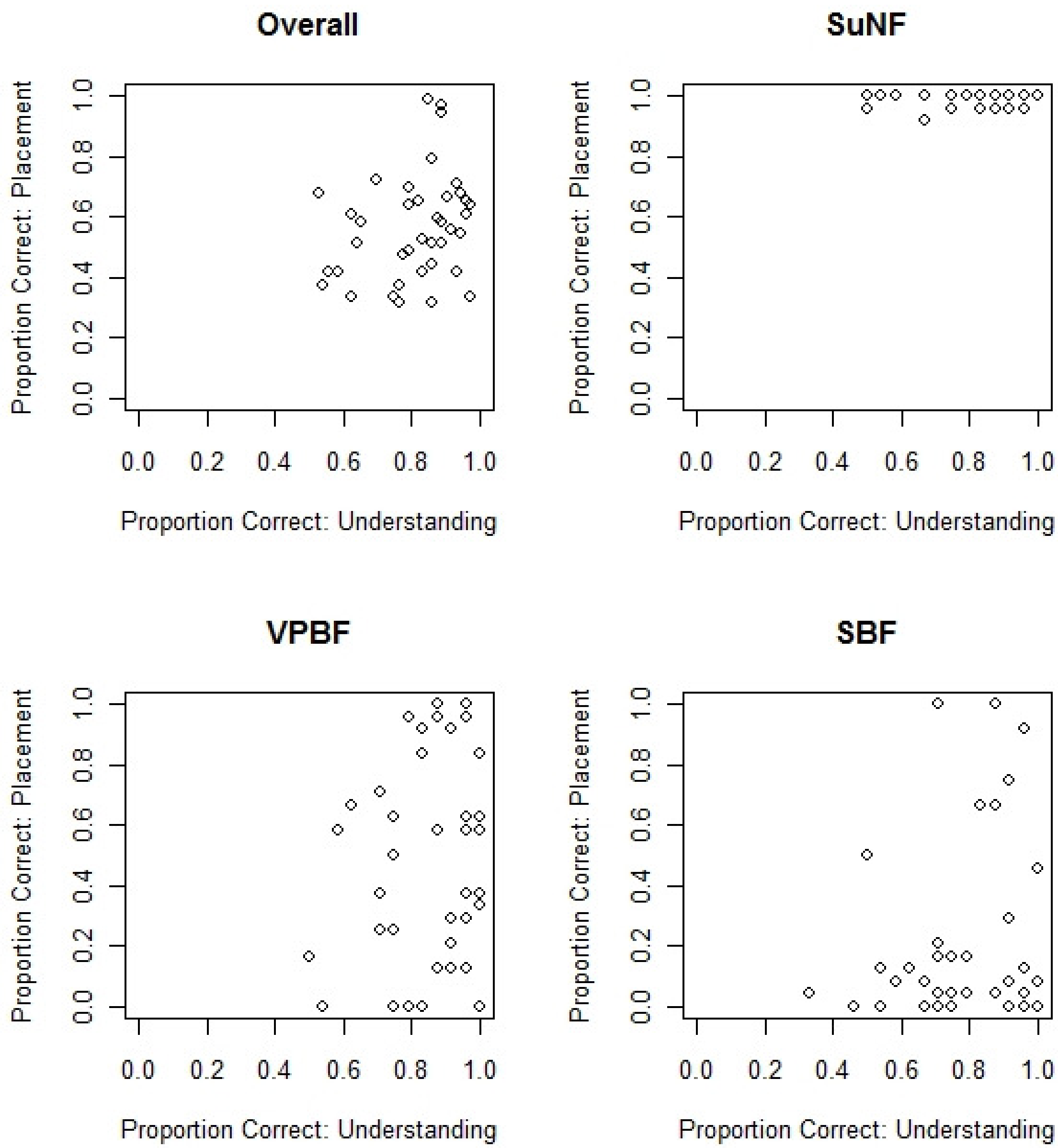

Figure 6.1. Scatterplots comparing individual English learners' accuracy at placing prominence (Exp. 4) to their accuracy at understanding the meaning of prominence location (Exp. 3), both overall and for the SuNF, VPBF, and SBF contexts

6.3 Correlations between Prominence Production (Exp. 1 and 5) and Understanding (Exp. 3)

Spearman correlations were used to compare non-native participants' accuracy on the spoken prominence production task (Exp. 1, as judged by native listener in Exp. 5) to their accuracy on the prominence understanding task (Exp. 3). Once again, participants' overall performance on these tasks was compared, as well as their performance for individual contexts (SuNF, VPBF, SBF). There were no significant correlations between an English learner's ability to produce prosody that was perceived as context-appropriate and their ability to understand the meaning of prominence placement (Overall: ρ=0.1320, S = 9252, p = 0.4167; SuNF: ρ=0.0375, S = 10260, p = 0.8184; VPBF: ρ= -0.1721, S = 12494, p = 0.2884; SBF: ρ=0.0511, S = 10115, p = 0.7541). These results can be seen in Figure 6.2.

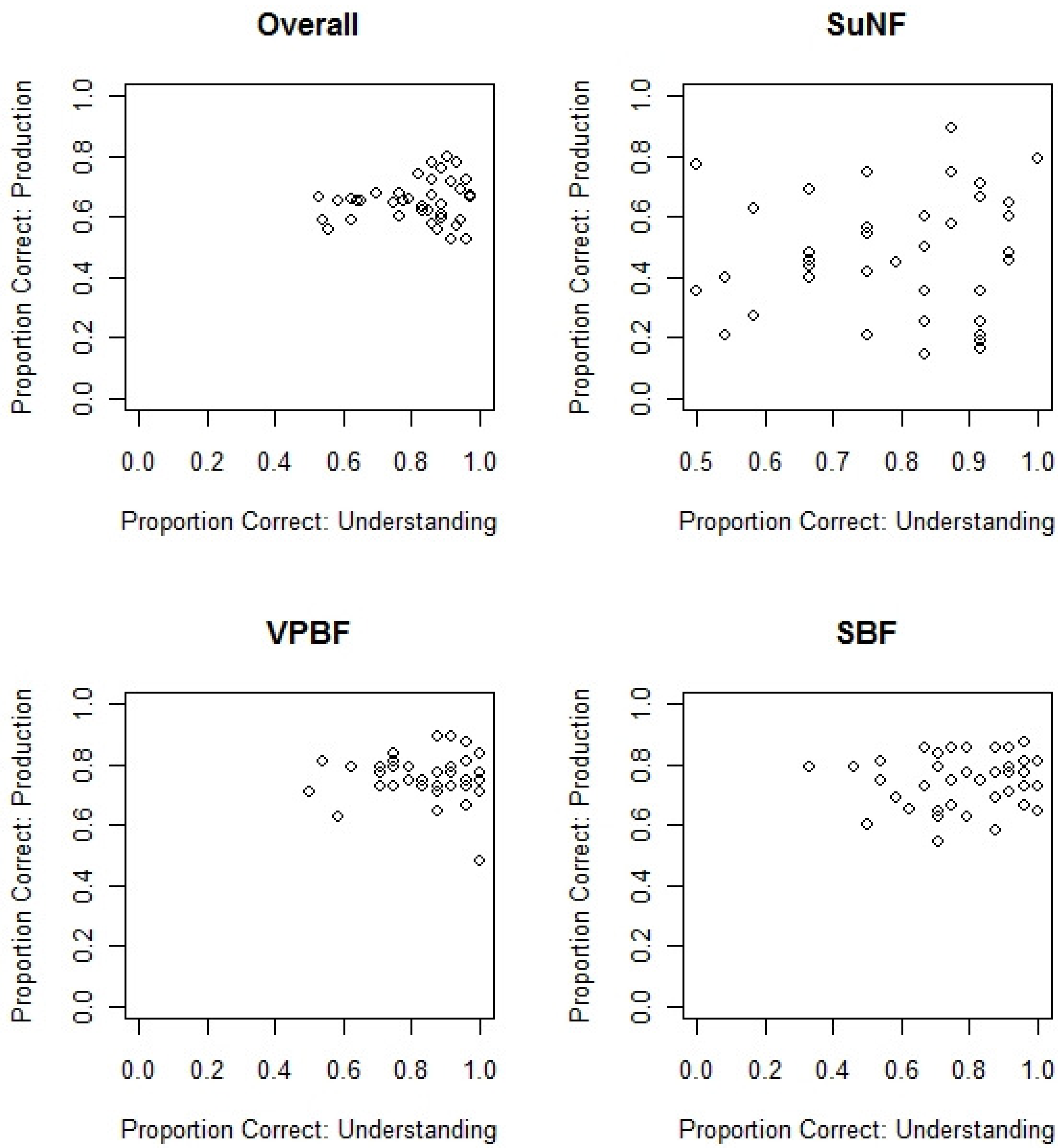

Figure 6.2. Scatterplots comparing individual English learners' accuracy at producing context-appropriate prosody (Exp. 1 and 5) to their accuracy understanding the meaning of prominence location (Exp. 3), both overall and for the SuNF, VPBF, and SBF contexts

