1-1-2015

Prosodic Focus Within and Across Languages

Yong-Cheol Lee
University of Pennsylvania, soongdora@gmail.com

Follow this and additional works at: http://repository.upenn.edu/edissertations
Part of the Linguistics Commons

Recommended Citation
http://repository.upenn.edu/edissertations/1831

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/1831
For more information, please contact libraryrepository@pobox.upenn.edu.
Prosodic Focus Within and Across Languages

Abstract
The fact that “purely” prosodic marking of focus may be weaker in some languages than in others, and that it varies in certain circumstances even within a single language, has not been commonly recognized. Therefore, this dissertation investigated whether and how purely prosodic marking of focus varies within and across languages. We conducted production and perception experiments using a paradigm of 10-digit phone-number strings in which the same material and discourse contexts were used in different languages.

The results demonstrated that prosodic marking of focus varied across languages. Speakers of American English, Mandarin Chinese, and Standard French clearly modulated duration, pitch, and intensity to indicate the position of corrective focus. Listeners of these languages recognized the focus position with high accuracy. Conversely, speakers of Seoul Korean, South Kyungsang Korean, Tokyo Japanese, and Suzhou Wu produced a weak and ambiguous modulation by focus, resulting in a poor identification performance.

This dissertation also revealed that prosodic marking of focus varied even within a single language. In Mandarin Chinese, a focused low/dipping tone (tone 3) received a relatively poor identification rate compared to other focused tones (about 77% vs. 91%). This lower identification performance was due to the smaller capacity of tone 3 for pitch range expansion and local dissimilatory effects around tone 3 focus. In Seoul Korean, prosodic marking of focus differed based on the tonal contrast (post-lexical low vs. high tones). The identification rate of high tones was twice as high than that of low tones (about 24% vs. 51%), the reason being that low tones had a smaller capacity for pitch range expansion than high tones.

All things considered, this dissertation demonstrates that prosodic focus is not always expressed by concomitant increased duration, pitch, and intensity. Accordingly, “purely” prosodic marking of focus is neither completely universal nor automatic, but rather is expressed through the prosodic structure of each language. Since the striking difference in focus-marking success does not seem to be determined by any previously-described typological feature, this must be regarded as an indicator of a new typological dimension, or as a function of a new typological space.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Linguistics

First Advisor
Mark Liberman

Keywords
Focus, Identification, Perception, Production, Prosody, Typology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/1831
Subject Categories
Linguistics

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/1831
PROSODIC FOCUS WITHIN AND ACROSS LANGUAGES

Yong-cheol Lee

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

__________________________
Mark Liberman
Christopher H. Browne Distinguished Professor of Linguistics

Graduate Group Chairperson

__________________________
Eugene Buckley
Associate Professor of Linguistics

Dissertation Committee

Florian Schwarz, Assistant Professor of Linguistics
Jianjing Kuang, Assistant Professor of Linguistics
PROSODIC FOCUS WITHIN AND ACROSS LANGUAGES

COPYRIGHT

2015

Yong-cheol Lee
By His grace
Acknowledgements

I have always waited for this moment to come – the moment dedicated to thanking all my teachers, colleagues, friends, and families for the great support and prayers they have poured into me during my PhD life. It has been a long, tough journey, but it would have been even harder without them. If I missed any names below, please kindly forgive me – you know you are greatly appreciated.

First and foremost, I would like to express my deepest gratitude to my advisor, Mark Liberman, for being a great source of inspiration and guidance. When I came to him with undeveloped ideas, he pointed me in the right direction and helped me formulate my raw ideas into coherent and articulate ones. During our meetings, he always provided me with brilliant ideas far beyond my capacity and constructive suggestions for this dissertation as well as other projects. Without him, my dissertation would not have been possible.

In addition, many special thanks go to committee members Florian Schwarz and Jianjing Kuang, who willingly dedicated time for our regular meetings and continued to provide me with constructive feedback and suggestions on my dissertation throughout the time I was working on it. They both challenged and stimulated me intellectually, pushing me to go beyond what I had prepared for, which ultimately led me to think more critically about this project.
I have had many great teachers over the years since I started my career in linguistics. Jiahong Yuan and Georgia Zellou were generous and kind to me and were always willing to share their knowledge. Their influence had a positive impact during my PhD study. Yi Xu and Sun-Ah Jun also deserve a mention. Although they are not Penn faculty members, they helped me outside of school through email correspondence. I am also very grateful to my Korean teachers and mentors, Kyutae Jung, Misun Seo, and Yong-hun Lee, who always inspired me to go beyond my limits and encouraged me with supportive words when I felt overwhelmed.

I also owe many thanks to my cohort-mates: Meredith Tamminga, Hilary Prichard, Kobey Shwayder, and Claire Crawford. As an international student, I was confronted with many obstacles both academically and socially when I first started studying at Penn. Through those tough times, they helped me overcome hardships both inside and outside of class. We share many memories of doing homework assignments and projects together and having regular meetings throughout the years.

I have been very fortunate to meet many wonderful people at Penn. I would like to single out Marielle Lerner for special thanks. She has played many roles for me throughout our friendship, including English teacher, proofreader, and amazing friend. I also want to express special thanks to Gayeon Son, Soohyun Kwon, and Sunghye Cho, who have provided me with endless encouragement, support, and love. Believe it or not, I often got the feeling that they bought me meals just to make sure that I wasn’t skipping any. It was a great joy to have such friends at Penn. Furthermore, I must also mention my co-workers in this line of appreciation: Satoshi Nambu and Ting Wang. By having regular
meetings, we discussed and shared many ideas on various projects and developed those initial ideas into presentable and publishable ones.

For the past six years, my colleagues and friends inside and outside the department provided me with a friendly and pleasant environment and also offered me so much help. The following people I would like to thank are alphabetically ordered by their first name: Aaron Ecay, Amy Forsyth, Anton Ingason, Aviad Eilam, Betsy Sneller, Blenda Im, Brittany McLaughlin, Catherine Lai, Chang Woon Jang, Chris Ahern, Chris Park, Constantine Lignos, Dimka Atanassov, Einar Freyr Sigurðsson, EunHae Grace Kim, Haitao Cai, Hannah Gajari, Ingrid Heidelberger, Jae Kim, Jane Kim, Jeehwan Cho, Ji Yea Kim, Jingjing Tan, Jinmin Lee, Jon Scott Stevens, Josep Fruehwald, Josep Folta, Joshua Tauberer, Karen Tseng, Katelyn HyunJin Cheon, Laurel MacKenzie, Mao-Hsu Chen, Marielle Lerner, Michael Falcetano, Rebecca Tomkinson Marsh, Rinho Kim, Robert Wilder, Sabriya Fisher, Saem Kim, SangHoon Chong, Yanyan Sui, Yixuan Guo, Young Ah Do, Young Hwang, and Yusuke Yoda.

I have had many unforgettable memories in Philadelphia. One of the best times came with my life at Brown Street. My landlady Doris Blendinger was always concerned about my career and success in the field of linguistics and in my personal life in general. She was also willing to proofread many pieces of my work on this project and others with greatest pleasure. Tyrell McCurbin and Ravi Akki, who were my house mates, also deserve special thanks. We had great times together with many interesting conversations and wonderful dinners. This precious time will last forever in my memory!

I also want to thank my family from the bottom of my heart. My wife, Sora Kang, who actually tops the list along with my advisor, supported me in everything from A to Z.
She was always there for me day in and day out. Words simply cannot describe the level of my gratitude towards her for her sacrifice and love for me. Thanks a lot honey! In addition, I sincerely thank both my parents and parents-in-law for their ongoing support and prayers for me.

Needless to say, special thanks go to both the Fulbright Fellowship, the Benjamin Franklin Fellowship, and the George L. Harrison Fellowship, through which I have been able to focus on my PhD courses, a series of research projects, and of course this dissertation without any financial hassles.

Last but not least, parts of this project are based on joint work with many other scholars. Without their help, this dissertation would never have reached the finish line. Therefore, I extend my sincere thanks to all my collaborators: Ting Wanag at Tongji University, Sunghye Cho at Penn, Bei Wang and Sisi Chen at Minzu University of China, Martine Adda-Decker and Angélique Amelot at Sorbonne Nouvelle, and Satoshi Nambu at National Institute for Japanese Language and Linguistics. A preliminary version of Chapter 4 was presented at ICASSP 2015 and is forthcoming in the proceedings. Earlier versions of Chapters 5 and 6 were presented at LSA 2015 and PLC 39, respectively. Parts of this dissertation were also presented at ETAP 3.
ABSTRACT

PROSODIC FOCUS WITHIN AND ACROSS LANGUAGES

Yong-cheol Lee
Mark Liberman

The fact that “purely” prosodic marking of focus may be weaker in some languages than in others, and that it varies in certain circumstances even within a single language, has not been commonly recognized. Therefore, this dissertation investigated whether and how purely prosodic marking of focus varies within and across languages. We conducted production and perception experiments using a paradigm of 10-digit phone-number strings in which the same material and discourse contexts were used in different languages.

The results demonstrated that prosodic marking of focus varied across languages. Speakers of American English, Mandarin Chinese, and Standard French clearly modulated duration, pitch, and intensity to indicate the position of corrective focus. Listeners of these languages recognized the focus position with high accuracy. Conversely, speakers of Seoul Korean, South Kyungsang Korean, Tokyo Japanese, and Suzhou Wu produced a weak and ambiguous modulation by focus, resulting in a poor identification performance.

This dissertation also revealed that prosodic marking of focus varied even within a single language. In Mandarin Chinese, a focused low/dipping tone (tone 3) received a relatively poor identification rate compared to other focused tones (about 77% vs. 91%).
This lower identification performance was due to the smaller capacity of tone 3 for pitch range expansion and local dissimilatory effects around tone 3 focus. In Seoul Korean, prosodic marking of focus differed based on the tonal contrast (post-lexical low vs. high tones). The identification rate of high tones was twice as high than that of low tones (about 24% vs. 51%), the reason being that low tones had a smaller capacity for pitch range expansion than high tones.

All things considered, this dissertation demonstrates that prosodic focus is not always expressed by concomitant increased duration, pitch, and intensity. Accordingly, “purely” prosodic marking of focus is neither completely universal nor automatic, but rather is expressed through the prosodic structure of each language. Since the striking difference in focus-marking success does not seem to be determined by any previously-described typological feature, this must be regarded as an indicator of a new typological dimension, or as a function of a new typological space.
## Contents

**Acknowledgements** ........................................ iv  

**Abstract** .................................................................. viii  

**Contents** ............................................................. x  

**List of Tables** .......................................................... xiv  

**List of Figures** ....................................................... xvii  

1. **Introduction** ....................................................... 1  
   1.1. Focus types .......................................................... 2  
   1.2. Some terminology issues ......................................... 4  
   1.3. Three basic approaches to focus marking ....................... 8  
   1.4. Current study .................................................. 13  
   1.5. Research questions and approaches ............................. 18  
   1.6. Outline of this dissertation ..................................... 20  

2. **Methods – Spoken telephone numbers** ....................... 22  
   2.1. Production ..................................................... 25  
   2.1.1. Speech materials ........................................... 25  
   2.1.2. Participants ............................................... 27  
   2.1.3. Recording procedure ...................................... 28  
   2.2. Perception ..................................................... 31  
   2.2.1. Audio stimuli .............................................. 31
2.2.2. Participants ...........................................................................................................32
2.2.3. Procedure ............................................................................................................32

3. Prosodic structures of languages under study .........................................................35
   3.1. Prosodic structures of each language .................................................................35
       3.1.1. American English .......................................................................................35
       3.1.2. Mandarin Chinese .......................................................................................36
       3.1.3. Seoul Korean ...............................................................................................37
       3.1.4. South Kyungsang Korean ..........................................................................39
       3.1.5. Tokyo Japanese .........................................................................................41
       3.1.6. Standard French .........................................................................................44
       3.1.7. Suzhou Wu .................................................................................................45
   3.2. Three types of prosodic typologies .................................................................48
       3.2.1. Rhythm-based typology ............................................................................48
       3.2.2. Accent-based typology ..............................................................................50
       3.2.3. Prominence-based typology .................................................................53

4. A cross-linguistic study of prosodic focus ..........................................................57
   4.1. Introduction .........................................................................................................57
   4.2. Production ..........................................................................................................60
       4.2.1. Data collection .........................................................................................60
       4.2.2. Acoustic measurements ............................................................................61
       4.2.3. Analyses ....................................................................................................61
       4.2.4. Results .....................................................................................................63
           4.2.4.1. American English and Seoul Korean .................................................63
           4.2.4.2. Mandarin Chinese and South Kyungsang Korean .........................65
           4.2.4.3. Standard French and Tokyo Japanese .............................................67
           4.2.4.4. Suzhou Wu .......................................................................................69
           4.2.4.5. The phonological unit that carries prosodic focus .......................71
   4.3. Perception .........................................................................................................74
4.3.1. Data collection and analyses .................................................................74
4.3.2. Results .............................................................................................................................74
  4.3.2.1. American English and Seoul Korean .................................................................74
  4.3.2.2. Mandarin Chinese and South Kyungsang Korean ............................................77
  4.3.2.3. Standard French and Tokyo Japanese .................................................................81
  4.3.2.4. Suzhou Wu ...........................................................................................................83
4.4. Discussion ......................................................................................................................84

5. Production and perception of tone 3 focus in Mandarin Chinese ........................................ 93
  5.1. Introduction .....................................................................................................................93
    5.1.1. A brief overview of four lexical tones in Mandarin Chinese .................................94
    5.1.2. Production and perception of focus in Mandarin Chinese .................................95
    5.1.3. The current study .....................................................................................................97
  5.2. Production ....................................................................................................................99
    5.2.1. Stimuli ...................................................................................................................99
    5.2.2. Subjects and recording procedure .......................................................................100
    5.2.3. A sketch of pitch contours ................................................................................100
    5.2.4. Acoustic measurements ......................................................................................103
    5.2.5. Analyses and results ............................................................................................104
  5.3. Perception ....................................................................................................................110
    5.3.1. Data collection ......................................................................................................110
    5.3.2. Results ................................................................................................................111
  5.4. Discussion ...................................................................................................................113

6. An interaction effect between tonogenesis and prosodic focus in Seoul Korean .............................. 117
  6.1. Introduction ..................................................................................................................117
  6.2. Production ..................................................................................................................121
    6.2.1. Stimuli ................................................................................................................121
    6.2.2. A sketch of pitch contours .................................................................................122
6.2.3. Analyses and results ............................................................. 124
6.3. Perception ............................................................................. 128
  6.3.1. Data collection ................................................................. 128
  6.3.2. Results ............................................................................. 129
6.4. Discussion ............................................................................. 132

7. General discussion, future directions, and conclusion 136
  7.1. Main findings ....................................................................... 136
  7.2. Limitations of existing prosodic typologies ......................... 141
  7.3. Implications ......................................................................... 144
  7.4. Future directions ................................................................. 147
  7.5. Conclusion ............................................................................ 153

Appendix A ..................................................................................... 154

Appendix B ..................................................................................... 155

Appendix C ..................................................................................... 157

Appendix D ..................................................................................... 159

Bibliography .................................................................................. 161
List of Tables

1.1. Focus morphemes in Chickasaw .................................................................11

2.1 Different pronunciations of the numerical digits from 0 to 9 in each language. We used the IPA for the two varieties of Korean, the Romaji system for TJ, the Pinyin Romanization system for Mandarin Chinese, and the IPA for Suzhou Wu. (AE: American English, MC: Mandarin Chinese, SF: Standard French, SK: Seoul Korean, SKK: South Kyungsang Korean, TJ: Tokyo Japanese, SW: Suzhou Wu)....

2.2 The number of speakers and the mean age (SD) of each language ..................28

2.3 The number of listeners and the mean age (SD) of each language ....................32

3.1. The possible tone patterns of monosyllabic, disyllabic, trisyllabic, and quadrisyllabic words in South Kyungsang Korean ......................................................40

3.2. The numerical digits of South Kyungsang Korean, together with IPA and the tone pattern of each digit ..............................................................................41

3.3. The numerical digits of Tokyo Japanese, together with Romaji and the mora type of each digit. A dash (-) indicates a mora boundary ........................................43

3.4. Seven citation tones and their tones names in Suzhou Wu. The pitch level is based on the actual normalized pitch contour in Lau (2002) ......................................46

3.5. The phonetic representations of seven citation tones and tone patterns of disyllables, trisyllables, and quadrisyllables (Based on Lau, 2002, p. 140) ..........47

3.6. The rhythmic unit of each language ..............................................................50

3.7. A typology based on the distinctions between stress vs. non-stress accent and between lexical vs. post-lexical pitch accent (Ladd, 1996, p. 156) ..........51

3.8. A typology of factors determining pitch accents, after Ladd (1996, p. 156), expanded to include the type represented by Wolof (Lindström & Remijsen, 2005, p. 847) ........................................................................................................52
3.9. The classification of the seven languages according to the accent-type typology (AE: American English, TJ: Tokyo Japanese, SKK: South Kyungsang Korean)...


4.1. Confusion matrix of corrective focus perception (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string. (Top panel: American English; Bottom panel: Seoul Korean)........................................................................................................76

4.2. The phrase-by-phrase confusion matrix for Seoul Korean................................77

4.3. Confusion matrix of corrective focus perception for Mandarin Chinese (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string ..................................................78

4.4. Position-by-position identification rates of tones 1-4........................................79

4.5. Confusion matrix of corrective focus perception for South Kyungsang Korean (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string ........................................80

4.6. The phrase-by-phrase confusion matrix for South Kyungsang Korean .............80

4.7. Confusion matrix of corrective focus perception (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string. (Top panel: Standard French, Bottom panel: Tokyo Japanese)..................................................................................................................82

4.8. The phrase-by-phrase confusion matrix for Tokyo Japanese ..............................83

4.9. Confusion matrix of corrective focus perception for Suzhou Wu (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string.................................................................84

5.1. The numerical digits of Mandarin and the lexical tone of each digit together with Pinyin Romanization .........................................................................................100
5.2. The duration, mean intensity, and target pitch values of the broad-focus and corrective-focus condition for each tone type (BF and CF are abbreviations for broad focus and corrective focus, respectively) ........................................107

5.3. Position-by-position identification rates of tones 1-4.................................111

5.4. Confusion matrix of corrective focus perception for tone 3 digits (percentage values). Cells highlighted in black indicate focused positions. Cells highlighted in gray refer to pre- or post-focus positions across the phrase boundary. Vertical lines refer to phrase boundaries.................................................................113

5.5. The mean intensity of tones 1-4. BF and CF are abbreviations for broad focus and corrective focus, respectively .................................................................115

6.1. The onset consonant type of each digit and the tone group depending on the tonal contrast that each digit shows.................................................................122

6.2. The output of the linear mixed effects model. Broad focus in the Low tone group is the reference category .................................................................127

6.3. Confusion matrix of corrective focus perception (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string. (Top panel: the Low tone group, Bottom panel: the High tone group)...............................................................................130

6.4. The phrase-by-phrase confusion matrix (Top panel: Low tone group, Bottom panel: High tone group)...............................................................................132

6.i. The mean identification rate (%) for each digit. The areas shaded in gray refer to the High tone group ...............................................................................129

7.1. The classification of languages based on rhythmic unit and type of prominence marking at the word level and at the phrase level. Languages above the dotted line produced clear prosodic marking of focus, whereas languages below that line did not produce such a clear prosodic marking of focus (AE: American English, MC: Mandarin Chinese, SF: Standard French, LPA: lexical pitch accent, Head: Head-prominent, Head/edge: Head/edge-prominent, Edge: Edge-prominent) .............143
List of Figures

1.1. Time-normalized average pitch contours of 36 repetitions of the sentence Jonathan remembered Jessica (left panel) and Minsuga manduril mskninda (‘Minswu is eating dumplings’) (right panel) by six speakers. The first area of each panel contains a target word (Raw data from Lee (2009) for the left panel and from Lee & Xu (2010) for the right panel) ..........................................................15

2.1. Screenshots of the perception experiment for AE (Left panel: broad-focus condition, Right panel: corrective-focus condition) ..........................................................30

2.2. Screenshots of the perception experiment for Tokyo Japanese (Left panel: broad-focus condition, Right panel: corrective-focus condition) ..................................................31