6.4 Correlation between Prominence Production (Exp. 1 and 5) and Perception (Exp. 2)

A Spearman correlation was used to compare non-native participants' accuracy at the spoken prominence production task (Exp. 1, as judged by native listeners in Exp. 5) to their accuracy at the prominence perception task (Exp. 2). For this correlation, only participants' overall performance on these tasks was compared, as participants were not made aware of the different discourse contexts in the prominence perception experiment (in this experiment they were presented with single sentences with no context question). There was no significant correlation between an English learner's ability to produce prosody that was perceived as context-appropriate and their ability to perceive the location of prosodic prominence (Overall: $\rho=0.0214$, $S = 10432$, p-value = 0.896). This result can be seen in Figure 6.3.

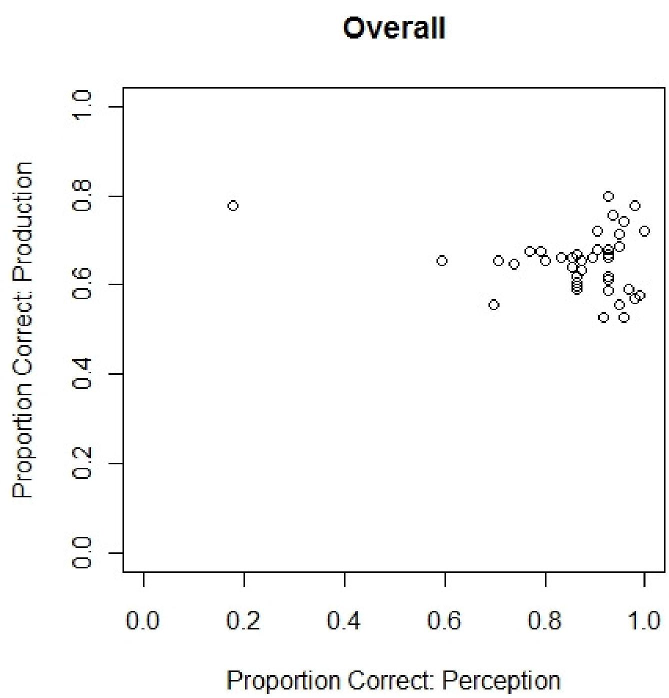

Figure 6.3. Scatterplot comparing individual English learners' overall accuracy at producing context-appropriate prosody (Exp. 1 and 5) to their accuracy at perceiving the location of prominence (Exp. 2)

6.5 Discussion of Prominence Placement, Production, Understanding, and Perception Correlations

The lack of correlations between any of the perceptual and production-related tasks provides some support for the Perception/Production Independence Model. This model assumes that different skills are required for pitch accent perception and understanding on the one hand, and for realization and placement on the other. Some language learners may find accurately placing pitch accents while producing speech more challenging than understanding pitch accents while listening to speech, because during speech production they have to focus on segment pronunciation in addition to prosody. Other language learners may find understanding pitch accents more challenging because of the unconstrained nature of the task: the speaker may use unfamiliar words or unusual segment pronunciations. If the language learner finds one of these tasks more challenging than the other, her pitch accent use in the more challenging task could degrade. Differences between pitch accent perception and realization abilities could arise because some language learners may have developed appropriate representations of pitch accents, but not the motor skills needed to produce them accurately (Bradlow, Pisoni et al. 1997). Other language learners may find realization easier because they have determined the articulatory movements necessary to produce pitch accents well enough to be understood by native speakers, but cannot use acoustic cues to correctly identify native productions of pitch accents.

The correlation results reported in this chapter should be interpreted with a certain amount of caution. For one thing, these are null results, and therefore could simply be due to inadequate power. The scatterplots in Figures 6.1-6.3 do not reveal striking relationships between perceptual and production-related tasks, although the non-natives who were most successful at placing and producing prominence did tend to be in the upper half of the non-native spectrum for understanding prominence (Figures 6.1 and 6.2).

Differences between the tasks and features of the tasks themselves also make these results challenging to interpret. For instance, the prominence production data are not directly based on the utterances produced by the non-native speakers (Exp. 1), but instead are based on the judgments of these recordings by native listeners (Exp. 5). The listeners may have been influenced by features of the non-natives' recordings apart from prosody, such as their segmental pronunciation, even though they were instructed to focus on the prosody. Listeners may also have made mistakes due to lack of attention during the relatively long judgment task. In addition, the production task (Exp. 1) combines prominence placement and realization in a way that makes it hard to determine the influence of each component in the data. Along the same lines, the prominence understanding task (Exp. 3) required listeners to both perceive prominence location and interpret it. Finally, the prominence placement task (Exp. 4) required meta-linguistic knowledge that even many native English speakers did not possess.

Some of the difficulties described in the previous paragraph are hard to overcome because of the nature of language use. Future research in this area should use experiments explicitly designed to be maximally comparable across perception and production and to distinguish as much as possible between the perception/realization aspects and the understanding/placement aspects of prominence. In the current study, the prominence understanding experiment (Exp. 3) is a good match for the prominence production experiment (Exp. 1) because both experiments require knowledge of prominence placement in different contexts and the acoustic cues used to mark prominence. The prominence placement experiment (Exp. 4) is being treated as a production experiment, because participants are asked what they would do if they were producing a sentence in a particular context. However, it is unclear what the perceptual correlate of this experiment could be. It is unlikely that language learners would provide different responses if they were asked to select the location of prominence that they would expect if they were listening to an answer to a particular question. One possible correlate would be an experiment in which they were shown a sentence with one of the words marked as prominent, and they would have to select the appropriate context from a set of options. The prominence perception experiment does not have a direct realization correlate in the current study. In order to focus on the perception/realization component, data from the prominence perception experiment (Exp. 2) should be compared to data from a realization only experiment. Such an experiment could involve participants reading sentences in which different words were underlined. The instructions would ask the participants to make the underlined words sound prominent or important. This experiment would require no knowledge of accent placement, and would therefore be a closer match to the prominence perception experiment.

Training studies have been used to examine the link between segmental perception and production (e.g. Bradlow, Pisoni et al. 1997; Bradlow, Akahane-Yamada et al. 1999; Rochet 1995). This type of study could also be informative when examining acquisition of prosodic prominence. By allowing researchers to compare a single participant's performance before and after training, this type of study may reveal relationships between perception and production that are not clear when comparing across participants, as the correlations in the current study do. While 'tHart and Collier (1975) and De Bot and Mailfert (1982) researched the effect of general prosody perception training on prosody production, a more focused study examining a particular aspect of prosody (in this case, prosodic prominence) would test the transfer of specific information from one domain to the other.

Multiple versions of the prosodic prominence training study are possible. One study could focus on the perception and realization component of pitch accent use. This study would include a pre-test and a post-test in which participants were asked to produce sentences with prominence on underlined words. During the training component, participants would hear productions of sentences with prominence on different words spoken by native English speakers. After hearing each production, they would see two representations of the sentence written on the screen, one with the prominent word underlined and one with a different word underlined. They would be instructed to select the version of the sentence with the prominent word underlined. The pre- and post-test productions could be evaluated based on acoustic measurements or perceptual judgments of pitch accent location. If the pitch accents in the post-test productions were clearer than those in the pre-test productions, it would indicate that the pitch accent perception training transferred to improved pitch accent realization ability. Another study might evaluate pitch accent understanding and placement skills. This experiment would also include a production pre-test and a post-test, which replicate the prominence production experiment in the current study. During the training portion of the experiment, participants would once again hear productions of sentences spoken by native English speakers with prominence on different words. They would see two questions on the screen, one providing the correct context for the spoken sentence, and the other providing the incorrect context. They would be instructed to select the question that provided the correct context for the sentence they heard. The pre- and post-test productions could be evaluated based on acoustic measurements or perceptual judgments of pitch accent location. If the pitch accents in the post-test productions were placed in the appropriate location for the context more often than those in the pre-test productions, it would show that the perceptual training transferred to production ability. Both experiments could vary whether the participants received feedback on their responses during the training phase. These types of training experiments can provide important information on the question of how pitch accent perception relates to pitch accent production.