2.3. A screenshot of the instructions for the survey in Qualtrics .............................................33

2.4. A screenshot of part of the survey in Qualtrics ..........................................................34

3.1. Time-normalized mean pitch contours of 36 repetitions by six speakers. The sentence is Minsuga manduril mskninda (Raw data from Lee & Xu, 2010) ........38

3.2. Time-normalized average pitch contours of five repetitions of the string (105)-(601)-(2318) by five speakers. The red fonts indicate a digit string. The dotted vertical lines demarcate phrases. The area shaded in gray represents a string position with an accentual peak ..................................................44

3.3. Time-normalized average pitch contours of the string (980)-(673)-(2518). The dotted vertical lines demarcate phrases. The red digits indicate a digit string and the green digits the citation tone of each digit ..................................................48

4.1. Prosodic differences between broad and corrective focus in American English when position 1 is focused. Character 1 represents focus position ..................62

4.2. Prosodic differences between focused digits and broad-focus counterparts in American English and Seoul Korean. Characters 1-A refer to positions 1 to 10 .....65

4.3. Prosodic differences between focused digits and broad-focus counterparts in
Mandarin Chinese and South Kyungsang Korean. Characters 1-A refer to positions 1 to 10 ........................................................................................................67

4.4. Prosodic differences between focused digits and broad-focus counterparts in Standard French and Tokyo Japanese. Characters 1-A refer to positions 1 to 10 ....69

4.5. Prosodic differences between focused digits and broad-focus counterparts in Suzhou Wu. Characters 1, 3, 4, 5, and A indicate positions 1, 3, 4, 5, and 10 ........70

4.6. The mean pitch contours of the broad- and corrective-focus conditions. Each pitch contour includes focus in position 1. The area shaded in gray refers to a prosodic phrase that includes the first three digits ..........................................................72

4.7. The mean intensity values of the broad-focus and corrective-focus conditions. Positions 5 and 10 are the target positions that include prosodic focus ..........88

4.8. The pitch and intensity values of the broad-focus and corrective focus conditions in two languages: American English and Mandarin Chinese. The target position is position 10 ........................................................................................................90

5.1. Sample pitch contours for tone 3 digits in two focus conditions. BF and CF are abbreviations for broad focus and corrective focus, respectively ................102

5.2. Pitch trajectories of tones 1-4 in two focus conditions. BF and CF are abbreviations for broad focus and corrective focus, respectively ................................103

5.3. Duration, mean intensity, and target pitch of all tone types in two focus conditions (BF: broad focus, CF: corrective focus; T1: Tone 1, T2: Tone 2, T3: Tone 3, T4: Tone 4). The downward arrow of the third panel indicates that tone 3 focus is expressed by lowering its pitch point ......................................................106

5.4. Duration, mean intensity, and mean pitch values of the pre-focus position of tone 3 digits, separated by final vs. non-final positions. BF and CF are abbreviations for broad focus and corrective focus, respectively.................................108

5.5. Duration, mean intensity, and mean pitch values of the post-focus position of tone 3 digits, separated by initial vs. non-initial positions. BF and CF are abbreviations for broad focus and corrective focus, respectively........................................110

6.1. Intonation model of Seoul Korean (Adapted from Jun & Cha, 2011) ............118

6.2. Time-normalized pitch contours sampled at ten equidistant points. The shaded area
indicates the focus position (Raw data from Lee & Xu, 2010). The sentence is 
\textit{minsuga manduril m\kappa\i\n\i\n} (Minsu is eating dumplings) ...........................................120

6.3. Sample pitch contours for the Low and High tone groups in two focus conditions. 
BF and CF are abbreviations for broad focus and corrective focus, respectively..124

6.4. Top panel: Points indicate mean values of each position, Bottom panel: Points 
indicate mean values and bars standard errors .................................................126

6.5. Means by corrective focus minus broad focus. Characters 1-A indicate focus 
position from 1 to 10 ........................................................................................................128

7.1. Time-normalized pitch contours, where the area in gray indicates a focus position. 
Figure 1a includes a natural sentence \textit{Minsuga manduril m\kappa\i\n\i\n} (‘Minswu is 
eating dumplings’); Figure 1b consists of a digit string 637-686-7664. 
Abbreviations are BF (broad focus), DF (discourse-new focus), and CF (corrective 
focus) ................................................................................................................................147

7.2. Time-normalized pitch contours, where the first area includes a focused digit and 
the dotted lines refer to a digit boundary. BF and CF abbreviate broad focus and 
corrective focus, respectively ........................................................................................................149
Chapter 1

Introduction

The primary purpose of communication is to convey information. The delivery of this information is expected to be systematic and understandable. A speaker makes assumptions about the hearer’s knowledge and behaves accordingly. He then structures sentences to integrate new information with given information he believes the listener is familiar with. This kind of structure is known as information structure (IS) (e.g., Halliday, 1967; Krifka, 2008; Lambrecht, 1994; Vallduví, 1990). Consider the brief dialogue in (1).

(1)  a. What does John drink?
    b. John drinks beer.

In (1b), John drinks is old information since it repeats content from the question, and beer – the answer to the question – is new information. In the context of the dialogue, the informative part (beer) is the focus, which indicates a discourse function that emphasizes a particular piece of information in a sentence (Ladd, 1984; Xu & Xu, 2005).

In this dissertation, our main goal is to gain a better understanding of the similarities and differences in prosodic marking of focus both within and across languages. However, the term focus has been used differently across contexts and
researchers, and even different languages employ different means of focus marking. Therefore, before setting up the research questions, it is necessary to untangle the terminology that surrounds focus. We first describe different focus types in Section 1.1. In Section 1.2, some of the issues regarding focus that still remain a matter of discussion are analyzed. Section 1.3 examines different focus-encoding mechanisms across languages, while Section 1.4 lays out the research questions. Finally, Section 1.5 outlines the remaining parts of the dissertation.

1.1. Focus types

Depending on the context where it is used, focus is divided into several types, such as discourse-new focus, contrastive focus, and corrective focus. We describe these three types of focus in turn.

(2)  a. What is your name?
    b. My name is \([x]_F\).

The variable \(x\) in (2b) represents new information, thus receiving *discourse-new focus*. This new information is not recoverable from prior context (Halliday, 1967; Kuno, 1972) and is not shared between speaker and hearer (Jackendoff, 1972).

Newness is not the only source of focus; focus can also derive from contrastiveness. This type of focus is called *contrastive focus*, meaning that it indicates an exhaustive choice among a set of alternatives (Rooth, 1992).
(3) a. Between pizza and noodles for lunch, which one do you prefer to have?
   b. I [VP prefer [pizza]F].

Although pizza (3b) is old information since it has already been stated in (3a), it contrasts with noodles. Following alternative semantics for focus (Rooth, 1992), the focus semantic value for the VP prefer pizza consists of the following set of alternatives: {prefer pizza, prefer noodles}. Since pizza is singled out in the response, it receives contrastive focus.

The constituent associated with only is also assured an exhaustive interpretation, as shown in (4).

(4) Eva only gave xerox copies to [the graduate students]F. (Partee, 1991, p. 179)

Sentence (4) means that Eva gave xerox copies to the graduate students and no one else. The constituent associated with only, the graduate students, is an exhaustive choice between many potential beneficiaries, and accordingly it receives contrastive focus by virtue of being associated with only.

The third type, corrective focus, is illustrated in (5). This focus is similar to contrastive focus in that it reflects contrast but differs in that it corrects wrong information. In (5b), the focused digit 9 (that contrasts with 7) is used to correct the wrong digit 7 from prior context.
(5) a. Mary’s phone number is 264-872-8618. Right?
   b. No, Mary’s phone number is 264-8[9]f2-8618.

1.2. Some terminology issues

Although the term “focus” has received considerable attention, the notion and definition of focus have not been coherent in the literature (Gundel, 1999; Levinson, 1983; Matić & Wedgwood, 2012; Schwarz, 2003). In this section, we address some of the issues that still remain a matter of discussion. These issues include: the oldness/newness distinction; a debate over the status of discourse-new focus; IS partitions; and different theoretical frameworks for focus. We will address these issues one by one.

First, the oldness/newness dichotomy is not simple to define. In Prince (1992), new information is labeled as discourse-new or hearer-new depending on whether it is new to a discourse or a hearer. Both discourse-new and hearer-new information represent purely new information since they have not been previously evoked or known to a hearer. A discourse-new entity like *Sandy Thompson* and *the moon* in (6) can be hearer-old since presumably they are already familiar or known to a hearer (Prince, 1992; Ward & Birner, 2003). Conventional wisdom is that hearer-old (e.g., inferable) information is not new despite being discourse-new at the time of speech (Birner, 2003; Hietaranta, 1984; Prince, 1981, 1992; Ward & Birner, 1995, 1997, 2003). However, Ward and Birner (1995, 1997) point out that hearer-old information can be marked as hearer-new, especially when the entity is used to recall information from a hearer’s memory. Overall, there is no clear-cut distinction between oldness and newness – instead it seems that the distinction depends
on many other factors, such as hearer’s mental representation, assumed familiarity between speaker and hearer, and others.

(6)  a. I’m waiting for it to be noon so I can call Sandy Thompson.  
     (Prince, 1992, p. 309)

     b. Last night the moon was so pretty that I called a friend on the phone and told him to go outside and look.  
     (Ward & Birner, 2003, p. 121)

Second, a debate has emerged recently over the status of discourse-new focus (aka information focus). It has been claimed that discourse-new information does not have a focus marking in syntax, and thus is not considered to be a type of focus (Katz & Selkirk, 2011; Rochemont, 2013; Selkirk, 2008b). The motivation for this is that in English, discourse-new information does not bear maximal prominence in a sentence, distinguishing it from contrastive focus in terms of prosodic prominence (Katz & Selkirk, 2011). Calhoun (2010b), however, argues that discourse-new information does receive focus, and discourse-new focus and contrastive focus are distinguished by different pitch-accent types, for example, a peak accent (H*) for discourse-new focus and a rising peak accent (L+H*) for contrastive focus in the ToBI system. The main difference is that contrastive focus is characterized by a sharp rise immediately preceded by a valley.

Third, IS has been partitioned in a number of different ways. There indeed exist terminological profusion and confusion in the IS partitions, such as topic-comment (Gundel, 1988, 1999), topic-focus (Sgall, Hajicová, & Panevová, 1986), presupposition-focus (Jackendoff, 1972), ground-focus (Vallduví, 1990), and theme-rheme (Halliday, 1967; Steedman, 2000). In the topic-comment framework, topic is what the sentence is
about and comment is what the speaker says about it (Gundel, 1988, 1999). Examples in (7) (Vallduví, 1990, p. 38) illustrate the topic-comment relation, where the vertical line (|) demarcates the topic from the comment. In (7a), the sentence is about John, so he is the topic and ran away is the comment. In (7b), the sentence is about that new book by Thomas Guernsey, and the speaker makes the comment by saying I haven’t read yet about the topic.

(7)  
   a. John | ran away.  
   b. That new book by Thomas Guernsey | I haven’t read yet.

In the topic-focus framework, topic refers to what has already been mentioned in prior context, and focus indicates new, important information. Consider the brief dialogue in (8). In (8b), the topic is his name is in that it has already been stated in the previous discourse, and the focus is Paul in that it is the informative part of the sentence.

(8)  
   a. What is his name?  
   b. His name is [Paul]F.

The definition of topic in the topic-comment articulation does not correspond to the one in the topic-focus structure. Example (9) demonstrates the difference, where John is the topic in (9bi), whereas John drinks is the topic in (9bii). In the topic-comment construction, the topic is simply the complement of the focus.
(9)  a. What does John drink?
    b.  i. John | drinks beer. (topic | comment)
        ii. John drinks | beer. (topic | focus)

In the remaining IS partitions, the ground-focus (Vallduví, 1990), presupposition-focus (Jackendoff, 1972), and theme-rheme (Firbas, 1964) relations are more or less equivalent to the topic-focus relations. Halliday’s (1967) and Steedman’s (2000) theme-rheme relations are similar to Gundel’s (1988, 1999) topic-comment relation.

Finally, focus has been treated in different theoretical frameworks. There are two main streams of focus theories: the structured meanings approach and alternative semantics. In the structured meanings (SM) approach, the meaning of a sentence is based on the meanings of its parts and how the parts are combined in a structured way (e.g., Jackendoff, 1972; Krifka, 1992). Examples (10aii) and (10bii) are truth-conditionally equivalent – they express the same proposition, introduce(John, Bill, Sue) – but they differ by the information status of each proposition. The SM approach assumes that focus partitions the sentence into two parts: presupposition and focus. For example, in (10aii), introduced Bill to Sue is the presupposition and John is the focus. The structure of this sentence has a presuppositional skeleton like (11ai) (Jackendoff, 1972), and it can be represented by lambda expressions in (11aii). The focus interpretation can be obtained by the property of lambda abstracting on the focus and the semantic value of the focus.
(10) a. i. Who introduced Bill to Sue?
   ii. [John]F introduced Bill to Sue.

b. i. Who did John introduce Bill to?
   ii. John introduced Bill to [Sue]F.

(11) a. i. x introduced Bill to Sue.
   ii. $<\lambda x [introduce (x, bill, Sue)], john>$

b. i. John introduced Bill to z.
   ii. $<\lambda z [introduce (john, bill, z)], sue>$

In the alternative semantics approach (Rooth, 1985, 1992), the semantic value of focus has a contrastive function to stand out among the set of alternatives. The semantic value of the sentence (12a) has a proposition, $like(John, Sue)$, whereas the focus semantic value of the sentence (12b) refers to the set of propositions, $\{like(John, y) \mid y \in E\}$, where $E$ indicates the domain of all individuals. This set of propositions contains a set of alternatives: $\{like(John, Sue), like(John, Mary), like(John, Amy), \ldots\}$. Since Sue is singled out among many alternatives, the meaning of (12b) becomes that John likes Sue and no one else.

(12) a. John likes Sue.

b. John likes [Sue]F.

1.3. Three basic approaches to focus marking
Different languages employ different means of focus marking. Focus is encoded largely by prosodic, syntactic, and/or morphological means (Lambrecht, 1994; Nakanishi, 2001; Zubizarreta, 1998). In stress-based languages, such as English (Birch & Clifton, 1995; Cohan, 2000; Gussenhoven, 2007; Ladd, 1996; Xu & Xu, 2005), Dutch (Swerts, Krahmer, & Avesani, 2002), and German (Baumann, Grice, & Steindamm, 2006), focus is primarily encoded by prosody using a nuclear pitch accent. Consider the following Q&A pairs.

<table>
<thead>
<tr>
<th>Context</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>(13) a.</td>
<td>What does John like?</td>
</tr>
<tr>
<td>(13) b.</td>
<td>John hates pizza.</td>
</tr>
<tr>
<td>(13) c.</td>
<td>Who likes pizza?</td>
</tr>
</tbody>
</table>

b. John likes [pizza].

(No,) John [likes] pizza.

[John] likes pizza.

The examples in (13) illustrate three different placements of focus depending on the context; (13a) exhibits focus on the object NP, (13b) on the verb, and (13c) on the subject NP. These examples demonstrate that focus can be conveyed purely by prosodic means by simply shifting the placement of prominence while keeping the syntactic structure the same.

With respect to prosodic focus, it is important to note that focus includes prosodic emphasis on the material that seems important and prosodic minimization of more-or-less redundant material. For example, in (13a) above, the focused item pizza becomes prosodically emphasized, and the redundant material John likes becomes prosodically minimized or deaccented. The latter phenomenon is widely known as anaphoric...
destressing (e.g., Calhoun, 2010a; Graf, 2012; Naylor, 1975; Wagner & McCurdy, 2010; Wagner, 2012; Williams, 1997).

Next, there are languages that employ syntactic means in encoding focus. Languages like Catalan (Estebas-Vilaplana, 2000; Vallduví, 1990) and Hungarian (Balogh, 2007; Brody, 1990; Onea, 2009) belong to this category. In Catalan, a syntactic rearrangement is required to get prominence that is specifically derived from clause-final position (Estebas-Vilaplana, 2000; Vallduví, 1990). Example (14) shows this syntactic rearrangement.

(14) L’amo l’[odia]F tì, el bròquilF

the boss Obj-cl 3s-hate the broccoli

‘The boss [hates]F broccoli.’  

(Vallduví, 1990, p. 6)

As shown in (14), when the verb odia is focused, the object NP el bròquil is detached from its base position, leaving the empty position (t) and a pronominal clitic (l’). The detached phrase is right-dislocated and occurs outside the “core” clause boundary. Although the focused element odia appears to be clause-medial, there is a disjuncture between odia and el bròquil; the comma indicates that el bròquil is right-dislocated from its base position (Vallduví, 1990), thereby the focused element occurs in clause-final position and receives prominence.

Hungarian also requires a syntactic rearrangement for focus marking. It differs from Catalan in that the syntactic rearrangement only relates to contrastive focus. Example (15b) demonstrates this syntactic operation – an object NP like Marit (15b) must move to a pre-verbal position to be contrastively focused in Hungarian.
(15)  a. Péter szereti  Marit.
    Peter loves Mary.Acc
    ‘Peter loves Mary.’

b. Péter [Marit]F szereti  t,.
    ‘Peter loves Mary (and no one else).’
    (Onea, 2009, p. 53)

Additionally, languages such as Buli (Schwarz, 2009), Chickasaw (Gordon, 2007),
Ewe (Fiedler & Jannedy, 2013), and Wolof (Rialland & Robert, 2001) mark focus by
means of morphology. Among them, we describe Chickasaw and Ewe as examples.
Chickasaw has focus morphemes suffixed to NPs (Gordon, 2007). The suffixation of
focus morphemes differs by the syntactic category of a focused element (subject vs.
object) and by the type of focus (discourse-new vs. contrastive). Table 1.1 displays the
focus morphemes, and examples in (16) display three different focus conditions (Gordon,
2007, p. 67).

<table>
<thead>
<tr>
<th>Focus Type</th>
<th>Subject</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discourse-new focus (DF)</td>
<td>-ho:,t</td>
<td>-ho</td>
</tr>
<tr>
<td>Contrastive focus (CF)</td>
<td>-akot</td>
<td>-akõ</td>
</tr>
</tbody>
</table>

Table 1.1. Focus morphemes in Chickasaw.
(16) a. hat:ak-at koni(ă) pisa. (Broad focus)
    man-Nom skunk.Acc sees
    ‘The man sees the skunk.’

b. [hat:ak-akot]F koni(ă) pisa. (Contrastive focus on the subject NP)
    man-FM.Nom skunk.Acc sees
    ‘[The man]F sees the skunk.’

c. hat:ak-at [koni-akõ]F pisa. (Contrastive focus on the object NP)
    man-Nom skunk-FM.Acc sees
    ‘The man sees [the skunk]CF.’

Sentence (16a) does not include a focus morpheme suffixed to either a subject or an object, whereas (16b) has a CF morpheme suffixed to a subject NP and (16c) to an object NP.

Let us now turn to Ewe. In this language, focus is encoded by word order variation and/or a focus morpheme. Ewe marks focus morpho-syntactically by placing a focus morpheme -(y)é after a subject or a preposed object. An object NP is focused in two ways: \textit{in situ} or \textit{ex situ}. The use of the focus morpheme -(y)é is obligatory for a focused subject NP but optional for a focused object NP. Example (17) shows \textit{in situ} focus on the object NP, (18) \textit{ex situ} focus on the object NP, and (19) focus on the subject NP.

(17) SV[O]F = \textit{in situ} focus on the object NP

Context: What did the woman buy?

é φlè [əɖɪbã]F.
3sg buy pawpaw

‘She bought a [pawpaw]F. ’

(Fiedler & Jannedy, 2013, p. 6)
(18) [O-(y)é]_{F}\text{SV} = \textit{ex situ} focus on the object NP
Context: The woman bought this banana.
áò, [àkòḍũ má (yé)]_{F} wò φlè.
no banana Dem (FM) 3sg.Dep buy
‘No, she bought [that banana]_{F}.’

(Fiedler & Jannedy, 2013, p. 7)

(19) [S-(y)é]_{F}\text{VO} = focus on the subject NP
Context: Mary bought a banana.
áò, [Súzánà yé]_{F} φlè àkòḍũ.
no Susan FM buy banana
‘No, [Susan]_{F} bought a banana.’