Chapter 7

Conclusions and Applications

7.1 Introduction

In the final chapter of this dissertation I will describe a predictive framework for future research into second language acquisition of prosodic focus marking. This framework is based on the results of the experiments covered in Chapters 3 and 4. I will also discuss how the results can be used to answer the questions posed in Chapter 1 regarding native listeners' perception of non-native pitch accent production, and the relationship between language learners' perception and production abilities. Finally, I will make some recommendations for how these results can be applied to improving second language prosody instruction.

The chapter will begin with a summary of the results from all experiments and analyses (Section 7.2). In Section 7.3, the predictive framework for the acquisition of prosodic focus marking will be laid out, with a discussion of how the results of the experiments support this framework. Native listeners' perception of non-native pitch accent production are covered in Section 7.4. The relationship between language learners' perception and production is discussed in Section 7.5. Section 7.6 has a discussion of how the results could be used to teach second language prosody more effectively. The chapter concludes with some brief statements about the implications of this research project (Section 7.7).

7.2. Summary of Results

7.2.1 Summary of Results for Experiment 2: The Prominence Perception Experiment

In the prominence perception experiment (Exp. 2), English and Korean participants were both more accurate at perceiving pitch accent location than Mandarin participants. There were no significant differences in accuracy for recordings produced in different contexts.

7.2.2 Summary of Results for Experiment 3: The Prominence Understanding Experiment

In the prominence understanding experiment (Exp. 3), English participants were more accurate at determining whether a sentence had context-appropriate prosody than Mandarin and Korean participants. Versant score was a significant predictor of non-native accuracy, with more proficient participants interpreting prosody more accurately than less proficient participants. In general, participants were more accurate on matched items than mismatched items, but they also showed different patterns of performance for the two item types. For matched items, non-native participants were more accurate in the SuNF condition than SBF or VPBF conditions. However, for mismatched items, non-native participants were most accurate in the VPBF condition, followed by the SBF condition, then the SuNF condition. Finally, participants were more accurate on productions with accented subjects than productions with accented objects.

7.2.3 Summary of Results for Experiment 4: The Prominence Placement Experiment

In the prominence placement experiment (Exp. 4), both English and Korean participants were more accurate than Mandarin participants at predicting the correct pitch accent location for the VPBF and SBF contexts. All three groups of participants were most accurate at placing pitch accents for sentences in SuNF contexts, followed by VPBF contexts, and were least accurate in SBF contexts.

7.2.4 Summary of Results for Experiment 1: The Prominence Production Experiment

The production experiment (Exp. 1) data were analyzed in two ways. The first was by having native English listeners judge the contextual-appropriateness of the productions in Exp. 5. The second was by labeling the accent locations and making acoustic measurements on the productions that were judged to be most and least appropriate. In the native perception of non-native production experiment (Exp. 5), broad focus Mandarin productions were judged to be more appropriate than English and Korean productions. However, in the SuNF context, English productions were judged to be most appropriate, followed by Korean productions, and Mandarin productions were judged least appropriate. Versant score was a significant predictor of non-native accuracy on this task, with more proficient participants producing more appropriate prosody than less proficient participants. Both groups of non-native speakers produced more appropriate prosody in the VPBF and SBF contexts than in the SuNF context.

All of the non-native productions that were judged to have inappropriate prosody by at least seven out of eight native listeners in Exp. 5 were in the SuNF context. The accent labeling and phonetic analysis indicated that these productions had prosody that was more appropriate for the VPBF or SBF context than the SuNF context. While none of the non-native SuNF productions that were judged appropriate by seven out of eight listeners had accented verbs or objects, 48% of inappropriate productions had accented verbs and 100% had accented objects. Inappropriate SuNF productions had subjects with lower RMS amplitudes, lower F0 maxima, and smaller F0 ranges, verbs with higher RMS amplitudes, and objects with longer durations, higher RMS amplitudes, higher F0 maxima, and larger F0 ranges than appropriate SuNF productions. However, the inappropriate SuNF productions also differed from the appropriate VPBF and SBF productions at the phonological and phonetic levels. The inappropriate SuNF productions were less likely to have accented verbs than appropriate broad focus productions. In addition, the inappropriate SuNF productions had subjects with longer durations and higher F0 maxima, and verbs with shorter durations and lower RMS amplitudes than appropriate VPBF productions. The inappropriate SuNF productions also had verbs with shorter durations and lower RMS amplitudes, and objects with smaller F0 ranges than appropriate SBF productions.

Non-native and native productions that were judged to have appropriate prosody by seven out of eight listeners in Exp. 5 had very similar patterns of accent placement within each context. However, there were some acoustic differences between native and non-native productions. Relative to native productions, non-native productions had longer object durations in VPBF and SBF contexts, and smaller F0 ranges on objects in SuNF contexts. These differences served to increase acoustic distinctions between accented and unaccented words.

7.2.5 Summary of Perception/Production Correlations

A number of correlation tests were carried out to determine whether there was a relationship between accuracy in perceptual tasks and accuracy in production-related tasks. Separate correlations compared participant accuracy in 1) the prominence placement (Exp. 4) and understanding (Exp. 3) experiments, 2) the prominence production (Exp. 1) and understanding (Exp. 3) experiments, and 3) the prominence production (Exp. 1) and perception (Exp. 2) experiments. None of these correlations were significant.

7.3. Predictive Framework for Studying Prosodic Prominence Acquisition

7.3.1 Predictive Framework for Studying Prosodic Prominence Acquisition: Perception and Realization

Two main types of language acquisition models were described in Chapter 1. Transfer Models are the most commonly studied type of phonology acquisition model. They focus on the effect that a language learner’s L1 has on their L2. For example, the SLM model (Flege, Munro et al. 1995) predicts that if a learner's L1 and L2 have identical categories, then positive transfer from the L1 to the L2 should lead to easy acquisition of these categories. If the two categories are similar, but not identical, in the L1 and L2 this should make it harder for the learner to acquire the L2 category. The Transfer Model for pitch accent perception and realization described in the introduction correctly predicted that Korean speakers would be better than Mandarin speakers at perceiving English pitch accents. It made this prediction because Korean, like English and unlike Mandarin, uses pitch only post-lexically. This means that native Korean speakers learning English may link intonational events with prosodic categories like pitch accents more easily than native Mandarin speakers.

L2 Challenge Models were the second type of model described in Chapter 1. The L2 Challenge Model for pitch accent perception and realization predicted that some types of pitch accents should be easier to perceive and realize than others because of their features within the L2, regardless of their relationships to categories in a learner's L1. It predicted that pitch accents in SuNF productions should be easier to perceive than pitch accents in broad focus, because pitch accents in narrow focus and earlier pitch accents tend to have stronger acoustic cues. If these differences between native speakers' productions of early and late pitch accents and between pitch accents in broad and narrow focus are replicated by non-native speakers, then non-native speakers' realizations of late pitch accents in broad focus may be less perceptible than their realizations of early pitch accents in narrow focus. The prominence perception experiment (Exp. 2) results did not support this prediction. However, the fact that in the prominence understanding experiment (Exp. 3), non-native participants were more accurate on items with final pitch accents on the subject than those with final pitch accents on the object does support this prediction. The final pitch accents on subjects were produced in narrow focus and were placed at the start of the utterance, while the final pitch accents on objects were produced in broad focus and were placed at the end of the utterance. It is possible that the differences in perceptibility between early narrow focus accents and late broad focus accents only affect performance during more challenging tasks, like the prominence understanding task (Exp. 3). This hypothesis can be tested by having participants do tasks with varying degrees of difficulty that require them to recognize early and late pitch accents produced in broad and narrow focus.