(Fiedler & Jannedy, 2013, p. 8)

1.4. Current study

It is well known that particular syntactic or morphological encoding of focus may be required in some languages, but optional or non-existent in other languages. It is also well known that a focused element is more prominent than its adjacent elements in the same domain (Samek-Lodovici, 2005; Truckenbrodt, 1995). This is formalized in terms of the Stress-Focus constraint below (Samek-Lodovici, 2005, p. 697).

(20) The Stress-Focus constraint

For any $XP_{F}$ and $YP$ in the focus domain of $XP_{F}$, $XP_{F}$ is prosodically more prominent than $YP$.

Previous studies have revealed the relationship between focus and prominence in many languages (e.g., English: Cooper, Eady, & Mueller, 1985; Xu & Xu, 2005, Korean:}
Jun & Lee, 1998; Lee & Xu, 2010, Mandarin: Xu, 1999, Japanese: Lee & Xu, 2012). In these studies, a focused element was more prominent than its adjacent elements by being realized with a longer duration, higher pitch, and higher intensity. Therefore, it has been long and widely believed that increased vocal efforts associated with focus are assumed to be a sort of universal phonetic symbolism – a focused element exhibiting the concomitant increased duration, intensity, and pitch range.

However, it is less commonly recognized that purely prosodic marking of focus may be weaker in some languages compared to others. A basic question arises here: is prosodic modulation by focus similar across languages? To pursue the answer, we examine to what extent modulation by focus varies between American English and Seoul Korean. Figure 1.1 displays broad focus and discourse-new focus conditions in American English (left panel) and Seoul Korean (right panel). These focus conditions were produced in an experimental setting, where six native speakers of each language read target sentences in isolation for broad focus and produced the same sentences in a Q&A dialogue for discourse-new focus.¹ The stimuli were repeated six times for both conditions.

---

¹ The questions used are *Who remembered Jessica?* (English) and *Who is eating dumplings?* (Korean), and the Korean sentence is Romanized using IPA.
The first area of each panel contains a target word (Raw data from Lee (2009) for the left panel and from Lee & Xu (2010) for the right panel).

In both languages, discourse-new focus produced a more expanded pitch range than broad focus. However, it seems like the modulation by focus is greater in American English than in Seoul Korean. To test this, we conducted a pairwise t-test analysis comparing the difference in maximum pitch between the focus conditions in each language (the peak difference: 2.52 st in American English, 1.18 st in Seoul Korean). The result demonstrated that American English employed a more expanded pitch range ($p < 0.05$). Here, an important point is that the pitch expansion via focus is fairly small in Seoul Korean – just 1.18 st extra pitch. The value of 1.18 st is quite similar to the interval between the notes C and C# on a musical scale. This implies that prosodic marking of

---

2 In this figure, the dotted vertical line demarcates each word of the sentence. Each word was time-normalized and averaged with ten equidistant points using Xu’s ProsodyPro Praat script (Xu, 2013).
focus in Seoul Korean may not be perceptually salient for listeners. As a consequence, we assume that prosodic marking of focus is neither completely universal nor automatic, but instead it differs according to the prosodic system of each language.

Furthermore, it is less recognized that prosodic marking of focus may differ even within the same language. Mandarin Chinese consists of four lexical tones: a high-level tone, a rising tone, a low/dipping tone, and a falling tone. Among these tones, a low/dipping tone may behave differently in marking prosodic focus since it is marked by lowering a pitch target (Chao, 1968; Chen & Gussenhoven, 2008; Xu, 1999). We assume that a focused low/dipping tone shows a smaller capacity for pitch expansion because lowering a pitch target seems physiologically more limited than raising a pitch target. The rationale behind this assumption is recognizing that a human’s pitch range is fairly small – within 100 Hz (Baken & Orlikoff, 2000; Keating & Kuo, 2012; Kuang, 2013), and a low/dipping tone is produced towards the low end of the pitch range. In addition, Seoul Korean is known for its tonogenesis-like sound change depending on the laryngeal articulations of consonants (Jun, 2005; Kang, 2014; Kingston, 2011; Silva, 2006; Wright, 2007): consonants with [+aspiration/tense] create a high pitch; those with [−aspiration/tense] create a low pitch. We speculate that consonants with [+aspiration/tense] would be more effective in marking prosodic focus than those with [−aspiration/tense]. Again, we continue to assume that prosodic marking of focus is neither completely universal nor automatic even within the same language.

Another area not fully understood is that of the phonological units which carry various prosodic focus effects across languages. Figure 1.1 above demonstrates that American English and Seoul Korean exhibit different prosodic focus effects. For example,
in American English, prosodic focus is expressed by a stressed syllable of Jonathan and post-focus compression occurs immediately after the stressed syllable, whereas in Seoul Korean, a focus effect spans over the entire phrase (Minsuga) and post-focus compression appears outside the focused phrase. On the basis of the prosodic patterns in Figure 1.1, it seems like a syllable is the smallest phonological unit that carries prosodic focus in American English, and a phrase is the smallest phonological unit that carries prosodic focus in Seoul Korean. However, this may not always be the case in Seoul Korean. Consider a situation in (21), in which the underlined digit 2 is used to correct the wrong digit 1.

(21) Q: Is the number 367-810-8717?
A: No, it is 367-820-8717.

In this case, it is not clear whether only the single digit 2 carries prosodic prominence or the focus effect spans over the entire phrase. If only the single digit carries prosodic prominence, the phonological unit of carrying prosodic focus would be a word (i.e., each digit) rather than a phrase.\(^3\) In comparison, if prominence spans over the entire phrase, the phonological unit carrying prosodic focus would be a phrase, and listeners would have difficulty identifying the corrected digit. Another ambiguous issue is where post-focus compression occurs. It may occur right after the focused digit 2 if the phonological unit carrying prosodic focus is a word. Alternatively, it may occur across the phrase if the

\(^3\) In a phone-number string, each digit is monosyllabic in Seoul Korean. However, since each digit is already a word (bigger than a syllable but smaller than a phrase), we cannot say if the smallest phonological unit is a syllable.
phonological unit carrying prosodic focus is a phrase.

Finally, another question that needs to be examined is whether or not prosodic marking of focus varies with different pitch-scaling conditions. Liberman and Pierrehumbert (1984) suggest that downstepping F0 patterns appear to show exponential decay. This is obviously not to 0 Hz, but rather to a value that we can refer to as a speaker’s baseline B. This suggests that the scaling of F0 values, rather than being purely a matter of ratios, as in the semitone scale, should be evaluated as “baseline units above the baseline,” i.e., the scaled intonational value “Int” would be:

\[ \text{Int} = \frac{(F0 - B)}{B} \]

or in the other direction

\[ F0 = \text{Int} \times B + B \]

Other considerations, such as the relationship between similar tone/intonational patterns in different pitch ranges or in the speech or different speakers, also suggest that F0 scaling has an additive as well as a multiplicative component. If emphasis or focus has a multiplicative effect on the “Int” value, then we predict that the impact should be much greater in terms of F0 ratios (as well as absolute amounts) for higher as opposed to lower pitches. Therefore, we assume that higher pitches in a speaker’s vocal range are more effective in marking prosodic focus than lower pitches in both production and perception.

1.5. Research questions and approaches
As previously stated, this dissertation aims to enhance our understanding of the nature of prosodic focus both cross- and intra-linguistically. Since we assume that prosodic marking of focus is expressed through the respective prosodic system of each language, it is ideal to collect data from prosodically diverse languages, which may range from lexical stress languages to phrase-final stress languages to lexical pitch-accent languages to tone languages to languages with none of these. Therefore, our comparison includes seven languages: American English, Mandarin Chinese, Seoul Korean, South Kyungsang Korean, Tokyo Japanese, Standard French, and Suzhou Wu. American English is a lexical stress language. Standard French is a phrase-final stress language. South Kyungsang Korean and Tokyo Japanese are lexical pitch-accent languages. Mandarin Chinese and Suzhou Wu are tone languages. Seoul Korean has none of these features. By comparing prosodically diverse languages, this dissertation raises and attempts to answer the following four questions.

Q1. Does prosodic marking of focus differ across languages both in production and perception? If so, how does it vary across languages?

Q2. Does prosodic marking of focus also differ even within the same language? If so, how does it vary within a given language?

Q3. What are the phonological units that carry various prosodic focus effects, such as prosodic focus marking and post-focus compression across languages?

Q4. Are higher pitches more effective in marking prosodic focus than lower pitches in both production and perception?

---

4 Suzhou Wu is spoken in the city of Suzhou in China’s Jiangsu province.
Reliable (quantitative) cross-linguistic comparisons of prosodic focus are difficult to quantify, given all of the many relevant ways that languages can differ. Therefore, in order to obtain comparable data across languages, we have developed a method based on corrective focus in phone-number strings, as shown in (22).

(22)  a. Mary’s phone number is 264-872-8618. Right?
     b. No, Mary’s phone number is [3]64-872-8618.

Although this method is not entirely free of the issue of anaphoric destressing given that a limited set of digits were used repeatedly in the stimuli, please note that we have limited consideration to the specific situation of corrective focus using randomly generated phone-number strings. This method actually includes several advantages over natural stimuli:

- Digit strings are possible in every (modern) language;
- Any order and phrasing of digit strings is possible;
- Syntactic and morphological revisions are completely ruled out;
- All of the positions are equally susceptible to focusing;
- Only prosodic modulation can be used to mark a focused item.

We have applied this method to seven languages for both production and perception experiments.

1.6. Outline of this dissertation
The remainder of this dissertation is organized into six chapters. Chapter 2 depicts how we conducted production and perception experiments using a phone-number string. Chapter 3 briefly describes the prosodic system of each language and introduces three types of prosodic typologies: 1) based on speech rhythm (stress-timed, syllable-timed, and mora-timed), 2) based on the distinctions between stress vs. non-stress accent and between lexical vs. post-lexical vs. no pitch accent, and 3) based on the types of prominence marking at the word level (stress, tone, or lexical pitch accent) and at the phrase level (head-prominent, head/edge-prominent, or edge-prominent). We then attempt to classify seven languages within the framework of each of these three prosodic typologies. Chapter 4 examines whether and how prosodic marking of focus differs across languages. Chapters 5 and 6 focus on whether and how prosodic marking of focus differs within a single language: Chapter 5 reanalyzes the production and perception data of Mandarin Chinese to illustrate the unique characteristics of tone 3 focus; next, Chapter 6 revisits the production and perception data of Seoul Korean to see whether high-tone inducing consonants are more effective and more identifiable in marking prosodic focus than low-tone inducing consonants. Finally, Chapter 7 discusses the main findings of this dissertation and addresses the potential limitations of existing prosodic typologies with respect to prosodic focus. The remaining sections of Chapter 7 explore the implications that this dissertation may have. They provide a series of open questions for future research and serve to conclude the dissertation.
Chapter 2

Method – Spoken telephone numbers

This chapter begins by reviewing five methods of eliciting prosodic focus.

- Q&A dialogue
- Templatic structure
- Discourse context
- Picture task
- Animation task

First, the Q&A dialogue is the most widely used method (e.g., Birch & Clifton, 1995; Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986; Heldner & Strangert, 2001; Xu & Xu, 2005; Xu, 1999). This is mainly because this method allows one to construct a simple dialogue by asking speakers a question that prompts an answer. As exemplified in (1), John – an answer to what – receives discourse-new focus since it is new, important information.

(1) a. What is your name?
    b. My name is [John]_{F}.
Second, a templatic structure elicits focus by providing contrast (Baltazani & Jun, 1999; Chen, 2006). For example, in the form of “X, not Y,” target words receive contrastive focus, since position X contrasts with position Y. In (2), tea is focused since it contrasts with coffee.

(2) I said [tea]$_F$, not coffee.

Third, a target sentence is embedded in a discourse context (Beaver, Clark, Flemming, Jaeger, & Wolters, 2007; Lee, 2012; Sudhoff, 2010), where focus is elicited through the context. In (3), Korean Air receives contrastive focus, since it contrasts with Delta Airlines in the context.

(3) I am attending the University of Pennsylvania in Philadelphia. However, since my parents live in Korea, I go to Korea to see them every school vacation. I have both Korean Air and Delta Airline membership cards. When I go to Korea, I only take [Korean Air]$_F$. (Lee, 2012, p. 98)

Fourth, in a picture task, simple pictures with different types and/or colors are used to trigger a focus reading (Fernald & Mazzie, 1991; Jannedy, 1997). In the case of Jannedy’s (1997) experiment, two pictures were always paired: the first picture was used for a neutral reading and the second for a focus reading. The experimenter showed the first picture to the speaker by asking “What is this?” and the speaker answered the question by saying “A red ball.” The experiment showed the second picture to the speaker by asking “And what is this?” and then the speaker answered the question by
saying “A red [eye]F.” In this context, speakers were involved in a natural dialogue setting and put a focus on the word “eye.”

Fifth, Hockey (1998) developed a computer animation task, where different objects (e.g., airplane, anchor, apple, ax, bed) or different geometric shapes (e.g., circle, square, rectangle, triangle) were paired with six different colors. In this task, two participants were paired for a spontaneous dialogue. For example, a participant asked a question (4a) by watching the static drawing that is missing important information, and the other participant answered the question (4b) by watching the animation that provides the missing information (Hockey, 1998, p. 78).

(4) a. Uh what does the black helicopter push?
   b. It pushes [the red triangle]F.

Although these methods can successfully elicit prosodic focus in most languages, they are not appropriate for a cross-linguistic comparison since different languages have different intrinsic word order and employ different means of focus marking. In an attempt to obtain comparable data across languages, as previously mentioned in Chapter 1, we have developed a method implementing 10-digit phone-number strings that are free of syntactic and morphological means of focus marking. This method has the advantage that only prosodic modulation can be used to signal the focused item. We have applied this method in both production and perception. In the remainder of this chapter, we illustrate how we conducted production and perception experiments for seven languages.
2.1. Production

2.1.1. Speech materials

Different languages allow different ways of reading phone-number strings. For example, in French, 10-digit number strings are typically divided into five clusters like (NN)-(NN)-(NN)-(NN)-(NN). In both Korean and English, the three-digit string 267 can be read as either “two hundred sixty-seven” or read as individual digits like “two six seven.” Japanese, Chinese, and Spanish speakers read 267 only as individual digits. In order to obtain comparable data across languages, therefore, we pursued the style of American telephone numbers – the string was connected as individual digits grouped in the frame of (NNN)-(NNN)-(NNNN).

By running a Python script, we created sets of 100 10-digit strings, designed so that i) every digit occurs equally often in every position, and ii) every pair of digits occur equally often across every pair of positions.¹ To elaborate on the second point, there are 9 pairs of adjacent digits in a string of 10 digits: 0-1, 1-2, 2-3, ..., 8-9. Therefore, in the datasets used in the experiment, each of these nine pairs of adjacent positions has every possible pair of digits in it. Thus, the first and second position and similarly the second and third positions, etc. will (in different digit strings) be as follows.

¹ The Python script is in Appendix A.
Speakers produced 100 10-digit strings in isolation for broad focus and in a Q&A dialogue below for corrective focus, where someone asks whether the phone-number is correct and the speaker answers the question with a string by correcting the wrong digit.²

(6) Q: Is Mary’s number 887-412-4699?
A: No, the number is 787-412-4699.

There are two caveats about the experiment. First, for Tokyo Japanese and Suzhou Wu, target strings were prompted by questions for both broad focus and corrective focus.³ Second, due to constraints on some Suzhou Wu speakers’ time, the production material was limited to 10 10-digit sequences, with each digit occurring ten times in each position, and each string read in six focus conditions, broad focus and focus on the 1ˢᵗ, 3ʳᵈ, 4ᵗʰ, 5ᵗʰ, and 1⁰ᵗʰ positions.⁴ In addition, the area code was presented in parentheses, such as (787) 412-4699.

² Digit strings used for the broad-focus and corrective-focus conditions are in Appendix B and C. The same strings were used for all the languages except Suzhou Wu.

³ The Q&A structure for broad focus was: Q: Did you get the number?: A: Yes, the number is 787-412-4699.

⁴ Digit strings used for Suzhou Wu are in Appendix D.
Before moving on, it is necessary to illustrate different pronunciations of the digits (0 to 9) in each language. As shown in Table 2.1, some languages have two forms of pronunciation for the same digit. For example, in American English, the digit 0 can be read as “O” or “zero.” In this case, speakers were instructed to choose the form of pronunciation with an asterisk (*).

Table 2.1. Different pronunciations of the numerical digits from 0 to 9 in each language. I used the IPA for the two varieties of Korean, the Romaji system for TJ, the Pinyin Romanization system for Mandarin Chinese, and the IPA for Suzhou Wu.  

<table>
<thead>
<tr>
<th>Digit</th>
<th>AE</th>
<th>MC</th>
<th>SF</th>
<th>SK/SKK</th>
<th>TJ</th>
<th>SW</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>*o/zero</td>
<td>ling2</td>
<td>zero</td>
<td>*goŋ/jʌŋ</td>
<td>*zero/rei</td>
<td>lin213</td>
</tr>
<tr>
<td>1</td>
<td>one</td>
<td>*yi1/iau1</td>
<td>un</td>
<td>il</td>
<td>ichi</td>
<td>iə5</td>
</tr>
<tr>
<td>2</td>
<td>two</td>
<td>er4</td>
<td>deux</td>
<td>i</td>
<td>ni</td>
<td>liá231</td>
</tr>
<tr>
<td>3</td>
<td>three</td>
<td>san1</td>
<td>trois</td>
<td>sam</td>
<td>san</td>
<td>se44</td>
</tr>
<tr>
<td>4</td>
<td>four</td>
<td>si4</td>
<td>quatre</td>
<td>sa</td>
<td>*yon/shi</td>
<td>sɪ522</td>
</tr>
<tr>
<td>5</td>
<td>five</td>
<td>wu3</td>
<td>cinq</td>
<td>o</td>
<td>go</td>
<td>η231</td>
</tr>
<tr>
<td>6</td>
<td>six</td>
<td>liu4</td>
<td>six</td>
<td>juk</td>
<td>roku</td>
<td>lo23</td>
</tr>
<tr>
<td>7</td>
<td>seven</td>
<td>qi1</td>
<td>sept</td>
<td>teʰi1</td>
<td>*nana/shichi</td>
<td>ts’iə5</td>
</tr>
<tr>
<td>8</td>
<td>eight</td>
<td>ba1</td>
<td>huit</td>
<td>pʰal</td>
<td>hachi</td>
<td>po5</td>
</tr>
<tr>
<td>9</td>
<td>nine</td>
<td>jiu3</td>
<td>neuf</td>
<td>gu</td>
<td>kyuu</td>
<td>teɪɪ51</td>
</tr>
</tbody>
</table>

2.1.2. Participants

---

5 Tones of Mandarin Chinese and Suzhou Wu will be described in detail in Chapter 3.
Table 2.2 presents how many speakers participated in the production experiment in each language. The mean age and the standard deviation (SD) were calculated based on the time of recording.

Table 2.2. The number of speakers and the mean age (SD) of each language.

<table>
<thead>
<tr>
<th>Language</th>
<th>Speakers</th>
<th>Mean age (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>American English</td>
<td>2 males, 3 females</td>
<td>27.8 years (4.5)</td>
</tr>
<tr>
<td>Mandarin Chinese</td>
<td>2 males, 3 females</td>
<td>25.2 years (1.1)</td>
</tr>
<tr>
<td>Standard French</td>
<td>2 males, 3 females</td>
<td>23.0 years (3.7)</td>
</tr>
<tr>
<td>Seoul Korean</td>
<td>2 males, 3 females</td>
<td>29.4 years (3.8)</td>
</tr>
<tr>
<td>South Kyungsang Korean</td>
<td>2 males, 3 females</td>
<td>29.2 years (5.4)</td>
</tr>
<tr>
<td>Tokyo Japanese</td>
<td>2 males, 3 females</td>
<td>28.4 years (7.0)</td>
</tr>
<tr>
<td>Suzhou Wu</td>
<td>3 males, 3 females</td>
<td>19.5 years (1.2)</td>
</tr>
</tbody>
</table>

Sets of speakers were recruited from four different locations. Speakers of American English, Mandarin Chinese, Seoul Korean, and South Kyungsang Korean were recruited at the University of Pennsylvania. Due to difficulties recruiting enough participants for the other languages, we asked colleagues in France, Japan, and China to find suitable speakers. Accordingly, Standard French speakers were recruited from the Sorbonne Nouvelle, Tokyo Japanese speakers from the National Institute for Japanese Language and Linguistics (NINJAL), and Suzhou Wu speakers from Suzhou, Gusu district.

2.1.3. Recording procedure
Each recording session was conducted in quiet conditions, where speakers were seated in front of a computer monitor. For the recording sessions at the University of Pennsylvania, speakers wore a Plantronics head-mounted microphone, and recordings were saved directly onto a laptop at a 44.1 kHz sampling rate with a 16-bit resolution. For the recording sessions at the Sorbonne Nouvelle, audio was recorded using an AKG C520L head-mounted condenser microphone and was saved using a Digidesign Audiobox and Pro Tools at a 44.1 kHz sampling rate with a 16-bit resolution. For the recording sessions at NINJAL, speakers’ speech was recorded at 44.1 kHz and 16 bps using a built-in Mac microphone and was saved directly onto a laptop. For the recording sessions in Suzhou, Gusu district, audio was recorded at 44.1 kHz and 16 bps by using a Shure BETA53 microphone connected to Steinberg USB Audio Interface, and was saved directly onto a MacBook Pro computer.

PowerPoint slides were used to present speech materials to speakers. As previously stated, for five of the seven languages (American English, Mandarin Chinese, Standard French, Seoul Korean, South Kyungsang Korean), only target strings were presented in isolation for the broad-focus condition (Figure 2.1a), and the same strings were produced in a Q&A structure for the corrective-focus condition (Figure 2.1b). For Suzhou Wu and Tokyo Japanese, target strings were prompted by questions for both focus conditions, as can be observed in Figure 2.2.

Following is a description of the experiment procedure. First, participants had a practice session of reading three digit strings displayed on the screen in isolation or in a Q&A structure for broad focus, followed by the experiment session of reading 100 actual experiment digit strings. Participants then took a five-minute break. Next, participants
had another practice session of reading three digit strings in a Q&A dialogue for corrective focus; participants read the target strings as corresponding answers after listening to pre-recorded prompt questions, played through headphones or speakers. Note that the experimenter did not force participants to answer the corrected digit in any biased way, but encouraged them to answer the question using the most appropriate prosody in their native language. Then the actual session of producing 100 target strings for corrective-focus began.

For American English, Mandarin Chinese, Seoul Korean, South Kyungsang Korean, Tokyo Japanese, and Standard French, a total of 1,000 strings of digits were produced (100 10-digit strings x 5 speakers x 2 focus conditions). For Suzhou Wu, a total of 360 digit strings were produced (10 10-digit strings x 6 speakers x 6 focus conditions (broad focus and corrective focus in the 1^{st}, 3^{rd}, 4^{th}, 5^{th}, and 10^{th} positions)).

Figure 2.1. Screenshots of the perception experiment for AE (Left panel: broad-focus condition, Right panel: corrective-focus condition).
a. Broad focus   

b. Corrective focus

Figure 2.2. Screenshots of the perception experiment for Tokyo Japanese (Left panel: broad-focus condition, Right panel: corrective-focus condition).

2.2. Perception experiment

2.2.1. Audio stimuli

From the production data, we chose a set of 100 phone-number strings produced with corrective focus from 5 speakers (20 strings per speaker) for American English, Seoul Korean, South Kyungsang Korean, Tokyo Japanese, and Standard French. The set of audio stimuli was designed in a way that every string position included 10 digits from 0-9 and each digit was equally focused in every string position. This design enabled us to counter-balance the distribution of focus tokens in every string position. For Mandarin Chinese, we chose the same number of digit strings in the same way above, and added 80 more strings produced with tone 3 focus to the set, designed for a case study in Chapter 5. For Suzhou Wu, we selected a set of 50 digit strings from 6 speakers (9 strings from 2
speakers and 8 strings from 4 speakers) for five focus positions (Positions 1, 3, 4, 5 and 10). This design also included an equal number of focus tokens in each focus location.

2.2.2. Participants

Table 2.3 presents the number of participants in the perception experiment in each language, and the mean age and standard deviation (SD) of listeners.

<table>
<thead>
<tr>
<th>Language</th>
<th>Listeners</th>
<th>Mean age (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>American English</td>
<td>67</td>
<td>19.5 years (1.1)</td>
</tr>
<tr>
<td>Mandarin Chinese</td>
<td>20</td>
<td>23.4 years (3.5)</td>
</tr>
<tr>
<td>Standard French</td>
<td>16</td>
<td>23.3 years (4.7)</td>
</tr>
<tr>
<td>Seoul Korean</td>
<td>34</td>
<td>25.6 years (4.6)</td>
</tr>
<tr>
<td>South Kyungsang Korean</td>
<td>20</td>
<td>30.3 years (4.7)</td>
</tr>
<tr>
<td>Tokyo Japanese</td>
<td>22</td>
<td>22.9 years (4.3)</td>
</tr>
<tr>
<td>Suzhou Wu</td>
<td>16</td>
<td>29.4 years (8.4)</td>
</tr>
</tbody>
</table>

2.2.3. Procedure

We set up the experiment using a web-browser (Qualtrics) in order to recruit listeners online and to enable ease of access for all the languages except Mandarin Chinese. Before the actual test began, participants were provided with a description of what corrective focus is and how the experiment proceeds, as shown in Figure 2.3.
Figure 2.3. A screenshot of the instructions for the survey in Qualtrics.

Furthermore, the survey asked basic demographic information such as participants’ name, age, and gender. Participants were provided with a consent form online and agreed to participate as a subject in the perception experiment.

Figure 2.4 shows a screenshot of part of the survey in Qualtrics. During the test, participants heard only the phrase with the correction by pressing a play button, and were asked to select the corrected digit in a ten-choice task, as indicated below, where numerical digits indicate each digit in a string. The system allowed participants to repeat the audio file as many times as desired.
Figure 2.4. A screenshot of part of the survey in Qualtrics.

For Mandarin Chinese, the perception experiment was conducted offline in a quiet room at Tongji University. The audio stimuli were presented to listeners through Sennheiser PC166 headset speakers using Paradigm software (Perception Research Systems, 2007). Before the actual test, we presented three practice trials to listeners to familiarize them with the procedure. In the actual sessions, participants first heard only the phrase with the correction, and were asked to choose which digit was corrected by using a computer mouse. Participants were able to listen to each stimulus as many times as desired. A set of 180 strings of digits was randomized and divided into two blocks of 90 strings each, and there was a short intermission between the two blocks.
Chapter 3

Prosodic structures of languages under study

In this chapter, we first briefly describe the prosodic structures of seven languages in the following order: American English, Mandarin Chinese, Seoul Korean, South Kyungsang Korean, Tokyo Japanese, Standard French, and Suzhou Wu. Subsequently, we introduce three types of prosodic typologies. The first typology is based on speech rhythm, whereas the second typology is based on two dimensions: 1) stress vs. non-stress accent and 2) lexical vs. post-lexical vs. no pitch accent. The third typology is based on word-level and phrase-level prominence marking. Hereafter, for the sake of simplicity, we label these typologies as rhythm-based, accent-based, and prominence-based. Finally, we classify the seven languages according to each of these prosodic typologies.

3.1. Prosodic structures of each language

3.1.1. American English

American English has lexical stress (Bolinger, 1958; Huss, 1978; Ladd, 1984; Liberman, 1975; Shin & Speer, 2012). Although stress patterning is not entirely predictable in
American English, unlike other languages such as Finnish (Suomi, Toivanen, & Ylitalo, 2003), a shared knowledge is that different placement of stress may cause a difference in meaning within a word (Huss, 1978; Shin & Speer, 2012), in a noun phrase (Ladd, 1984; Liberman, 1975), and even in a sentence (Bolinger, 1958; Liberman, 1975; Shin & Speer, 2012). For example, the word desert differs in meaning depending on the placement of stress, e.g. DESert vs. deSERT, where capitalization represents a stressed syllable. A noun phrase English teacher may also have two different meanings. When primary stress falls on the first element, it means “a teacher of English;” otherwise, it means “a teacher who is English” (Liberman, 1975, p. 10). Even the same sentence structure differs in meaning depending on the location of stress, e.g. I want CHOcolate pie (not vanilla) vs. I want chocolate PIE (not cake) (Shin & Speer, 2012). In each of these sentences, the most prominent syllable (CHO in CHOcolate and PIE, respectively) is known to be expressed through a nuclear pitch accent assigned at the phrase level (Beckman & Pierrehumbert, 1986).

3.1.2. Mandarin Chinese

Mandarin Chinese is a tone language.¹ It consists of four lexical tones: a high level tone (tone 1), a rising tone (tone 2), a low/dipping tone (tone 3), and a falling tone (tone 4).

¹ The nature of stress in Mandarin Chinese has not clearly been established. Various studies have determined that Mandarin Chinese has stress. However, no exact consensus on the acoustic correlates of stress has been made in the literature. Lai, Sui, and Yuan (2010) revealed that pitch is a more reliable acoustic cue for stress than duration. Shen (1993) found that stress can be identified by cues of duration and intensity. Duanmu (2004) argued that duration (i.e., lengthening) distinguishes stressed syllables from
Depending on the tone shape, tones 1-4 are conventionally labeled as [55], [35], [214], and [51] in a 5-point scale, where the numbers represent pitch levels ([1]: lowest, [5]: highest) (Chao, 1968).

In Mandarin Chinese, lexical tones primarily shape the pitch pattern of a sentence since the main role of pitch is to deliver tonal information (Wang & Lee, 2015). For example, when a sentence consists of all high level tones like (1a), the pitch pattern remains (quasi) static in order to convey its tonal information. When a sentence contains all rising tones like (1b), the pitch pattern of a sentence repeats the rising tonal patterns over the course of a sentence. When a sentence includes mixed tone sequences like (1c), the pitch pattern reflects the tone shape of each syllable.

(1) a. gu1 ma1 jin1 tian1 ca1 chu1 zu1 che1.
   ‘My auntie cleaned the taxi today.’
b. wang2 ting2 ming2 nian2 lai2 ji2 lin2.
   ‘Ting Wang will come to Jilin Province next year.’
c. wo3 bu4 gan3 xiang1 xin4 zhe4 shi4 zhen1 de0.
   ‘I can’t believe this is true.’

(Wang & Lee, 2015, p. EL119)

3.1.3. Seoul Korean

Seoul Korean has neither lexical stress nor lexical pitch accents (Jun, 1998, 2005; Song, 2005) – different stress patterns (e.g., GAL.bi vs. gal.BI) do not indicate a difference in unstressed ones. In comparison, a group of researchers (e.g., Altmann, 2006; Chen, 2007) have argued that Mandarin has no stress mainly because there are no clear phonetic reflexes of stress.
meaning (Song, 2005, p. 40). The Seoul Korean tonal pattern instead comes from a combination of phrasal and boundary tones. In default prosodic phrasing, each content word can form a small prosodic unit, an Accentual Phrase (AP) that is post-lexically marked. In normal speech, the sentence *Minsuga manduril mskninda* ‘Minsu is eating dumplings’ consists of three APs, *(Minsuga)*(manduril)*(mskninda)*, where parentheses represent each AP. As Figure 3.1 shows, each AP, except the sentence-final one, exhibits a rising pitch contour towards the edge of the phrase. The rising tonal pattern marks an AP boundary tone in Seoul Korean. The sentence-final falling tonal pattern marks a declarative sentence.

![Figure 3.1. Time-normalized mean pitch contours of 36 repetitions by six speakers. The sentence is *Minsuga manduril mskninda* (Raw data from Lee & Xu, 2010).](image)

The AP’s basic melody is typically THLH and is fixed at the phrase level. The initial tone (T) differs by the laryngeal feature of the AP-initial segment (Jun, 1993, 1998, 2005, 2006). When the initial consonant is aspirated/tensed, the AP begins with H, and
elsewhere with L. Furthermore, a recent sound change has revealed that the AP begins with H when a digit 1 [il] is in AP-initial position (Jun & Cha, 2011). When the AP has fewer than four syllables, the second or third tone, or both, may not be realized.

3.1.4. South Kyungsang Korean

South Kyungsang Korean is different from Seoul Korean in that different tonal patterns cause a difference in meaning. The examples in (2) show three tonal patterns as a minimal triplet (Kim & Jun, 2009, p. 44).

(2) a. ga.dzi  HL ‘type’  
b. ga.dzi  HH ‘branch’  
c. ga.dzi  LH ‘eggplant’

Unlike a tonal language like Mandarin Chinese, South Kyungsang Korean does not allow all the possible tonal patterns over each syllable of a word, so it is considered a lexical pitch accent language (Kim & Jun, 2009; Lee & Davis, 2009; Lee & Zhang, 2014). The initial syllable of a prosodic word begins with either L or H, but a LL sequence cannot occur word-initially. Once a falling pitch contour occurs from H to L, another H tone is not allowed within the same word or phrase. In addition, three consecutive H or L tones (i.e., HHH, LLL) are not allowed either in this language (Lee & Davis, 2009; Lee & Zhang, 2014). Table 3.1 displays the possible tonal patterns over monosyllabic, disyllabic, trisyllabic, and quadrisyllabic words (Lee & Davis, 2009, p. 6).
Table 3.1. The possible tone patterns of monosyllabic, disyllabic, trisyllabic, and quadrisyllabic words in South Kyungsang Korean.

<table>
<thead>
<tr>
<th>Monosyllabic</th>
<th>Disyllabic</th>
<th>Trisyllabic</th>
<th>Quadrisyllabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>HH</td>
<td>HHL</td>
<td>HHLL</td>
</tr>
<tr>
<td>L</td>
<td>HL</td>
<td>HLL</td>
<td>LHHL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LHL</td>
<td>LHH</td>
</tr>
</tbody>
</table>

An Accentual Phrase (AP) is the lowest tonally defined prosodic unit and is marked by a low boundary tone at its initial edge (Kim & Jun, 2009; Kim, 2009). Although the tonal pattern of an AP depends on the number of syllables and the combination of tones, the surface tonal patterns of South Kyungsang Korean are generally predictable within a phrase. As previously stated, South Kyungsang Korean does not allow three consecutive H tones. When a three-digit string 333 forms a phrase, it is realized with HHL, not with HHH, although the underlying tone of each digit is H. When a string 555 forms a phrase, it is realized with LHL, not LLL, since three consecutive L tones cannot occur in this language. When there is a three-digit string like 329, the surface tonal pattern of this string is HLL, not HLH, since a H tone cannot appear again within the same phrase once there occurs a falling pitch contour from H to L.
Table 3.2. The numerical digits of South Kyungsang Korean, together with IPA and the tone pattern of each digit.

<table>
<thead>
<tr>
<th>Digit</th>
<th>IPA</th>
<th>Tone Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>/goŋ/</td>
<td>High</td>
</tr>
<tr>
<td>1</td>
<td>/il/</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>/i/</td>
<td>Low</td>
</tr>
<tr>
<td>3</td>
<td>/sam/</td>
<td>High</td>
</tr>
<tr>
<td>4</td>
<td>/sa/</td>
<td>Low</td>
</tr>
<tr>
<td>5</td>
<td>/o/</td>
<td>Low</td>
</tr>
<tr>
<td>6</td>
<td>/juk/</td>
<td>High</td>
</tr>
<tr>
<td>7</td>
<td>/teʰil/</td>
<td>High</td>
</tr>
<tr>
<td>8</td>
<td>/pʰal/</td>
<td>High</td>
</tr>
<tr>
<td>9</td>
<td>/gu/</td>
<td>High</td>
</tr>
</tbody>
</table>

3.1.5. Tokyo Japanese

Tokyo Japanese is a mora-timed language with lexical pitch accent (Amino & Osanai, 2014; Sugiyama, 2006). Words are lexically specified as accented and unaccented. The examples in (3) show a minimal pair (Venditti, 2005, p. 173) differing only by pitch accent, where an apostrophe (’) marks accentuation and a dash (⁻) indicates a mora boundary.

\[(3)\]
\[
a. \text{ u-e-ru-mo-no } \quad \text{LHHHH} \quad \text{‘something to plant’} \\
b. \text{ u-e’-ru-mo-no } \quad \text{LHLLL} \quad \text{‘the ones who are starved’}
\]
(3a) has phrase-initial rise to the second mora, and the pitch continues to be high towards the end. In (3b), phrase-initial rise occurs towards the accented mora, followed by a sharp fall, and the low pitch continues to the end.

Tokyo Japanese exhibits limited pitch patterns, as in (4) (Noguchi et al., 1999). The tonal pattern is dependent on the presence or the locus of pitch accent.

(4)  a. ya-ma-za’-ku-ra    LHHLL  ‘wild cherry’
     b. ka’-ge-bo-o-si     HLLLL  ‘shadow’
     c. mu-ra-sa-ki-i-ro    LHHHH  ‘purple’

(4a) shows initial rise to the second mora and remains high on the accented mora. After the accented mora, a sharp fall follows. (4b) begins with high pitch on the accented mora, followed by a sharp fall, and the low pitch continues to the end. (4c) shows a gradual initial rise to the second mora, and the high pitch remains to the end.

The lowest tonally-delimited prosodic grouping is an Accentual Phrase (AP) (Venditti, 2005). An AP normally includes one or two words. A low boundary tone typically marks the beginning and end of an AP (%L for the beginning and L% for the end). Within an AP, accentuation only occurs aligned with the leftmost high-toned mora, and the remaining morae exhibit suppressed tonal patterns towards the right edge of an AP (Oshima, 2005).

Tokyo Japanese has a special prosodic pattern for phone-number strings (Amino & Osanai, 2014). In producing phone-number strings, all the digits (0 to 9) become two morae (Amino & Osanai, 2014; Ito, 1990), although digits 2 and 5 are phonologically one mora, as observed in Table 3.3. In a phone-number string, every four morae (i.e., two
digits) belong together to form a prosodic phrase, called a bipodic template (BT) (Amino & Osanai, 2014). For example, a four-digit string is divided into two BTs – the first two digits belong to the first BT and the last two digits belong to the second BT. In a three-digit string, the first two digits are grouped together in one BT, and the last digit forms another BT. Accordingly, as the gray area of Figure 3.2 represents below, Tokyo Japanese has an accentual peak every two digits (i.e., every four morae). Put differently, there is only one accentuation peak in a three-digit phrase, whereas there are two accentuation peaks in a four-digit phrase.

Table 3.3. The numerical digits of Tokyo Japanese, together with Romaji and the mora type of each digit. A dash (-) indicates a mora boundary.

<table>
<thead>
<tr>
<th>Digit</th>
<th>Romaji</th>
<th>Mora type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>ze-ro</td>
<td>bimoraic</td>
</tr>
<tr>
<td>1</td>
<td>i-chi</td>
<td>bimoraic</td>
</tr>
<tr>
<td>2</td>
<td>ni</td>
<td>monomoraic</td>
</tr>
<tr>
<td>3</td>
<td>sa-n</td>
<td>bimoraic</td>
</tr>
<tr>
<td>4</td>
<td>yo-n</td>
<td>bimoraic</td>
</tr>
<tr>
<td>5</td>
<td>go</td>
<td>monomoraic</td>
</tr>
<tr>
<td>6</td>
<td>ro-ku</td>
<td>bimoraic</td>
</tr>
<tr>
<td>7</td>
<td>na-na</td>
<td>bimoraic</td>
</tr>
<tr>
<td>8</td>
<td>ha-chi</td>
<td>bimoraic</td>
</tr>
<tr>
<td>9</td>
<td>kyu-u</td>
<td>bimoraic</td>
</tr>
</tbody>
</table>
Figure 3.2. Time-normalized average pitch contours of five repetitions of the string \((105)-(601)-(2318)\) by five speakers. The red fonts indicate a digit string. The dotted vertical lines demarcate phrases. The area shaded in gray represents a string position with an accentual peak.

3.1.6. Standard French

Standard French does not have a lexically specified head, so prominence is determined at the phrase level (Beyssade & Marandin, 2006; Jun & Fougeron, 2002). In Standard French, two rising tones appear at the phrase level (Jun & Fougeron, 2002). The initial and final rises (LHi and LH*) mark the left and right edges of a phrase. In the model of French intonation (Jun & Fougeron, 2002), an Accentual Phrase (AP) is the lowest tonally defined prosodic constituent and the basic melody of an AP is LHLH*. The initial rise (LH) is optional, meaning that either L or H is not always realized. When the initial rise is fully realized, the first L is linked to the beginning of an AP, and H is associated with the first syllable of an AP-initial content word. The final rise (LH*) is obligatory, where L appears before the H-toned syllable, and H* is associated with the phrase-final
full vowel of the AP. When the AP begins with a content word, the first L is not often
realized, thus having HLH* (Jun & Fougeron, 2002). This suggests that HLH* is the
basic tonal melody of a three-digit string.