7.3.2 Predictive Framework for Studying Prosodic Prominence Acquisition: Understanding and Placement

Two Transfer Models and three L2 Challenge Models were described in Chapter 1, each with specific predictions for non-native acquisition of pitch accent understanding and placement. The Any Prominence Location Helps Transfer Model proposes that the abstract knowledge of a category, based on experience with this category in their L1, will help a language learner to acquire similar categories in their L2. The Different Prominence Locations Hurt Transfer Model proposes that having categories that are similar, but slightly different, in a language learner's L1 and L2 will lead to the L1 category interfering with the L2 category, making it harder to learn.

The L2 Challenge Models focus on features of a second language that might be particularly difficult to learn, regardless of the learner's L1. Two factors were proposed which may influence the ease with which knowledge of prosodic focus marking is acquired. The Relationship factor states that it is easier to acquire prosodic focus marking for information structures in which there is a more direct relationship between focus and prominence. This factor predicts that it should be easiest to acquire prosodic focus marking for narrow focus sentences because only one word in these sentences is focused, and it is also accented, leading to a direct relationship between focus and accent. The Relationship factor also predicts that it should be easier to acquire prosodic focus marking for VPBF sentences than SBF sentences, because in the VPBF sentences the focused constituent is smaller. In contrast, the Frequency factor states that it is easier to acquire prosodic focus marking for information structures that use common accent patterns, such as putting a final pitch accent on the object in English. Therefore, this factor predicts that prosodic marking of broad focus should be easier to acquire than narrow focus, because broad focus (on sentences with transitive verbs) is marked with a final pitch accent on the object.

In the Relationship L2 Challenge Model, only the Relationship factor affects performance. This would make narrow focus the easiest information structure and sentence broad focus the hardest information structure for non-natives to use in all situations. In the Frequency L2 Challenge Model, only the Frequency factor affects performance, making broad focus easiest in all situations because the prosody used to mark it is so common. In the Hybrid L2 Challenge Model, the two factors interact, so narrow focus is easiest in some contexts and broad focus is easiest in others.

The two main types of models (Transfer Models and L2 Challenge Models) are not contradictory because they make different types of predictions. Transfer Models make predictions about the relative ease with which speakers of different L1s will acquire particular features of an L2. L2 Challenge Models make predictions about which structures in an L2 will be acquired most easily by all language learners. However, different versions of the two model types do make contradictory predictions. The results of the experiments in this dissertation were used to see whether both types of model made accurate predictions, and which version of each model was supported by the data.

I propose that the acquisition of prosodic prominence is best explained by a combination of the Any Prominence Location Helps Transfer Model and the Hybrid L2 Challenge model. A number of the Any Prominence Location Helps Model's predictions were realized in the experimental results. This model correctly predicted that Korean participants would be more accurate than Mandarin participants at determining the correct accent placement for sentences in VPBF and SBF contexts (Exp. 4). The model makes this prediction because broad focus is marked prosodically in Korean, but not in Mandarin. Both Transfer Models also correctly predicted that Korean participants would be more accurate than Mandarin participants at producing pitch accents on the appropriate word for sentences in the SuNF context (Exp. 1). This is because Korean always marks narrow focus prosodically, while prosodic narrow focus marking is optional in Mandarin.

The Hybrid L2 Challenge Model is supported by the fact that there were some tasks for which both groups of non-native participants performed better in the SuNF context than the broad focus contexts and better in the VPBF context than the SBF context, and other tasks for which both groups performed best in the broad focus contexts. I propose that task difficulty is the criterion that determines whether prosodic marking of narrow focus will be easier or harder than broad focus. As predicted by the Relationship factor, non-native participants were more accurate in the SuNF condition than the broad focus conditions in the prominence placement experiment (Exp. 4) and for matched items in the prominence understanding experiment (Exp. 3). They were also more accurate in the VPBF condition than the SBF condition in the prominence placement experiment (Exp. 4). Both the prominence understanding experiment for matched items (Exp. 3) and the prominence placement experiment (Exp. 4) could be considered 'easy' tasks. The non-native participants preferred to accept prosody as appropriate than to reject it as inappropriate in the prominence understanding task (Exp. 3), so the matched items (in which accepting the prosody was the correct response) could be considered easier than the mismatched items. In the prominence placement experiment (Exp. 4), participants had unlimited time to concentrate only on prosodic prominence location, freeing them from the other aspects of language, such as segment pronunciation, which may distract them in more naturalistic tasks.

In contrast to the results discussed in the preceding paragraph, results from harder tasks supported the predictions of the Frequency factor. Non-native participants were more successful in the broad focus conditions than the SuNF condition in the prominence production experiment (Exp. 1) and for mismatched items in the prominence understanding experiment (Exp. 3). Both of these could be considered ‘hard’ tasks. Non-native participants dispreferred rejecting prosody in the prominence understanding task (Exp. 3), perhaps because it was like saying that a native English speaker was making a mistake. This made the mismatched items harder to respond to than the matched items. The prominence production experiment (Exp. 1) required participants to read the answers in question-answer pairs as if they were having an actual conversation. This task was difficult because participants not only have to make accent placement decisions, but also realize that accent placement with suprasegmental features while simultaneously trying to correctly produce the segments that make up each word. Adding to the challenge was the fact that some of the words may have been unfamiliar before the short training session. They also had to role-play half of a conversation with an invisible partner, which is hard even for native speakers. It seems that when a task is simple, learners of English are able to make the connection between prominence location and focus, which is necessary for correct performance on SuNF items. However, when the task is more challenging, learners of English fall back on a commonly-used prominence patterns, and treat it as acceptable in an SuNF context.

One result did not fit into this pattern of the Relationship factor predicting behavior for easy tasks and the Frequency factor predicting behavior for hard tasks. As predicted by the Relationship factor, participants were more accurate in the VPBF condition than the SBF condition for mismatched items in the prominence understanding experiment (Exp. 3), which has been classified as a hard task. One explanation for this anomaly is that the Relationship and Frequency factors do not have conflicting predictions about the relative ease with which participants will acquire prosodic focus marking for VPBF and SBF contexts. The Relationship factor predicts that VPBF will be easier than SBF, but the Frequency factor makes no prediction either way. As a result, participants may generally find focus marking in VPBF contexts easier than SBF contexts, as both VPBF and SBF contexts offer the advantage of a commonly used pitch accent pattern, but VPBF contexts offer the added advantage of a closer relationship between focus and accent placement.

One final predicted result needs to be considered in light of related results that were not predicted by any of the proposed models. Mandarin participants produced more appropriate prosody than Korean participants in the VPBF context. This was predicted by the Different Prominence Locations Hurt Transfer Model because Korean and English make different words prominent in VP broad focus. However, this result may not be reliable because Mandarin participants also produced more appropriate prosody than Korean participants in the SBF context. The SBF result was not predicted by the Different Prominence Locations Hurt Model or any other model. Another surprise was that Mandarin participants produced more appropriate prosody than English participants in the VPBF and SBF contexts. These results, combined with the fact that Mandarin participants produced less appropriate prosody than Korean and English participants in the SuNF context, suggest that Mandarin speakers may be producing such appropriate prosody in the broad focus contexts because they were switching between different patterns of pitch accents less than the Korean and English participants. As a result, they may have produced the broad focus pitch accent pattern more strongly or consistently than the Korean and English participants. This explanation could be tested with an experiment that compared the prosody produced by speakers who were only asked to produce one pitch accent pattern in a recording session to the prosody of speakers who were asked to produce two or three different patterns.

As noted above, Mandarin participants in this study struggled with prominence perception, placement, and production more than Korean participants, even after controlling for proficiency using individuals' Versant scores. There are two possible explanations for this difference. The first, discussed extensively in this dissertation, is that there are some salient differences between Mandarin and Korean, such that transfer of Mandarin features into English is more detrimental than transfer of Korean features. However, another possible explanation might be found in differences between the Chinese and Korean educational systems. This confound is inherent in almost all research comparing groups of language learners with different L1s. Still, it would be interesting to investigate what, if anything, English learners are taught about prosodic prominence in China and Korea. Of course, even if differences in education are found, L1 and educational factors may both have influenced the performance of the participants in this study.