3.1.7. Suzhou Wu

Suzhou Wu is a language spoken in Suzhou in the southern part of Jiangsu Province,
China. It has seven citation tones, as observed in Table 3.4 (Lau, 2002): tones 1 and 6 are
high-level tones, mainly differing in duration; tone 2 is a dipping tone; tone 3 is a falling
tone; tone 4 is a falling-level tone; tone 5 is a convex tone; tone 7 is a rising tone.²

² There has been a long-standing debate over the exact pitch level of each tone (see Lau, 2002, p. 5 for
detailed information).
Table 3.4. Seven citation tones and their tones names in Suzhou Wu. The pitch level is based on the actual normalized pitch contour in Lau (2002).

<table>
<thead>
<tr>
<th>Tone number</th>
<th>Tone name</th>
<th>Pitch level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td>yinping</td>
<td>44</td>
</tr>
<tr>
<td>Tone 2</td>
<td>yangping</td>
<td>213</td>
</tr>
<tr>
<td>Tone 3</td>
<td>shangsheng</td>
<td>51</td>
</tr>
<tr>
<td>Tone 4</td>
<td>yinqu</td>
<td>522</td>
</tr>
<tr>
<td>Tone 5</td>
<td>yangqu</td>
<td>231</td>
</tr>
<tr>
<td>Tone 6</td>
<td>yinru</td>
<td>5</td>
</tr>
<tr>
<td>Tone 7</td>
<td>yangru</td>
<td>23</td>
</tr>
</tbody>
</table>

Suzhou Wu is known for its interesting tone sandhi phenomena. The tone pattern of a phrase is basically determined by the tone of a phrase-initial syllable. For example, as in Table 3.5, when the tone of the phrase-initial syllable is [44], the tone pattern becomes [4-31] for disyllables, [4-4-31] for trisyllables, and [4-4-4-31] for quadrisyllables, irrespective of the tones of the remaining syllables. When the tone of the phrase-initial syllable is either [213] or [231], the tone pattern merges into [2-23], regardless of the tones of the rest the syllable. Syllables beginning with the *yinqu* tone [522] yield two kinds of tone patterns, depending on the grammatical construction of syllables. When syllables form a verb-object phrase, the surface tone pattern becomes [4-22]. On the other hand, when they form a compound, it becomes [4-31]. In addition, the *yinru* and *yangru*

---

3 The pitch level here is based on the actual phonetic data found in Lau (2002). Please note that there has been a long-standing debate over the exact pitch level of each tone (See Lau, 2002, p. 5 for detailed information). In resolving the issue and in determining the exact nature of tone and tone sandhi rules in Suzhou Wu, Lau (2002) conducted a thorough production experiment using disyllabic, trisyllabic, and quadrisyllabic stimuli.
tones exhibit different tone patterns depending on whether the tone of the second syllable is long or short.

Table 3.5. The phonetic representations of seven citation tones and tone patterns of disyllables, trisyllables, and quadrisyllables (Based on Lau, 2002, p. 140).

<table>
<thead>
<tr>
<th>Pitch level (Tone name)</th>
<th>Tone patterns (2syll)</th>
<th>Tone patterns (3syll)</th>
<th>Tone patterns (4syll)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[44] (yingping)</td>
<td>[4-31]</td>
<td>[4-4-31]</td>
<td>[4-4-4-31]</td>
</tr>
<tr>
<td>[213] (yangping)</td>
<td>[2-23]</td>
<td>[2-23-21]</td>
<td>[2-23-3-21]</td>
</tr>
<tr>
<td>[51] (shangsheng)</td>
<td>[42-2]</td>
<td>[42-2-21]</td>
<td>[42-2-2-21]</td>
</tr>
<tr>
<td>[522] (yinqu) V-O Phrase</td>
<td>[42-2]</td>
<td>[42-2-21]</td>
<td>[42-2-2-21]</td>
</tr>
<tr>
<td></td>
<td>Compound</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[4-31]</td>
<td>[4-4-31]</td>
<td>[4-4-4-31]</td>
</tr>
<tr>
<td>[231] (yangqu)</td>
<td>[2-23]</td>
<td>[2-23-21]</td>
<td>[2-23-3-21]</td>
</tr>
<tr>
<td>[5] (yinru) Long</td>
<td>[4-2]</td>
<td>[4-2-21]</td>
<td>[4-2-2-21]</td>
</tr>
<tr>
<td></td>
<td>Short</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[4-4]</td>
<td>[4-4-21]</td>
<td>[4-4-4-21]</td>
</tr>
<tr>
<td>[23] (yangru) Long</td>
<td>[2-31]</td>
<td>[2-31-1]</td>
<td>[2-32-2-21]</td>
</tr>
<tr>
<td></td>
<td>Short</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[2-3]</td>
<td>[2-3-1]</td>
<td>[2-3-2-21]</td>
</tr>
</tbody>
</table>

However, each digit seems to retain its citation tone when a phone-number string is read. Figure 3.3 displays the pitch contour of a digit string 980-673-2514, which is divided into three phrases like (980)-(673)-(2514). In this figure, the red digits indicate each digit of the string and the green digits indicate each citation tone of each digit. In the first phrase, since the phrase-initial tone is [51], the tone pattern should come out to [42-2-21]. However, the tone pattern does not follow Suzhou Wu’s tone sandhi rule; instead the pitch pattern basically reflects the citation tone of each digit. As clearly indicated by Figure 3.3, the pitch pattern of the first phrase 980 is close to the citation tone of each
digit; the first digit 9 shows a (high-)falling tone, the second digit 5 a high-level tone, and the third digit 0 a dipping tone. The second and third phrases of this string do not follow the tone sandhi rule, either. Instead, the pitch patterns are expressed through the citation tone of each digit.

Figure 3.3. Time-normalized average pitch contours of the string (980)-(673)-(2518). The dotted vertical lines demarcate phrases. The red digits indicate a digit string and the green digits the citation tone of each digit.

3.2. Three types of prosodic typologies

3.2.1. Rhythm-based typology

One of the most well-known prosodic typologies is to classify languages into either stress-timing or syllable-timing categories based on the rhythmic unit of the language (e.g., Abercrombie, 1967; Pike, 1945). Stress-timing means that the temporal duration between the intervals of stressed syllables is equal. Syllable-timing means that the
duration of every syllable is equidistant. There is also another kind of rhythmic unit: mora-timing (e.g., Bloch, 1950), in which the duration of every mora is equidistant. According to their rhythmic units, English and German are typical stress-timed languages, French, Spanish, and Italian are typical syllable-timed languages, and Japanese is a typical mora-timed language (Mok, 2009).

Within the continuum of speech rhythm, a language like English falls towards one end as stress-timed, while a language like French falls towards the other end as syllable-timed (Dauer, 1987). Although some languages can be clustered in this way, the speech rhythm of many languages is uncertain. For example, Seoul Korean has been regarded in some instances as syllable-timed (Kim, Davis, & Cutler, 2008), in others as stress-timed for older speakers (Lee, Jin, Seong, Jung, & Lee, 1994), and in still other cases as mora-timed (Moon-Hwan, 2004). However, works by Mok and Lee (2008) and Arvaniti (2012) suggest that Seoul Korean is closer to being a syllable-timed language despite showing characteristics of a middle stage between the two classes. This supports the view that many languages actually fall somewhere within the continuum of speech rhythm (Nespor, 1990).

Although the rhythmic status of some languages is more or less intermediate between stress-timing and syllable-timing, we have attempted to cluster seven languages into three rhythmic classes. This grouping was made based on published data. For languages for which there is no published data, we made a decision based on consultation with researchers working on that language. Please note that the grouping in Table 3.6 may be debatable among researchers. When the rhythmic status of a certain language was not clear, (√) was given.
Table 3.6. The rhythmic unit of each language.

<table>
<thead>
<tr>
<th>Language</th>
<th>Rhythmic unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stress-timed</td>
</tr>
<tr>
<td>American English</td>
<td>√</td>
</tr>
<tr>
<td>Mandarin Chinese</td>
<td></td>
</tr>
<tr>
<td>Standard French</td>
<td></td>
</tr>
<tr>
<td>Seoul Korean</td>
<td></td>
</tr>
<tr>
<td>South Kyungsang Korean</td>
<td></td>
</tr>
<tr>
<td>Tokyo Japanese</td>
<td></td>
</tr>
<tr>
<td>Suzhou Wu</td>
<td></td>
</tr>
</tbody>
</table>

Notes: American English data is from Ramus, Nespor, and Mehler (1999); Mandarin Chinese is from Mok (2009); Standard French is from Ramus, Nespor, and Mehler (1999); South Kyungsang Korean data is from Jun (p.c.); Seoul Korean data is from Mok and Lee (2008); and Suzhou Wu data is from Kuang (p.c.).

3.2.2. Accent-based typology

Another prosodic typology is through clustering languages based on the distinctions: 1) between stress vs. non-stress accent, and 2) between lexical vs. post-lexical vs. no pitch accent. Beckman (1986) first used the terms stress accent and non-stress accent while comparing English with Japanese. She defined a stress-accent language (e.g., English) as a language that uses metrically prominent positions (i.e., stressed positions) derived from a combination of increased duration, pitch, and intensity, while a non-stress accent language (e.g., Japanese) is one that only uses pitch to make lexical contrasts. Ladd (1996) further expanded Beckman’s idea and included “lexical pitch accent features.” His
approach stemmed from the fact that languages like Swedish have lexical specifications of pitch together with stress accent while other languages such as Bengali have no stress accent but have post-lexical pitch accent derived from a word. Therefore, as Table 3.7 indicates, stress-accented languages such as Swedish and English differ from each other: Swedish has a lexically specified pitch feature, whereas English has post-lexical pitch accent that is assigned to the most prominent syllable of a word within a phrase. Non-stress-accented languages such as Japanese and Bengali also differ from each other: Japanese has lexical pitch accent, while Bengali has post-lexical pitch accent.

Table 3.7. A typology based on the distinctions between stress vs. non-stress accent and between lexical vs. post-lexical pitch accent (Ladd, 1996, p. 156).

<table>
<thead>
<tr>
<th>Lexical typology</th>
<th>Phonetic typology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stress accent</td>
</tr>
<tr>
<td>Lexical pitch accent</td>
<td>Swedish</td>
</tr>
<tr>
<td>Post-lexical pitch accent</td>
<td>English</td>
</tr>
</tbody>
</table>

Furthermore, Lindström and Remijsen (2005) highlighted that some languages like Wolof and Kuot do not use pitch to mark stress, rather rely on other parameters such as duration, intensity, vowel quality or a combination of those elements. Therefore, another column entitled “no pitch accent” was added to accommodate languages comparable to Wolof and Kuot, as shown in Table 3.8. This table has a cell labeled “impossible” since there exists no language with neither pitch accent nor stress accent simultaneously (Lindström & Remijsen, 2005).
Table 3.8. A typology of factors determining pitch accents, after Ladd (1996, p. 156), expanded to include the type represented by Wolof (Lindström & Remijsen, 2005, p. 847).

<table>
<thead>
<tr>
<th>Lexical typology</th>
<th>Phonetic typology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stress accent</td>
</tr>
<tr>
<td>Lexical pitch accent</td>
<td>Swedish</td>
</tr>
<tr>
<td>Post-lexical pitch accent</td>
<td>English</td>
</tr>
<tr>
<td>No pitch accent</td>
<td>Wolof, Kuot</td>
</tr>
</tbody>
</table>

Nevertheless, as Table 3.9 indicates, we have not been able to cluster all of the seven languages even based on the extended typology shown in Table 3.8. First, tonal languages like Mandarin Chinese and Suzhou Wu do not conform to the framework. The reason is that lexical pitch accents and lexical tones differ from each other. Lexical pitch accents are assigned to one syllable in a word, whereas lexical tones are assigned to every syllable in a word. Therefore, Mandarin Chinese and Suzhou Wu cannot share a cell with Tokyo Japanese and South Kyungsang Korean. Second, Tokyo Japanese also may not fit into this typology in a certain circumstance, particularly when it is read in a phone-number string, the reason being that its tonal patterns are shaped by a bipodic template rather than lexical pitch accents. Third, Standard French and Seoul Korean cannot be included in this table since in these languages pitch is not used to make a certain syllable more prominent than its surrounding syllables at the word level. Rather, pitch events are “associated with a particular syllable of a phrase rather than a word” (Lindström & Remijsen, 2005, p. 846). As a consequence, this prosodic typology shows a limitation since it only covers languages marking prosodic features at the word level excluding lexical tones.
Table 3.9. The classification of the seven languages according to the accent-type typology (AE: American English, TJ: Tokyo Japanese, SKK: South Kyungsang Korean).

<table>
<thead>
<tr>
<th>Phonetic typology</th>
<th>Stress accent</th>
<th>Non-stress accent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical pitch accent</td>
<td>TJ, SKK</td>
<td></td>
</tr>
<tr>
<td>Lexical typology</td>
<td>Post-lexical pitch accent</td>
<td>AE</td>
</tr>
<tr>
<td>No pitch accent</td>
<td>“Impossible”</td>
<td></td>
</tr>
</tbody>
</table>

*Note: In Tokyo Japanese, only natural stimuli conform to this typology.

3.2.3. Prominence-based typology

The next and most recent prosodic typology is to classify languages according to their types of word-level and phrase-level prominence marking (Jun, 2005, 2012, 2014). The word-level (or lexical) prominence marking includes lexical stress, lexical tone, and lexical pitch accent. The phrase-level (post-lexical) prominence marking consists of three categories: head-prominent, head/edge-prominent, and edge-prominent. In the head-prominence languages, prominence is derived from a head (designated syllable) of a word. This type of prominence appears in languages with lexical stress, lexical tone, and lexical pitch accent. In the head/edge-prominence languages, prominence is determined by the head or the edge of a word. For example, Tokyo Japanese belongs to this category since prominence is derived from both lexical pitch accent and boundary tones at the...

---

4 Here, we did not consider a macro-rhythm component, which has been proposed since Jun (2012, 2014). The reason is that the number of languages under study seems insufficient to include that component.
edge of a phrase.\textsuperscript{5} Edge-prominence languages do not have a lexically specified head at the word level; instead, prominence is derived at the initial or final edge of a phrase.

Unlike the accent-based typology, we have been able to classify all of the seven languages within the prominence-based framework. As Table 3.10 displays, word-level prominence marking consists of three categories which capture the features of word prosody in a given language. Phrase-level prominence marking also includes three categories based on the realization of prominence in a phrase. Please note that when a given category is not agreed upon among researchers, (✓) is given.

American English and Mandarin Chinese are head-prominence languages given that prominence marking is cued by a head of a word via lexical stress (American English) and via lexical tone (Mandarin Chinese). Seoul Korean is an edge-prominence language since it has neither lexical stress nor lexical pitch accents. Prominence marking is basically cued by a boundary tone marking the right edge of an AP, or it optionally occurs at the initial edge of an AP by aspirated/tense consonants or a high tone digit [il]. In addition, Tokyo Japanese belongs to the edge-prominent category when it is read in a phone-number string, the reason being that phone-number strings produce fixed tonal patterns through a bipodic template. South Kyungsang Korean and Tokyo Japanese (for natural stimuli) are head/edge-prominence languages because prominence is cued by a head of a word via lexical pitch accent, and a tonal pattern is mainly determined at the phrase level. Standard French is also a head/edge-prominence language because prominence is derived from “a rising pitch accent (LH*) simultaneously marking the edge of an AP” (Jun, 2012, p. 537). Finally, Suzhou Wu can be either head-prominent or

\textsuperscript{5} The term “tone” is used not as lexical but post-lexical, as conventionally used in the ToBI framework.
head/edge-prominent depending on the style. When reading a phone-number string, Suzhou Wu is considered a head-prominence language given that the citation tone of each digit primarily shapes the pitch contour of a digit string. In natural stimuli, Suzhou Wu is deemed a head/edge-prominence language since the pitch pattern is mainly affected by the tone of a phrase-initial syllable.


<table>
<thead>
<tr>
<th>Language</th>
<th>Style</th>
<th>Prominence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Word level</td>
<td>Phrase level</td>
</tr>
<tr>
<td></td>
<td>Tone</td>
<td>Stress</td>
</tr>
<tr>
<td>AE</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>SKK</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>TJ</td>
<td>Digit string</td>
<td>√</td>
</tr>
<tr>
<td>Natural</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Stimuli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW</td>
<td>Digit string</td>
<td>√</td>
</tr>
<tr>
<td>Natural</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Stimuli</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

So far, we have described the prosodic structures of each language, and we have also introduced three types of prosodic typologies. We believe that prosodic marking of focus is expressed through each language’s prosodic structure. For this reason, it will be
interesting to see what kinds of rhythmic units and/or prosodic features are main factors contributing to the prominence of prosodic focus. Furthermore, our interest lies in what kinds of prosodic typologies may line up with different characteristics of prosodic focus across languages. We will come back to this point in Chapter 7.
Chapter 4

A cross-linguistic study of prosodic focus

This chapter presents production and perception experiments that examine not only whether prosodic marking of focus varies across languages but also how it varies. Although prosodic focus has received a great deal of attention in the literature, there is no clear picture of its exact nature across languages. An important source of discrepancy is methodological since there have been no comparable data collection procedures. Therefore, in order to compare data with the same methodology, phone-number strings in the style of American English are implemented and analyzed based on the collected data of seven languages, as previously described in Chapter 2.

4.1. Introduction

Conventional wisdom about the prosodic reflex of focus is that a focused element attracts prominence-related effects by suprasegmental features. However, languages display various prosodic properties of focus cross-linguistically (Jun, 2011; Selkirk, 2008a; Zerbian, 2006). For example, languages like English (Cooper et al., 1985; Xu & Xu, 2005), German (Baumann et al., 2006), and Dutch (Swerts et al., 2002) mark prosodic
focus with a nuclear pitch accent aligning it with a primary stressed syllable. Languages like Korean (Jun & Lee, 1998; Lee & Xu, 2010) and Japanese (Pierrehumbert & Beckman, 1988) use prosodic phrasing to express prosodic focus by manipulating an accentual phrase (AP). Languages like European Portuguese (Frota, 2002) and Bengali (Selkirk, 2008a) employ a combination of pitch accent and prosodic phrasing to signal prosodic focus.

Although languages use various means for prosodic focus, the widely accepted assumption is that a focused element is “maximally prominent” in a sentence (Büring, 2010; Samek-Lodovici, 2005; Truckenbrodt, 1995), as reflected by longer duration, greater intensity, and higher pitch in the phonetic implementation. As discussed in Chapter 1, previous studies have examined the prosodic effects of focus in many languages to prove such a focus-prominence relationship. Some representative work includes: Cooper, Eady, and Mueller (1985) and Xu and Xu (2005) for English; Jun and Lee (1998) and Lee and Xu (2010) for Korean; Xu (1999) for Mandarin; Lee and Xu (2012) for Japanese; and Dohen and Lœvenbruck (2004) for French.

Recently, however, a growing body of evidence demonstrates that focus is not always correlated with maximal prominence (e.g., Downing, 2008; Fiedler & Jannedy, 2013; Gordon, 2007). For example, Gordon (2007) found that focus is not primarily encoded by prosody in the American Indian Language of Chickasaw, but rather it is the morphology that plays a primary role due to the presence of focus morphemes. Downing (2008) analyzed three Bantu languages (Chichewa, Durban Zulu and Chitumbuka), in which prominence is conditioned by position in a sentence – phrasal prominence occurs on the last word of a phrase and sentence prominence occurs on the last word of a
sentence. Downing found a mismatch between prominence and the position of prosodic focus in these languages. Although a phrase-initial or phrase-medial word was focused, these focused words did not receive prominence. Instead, prominence occurred on the phrase-final word since phrasal prominence is fixed at the last word of a phrase. These findings suggest that there is no direct relationship between prominence and the position of prosodic focus. In order to assess the cross-linguistic validity of this possibility, we need to raise the question of whether or not focused elements are actually prominent and clearly marked in both production and perception.

Reliable quantitative cross-linguistic comparisons of prosodic focus are challenging. This is primarily because languages use different levels of linguistic structure to convey prosodic focus. As previously discussed in Chapter 1, syntactic rearrangements are required for getting prominence in some languages like Hungarian (Balogh, 2007; Brody, 1990; Onea, 2009) and Catalan (Estebas-Vilaplana, 2000; Vallduví, 1990). In contrast, without changing the syntactic structure, prosody alone is sufficient to encode prosodic focus in other languages like English (Ladd, 1996; Xu & Xu, 2005) and Dutch (Swerts et al., 2002). Furthermore, in the case of Seoul Korean using prosodic phrasing to express prosodic focus, an obstacle to quantifying prosodic effects comes with focus particles, such as -man meaning “only.” In this case, the focus particle occurs at the end of an accentual phrase (AP), where phrasal prominence is derived. Therefore, Seoul Korean shows a confounding prosodic effect in a situation where an NP is contrastively focused, meaning that it is not certain whether prominence is directly related to the focus particle -man or it is just phrasal prominence. Given that different
languages use different mechanisms to express prosodic focus, a single platform is required to quantify prosodic focus effects in a comparable way.

In this project, the same platform was built with a phone-number string in the style of American English, as described in Chapter 2. This method enables us to quantify various prosodic focus effects, such as pitch expansion and post-focus compression, in a comparable way within the framework of a large-scale cross-linguistic project. This project will allow us to directly compare the different levels of pitch expansion, post-focus compression, and/or prosodic phrasing across languages. In addition, we can examine whether or not prosodic focus of each language is well identifiable and further analyze which language produces better identification performance.

To summarize, this chapter aims to examine whether and how “purely” prosodic marking of focus – as a general mechanism for communication of information structure – varies across languages in both production and perception. Our examination expects to provide further details regarding 1) to what extent different languages produce various prosodic focus effects in production; and 2) which language allows a better identification rate in perception.

4.2. Production

4.2.1. Data collection

The stimuli, subject recruitment, and recording procedure were the same as described in Chapter 2.
4.2.2. Acoustic measurements

Digit boundaries were manually labeled. Based on those boundaries, we obtained duration in milliseconds, mean intensity in dB, and mean pitch (or pitch range for Mandarin Chinese and Suzhou Wu contour tones) in Hz by using a Praat script (Xu, 2013). Each pitch contour was carefully examined to spot pitch tracking errors (i.e., pitch halving or pitch doubling). Any pitch target or point strongly deviating from the rest was determined to be an error, and it was corrected by hand. In this chapter, the duration values in milliseconds are expressed as percentages: the duration of the corrective-focus was calculated as a percentage relative to that of the broad-focus. In addition, the obtained pitch values were converted to semitones. A semitone refers to the musical interval between two neighboring notes logarithmically measured on a 12-tone musical scale. Pitch is a major auditory feature of musical tones (Plack, Oxenham, & Fay, 2005), and it is known as logarithmic in nature both in terms of production (Fujisaki, 2003) and perception (Nolan, 2003). This is the motivation for using semitones in this study. The conversion was done by applying the following equation (Xu & Wang, 2009): 

\[ st = 12 \log_2 F0. \]

4.2.3. Analyses

The phone-number strings in the broad-focus condition were directly compared with the same sequences in the corrective-focus condition by the aggregate measures of mean
pitch (or pitch range for Mandarin Chinese and Suzhou Wu contour tones), duration, and mean intensity. Hereafter, we will refer to each respective parameter as pitch, duration, and intensity, for simplicity’s sake.

For simple graphical exploration, we calculated 10 10-element vectors of the mean values aggregated by all digits of each position for corrective focus. We refer to them as CF1mean, CF2mean, ... CF10mean. For example, CF1mean consists of a 10-element vector, in which the first element is the mean value for corrective focus for position 1, and the other elements are the mean values of position 2, position 3, ... position 10. We also calculated 10 similar vectors, one for each broad-focus position, calling them BF1mean, BF2mean, ... BF10mean. Figure 4.1 presents a simple plot subtracting BF1mean from CF1mean. This plot delivers a simple and clear message about prosodic focus effects – focus in position 1 is clearly marked by a higher pitch and post-focus compression is clearly shown over the post-focus positions.

Figure 4.1. Prosodic differences between broad and corrective focus in American English when position 1 is focused. Character 1 represents focus position.
In this project, we plotted 10 10-element vectors for each respective parameter to capture the prosodic differences between broad focus and corrective focus in the 10-digit phone-number strings.

\[
\begin{align*}
\text{CF1mean-BF1mean} \\
\text{CF2mean-BF2mean} \\
\vdots \\
\text{CF10mean-BF10mean}.
\end{align*}
\]

4.2.4. Results

It is not ideal to examine all seven languages simultaneously. This is mainly because plotting the production data from all the languages at once would literally require too much space, rendering it difficult to interpret the data. In order to emphasize the different behaviors in prosodic marking of focus, two languages were chosen as a pair: one language that shows a clear prosodic marking of focus and another language that shows a weak and ambiguous pattern. In addition, Suzhou Wu was analyzed by itself since it has only five focus positions unlike the other languages, which have ten focus positions. Therefore, we analyzed the production data in the following four groups: American English and Seoul Korean, Mandarin Chinese and South Kyungsang Korean, Standard French and Tokyo Japanese, and Suzhou Wu.

4.2.4.1. American English and Seoul Korean
Figure 4.2 shows clear differences in prosodic focus effects between American English and Seoul Korean. In American English, there was a clear and consistent effect of focus marking in all the parameters. Focus positions were clearly indicated by greater pitch, duration, and intensity. American English also showed clear post-focus compression with reduced pitch, duration, and intensity in post-focus positions. In comparison, Seoul Korean exhibited no clear prosodic focus effects in focus positions – prosodic modulation by focus was weak and ambiguous. The amount of modulation by focus for pitch was close to or less than a third of the size of American English (0.96 st vs. 2.80 st). The durational cues of focus positions were very small – on average, about 3.8% extra duration in focus positions. The ambiguity of focus modulation was clearly illustrated by pitch: when position 1 was focused, position 2 was even higher; when position 4 was focused, position 5 was even higher; when position 7 was focused, position 8 was even higher. The intensity effects of prosodic focus were not clear at all. It seems that there was no clear indication of focus positions.
Figure 4.2. Prosodic differences between focused digits and broad-focus counterparts in American English and Seoul Korean. Characters 1-A refer to positions 1 to 10.

4.2.4.2. Mandarin Chinese and South Kyungsang Korean

Figure 4.3 displays the prosodic focus effects in both Mandarin Chinese and South Kyungsang Korean. As Figure 4.3 illustrates, focus positions in Mandarin Chinese were clearly marked by higher pitch, longer duration, and greater intensity. In addition, post-focus positions exhibited clear post-focus compression effects with considerably reduced
pitch, duration, and intensity. In contrast, South Kyungsang Korean (which is similar to Seoul Korean) showed relatively weak and ambiguous modulation by focus. Focus positions were not clearly indicated by increased pitch, duration and intensity. Instead, an increase in pitch was often found in neighboring positions: when position 4 was focused, position 5 was even higher; focus on 7 made 8 higher; focus on 9 made 10 higher. Although the durational cues of the focus positions were fairly clear with an increase of about 12% on average, increased duration was also found in adjacent positions. For example, when position 2 was focused, position 1 was even longer, and when positions 5 and 8 were focused, positions 4 and 7 were also increased to a similar extent. For the intensity effects, South Kyungsang Korean showed no clear focus markings in the focus positions.
Figure 4.3. Prosodic differences between focused digits and broad-focus counterparts in Mandarin Chinese and South Kyungsang Korean. Characters 1-A refer to positions 1 to 10.

4.2.4.3 Standard French and Tokyo Japanese

Figure 4.4 illustrates clearly different patterns of prosodic focus between Standard French and Tokyo Japanese. Standard French showed a clear prosodic marking of focus in each position of a corrected digit that was expressed by a considerable increase in duration,
pitch, and intensity. Furthermore, immediately following the focus position, sharp drops in duration, pitch, and intensity characterized the post-focus positions. On the contrary, very similar to Seoul Korean and South Kyungsang Korean, Tokyo Japanese produced relatively weak, ambiguous, and unclear prosodic modulation by focus. Pitch was not considerably increased – on average, an increase of only 1.23 st was observed in focus positions. Instead, an increase in pitch was often found in adjacent positions: when position 1 was focused, position 2 was even higher; when position 4 was focused, position 5 was even higher; when position 7 was focused, position 8 was even higher; when position 9 was focused, position 10 was even higher. Although post-focus compression was shown in some positions, it was not clearly visualized over the post-focus positions. For example, although the pitch value continued to drop until position 5 when position 1 was focused, a pitch rebound occurred at position 6. A similar trend was also observed for the first and second focus positions. In addition, the durational cues of the focus positions were not clear enough, often seeming ambiguous; when position 2 was focused, position 1 was even longer; when position 5 was focused, position 4 was even longer; when position 10 was focused, position 9 was even longer. Finally, the intensity effects of prosodic focus were not clear at all in Tokyo Japanese. In this case, all focus positions showed an intensity value of no greater than 0.
Figure 4.4. Prosodic differences between focused digits and broad-focus counterparts in Standard French and Tokyo Japanese. Characters 1-A refer to positions 1 to 10.

4.2.4.4. Suzhou Wu

In Suzhou Wu, the overall prosodic marking of focus was fairly clear based on higher pitch, longer duration, and greater intensity. In addition, post-focus compression was clearly seen over the post-focus positions through reduced pitch, duration, and intensity. However, certain focus positions were not clearly marked by increased duration and
intensity, especially compared to surrounding positions. For example, when position 5 was focused, positions 4 and 6 were also lengthened with an increase of 6.1% and 4.3%, respectively. When position 10 was focused, position 9 was also lengthened with an increase of 5.9%. Furthermore, with respect to the intensity effects of prosodic focus, when position 5 was focused, position 4 had an even greater intensity value. Similarly, but to a lesser extent, when position 10 was focused, an increase in intensity was also observed for position 9.

Figure 4.5. Prosodic differences between focused digits and broad-focus counterparts in Suzhou Wu. Characters 1, 3, 4, 5, and A indicate focus positions 1, 3, 4, 5, and 10.
4.2.4.5 The phonological unit that carries prosodic focus

Although the production data enabled us to get an idea of what the phonological unit carrying prosodic focus is in each language, we were not able to determine the exact phonological unit carrying prosodic focus, especially for the languages of Seoul Korean, South Kyungsang Korea, and Tokyo Japanese. This was mainly because, in some cases, the “messy” distribution of prosodic marking of focus was somewhat difficult to analyze. Therefore, in an attempt to get a better picture of the phonological unit carrying prosodic focus in each language, we simply plotted a comparison between the broad- and corrective-focus conditions.
Figure 4.6. The mean pitch contours of the broad- and corrective-focus conditions. Each pitch contour includes focus in position 1. The area shaded in gray refers to a prosodic phrase that includes the first three digits.
Figure 4.6 illustrates the prosodic differences between broad and corrective focus when position 1 was focused.\(^1\) With respect to the phonological unit carrying prosodic focus, the languages under study can be classified into three groups. One group includes the languages of American English, Mandarin Chinese, and Standard French. Although post-focus positions show some variance between these languages, prosodic marking of focus was clearly marked in the focus position, and post-focus positions featured a reduced pitch range beginning right after the focus position. Accordingly, we can say that the phonological unit of carrying prosodic focus is a word (i.e., each digit) and that post-focus compression starts right after the focused digit in these languages. The other group consists of the languages Tokyo Japanese, Seoul Korean, and South Kyungsang Korean, in which prosodic marking of focus was not clear. As the middle panel of Figure 4.6 clearly shows, when position 1 was focused, position 2 also showed an increase in pitch in all three languages. Furthermore, post-focus positions were produced with a minimal level of post-focus compression, and more importantly, post-focus compression did not occur right after the focus position; instead it was observed in the next phrase. Therefore, we can say that the phonological unit carrying prosodic focus is a phrase and that post-focus compression appears across the phrase boundary. The third group includes the language of Suzhou Wu. As Figure 4.6 shows, it is likely that the phonological unit carrying prosodic focus is a word (i.e., each digit). Yet, since the effects of prosodic focus turned out to be rather mixed in certain positions (as demonstrated in Figure 4.6), we are

\(^1\) Although Figure 4.6 cannot depict the whole story about the other focus positions, a similar trend was observed throughout all focus positions.
not able to make a straightforward interpretation of the phonological unit carrying prosodic focus in Suzhou Wu.

4.3. Perception

Given these striking differences in production, we were not surprised to see equally striking differences in perception among these seven languages. Again, we simply divided the languages into four groups in the same way as was done for production.

4.3.1. Data collection and analyses

The stimuli, design, and procedure were the same as described in Chapter 2. The perception data were classified in a confusion matrix to see how accurately each focus position was identified.

4.3.2. Results

4.3.2.1. American English and Seoul Korean

Table 4.1 below shows a confusion matrix for the identification of focus positions in the two languages, American English and Seoul Korean. We observed a striking difference between these languages. Listeners identified the focus positions 97.2% of the time in American English – each focus position was clearly identified. In comparison, focus
positions were identified at a rate of 37.3% in Seoul Korean. The confusion matrix of Seoul Korean demonstrated that incorrect answers usually occurred within the same phrase (demarcated by dotted lines) before or after focus positions. For example, when position 1 was focused, positions 2 and 3 were identified at a rate of 16.8% and 22.9%, respectively. When position 2 was focused, positions 1 and 3 were identified at a rate of 16.5% and 22.2%, respectively. Given that the chance level is 10% (=100/10), the rate of incorrect answers is neither random nor negligible since it was actually above the level of chance, confirmed by a binary logistic regression analysis ($\chi^2 = 5.99$, $df = 1$, $p = 0.014$). Other focus positions also showed a similar trend, meaning that prosodic marking of focus was actually ambiguous in Seoul Korean.
Table 4.1. Confusion matrix of corrective focus perception (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string. (Top panel: American English; Bottom panel: Seoul Korean)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target</strong></td>
<td>1</td>
<td>16.5</td>
<td>36.5</td>
<td>22.1</td>
<td>5.9</td>
<td>3.2</td>
<td>4.7</td>
<td>6.2</td>
<td>1.2</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10.6</td>
<td>4.4</td>
<td>51.8</td>
<td>5.6</td>
<td>1.8</td>
<td>7.9</td>
<td>5.6</td>
<td>3.2</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7.9</td>
<td>5.0</td>
<td>12.9</td>
<td>35.9</td>
<td>5.0</td>
<td>12.6</td>
<td>12.9</td>
<td>1.5</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8.5</td>
<td>9.1</td>
<td>5.6</td>
<td>18.2</td>
<td>38.5</td>
<td>12.9</td>
<td>4.4</td>
<td>1.2</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5.3</td>
<td>2.6</td>
<td>7.9</td>
<td>10.9</td>
<td>3.8</td>
<td>45.9</td>
<td>14.4</td>
<td>6.5</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>8.8</td>
<td>5.9</td>
<td>8.2</td>
<td>4.1</td>
<td>0.9</td>
<td>4.4</td>
<td>41.2</td>
<td>7.9</td>
<td>14.4</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>15.9</td>
<td>9.1</td>
<td>9.7</td>
<td>7.6</td>
<td>4.1</td>
<td>7.1</td>
<td>20.0</td>
<td>17.9</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>7.1</td>
<td>2.9</td>
<td>8.5</td>
<td>14.4</td>
<td>2.4</td>
<td>5.3</td>
<td>13.5</td>
<td>6.5</td>
<td>36.8</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.9</td>
<td>4.1</td>
<td>12.6</td>
<td>5.9</td>
<td>2.4</td>
<td>6.2</td>
<td>13.5</td>
<td>2.6</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>31.5</td>
<td>16.8</td>
<td>22.9</td>
<td>8.2</td>
<td>4.4</td>
<td>4.7</td>
<td>6.2</td>
<td>1.2</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Even if we score by phrase rather than by position, the overall identification rate would become 63.4% for Seoul Korean as indicated by Table 4.2. It should be noted that
this identification rate is still much lower than that of American English, suggesting that focus marking by prosodic modulation was actually weak in Seoul Korean.

<table>
<thead>
<tr>
<th>Target</th>
<th>Perceived</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st phrase</td>
</tr>
<tr>
<td>1st phrase</td>
<td>71.0</td>
</tr>
<tr>
<td>2nd phrase</td>
<td>21.7</td>
</tr>
<tr>
<td>3rd phrase</td>
<td>25.4</td>
</tr>
</tbody>
</table>

### 4.3.2.2. Mandarin Chinese and South Kyungsang Korean

A confusion matrix in Table 4.3 contains information about the identification of focus positions in Mandarin Chinese. We observed that Mandarin Chinese listeners identified the focus positions at a rate of 87.7%. In contrast to our expectations, this identification rate seemed a bit low, given that each focus position showed a clear prosodic marking of focus by having greater pitch, duration, and intensity. In order to examine whether or not this low identification performance was related to each tone type, we analyzed the perception data on a tone-by-tone basis.
Table 4.3. Confusion matrix of corrective focus perception for Mandarin Chinese (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>85.5</td>
<td>7.7</td>
<td>1.9</td>
<td>1.5</td>
<td>0.4</td>
<td>0.8</td>
<td>1.6</td>
<td>0.0</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>86.4</td>
<td>5.2</td>
<td>1.3</td>
<td>1.2</td>
<td>0.1</td>
<td>1.4</td>
<td>0.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>2.2</td>
<td>5.1</td>
<td>87.7</td>
<td>3.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.8</td>
<td>0.4</td>
<td>0.7</td>
<td>88.8</td>
<td>5.6</td>
<td>2.4</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>5</td>
<td>0.4</td>
<td>0.0</td>
<td>1.5</td>
<td>6.0</td>
<td>87.1</td>
<td>4.4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>1.2</td>
<td>0.5</td>
<td>0.8</td>
<td>0.7</td>
<td>4.2</td>
<td>90.7</td>
<td>1.5</td>
<td>0.3</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>7</td>
<td>1.3</td>
<td>0.4</td>
<td>0.9</td>
<td>1.6</td>
<td>0.8</td>
<td>1.4</td>
<td>90.9</td>
<td>2.4</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
<td>7.6</td>
<td>90.5</td>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>9</td>
<td>0.6</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
<td>1.9</td>
<td>95.1</td>
<td>1.3</td>
</tr>
<tr>
<td>10</td>
<td>0.9</td>
<td>2.6</td>
<td>1.3</td>
<td>0.8</td>
<td>1.3</td>
<td>1.5</td>
<td>4.3</td>
<td>1.7</td>
<td>11.5</td>
<td>74.2</td>
</tr>
</tbody>
</table>

Table 4.4 illustrates the tone-by-tone identification rate in each focus position, where the columns refer to positions in a string and the rows to tones 1-4. Table 4.4 demonstrates that although focus positions were well identified for other focused tones (91.3% overall), focused tone 3 digits received a relatively poor identification rate (77.1% overall). Another important observation was that position 10 was not well identified, except for tone 2. The reason for this low identification rate will be explored in the Discussion section with a comparison of American English results.

---

2 The reasons for the (relatively) poor identification performance are due to smaller opportunity for pitch range expansion and confusion from local dissimilatory effects. We will explore these prosodic behaviors of focused tone 3 syllables in Chapter 5.
Table 4.4. Position-by-position identification rates of tones 1-4.