In light of the possible effects of education and life experience on participant performance, future research should test the Transfer Model of perception and realization, and the Any Prominence Location Helps Transfer Model of placement and understanding, with a wider variety of L1s and L2s. For example, the Transfer Model of perception and realization predicts that speakers of languages without lexical tone will more easily perceive and realize English pitch accents than speakers of languages with lexical tone. This prediction could be tested by comparing the performance of English learners with tone language L1s (e.g. Mandarin, Cantonese, Thai, Vietnamese, Ewe, and Igbo) to English learners with non-tone language L1s (e.g. Hindi, Turkish, Finnish, Hungarian, Hebrew, and Persian) on pitch accent perception and realization tasks. The Any Prominence Location Helps Model of placement and understanding predicts that speakers of languages that mark focus prosodically for a particular structure will more easily learn to prosodically mark focus for that structure in English than speakers of languages that do not prosodically mark focus for the structure. Therefore, this prediction could be tested by comparing the performance of English learners with L1s that do not use prosody to mark focus (e.g. Hungarian, Wolof, Buli, Hausa, Sotho-Tswana, and Zulu) to learners with L1s that do use prosody to mark focus (e.g. Akan, Japanese, Romani, and Greek) on pitch accent understanding and placement tasks.

Confirmations of the L2 Challenge Model of perception and realization and the Hybrid L2 Challenge Model of placement and understanding do not require the careful selection of speakers with particular L1s. This is because the models predict that all language learners should be affected by features of L2 prosodic focus marking like the acoustic salience of pitch accent cues, the relationship between focus and prominence, and the frequency with which prominence patterns are used. Therefore, re-running the experiments from the current study using participants with a wider variety of L1s would test the predictions of these models. An interesting extension of this research would examine whether L1 prosodic focus marking acquisition is affected by the same features as L2 acquisition. This could be investigated by having native English speaking children of various ages participate in these experiments.

7.4. Native Perception of Non-Native Pitch Accent Productions

The following question was posed in the introduction: Can an English learner deviate from native-like pitch accent realization and still have his pitch accents accurately perceived by native listeners? The results of the acoustic analysis, combined with native listener judgments, show that the answer is yes. Both native and non-native speakers had productions that were judged to have context-appropriate prosody by at least seven out of eight native listeners. These productions had very similar patterns of pitch accents across native and non-native speakers, indicating that correct pitch accent placement is important for prosodic appropriateness. However, there were some phonetic differences between the prosodically appropriate native and non-native productions. Non-native speakers produced utterances with longer durations, higher F0 maxima, and larger F0 ranges on verbs, and greater RMS amplitudes and lower F0 maxima on objects. They also produced some stronger pitch accent cues than native speakers (longer object durations in VPBF and SBF contexts), and more reduction on unaccented words (smaller pitch ranges on objects in SuNF contexts).

The context-general phonetic differences in the prosody produced by native and non-native speakers may not be perceived by native listeners, or may not affect listener judgments of the context-appropriateness of their prosody. The context-specific differences between native and non-native prosody could make the non-native productions sound more appropriate by making some prosodic distinctions clearer (i.e. the difference between accented and unaccented words). This raises the question of whether non-native productions have to have stronger acoustic cues marking pitch accent than native productions in order to be judged equally appropriate. General prosodic and segmental differences between native and non-native speech may result in fewer listeners judging these productions appropriate. The other possibility is that non-native speakers generally provide stronger prosodic cues marking pitch accents. Further research is required to distinguish between these two alternatives.

The combined results of the native perception of non-native production experiment and the acoustic analyses illustrate an exciting new avenue of research into non-native prosody production. This research program adds the element of listener judgments to the commonly used phonetic measurements and phonological prosodic labels. Listener judgments can help researchers determine which differences between native and non-native productions are important, and which are unimportant.

7.5. Perception and Production in Prosodic Prominence Acquisition

The introduction also posed the question: Does an English learner's ability to accent the appropriate word for a particular information structure and to produce this accent in a native-like manner depend on the ability to accurately perceive and understand such accents? The non-significant perception/production correlations carried out in Chapter 5 did not indicate that pitch accent placement and realization abilities depend on understanding and perception abilities. However, as a null result, this finding is inconclusive. The discussion at the end of Chapter 5 describes a number of experiments that might shed more light on this complex question.

7.6 General Implications for English Prosodic Prominence Education

7.6.1 The Role of L1 in Prosodic Prominence Education

Assuming that L1 transfer plays a part in the differences between native Korean and Mandarin speakers, the most likely cause is the difference between the two languages in their use of pitch and their methods of focus marking. Mandarin uses pitch both lexically and post-lexically, while Korean, like English, uses pitch only post-lexically. Mandarin can mark only narrow focus prosodically, while Korean, like English, can mark both narrow focus and VP broad focus prosodically.

The generally poorer performance of Mandarin participants on the prominence perception task seems most likely to be related to the different uses of pitch in Mandarin, Korean, and English. This suggests that native speakers of tone languages, in which pitch is used lexically, will have greater difficulty perceiving English pitch accents than native speakers of non-tone languages. From a pedagogical perspective, it means that English teachers should focus more on low-level pitch accent perception when teaching students who speak tone languages, such as Mandarin, Cantonese, Thai, Vietnamese, Ewe, and Igbo, to name just a few.

The generally poorer performance of Mandarin speakers on the prominence placement task and their particular difficulty with broad focus items in this task are likely to be due to the different scopes of prosodic focus marking in Mandarin, Korean, and English. This suggests that if a language has the means to mark broad focus prosodically, native speakers of this language will have an advantage when learning to use English prosodic marking of broad focus. Interestingly, this seems to be true even if the language learners would place prominence in a different part of the focused phrase in their L1. From a pedagogical perspective, English teachers whose students speak languages, like Mandarin, that do not mark broad focus prosodically should spend more time teaching the patterns of prosodic prominence used to mark broad focus in English.

The particularly poor performance of Mandarin speakers producing sentences in the SuNF context may be due to the fact that prosodic marking of narrow focus is not required in Mandarin, as it is in Korean and English. As a result, the Korean speakers had a slight advantage when asked to produce a less common pattern of prominence to indicate narrow focus. From a pedagogical perspective, English teachers whose students speak languages that do not require or do not allow prosodic marking of focus (i.e. focus can be marked through syntactic or morphological means alone) should spend more time teaching all levels of focus marking (narrow and broad). This includes students with Mandarin, Hungarian, Wolof, Buli, Hausa, Sotho-Tswana, and Zulu, as native languages.

The one significant difference between the non-native language groups that was not explained in the preceding discussion is Mandarin speakers' slightly more acceptable productions of sentences in the two broad focus conditions. I hypothesized that this result is due to Mandarin speakers being more comfortable producing a commonly used accent pattern across all contexts, because they are not required to prosodically mark narrow focus in their native language. This accent pattern happened to be acceptable for sentences in VP and sentence broad focus. In contrast, Korean speakers may have made slightly more of an effort to produce different prosodic patterns in the different contexts, because they consistently mark narrow focus and VP broad focus prosodically in their native language. This could have led to some productions in the broad focus conditions with prominence incorrectly placed on the subject or verb. From a pedagogical perspective, this highlights the importance of ensuring that learning appropriate prosody for one context does not have a degrading effect on the prosody used in a different context.

7.6.2 The Relationship between Perception and Production in English Prominence Education

The results of these experiments suggest that the ability to produce context-appropriate prosodic prominence in English and the ability to understand it are not closely linked. These results do not necessarily mean that perceptual training could not improve production of prosodic prominence. As this project did not involve any training studies, this remains an open question. Therefore, future work is needed to determine the extent to which such training could improve the use of prosodic prominence across both modalities.

7.6.3 Teaching English Prominence Understanding

Non-native participants in general had less difficulty determining the location of English final pitch accents than understanding the meaning of those accents. This means that education should focus on the connection between prosody and meaning more than the acoustic cues signaling accent location. Non-native participants had difficulty understanding accent placement in broad focus (for the matched items). They also tended to accept accent placement on the object in the SuNF context and accept accents placed anywhere within a focused constituent (for the mismatched items). To correct these mistakes, English learners need to be explicitly taught the correct accent placement for both broad and narrow focus contexts. The two important points to communicate are: 1) When the VP or sentence provides new information, the last pitch accent should go on the object (for sentences with transitive verbs); 2) When only one word provides new information, the last pitch accent should go on that word. Once these two maxims are mastered, English learners should be able to correctly interpret the prosody of sentences in both broad and narrow focus. Practice exercises should involve listening to native speakers produce dialogues that have different types of focus on the target sentence.