<table>
<thead>
<tr>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td>95.0</td>
<td>82.5</td>
<td>88.8</td>
<td>97.5</td>
<td>98.8</td>
<td>88.8</td>
<td>98.8</td>
<td>95.0</td>
<td>98.8</td>
<td>63.8</td>
<td><strong>90.8</strong></td>
</tr>
<tr>
<td>Tone 2</td>
<td>85.0</td>
<td>95.0</td>
<td>90.0</td>
<td>90.0</td>
<td>75.0</td>
<td>100.0</td>
<td>80.0</td>
<td>100.0</td>
<td>95.0</td>
<td>95.0</td>
<td><strong>90.5</strong></td>
</tr>
<tr>
<td>Tone 3</td>
<td>70.5</td>
<td>76.5</td>
<td>75.5</td>
<td>76.0</td>
<td>78.0</td>
<td>79.0</td>
<td>88.0</td>
<td>75.5</td>
<td>88.5</td>
<td>63.0</td>
<td><strong>77.1</strong></td>
</tr>
<tr>
<td>Tone 4</td>
<td>91.7</td>
<td>91.7</td>
<td>96.7</td>
<td>91.7</td>
<td>96.7</td>
<td>95.0</td>
<td>96.7</td>
<td>91.7</td>
<td>98.3</td>
<td>75.0</td>
<td><strong>92.5</strong></td>
</tr>
</tbody>
</table>

Table 4.5 below represents a confusion matrix of corrective focus perception in South Kyungsang Korean. We observed that focused positions were not clearly identified – the overall identification rate was just 48.2%. Similar to Seoul Korean, incorrect answers often appeared within the same phrase before or after focus positions. For example, when position 1 was focused, position 2 was identified at a rate of 23.0%. When position 2 was focused, position 1 was identified about 22.5% of the time. When position 5 was focused, position 4 was identified 29.0% of the time. This trend was also observed in other focus positions, suggesting that prosodic marking of focus was actually ambiguous in South Kyungsang Korean.
Table 4.5. Confusion matrix of corrective focus perception for South Kyungsang Korean (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string.

<table>
<thead>
<tr>
<th>Target</th>
<th>1st phrase</th>
<th>2nd phrase</th>
<th>3rd phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st phrase</td>
<td>82.0</td>
<td>7.5</td>
<td>10.3</td>
</tr>
<tr>
<td>2nd phrase</td>
<td>18.0</td>
<td>67.0</td>
<td>14.5</td>
</tr>
<tr>
<td>3rd phrase</td>
<td>14.9</td>
<td>7.5</td>
<td>77.6</td>
</tr>
</tbody>
</table>

Even if we score by phrase, as shown in Table 4.6, the overall identification rate would turn out to be 75.5% for South Kyungsang Korean. This identification rate is still lower than that of American English and Mandarin Chinese, meaning that South Kyungsang Korean produced weak prosodic marking for focus.
4.3.2.3. Standard French and Tokyo Japanese

Table 4.7 below shows a confusion matrix for the identification of focus positions in Standard French and Tokyo Japanese. There was also a striking difference between these languages. Focus positions were very well identified in Standard French (95.9% overall). In general, each focus position received a high identification rate, except for position 10, which had an identification rate of 82.1%. In contrast, focus positions were not very well identified in Tokyo Japanese (40.9% overall). Similar to Seoul Korean and South Kyungsang Korean, incorrect answers were usually given for Tokyo Japanese within the same phrase. To illustrate the confusion matrix of the first phrase (i.e., the first three positions), when position 1 was focused, position 2 was chosen at a rate of 40.9%; when position 2 was focused, position 1 was chosen at a rate of 15.0%; when position 3 was focused, position 2 was chosen at a rate of 15.9%. A similar trend continued to be observed throughout the second and third phrases. This suggests that Tokyo Japanese actually produced ambiguous prosodic marking of focus.
Table 4.7. Confusion matrix of corrective focus perception (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string. (Top panel: Standard French, Bottom panel: Tokyo Japanese).

<table>
<thead>
<tr>
<th>Perceived</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>98.6</td>
<td>0.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>97.9</td>
<td>2.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
<td>91.4</td>
<td>2.9</td>
<td>0.0</td>
<td>0.0</td>
<td>5.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>1.4</td>
<td>0.0</td>
<td>0.0</td>
<td>98.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>2.9</td>
<td>96.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.7</td>
<td>97.1</td>
<td>0.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>100</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>98.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.4</td>
<td>0.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>97.9</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>2.1</td>
<td>0.7</td>
<td>10.0</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.0</td>
<td>2.1</td>
<td>0.7</td>
<td>82.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38.2</td>
<td>40.9</td>
<td>4.1</td>
<td>0.9</td>
<td>2.3</td>
<td>2.3</td>
<td>3.2</td>
<td>4.1</td>
<td>3.2</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>15.0</td>
<td>53.6</td>
<td>1.8</td>
<td>1.4</td>
<td>3.6</td>
<td>0.5</td>
<td>6.8</td>
<td>5.0</td>
<td>5.5</td>
<td>6.8</td>
</tr>
<tr>
<td>3</td>
<td>10.9</td>
<td>15.9</td>
<td>47.7</td>
<td>0.9</td>
<td>1.8</td>
<td>13.2</td>
<td>1.8</td>
<td>4.1</td>
<td>3.2</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>16.4</td>
<td>8.2</td>
<td>2.7</td>
<td>22.7</td>
<td>20.5</td>
<td>9.1</td>
<td>5.0</td>
<td>5.5</td>
<td>7.7</td>
<td>2.3</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>3.2</td>
<td>0.5</td>
<td>24.5</td>
<td>55.5</td>
<td>5.9</td>
<td>1.8</td>
<td>0.5</td>
<td>1.4</td>
<td>1.8</td>
</tr>
<tr>
<td>6</td>
<td>7.3</td>
<td>10.5</td>
<td>5.5</td>
<td>10.9</td>
<td>24.1</td>
<td>32.3</td>
<td>4.1</td>
<td>0.9</td>
<td>4.1</td>
<td>0.5</td>
</tr>
<tr>
<td>7</td>
<td>4.1</td>
<td>12.7</td>
<td>2.3</td>
<td>0.5</td>
<td>1.4</td>
<td>0.9</td>
<td>39.1</td>
<td>23.2</td>
<td>9.5</td>
<td>6.4</td>
</tr>
<tr>
<td>8</td>
<td>4.5</td>
<td>7.3</td>
<td>5.0</td>
<td>5.5</td>
<td>6.8</td>
<td>1.4</td>
<td>25.0</td>
<td>43.6</td>
<td>0.9</td>
<td>0.0</td>
</tr>
<tr>
<td>9</td>
<td>10.9</td>
<td>8.6</td>
<td>3.2</td>
<td>1.4</td>
<td>0.5</td>
<td>0.0</td>
<td>14.5</td>
<td>12.7</td>
<td>38.6</td>
<td>9.5</td>
</tr>
<tr>
<td>10</td>
<td>3.2</td>
<td>10.5</td>
<td>3.2</td>
<td>4.5</td>
<td>1.4</td>
<td>2.3</td>
<td>9.5</td>
<td>9.1</td>
<td>18.6</td>
<td>37.7</td>
</tr>
</tbody>
</table>

Table 4.8 presents the phrase-by-phrase confusion matrix for Tokyo Japanese. As Table 4.8 shows, even if we score by phrase, the overall identification rate would increase from 40.9% to 73.0%. Nevertheless, the overall identification rate is still lower than that

Table 4.8. The phrase-by-phrase confusion matrix for Tokyo Japanese.

<table>
<thead>
<tr>
<th>Target</th>
<th>1st phrase</th>
<th>2nd phrase</th>
<th>3rd phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st phrase</td>
<td>76.1</td>
<td>8.9</td>
<td>15.0</td>
</tr>
<tr>
<td>2nd phrase</td>
<td>19.7</td>
<td>68.5</td>
<td>11.8</td>
</tr>
<tr>
<td>3rd phrase</td>
<td>19.5</td>
<td>6.1</td>
<td>74.4</td>
</tr>
</tbody>
</table>

4.3.2.4. Suzhou Wu

Table 4.9 displays a confusion matrix for the identification of five focus positions in Suzhou Wu. As a reminder, unlike other languages, Suzhou Wu consists of five focus positions: positions 1, 3, 4, 5, and 10. The overall identification rate was 55.9%, suggesting that prosodic marking of focus in Suzhou Wu was not clear compared to that of American English, Mandarin Chinese, and Standard French. Table 4.9 presents the following three interesting results. First, identification performance was indeed position-dependent. Positions 1, 3, and 4 received a relatively good identification rate of over 60%, whereas positions 5 and 10 received a relatively poor identification rate. Second, when position 5 was focused, listeners chose position 4 at a rate of 27.1%. This trend also appeared when position 10 was focused. In this case, position 4 was misidentified at a rate of 33.6%, which was actually greater than the identification rate of position 10. Third, some focus positions featured an ambiguous marking of prosodic focus. When positions
1, 4, or 5 were focused, incorrect answers were likely to appear around the focus position within the same phrase.

To sum up, Suzhou Wu exhibited several unique characteristics of prosodic focus. Prosodic marking of focus: 1) was weaker when compared to that of American English, Mandarin Chinese, and Standard French; 2) was better identified under certain positions; 3) showed some degree of ambiguity.

Table 4.9. Confusion matrix of corrective focus perception for Suzhou Wu (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string.

<table>
<thead>
<tr>
<th>Target</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60.0</td>
<td>15.7</td>
<td>10.0</td>
<td>7.9</td>
<td>2.9</td>
<td>1.4</td>
<td>2.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>6.4</td>
<td>3.6</td>
<td>73.6</td>
<td>7.1</td>
<td>2.9</td>
<td>0.7</td>
<td>3.6</td>
<td>1.4</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>4.3</td>
<td>2.9</td>
<td>0.7</td>
<td>75.0</td>
<td>8.6</td>
<td>7.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td>5</td>
<td>5.0</td>
<td>0.0</td>
<td>0.7</td>
<td>27.1</td>
<td>45.0</td>
<td>17.9</td>
<td>2.1</td>
<td>1.4</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>5.7</td>
<td>5.0</td>
<td>5.0</td>
<td>33.6</td>
<td>7.9</td>
<td>2.1</td>
<td>7.1</td>
<td>3.6</td>
<td>4.3</td>
<td>25.7</td>
</tr>
</tbody>
</table>

4.4. Discussion

This chapter used production and perception experiments to investigate whether and how prosodic marking of focus varies across languages. The technique described in this paper allowed a systematic and quantitative comparison of languages in terms of the prosodic marking of corrective focus. We have shown that this method can be used for the study of
perception as well as production, and that the perception and production results were generally congruent.

The experiments done so far established clearly that languages differ greatly in how well their speakers communicated the location of corrective focus by purely prosodic means. In particular, speakers of American English, Mandarin Chinese, and Standard French modulated duration, pitch, and intensity in a clear way to signal the location of corrective focus, and listeners in those languages recognized the intended location with high accuracy. In contrast, speakers of Seoul Korean, South Kyungsang Korean, and Tokyo Japanese did not clearly mark corrective focus by prosodic changes. Listeners of these languages had a much harder time correctly locating the corrected digits from prosodic cues since prosodic marking of focus was actually weak and ambiguous. In Suzhou Wu, focus positions were fairly well marked by increases in duration, pitch, and intensity. However, at the same time, increases in duration and intensity were also found in adjacent positions. Therefore, Suzhou Wu listeners did not always identify the focus position with high accuracy as demonstrated in the languages of American English, Mandarin Chinese, and Standard French.

The results of the production and perception experiments clearly demonstrated the phonological unit carrying prosodic focus in all of the languages under study, except for Suzhou Wu. Given that each focus position (or digit) showed a clear prosodic marking of focus and post-focus compression starting right after the focus position in the languages of American English, Mandarin Chinese, and Standard French, the phonological unit carrying prosodic focus was a word (or digit) in these languages. Since speakers of Seoul Korean, South Kyungsang Korean, and Tokyo Japanese often produced an ambiguous
prosodic modulation by focus within a phrase, and post-focus compression appeared across the phrase boundary, the phonological unit carrying prosodic focus was a phrase in these languages. Finally, Suzhou Wu showed ambiguous results for prosodic marking of focus. In certain circumstances, prosodic focus was clearly marked in production and accurately recognized in perception, while at other times it was less clearly marked in production, thus resulting in poorer identification performance. Since Suzhou Wu certainly generated mixed results for prosodic focus, it would be premature to make a definitive statement about the exact phonological unit carrying prosodic focus in this language; instead, additional research with a larger range of data will be important for identifying its exact nature.

An immediate question arises as to why the languages of American English, Mandarin Chinese, and Standard French showed a clear prosodic marking of focus, and the languages of Seoul Korean, South Kyungsang Korean, and Tokyo Japanese did not. Based on the findings, we speculate that the “clear” distribution of prosodic focus was related to the higher degree of freedom in the languages of American English, Mandarin Chinese, and Standard French. The production and perception results suggest that these languages had enough room for variability in the distribution of prosodic focus, enabling any position (or digit) to be prosodically prominent in a string. However, the “messy” distribution of prosodic focus was related to the lower degree of freedom in the other languages, which have little room for variability in the distribution of prosodic focus. As described in Chapter 3, the prosodic patterns are fixed at the phrase level in those languages. Seoul Korean shows only two prosodic patterns (LHLH or HHLH) within a phrase, where the initial tone depends on the laryngeal feature of an AP-initial segment.
South Kyungsang Korean also shows limited prosodic patterns within a phrase. The possible prosodic patterns include HHL, HLL, LHL, and LHH for the trisyllabic phrase and HHLL and LHHL for the quadrisyllabic phrase. Tokyo Japanese forms a special prosodic pattern, called a bipodic template, for reading phone-number strings, where an accentual peak appears every two digits within a phrase. Therefore, speakers of these languages did not seem to control their vocal efforts easily for each focus position (or digit).

We have also learned that in Standard French, prominence is determined at the phrase level (Beyssade & Marandin, 2006; Jun & Fougeron, 2002), mainly showing a phrase-initial rise and a phrase-final rise (Jun & Fougeron, 2002). Therefore, it has been widely suggested that prosodic focus is marked by phrasing (D’Imperio, German, & Michelas, 2012; Féry, 2001). In particular, Beyssade et al. (2004, p. 477) claimed that “focus is not marked by a specific tone or accent associated with the focalized constituent, but by a boundary tone.” However, the findings of this chapter demonstrated that Standard French used clear and consistent prosodic marking of focus, leading to a high identification rate of 95.9% by demonstrating a higher degree of freedom in the distribution of prosodic marking of focus. We posit that the clear prosodic marking of focus was made by the assignment of emphatic accent (or l’accent d’insistance in French). In Dahan and Bernard (1996), the emphatic accent clearly marked the contrast between the target item and adjacent items by increased duration, intensity, and pitch. Therefore, as opposed to previous work, we can say that in Standard French, focus can be encoded by purely prosodic means, directly associated with the focus position.
Another question concerns why certain focus positions (i.e., positions 5 and 10) produced poor identification performance in Suzhou Wu. Out of five focus positions, three focus positions (positions 1, 3, and 4) were relatively well identified, whereas positions 5 and 10 were not very well identified. We suggest that the main reason for the poor identification performance was the relatively weak intensity of these positions. Consider Figure 4.7.

![Figure 4.7](image)

Figure 4.7. The mean intensity values of the broad-focus and corrective-focus conditions. Positions 5 and 10 are the target positions that include prosodic focus.

Figure 4.7 compares the mean intensity values between broad and corrective focus when positions 5 and 10 were focused. As the left panel of Figure 4.7 shows, when position 5 was focused, it certainly produced a greater intensity value than that of the broad-focus counterpart. However, position 4 also exhibited an increase in intensity. This indicates that although position 5 made a paradigmatic (or vertical) contrast between broad and corrective focus, it did not make a syntagmatic (or horizontal) contrast with position 4 in
the corrective-focus condition. Position 10 also displayed a similar trend. Although position 10 produced a greater intensity value in the corrective-focus condition than in the broad-focus one, the intensity value of position 10 was actually lower than that of the preceding positions. In a nutshell, Suzhou Wu did not make a syntagmatic contrast in marking prosodic focus in certain focus conditions, which was indeed linked to a poor identification rate.3

We observed that American English and Mandarin Chinese produced different identification rates in sentence-final focus position. The identification rate of the final focus position was 98.7% for American English and 74.2% for Mandarin Chinese. It has been argued that the reason for the relatively low identification rate of this position for Mandarin Chinese is due to the fact that post-focus compression is not possible in the sentence-final focus position (Chen et al., 2009; Liu & Xu, 2005; Xu et al., 2004; Xu et al., 2012). This argument, however, seems less plausible given that the language of American English yielded an identification rate of 98.7% in the final focus position even without post-focus compression. Rather, we believe that Mandarin Chinese produced a weaker prosodic marking of focus in the sentence-final focus position than American English. Figure 4.8 illustrates such a feature with pitch and intensity values.

---

3 The ambiguity of position 10 may be related to the low intensity typically occurring in utterance-final position. One may wonder why positions 5 produced such ambiguous prosodic marking of focus. We speculate that this is related to tone sandhi in Suzhou Wu, which includes rightward spreading of the initial tone within a phrase. Since the phrase-initial tone plays a dominant role in shaping the pitch contour of a phrase, the phrase-initial position seemingly gains prominent intensity, as well. Additional research is necessary to verify this assumption.
Figure 4.8. The pitch and intensity values of the broad-focus and corrective focus conditions in two languages: American English and Mandarin Chinese. The target position is position 10.

In Figure 4.8, American English clearly showed a remarkable increase in both pitch and intensity in the sentence-final focus position (i.e., position 10) between broad and corrective focus. In comparison, Mandarin Chinese indicated a relatively smaller difference in this position. This means that speakers of American English produced a more prominent prosodic modulation by focus than those of Mandarin Chinese. This seems to be the main factor which resulted in a different identification rate of the final focus position of the two languages.

The two varieties of Korean produced different identification rates over the focus positions. Seoul Korean received an identification rate of about 37%, whereas South
Kyungsang Korean received an identification rate of about 48%. From the production data, we observed that South Kyungsang Korean produced relatively longer durations than Seoul Korean. While Seoul Korean showed only a slight increase of, on average, 8.6 ms of extra duration for prosodic marking of focus, South Kyungsang Korean actually showed about 32 ms of extra duration for focus marking. A linear mixed effect model confirmed that this difference is statistically significant \([t = 8.407, p < 0.001]\). Therefore, we can say that the difference in durational cues serves as the main factor accounting for the better identification performance of South Kyungsang Korean.

An important question still remains about the relationship of our results to variation on other prosodic typological dimensions. Why did the languages of American English, Mandarin Chinese, and Standard French produce a clear prosodic marking of focus, while also showing a higher degree of freedom in the distribution of that marking? As is well known, these languages are prosodically different. American English has a lexical stress that the other languages lack. Mandarin Chinese has a lexical tone that the other languages do not. Standard French is known to show a fixed intonational melody at the phrase level, which certainly distinguishes it from the others. Additionally, according to prominence-based typology (Jun, 2005, 2012, 2014), American English and Mandarin Chinese are head-prominence languages, whereas Standard French is a head/edge-prominence language. Yet, this typology also cannot provide a clear answer to the question with which we are concerned. This unresolved issue re-opens the question of the

---

4 The linear mixed effect model was conducted by implementing the \texttt{lmerTest} package (Kuznetsova et al., 2013) in R (R Core Team, 2015). The duration values were fitted into a model, in which language (Seoul Korean vs. South Kyungsang Korean) was used as a fixed effect and speakers (5 speakers), digit string positions (10 positions), and individual digits (0-9) were used as random effects.
typology of prosodic focus: why do certain languages display clear prosodic marking of focus and others do not?\textsuperscript{5} Since our data collection, even though it includes seven languages, may not be sufficient in developing a clear picture of the question, the need for more comprehensive data motivates us to collect a larger body of prosodically similar and dissimilar languages.

To summarize, this chapter investigated whether and how prosodic marking of focus varies across languages. The results demonstrated that in the languages of American English, Mandarin Chinese, and Standard French, there were clear indications of prosodic marking of focus and post-focus compression, which produced a high identification rate in perception. In comparison, the languages of Seoul Korean, South Kyungsang Korean, and Tokyo Japanese showed weak and ambiguous prosodic marking of focus and a minimal level of post-focus compression. Therefore, the focus positions were not very well identified in these languages. Lastly, Suzhou Wu showed an intermediate level of prosodic marking of focus in both production and perception. All things considered, we claim that prosodic marking of focus is neither completely universal nor automatic; instead it actually varies according to the prosodic system of each language.

\textsuperscript{5} In Chapter 7, we will further analyze whether different prosodic focus effects between languages conform to exiting prosodic typologies. We will also point out the potential limitations of the existing prosodic typologies considered in greater detail with regards to prosodic focus.
Chapter 5

Production and perception of tone 3 focus in Mandarin Chinese

In Chapter 4, we found that although prosodic marking of focus was clear and consistent in each corrected position, the location of the focus positions was identified differently for each lexical tone. In particular, tone 3 focus received a relatively poorer identification rate (77.1% overall). Therefore, we reanalyzed the production and perception data to determine the main factors affecting the lower identification performance of tone 3 focus.