7.6.4 Teaching English Prominence Production

Non-native participants, despite their apparent lack of explicit knowledge about pitch accent placement in broad focus contexts, managed to produce broad focus sentences with acceptable prosody at least as often as native English speaking participants. This suggests that the default prosody they were using is acceptable for broad focus sentences with transative verbs. Their performance on this task shows that the bulk of prominence production education should concentrate on narrow focus.

7.7 Conclusions

This dissertation explored second language acquisition of English prosodic focus marking by studying the complete communicative chain, from non-native perception of native production, to non-native production, to native perception of non-native production. The results of five experiments support a predictive framework that combines two types of second language acquisition models. This framework includes a Transfer Model and an L2 Challenge Model for both pitch accent perception and realization, and pitch accent understanding and placement. The Transfer Model for perception and realization predicts that native speakers of non-tone languages will perceive and realize pitch accents more accurately than native speakers of tone languages. The L2 Challenge Model for perception and realization predicts that non-native speakers more easily perceive and realize early pitch accents in narrow focus than late pitch accents in broad focus, at least during difficult tasks. The Transfer Model for understanding and placement predicts that prosodic structures in the L2 will be more easily acquired by language learners that have similar structures in their L1 than those who do not, even if there are differences between the L1 and L2 in how the structures are realized. The L2 Challenge Model for understanding and placement predicts that for hard tasks, language learners will rely on common prosodic patterns, making them more successful at prosodically marking broad focus than narrow focus. However, for easy tasks, language learners will more successfully mark information structures that have a more direct relationship between focus and accent placement, such as narrow focus.

The next step is to expand and clarify this framework by testing it on language learners with a wider variety of L1s and L2s, and on children acquiring their L1. It is also important to see whether the framework can be modified to predict the acquisition of other prosodic features, such as phrasing, and the meanings associated with different types of pitch events (e.g. different pitch accent types in English). More research is needed to determine how prosody perception relates to prosody production, and whether this relationship can also play a role in a more complete framework.

References

'tHart, J. and R. Collier (1975). "Integrating different levels of intonation analysis." Journal of Phonetics **3**: 235-255.

Aiken, L. S. and S. G. West (1996). Multiple regression: Testing and interpreting interactions. Thousand Oaks, CA: Sage.

Akker, E. and A. Cutler (2003). "Prosodic cues to semantic structure in native and nonnative listening." Bilingualism: Language and Cognition **6**(2): 81-96.

Aoyama, K. and S. G. Guion (2007). Prosody in second language acquisition: Acoustic analysis of duration and F0 range. Language Experience in Second Language Speech Learning: In honor of James Emil Flege. O.-S. Bohn. Amsterdam, John Benjamins Publishing Co.**:** 281-297.

Bailey, N., C. Madden, and S. D. Krashen (1974). Is there a "natural sequence" in adult second language learning? *Language Learning,* **24**: 235–243.

Baumann, S. and M. Grice (2006). The intonation of accessibility. *Journal of Pragmatics* **38**: 1636–1657.

Beckman, M. (1986). Stress and non-stress accent. Dordrecht, Foris.

Beckman, M. and G. Ayers Elam (1997). Guidelines for ToBI labeling, Version 3. Ohio State University.

Beckman, M. and J. Pierrehumbert (1986). "Intonational structure in English and Japanese." Phonology Yearbook **3**: 255-310.

Bernstein, J. and J. Cheng (2008). Logic and validation of a fully automatic spoken English test. The Path of Speech Technologies in Computer Assisted Language Learning. V. M. Holland and F. P. Fisher. New York, Routledge: 174-194.

Best, C. T., G. W. McRoberts, et al. (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system." The Journal of the Acoustical Society of America **109**(2): 775-794.

Birch, S. and C. Clifton (1995). Focus, accent, and argument structure: Effects on language comprehension. Language and Speech **38**(4): 365-391.

Boersma, P. and D. Weenink (2009). Praat: doing phonetics by computer (Version 5.1.20) [Computer Program]. Retrieved 10/31/2009, from http://www.praat.org/

Bradlow, A. R., R. Akahane-Yamada, et al. (1999). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production."

Perception and Psychophysics **61**(5): 977-985.

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

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

Büring, D. (2003). "On D-Trees, Beans, and B-Accents." Linguistics and Philosophy **26**(5): 511-545.

Büring, D. (2006). "Focus projection and default prominence." In Valéria Molnár and Susanne Winkler (eds) *The Architecture of Focus*. Berlin/New York: Mouton De Gruyter.

Büring, D. (2009). "Towards a Typology of Focus Realization". In Malte Zimmermann & Caroline Féry (eds) *Information Structure*. OUP. 177-205.

Celik, M. (2005). Teaching English intonation to EFL/ESL students. In Meenakshi Raman (ed) *English Language Teaching*. New Dehli: Atlantic Publishers and Distributers.

Chafe, W. (1994). Discourse, Consciousness and Time. University of Chicago Press, Chicago.

Chun, D., M. (2002). Discourse Intonation in L2: From Theory and Research to Practice, John Benjamins Publishing Co.

Coburn, C. V. (2000). Individual differences in intonation use and production-perception link. Linguistics, Northwestern University. **M.A.**

de Bot, K. and K. Mailfert (1982). "The teaching of intonation: Fundamental research and classroom applications." TESOL Quarterly **16**(1): 71-77.

Derwing, T. M. and M. J. Munro (1997). "Accent, intelligibility, and comprehensibility: Evidence from four L1s." Studies in Second Language Acquisition **19**: 1-16.

Dulay, H. C., and Burt, M. K. (1973). Should we teach children syntax? Language Learning **23**: 245–258.

Dulay, H. C., and M. K. Burt (1974). Natural sequences in child second language acquisition. *Language Learning,* **24**: 37–53.

Dupoux, E., N. Sebastian-Galles, et al. (2007). "Persistent stress 'deafness': The case of French learners of Spanish." Cognition **106**(2): 682-706.

Eady, S. J., W. E. Cooper, G. V. Klouda, P. R. Mueller, and D. W. Lotts (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. Language and Speech **29**(3): 233-251.

Fathman, A. (1975). Language background, age and the order of acquisition of English structures. In M. K. Burt and H. C. Dulay (Eds.), *On TESOL '75: New directions in second language learning, teaching and bilingual education* (pp. 33–43).Washington, DC: TESOL.

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

Flege, J. E., M. J. Munro, et al. (1995). "Effects of age of second language learning on the production of English consonants." Speech Communication **16**(1): 1-26.

Francis, A. L. and H. C. Nusbaum (2002). "Selective attention and the acquisition of new phonetic categories." Journal of Experiment Psychology: Human Perception and Performance **28**: 349-366.

German, J., J. Pierrehumbert and S. Kaufmann (2006). Evidence for phonological constraints on nuclear accent placement. *Language* **82**(1): 151-168.

German, J. (2009). Prosodic strategies for negotiating reference in discourse. Evanston, Northwestern. **Ph.D.**

Goldsmith, J. A. (1976). Autosegmental phonology. Linguistics. Cambridge, MIT. **Ph.D**.

Goldscheider, J. M. and R. M. DeKeyser (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. Language Learning **51**(1): 1–50.

Gordon, M. (2008). The intonational realization of contrastive focus in Chickasaw. In Chungmin Lee, Matthew Gordon, and Daniel Büring (eds) *Topic and Focus: Cross-linguistic Perspective on Meaning and Intonation*. Dordrecht: Springer.

Gussenhoven, C. (1983a). Focus, mode, and the nucleus. *Journal of Linguistics* **19**(2): 377-417.

Gussenhoven, C. (1983b). Testing the reality of focus domains. *Language and Speech* **26**(1): 61-80.

Gussenhoven, C. (1999). On the limits of focus projection in English. in Peter Bosch and Rob van der Sandt (eds) *Focus: Linguistic, Cognitive, and Computational Perspectives.* Cambridge: Cambridge University Press, 43-55.

Hammarberg, B. (1990). Conditions on transfer in second language acquisition. New Sounds 90, Proc. of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech., University of Amsterdam.

Hoskins, S. (1996). A phonetic study of focus in intransitive verb sentences. Proceedings ICSLP'96. Philadelphia, USA.

Jaeger, T. F. (2008). "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models." Journal of Memory and Language **59**: 434–446.

Jun, S.-A. (1998). "The accentual phrase in the Korean prosodic hierarchy*." Phonology **15**(2): 189-226.

Jun, S.-A. (2002). Syntax over focus. Seventh International Conference on Spoken Language Processing**:** 2281-2289.

Jun, S.-A. (2005). Prosody in sentence processing: Korean vs. English. UCLA Working Papers in Phonetics. **104:** 26-45.

Jun, S.-A., H.-S. Kim, et al. (2006). "An experimental study on the effect of argument structure on VP focus." Korean Linguistics **13**: 89-112.

Jun, S.-A. and H.-J. Lee (1998). Phonetic and phonological markers of contrastive focus in Korean. Fifth International Conference on Spoken Language Processing. Sydney, Australia.

Jun, S.-A. and M. Oh (2000). "Acquisition of second language intonation." Proceedings of the International Conference on Spoken Language Processing, Vol. 4: 76-79. Beijing, China.

Kenesei, I. (2009). "Quantifiers, negation, and focus on the left periphery in Hungarian." Lingua **119**: 564-591.

Krashen, S. D., V. Sferlazza, L. Feldman, and A.K. Fathman (1976). Adult performance on the SLOPE test: More evidence for a natural sequence in adult second language acquisition. *Language Learning,* **26**: 145–151.

Ladd, D. R. (1996). Intonational Phonology. Cambridge, UK, Cambridge University Press.

Lambrecht, K. (1994). Information Structure and Sentence Form. Cambridge University Press, Cambridge.

Laures, J. S. and G. Weismer (1999). "The effects of a flattened fundamental frequency on intelligibility at the sentence level." Journal of Speech, Language, and Hearing Research **42**: 1148-1156.

Lee, I. and S. R. Ramsey (2000). The Korean Language. Albany, NY, State University of New York Press.

Lee, Y.-C. and Y. Xu (2010). Phonetic realization of contrastive focus in Korean. Speech

Prosody 2010, Chicago, USA.

Li, C. N. and S. A. Thompson (1989). Mandarin Chinese: A Functional Reference Grammar. Berkeley, University of California Press.

Liberman, M. and A. Prince (1977). "On stress and linguistic rhythm." Linguistic Inquiry **8**: 249-336.

Liu, F. and Y. Xu (2005). "Parallel encoding of focus and interrogative meaning in Mandarin intonation." Phonetica **62**: 70-87.

McGory, J. T. (1997). The acquisition of intonation patterns in English by native speakers of Korean and Mandarin. Columbus, OH, The Ohio State University. **ph. D Dissertation**.

Mennen, I. (1998). Second language acquisition of intonation: The case of peak alignment. Chicago Linguistic Society 34, vol. II: The panels.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. Journal of Phonetics **32**: 543-563.

Munro, M. J. and T. M. Derwing (1995). "Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech." Language and Speech **38**: 289-306.

Nava, E. (2007). Prosody in L2 acquisition. 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007), Iowa City, Iowa, Cascadilla Proceeding Project.

Nava, E. and M. L. Zubizarreta (2010). Deconstructing the Nuclear Stress Algorithm: Evidence from second language speech. The Sound Patterns of Syntax. N. Erteschik-Shir and L. Rochman. Oxford, Oxford University Press.

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. Linguistics. Cambridge, MIT. **PhD**.

Pierrehumbert, J. and J. Hirschberg (1990). The Meaning of Intonation in the Interpretation of Discourse, in P. Cohen, J. Morgan, and M. Pollack, (eds) Intentions in Communication, MIT Press, Cambridge MA. 271-311.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. Speech Perception and Linguistic Experience: Issues in Cross-Language Research. W. Strange. Baltimore, York Press**:** 379-410.

Rooth, M. (1992). "A Theory of Focus Interpretation." Natural Language Semantics **1**(1): 75-116.

Rosenberg, A., J. Hirschberg, and K. Manis (2010). Perception of English prominence by native Mandarin Chinese speakers. Speech Prosody 2010, Chicago, USA.

Schack, K. (2000). Comparison of intonation patterns in Mandarin and English for a particular speaker. U. of Rochester Working Papers in the Language Sciences. Rochester, NY.

Schwarzschild, R. (1999). "GIVENness, AvoidF and other Constraints on the Placement of Accent." Natural Language Semantics **7**(2): 141-177.

Selkirk, E. (1996). Sentence Prosody: Intonation, Stress and Phrasing. In The Handbook of Phonological Theory. J. A. Goldsmith. London, Basil Blackwell.

Shue, Y.-L., S. Shattuck-Hufnagel, M. Iseli, S.-A. Jun, N. Veilleux, and A. Alwan (2010) On the acoustic correlates of high and low nuclear pitch accents in American English. Speech Communication **52**: 106: 122.

Shue, Y.-L., M. Iseli, N. Veilleux, and A. Alwan (2007). Pitch accent versus lexical stress: Quantifying acoustic measures related to the voice source. Proceedings of Interspeech 2007: 2625-2628. Antwerp, Belgium.

Shyu, S.-I. (1995). The syntax of focus and topic in Mandarin Chinese. Linguistics, University of Southern California.

Sohn, H.-M. (1999). The Korean Language. Cambridge, Cambridge University Press.

Stanislaw, H. and N. Todorov (1999). Calculation of signal detection theory measures. Behavior Research Methods, Instruments, & Computers **31**(1): 137-149.

Steedman, M. (2007). Information-Structural Semantics for English Intonation. Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation. C. Lee, M. Gordon and D. Büring. Dordrecht, the Netherlands, Springer **82:** 245-264.

Swan, M. and B. Smith (2001). Learner English: A teacher's guide to interference and other problems. Cambridge, Cambridge University Press.

Swerts, M. and E. Krahmer (2008). "Facial expressions and prosodic prominence: Comparing modalities and facial areas." Journal of Phonetics **36** (2): 219-238.

Terken, J. M. B. and S. Nooteboom (1988). "Opposite effects of accentuation and deaccentuation on verification latencies for given and new information." Language and Cognitive Processes **2**: 145-163.

Truckenbodt, H. (1995). Phonological Phrases: Their relation to syntax, focus, and prominence, MIT.

Verdugo, D. R. (2006). "Prosodic realization of focus in the discourse of Spanish learners and English native speakers." World Englishes **14**: 9-32.

Wang, B. and Y. Xu (2006). Prosodic encoding of topic and focus in Mandarin. Speech Prosody. Dresden.

Wang, J. (2003). "Intonation and communication of meaning: A study of the effect of intonation on the speakers' communicative performance." US-China Foreign Language: 20-28.

Wei, L. (2000). Unequal election of morphemes in adult second language acquisition. Applied Linguistics **21**(1): 106-140.

Welby, P. (2003). "Effects of pitch accent type and status on focus projection." Language and Speech **46**(1): 53-81.

Wennerstrom, A. (1994). "Intonational meaning in English discourse: A study of non-native speakers." Applied Linguistics **15**: 399-420.

Xu, L. (2004). "Manifestation of informational focus." Lingua **114**(3): 277-299.

Xu, Y. (1999). "Effects of tone and focus on the formation and alignment of f0 contours." Journal of Phonetics **27**: 55-105.

Xu, Y., C. X. Xu, et al. (2004). On the temporal domain of focus. Speech Prosody 2004. Nara, Japan.

Xu, Y. and C. X. Xu (2005). Phonetic realization of focus in English declarative intonation. Journal of Phonetics **33** (2): 159-197.

Yeou, M. (2004). The realization of accentual focus in Moroccan learners of English. Speech Prosody 2004. Nara, Japan.

Zhang, S., K. Li, W.-K. Lo, and H. Meng (2010). Perception of English suprasegmental features by non-native Chinese learners. Speech Prosody 2010. Chicago, USA.

Appendix A: Experiment Stimulus Sentences

Prominence Perception Experiment:
1 syllable -
John tried some wine.
Luke watched the moon.
Ben rubbed his leg.
Dan greeted Lloyd.
Kim bought a fan.
Russ mowed his lawn.
Lon hid a key.
Lynn hugged Don.
Fay fixed a chain.
Jean carried a chair.
Tim cut a lime.
Lou punished Mo.

3 syllable -
Diana played the viola.
Natasha mentioned Korea.
Melinda served some salami.
Siena drank her Chianti.
Tobias wrote a sonata.
Teresa purchased a casino.
Evita cleaned a marina.
Anita left Chicago.
Latisha sold a kimono.
Miranda paid her admission.
Domingo baked a potato.
Melissa bypassed Toledo.

Prominence Understanding Experiment
1 syllable -
Dean grew a vine.
Wayne stole a car.
Jane cooked some rice.
Dawn took a ring.
Rick hit Wade.
Nieve painted a lake.
Reese opened a can.
Ross wiped his chin.
Ray called Joan.
Shane bent his knee.
Mac studied the law.
Roy hired Ken.

3 syllable -
Selina fed her canary.
Adina helped Vanessa.
Katinka followed the tornado.
Serena solved a dilemma.
Sophia picked a papaya.
Naomi saved a koala.
Alisha sliced a tomato.
Bianca did some addition.
Joanna planned her agenda.
Maria pictured Kentucky.
Alanis toured Morocco.
Diego rented a tuxedo.

Prominence Production and Placement Experiments
1 syllable -
Ron shouted his name.
Ann crossed the road.
Lee chose Maine.
Jan drew a line.
Jeff dropped a knife.
Rod measured a lane.
Rob made some jam.
Len scratched his shin.
Sam phoned Maud.
Ned told a lie.
May answered Sue.
Dane moved a rug.

3 syllable -
Nikita smoked some tobacco.
Elena ate her baloney.
Ramona mixed a martini.
Tameka sued her attorney.
Jemima filmed a volcano.
Oksana entered Miami.
Elijah skipped the audition.
Fiona led a committee.
Amanda researched Jamaica.
Alexis wore a bikini.
Rebecca damaged a pagoda.
Dakota fled Malawi.

Appendix B: Experiment Instructions

Prominence Production Experiment

In this experiment, you will be reading the answers to a series of questions. You will see both the question and the answer on your screen. You will be able to hear someone asking the question by clicking on a button (that looks like this:) next to it.

When you have heard the question, you should read the answer out loud, as naturally and fluently as possible, as if you were having a real conversation. Try to keep the question you are answering in mind as you read the answer. If you make a mistake, just say the answer again. If you don’t know how to pronounce a name or word, that’s OK, just make a guess.

To move on to the next slide, hit the space bar on the keyboard. If you accidentally skip a slide, hit the ‘page up’ button to go back.

If you have any questions, ask the experimenter now. Otherwise hit the space bar to see a practice slide.

Practice Slide:
What did Robert do with the apple?
Robert sold the apple.

Prominence Perception Experiment

In English, some words are pronounced in a way that makes them sound more important or prominent than others. A ‘prominent’ word stands out when you hear it. It may have noticeable intonation or may be especially long. A sentence can have more than one prominent word in it, but today we’re interested in the LAST PROMINENT word in a sentence.

Click on the button labeled ‘PLAY Recording 1’ to hear an example sentence with one word that is more prominent than the others. Click on the button labeled 'PLAY Recording 2' to hear an example sentence with two prominent words. You can listen to them a couple of times if you like.[12]

In Recording 1, the only prominent word is ‘wrote’, so 'wrote' is the last prominent word.
In Recording 2, both 'Adam' and 'book' are prominent, but 'book' is the last prominent word.

[12] Both recordings were of the sentence “Adam wrote a book.” Recording 1 had a pitch accent on *wrote*. Recording 2 had pitch accents on *Adam* and *book*.

In this experiment, you'll see a sentence on your screen. You can click on the 'PLAY' button to hear someone saying the sentence. Below the sentence you'll see a question asking whether a particular word in the sentence is the last prominent word. When you have decided, click 'Yes' or 'No'. If you are not sure, make your best guess. You can play the recording more than once if that will help you.

You will be hearing 96 sentences in all. Work at your own pace, and feel free to take a break at any point.

Prominence Understanding Experiment

The word 'prosody' refers to the way that sentences are spoken. This includes things like the intonation and rhythm of words in a sentence. The prosody of a sentence can give information about what the sentence means, and different prosodies are appropriate in different contexts.
If someone asked you a question, you could produce the same answer with different prosodies. Recording 1 and Recording 2 both contain the question "How did Eric get home?" In both recordings the answer to this question is "Eric flew home." but the answers have different prosodies in the two recordings. Click on the buttons labeled 'PLAY Recording 1' and 'PLAY Recording 2' to hear the differences. You can listen to each of them a couple of times if you like.[13]

The answer in Recording 1 has appropriate prosody for a sentence answering the question "How did Eric get home?", while the answer in Recording 2 has inappropriate prosody for answering this question.

In this experiment you'll see a series of question-answer pairs on your screen. You should click on the 'PLAY' button to hear the question-answer pair being spoken. The prosody of the answer may be appropriate for an answer to that question, or it may be inappropriate. You'll be asked whether you think the prosody is appropriate for answering the question. When you have decided, click 'Yes' or 'No'. If you are not sure, make your best guess. You can play the recording more than once if that will help you.

You will be hearing 72 question-answer pairs in all. Work at your own pace, and feel free to take a break at any point.

Prominence Placement Experiment

In an earlier experiment, you learned about prominent words in English. Recall that, in English, some words are pronounced in a way that makes them sound more important or prominent than

[13] Both recordings were of the dialogue "How did Eric get home? Eric flew home." The answer in Recording 1 had a pitch accent on *flew*. The answer in Recording 2 had a pitch accent on *Eric*.

others. A ‘prominent’ word stands out when you hear it. It may have noticeable intonation or may be especially long.

In this experiment you’ll see a series of question-answer pairs on your screen. You won't be listening to anything. Instead, you’ll be asked about how you would produce the answer to the question. Specifically, you will be asked which word in the answer, if any, you would make the most prominent. When you have decided, click on that word. If you would make all the words equally prominent, click on 'NONE'. If you are not sure, make your best guess.

You will be reading 72 question-answer pairs in all. Work at your own pace, and feel free to take a break at any point.

Native Perception of Non-Native Production Experiment

The word ‘prosody’ refers to the way that sentences are spoken. This includes things like the intonation and rhythm of words in a sentence. The prosody of a sentence can give information about what the sentence means, and different prosodies are appropriate in different contexts.
If someone asked you a question, you could produce the same answer with different prosodies. Recording 1 and Recording 2 both contain the question "How did Eric get home?" In both recordings the answer to this question is "Eric flew home." but the answers have different prosodies in the two recordings. Click on the buttons labeled ‘PLAY Recording 1’ and ‘PLAY Recording 2’ to hear the differences. You can listen to each of them a couple of times if you like.[14]

The answer in Recording 1 has appropriate prosody for a sentence answering the question “How did Eric get home?”, while the answer in Recording 2 has inappropriate prosody for answering this question.

In this experiment you’ll see a series of question-answer pairs on your screen. You should click on the ‘PLAY’ button to hear the question-answer pair being spoken. The prosody of the answer may be appropriate for an answer to that question, or it may be inappropriate. You’ll be asked whether you think the prosody is appropriate for answering the question. When you have decided, click ‘Yes’ or ‘No’. If you are not sure, make your best guess. You can play the recording more than once if that will help you.

You will see the same question-answer pairs repeated many times, sometimes twice in a row. The second one is a different item, with an answer spoken by a different speaker. This experiment includes sentences produced by both native and non-native English speakers.

---

[14] Both recordings were of the dialogue “How did Eric get home? Eric flew home.” The answer in Recording 1 had a pitch accent on *flew*. The answer in Recording 2 had a pitch accent on *Eric*.

Sometimes these speakers make mistakes when pronouncing words, but please try to overlook these mistakes and focus on the appropriateness of the sentence prosody.

You will be hearing 360 question-answer pairs in all. Work at your own pace, and please take a break at any point if you are getting tired.