5. 1. Introduction

In Mandarin Chinese, focus can be conveyed clearly by purely prosodic means, as described in Chapter 4, but as a tone language, the prosodic realization of focus actually differs by the pitch target of each lexical tone (Shih, 1988; Xu, 1999). For example, focus raises the pitch of a high tone (tone 1) but lowers the pitch of a low tone (tone 3) (Cao, 2002, 2012; Xu, 1999), indicating that unlike the universal phonetic symbolism of focus that raises pitch, tone 3 focus is characterized by a low pitch target. An important issue here is whether such a downward pitch movement is sufficient in cueing focus. If it is not
sufficient, which parameters then play key roles in tone 3 focus? Before setting up the research goals, let us first begin with a brief overview of lexical tones in Mandarin Chinese, which is important in understanding the details of the study. We then review the literature on the production and perception of prosodic focus in Mandarin Chinese. Finally, the research goals of this study will be presented based on the review of relevant literature.

5.1.1. A brief overview of four lexical tones in Mandarin Chinese

We have learned that Mandarin Chinese consists of four lexical tones: a high level tone (tone 1), a rising tone (tone 2), a low/dipping tone (tone 3), and a falling tone (tone 4). These tones are used to contrast homophonic morphemes, as illustrated in (1). As a reminder, tones 1-4 are conventionally labeled as [55], [35], [214], and [51] depending on the pitch level, where [1] represents the lowest pitch level, and [5] the highest pitch level (Chao, 1968).

(1) a. /ma/ with tone 1 ⇒ “mother”
   b. /ma/ with tone 2 ⇒ “hemp”
   c. /ma/ with tone 3 ⇒ “horse”
   d. /ma/ with tone 4 ⇒ “to scold”

(Xu, 1997, p. 64)

Furthermore, Mandarin Chinese includes two basic kinds of pitch targets associated with tones: static and dynamic (Xu & Wang, 2001). There are two static (low
and high) and two dynamic (rising and falling) pitch targets: tone 1 has a high pitch target; tone 2 has a rising pitch target from low to high; tone 3 has a low pitch target; tone 4 has a falling pitch target from high to low. Based on their pitch target, tones 1–4 can be broadly classified into two groups, one in which tones 1, 2, and 4 have a high pitch point, the other in which tone 3 lacks a high pitch point.

5.1.2 Production and perception of focus in Mandarin Chinese

It has been observed that focus involves different kinds of prosodic adjustments that differ depending on whether they occur in focus, post-focus, or pre-focus position. In the focus position, focus increases duration, intensity, and pitch (Chen & Gussenhoven, 2008; Liu & Xu, 2005; Wang et al., 2002; Xu, 1999; Yuan, 2004). In post-focus positions, duration, intensity, and pitch range are considerably compressed (Liu & Xu, 2005; Xu, 1999; Yuan, 2004), known as post-focus compression (Chen et al., 2009; Xu et al., 2012). Yet these parameters show no significant changes in the pre-focus positions (Liu & Xu, 2005; Xu, 1999; Yuan, 2004).

As previously stated, tone 3 focus is expressed in a unique fashion by lowering a pitch target. Several studies have attempted to ascertain the prosodic characteristics of tone 3 focus, but no clear picture has been obtained of its exact nature. Shih (1988) argues that it is unclear whether the low pitch target is actually lowered under focus. On the other hand, other studies claim that focus lowers the low pitch target of a tone 3 syllable (Cao, 2012; Chao, 1968; Chen & Gussenhoven, 2008; Xu, 1999). Another different view is that a long duration plays an important role in cueing tone 3 focus.
Regarding pre- and post-focus effects, tone 3 focus involves (unique) local dissimilatory effects: pitch becomes raised immediately before focus, known as pre-low raising (Liu & Xu, 2007; Xu & Wang, 2001); and pitch bounces back immediately after focus, known as post-low bouncing (Liu & Xu, 2007; Prom-on et al., 2012). It should be noted that post-focus compression is absent where post-low bouncing is present.

Moving onto perception, it has been attested that focus identification does not differ by tone but does differ by position (Liu, 2009; Yuan, 2004). Yuan (2004) found that the ordering of identification rates from highest to lowest was sentence-medial (92.9%) > sentence-initial (87.2%) > sentence-final (75.5%), where the symbol “>” indicates a significant difference. In Liu (2009), the results revealed a similar ordering: sentence-medial (97.2%), sentence-initial (95.3%) > sentence-final (82.6%). Regarding the perceptual cues for focus, shifting pitch contours and raising a high pitch target are important perceptual cues, although the latter plays a more important role (Wang et al., 2002). Moreover, a large body of evidence has demonstrated that post-focus compression serves as a highly effective perceptual cue to focus (Chen et al., 2009; Liu & Xu, 2005; Xu et al., 2004; Xu et al., 2012).

In contrast to other focused tones, tone 3 focus draws relatively little attention in perception – to our knowledge, there have been only two studies that attempted to examine the perception of tone 3 focus (Cao & Zhang, 2008; Yuan, 2004). In the case of Cao and Zhang’s (2008) experiment, stimuli were synthesized in three separate positions (sentence-initial, sentence-medial, sentence-final), where duration, creakiness, and pitch range were incremented during each step in order to approximate the natural prosody of tone 3 focus. The findings indicated that creakiness is important in sentence-initial
position, whereas lengthening is important in sentence-final position. However, they concluded that the most important cue to tone 3 focus is a mid-sized pitch drop (6 semitones), although creakiness and lengthening improve identification in some positions. Using natural stimuli, Yuan (2004) reported that the ordering of identification rates of tone 3 focus is congruent with the identification of other focused tones, i.e., sentence-medial > sentence-initial > sentence-final.

5.1.3 The current study

From the literature review, we have observed several limitations in the stimuli of the previous work. First, in many studies, a tone 3 syllable was excluded from the stimuli (e.g., Kabagema-Bilan et al., 2011; Ouyang & Kaiser, 2015; Wang & Xu, 2011; Xu, 1999), presumably due to tone sandhi – tone 3 becomes tone 2 when followed by another tone 3. Second, although some studies indeed included a tone 3 syllable in the stimuli (e.g. Greif, 2010; Liu, 2009), it seems that the full scale of tone 3 focus was (largely) masked by the structural limitations inherent in the stimuli. For example, in the stimuli of Greif (2010), the name Ma3 Long2 was designed to be contrastively focused, as shown in (2), where the subscript F refers to focus. In this case, although the entire sequence was in the domain of semantic focus, the tone 2 syllable seemed to carry main prominence (via focus).
(2) Q: Has Tom got two or three watermelons?\(^1\)
‘No, Marlon has the watermelons.’

In Liu (2009), two consecutive tone 3 syllables (e.g., Li3 Min3) were designed to receive focus, but the sequence changed to Li2 Min3 due to the tone sandhi rule. As a result, similar to the case of Grief (2010), the tone 2 syllable seemed to carry main prominence. The phenomena described here are similar to the case where focus is encoded by a primary stressed syllable in a (multisyllabic) word in English (Cohan, 2000; Ladd, 1996), although the whole word is in the focus domain. Therefore, we need a design where tone sandhi is avoided, and at the same time a tone 3 syllable is encoded by prosodic prominence. Third, although Liu and Xu (2007) discovered the local dissimilatory effects of tone 3 focus, the distribution of tone 3 focus was fairly restricted in their stimuli – only the second and third position alternately contained tone 3 focus in a sentence, which leads us to explore the local dissimilatory effects in a full scale. Finally and most importantly, the local dissimilatory effects have not yet been studied in perception.

Due to the limited distributions of tone 3 focus in both production and perception, our understanding of tone 3 focus is far from complete. There are some important issues that need to be considered. First, it is unclear whether a downward pitch movement of tone 3 focus is a perceptually sufficient cue for listeners. Second, we do not know yet whether pre-low raising and/or post-low bouncing are independent of focus positions: do they only appear within the same prosodic phrasing or are they still visible across the

\(^1\) In the stimuli of Greif (2010), the question part was not provided in Mandarin.
phrase boundary? Given that pitch normally resets after the phrase boundary, it is likely that the local dissimilatory effects appear within the same phrase. The third issue concerns the role of local dissimilatory effects in perception: do the local dissimilatory effects help listeners perceive tone 3 focus or hinder listeners’ perception? With these issues in mind, this chapter attempts to achieve two research goals: a) to determine the nature of tone 3 focus and its local dissimilatory effects; and b) to examine whether listeners can successfully identify tone 3 focus or whether the local dissimilatory effects hinder the recognition of tone 3 focus. We achieve these goals through production and perception experiments using 10-digit phone-number strings.

5.2. Production

5.2.1. Stimuli

We used the same production data described in Chapter 2. Among the ten digits, there are four tone 1 digits (1, 3, 7, 8), one tone 2 digit (2), two tone 3 digits (5, 9), and three tone 4 digits (2, 4, 6), as Table 5.1 indicates. In total, we collected 1000 10-digit strings (100 strings x 5 speakers x 2 focus conditions). Divided by each tone, we collected 400 strings for tone 1, 100 strings for tone 2, 200 strings for tone 3, and 300 strings for tone 4.
Table 5.1. The numerical digits of Mandarin and the lexical tone of each digit together with Pinyin Romanization.

<table>
<thead>
<tr>
<th>Digit</th>
<th>Pinyin Romanization</th>
<th>Lexical tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>ling</td>
<td>Tone 2</td>
</tr>
<tr>
<td>1</td>
<td>yi</td>
<td>Tone 1</td>
</tr>
<tr>
<td>2</td>
<td>er</td>
<td>Tone 4</td>
</tr>
<tr>
<td>3</td>
<td>san</td>
<td>Tone 1</td>
</tr>
<tr>
<td>4</td>
<td>si</td>
<td>Tone 4</td>
</tr>
<tr>
<td>5</td>
<td>wu</td>
<td>Tone 3</td>
</tr>
<tr>
<td>6</td>
<td>liu</td>
<td>Tone 4</td>
</tr>
<tr>
<td>7</td>
<td>qi</td>
<td>Tone 1</td>
</tr>
<tr>
<td>8</td>
<td>ba</td>
<td>Tone 1</td>
</tr>
<tr>
<td>9</td>
<td>jiu</td>
<td>Tone 3</td>
</tr>
</tbody>
</table>

5.2.2. Subjects and recording procedure

The subject recruitment and recording procedure was identical to the one that was described in Chapter 2.

5.2.3. A sketch of pitch contours

Before analyzing the data, we illustrate some of the pitch contours that enable us to capture the overall picture of tone 3 focus and then move onto the pitch contours that portray prosodic characteristics of each focused tone. In this study, the pitch contours were sampled at 10 equidistant points of each labeled digit using ProsodyPro (Xu, 2013).
Figure 5.1 displays time-normalized pitch contours averaged by five speakers for tone 3 digits, where the shaded area in gray represents the focus position and vertical lines refer to phrase boundaries. From Figure 5.1, we can observe that the corrective-focus condition shows a more expanded pitch range than its broad-focus counterpart in the focus position. At the same time, we can observe noticeable differences in the pre- and post-focus positions. In Figures 5.1a and 5.1c, the corrective-focus condition shows a higher level of pitch in the pre-focus position (i.e., pre-low raising), whereas Figures 5.1b and 5.1d show no such thing in the same position. Regarding the post-focus effect, only Figure 5.1a shows clearly compressed pitch contours (i.e., post-focus compression) right after focus. In Figures 5.1b-d, post-focus compression is not visible right after focus; rather, the pitch bounces up after a very low pitch (i.e., post-low bouncing), indicating that post-focus compression is absent where post-low bouncing is present. It seems that the local dissimilatory effects like pre-low raising and post-low bouncing are thought to be position-dependent – they are present within the same phrase but absent across the phrase boundary.
Figure 5.1. Sample pitch contours for tone 3 digits in two focus conditions. BF and CF are abbreviations for broad focus and corrective focus, respectively.

Figure 5.2 exhibits the pitch trajectories of tones 1-4 in two focus conditions aggregated by 10 string positions. As demonstrated in Figure 5.2, corrective focus is marked similarly by greater pitch expansion for all tone types, which is captured by the rise/fall size, subtracting the high/low pitch point of broad focus from that of corrective focus. However, there appear two noticeable differences between tone 3 and other tones. One is that tone 3 focus is characterized by lowering a low pitch point, but other focused
tones are realized by raising a high pitch point. The other is that unlike other focused tones, tone 3 focus seems to have a smaller opportunity for pitch expansion.

Figure 5.2. Pitch trajectories of tones 1-4 in two focus conditions. BF and CF are abbreviations for broad focus and corrective focus, respectively.

5.2.4. Acoustic measurements

Based on the visual observations of Figures 5.1 and 5.2, the acoustic measurements were conducted from three different areas: focus, pre-focus, and post-focus. In the focus position, we measured duration (st) and mean intensity (dB) of each labeled digit to
directly compare the two focus conditions. Furthermore, in order to estimate the size of pitch expansion between broad and corrective focus, we measured maximum pitch (st) for tones 1, 2, and 4 and minimum pitch (st) for tone 3, which are assumed to best reflect the underlying pitch target of each tone. In this study, we label the maximum and minimum pitch as “target pitch” for simplicity’s sake. In pre- and post-focus positions, we measured duration, mean intensity, and mean pitch (st) from the positions immediately preceding and following focus to analyze the local dissimilatory effects of tone 3 focus. We automatically obtained these measurements by implementing ProsodyPro (Xu, 2013), based on hand-labeled digit boundaries.

5.2.5. Analyses and results

The basic analysis strategy was to make a direct comparison between the broad-focus and corrective-focus conditions, which are separated by focus position: focus, pre-focus, and post-focus. In the focus position, we examined the tone 3 digits by the aggregate measures of duration, mean intensity, and target pitch, and also included the other tones for reference data. In the pre-focus positions, given that pre-low raising seems to occur only within the phrase, we divided the string position into two parts: final vs. non-final positions. In the 10-digit string (NNN)-(NNN)-(NNNN), “N” refers to non-final position; “N” to final position; and “N” to non-applicable position for pre-focus. Similarly, in the post-focus positions, since post-low bouncing also seems position-dependent, we divided the string position into two parts: initial vs. non-initial positions. In this string (NNN)-
(NNN)-(NNNN), “N” refers to non-applicable position for post-focus; “N” to non-initial position; and “N” to initial position.

For statistical analysis, we built a linear mixed model implementing the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2013) in R (R Core Team, 2015). In the focus position, duration, mean intensity, and target pitch were regressed against a model for each tone, where focus was used as a fixed effect and speaker (5 speakers), string position (10 positions in a digit string), and digit (different digits for each tone, except tone 2) were used as random effects. In the pre- and post-focus positions, duration, mean intensity, and mean pitch were regressed against a model with position as a fixed effect and speaker and string position as random effects. We present the results in the order of focus, pre-focus, and post-focus position.

Figure 5.3 plots duration, mean intensity, and target pitch of tones 1-4 in two focus conditions. The message seems simple and clear about the prosodic marking of focus – corrective focus shows longer duration, higher intensity, and greater pitch expansion than its broad-focus counterpart for all tone types. Statistical analyses confirmed this visual impression. There was a significant effect of focus on duration for all tone types, such that corrective focus induced a longer duration than broad focus [T1: \( t = 13.32, p < 0.001 \), T2: \( t = 8.19, p < 0.001 \), T3: \( t = 10.56, p < 0.001 \), T4: \( t = 10.47, p < 0.001 \)]. The effect of focus on mean intensity was also significant for all tone types, indicating that corrective-focus conditions showed a higher intensity than their broad-focus counterparts [T1: \( t = 5.68, p < 0.001 \), T2: \( t = 4.04, p < 0.001 \), T3: \( t = 3.43, p < 0.001 \), T4: \( t = 4.68, p < 0.001 \)].

---

2 For the pre-focus and post-focus positions, we will only provide the results of tone 3 focus since it produced a relatively poorer identification rate than other focused tones.
Finally, focus produced a significant effect on target pitch for all tone types, such that corrective focus used greater pitch expansion than broad focus [T1: $t = 22.15$, $p < 0.001$, T2: $t = 8.15$, $p < 0.001$, T3: $t = -3.37$, $p < 0.001$, T4: $t = 19.95$, $p < 0.001$].

![Graph showing duration, mean intensity, and target pitch of all tone types in two focus conditions (BF: broad focus, CF: corrective focus; T1: Tone 1, T2: Tone 2, T3: Tone 3, T4: Tone 4). The downward arrow of the third panel indicates that tone 3 focus is expressed by lowering its pitch point.](image)

Figure 5.3. Duration, mean intensity, and target pitch of all tone types in two focus conditions (BF: broad focus, CF: corrective focus; T1: Tone 1, T2: Tone 2, T3: Tone 3, T4: Tone 4). The downward arrow of the third panel indicates that tone 3 focus is expressed by lowering its pitch point.

Although the corrective-focus condition used greater pitch expansion than the broad-focus one for all tone types, we need to pay attention to the smaller size of the pitch expansion for tone 3 focus. As Table 5.2 indicates, the size of the pitch expansion was only 1.56 st for tone 3 focus, whereas that of the pitch expansion was at least 2.37 st for other focused tones. This difference indicates that tone 3 digits actually produced a relatively smaller pitch expansion in marking prosodic focus. However, this kind of difference was not reflected in the other parameters of duration and intensity – all tone types showed quite similar increases in encoding focus. Therefore, we speculate that the smaller size of the pitch expansion for tone 3 focus was due to its unique prosodic
structure: tone 3 focus was expressed by lowering the pitch target, which distinguishes it from other focused tones.

Table 5.2. The duration, mean intensity, and target pitch values of the broad-focus and corrective-focus condition for each tone type (BF and CF are abbreviations for broad focus and corrective focus, respectively).

<table>
<thead>
<tr>
<th></th>
<th>BF (A)</th>
<th>CF (B)</th>
<th>B – A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Duration (ms)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 1</td>
<td>279.27</td>
<td>337.75</td>
<td>58.48</td>
</tr>
<tr>
<td>Tone 2</td>
<td>251.77</td>
<td>319.83</td>
<td>68.06</td>
</tr>
<tr>
<td>Tone 3</td>
<td>231.65</td>
<td>305.14</td>
<td>73.49</td>
</tr>
<tr>
<td>Tone 4</td>
<td>258.76</td>
<td>311.76</td>
<td>53.00</td>
</tr>
<tr>
<td><strong>Mean intensity (dB)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 1</td>
<td>73.41</td>
<td>74.76</td>
<td>1.35</td>
</tr>
<tr>
<td>Tone 2</td>
<td>70.42</td>
<td>72.11</td>
<td>1.69</td>
</tr>
<tr>
<td>Tone 3</td>
<td>68.97</td>
<td>70.27</td>
<td>1.30</td>
</tr>
<tr>
<td>Tone 4</td>
<td>74.65</td>
<td>75.90</td>
<td>1.25</td>
</tr>
<tr>
<td><strong>Target pitch (st)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tone 1</td>
<td>90.59</td>
<td>92.96</td>
<td>2.37</td>
</tr>
<tr>
<td>Tone 2</td>
<td>88.63</td>
<td>91.15</td>
<td>2.52</td>
</tr>
<tr>
<td>Tone 3</td>
<td>85.07</td>
<td>83.51</td>
<td>-1.56</td>
</tr>
<tr>
<td>Tone 4</td>
<td>90.70</td>
<td>94.02</td>
<td>3.32</td>
</tr>
</tbody>
</table>

Moving on to consider the pre-focus position of tone 3 digits, Figure 5.4 plots a simple comparison of the differences in duration, mean intensity, and mean pitch of the two focus conditions, which are separated by final vs. non-final positions. In this figure, the average of each point was determined by subtracting the paired values between broad
focus and corrective focus. As shown in Figure 5.4, the non-final positions had longer duration, higher intensity, and a higher mean pitch than the final positions. In other words, focused tone 3 digits increased the duration, mean intensity, and mean pitch values of their pre-focus position in the non-final position. This indicates that the pre-low raising effect was contingent on string position; it only occurred within the same phrase. The results from the linear mixed models partly supported this observation. There were significant effects of string position on both duration \( t = 4.48, p < 0.001 \) and mean intensity \( t = 2.49, p < 0.05 \). Although non-final positions had greater mean pitch values than final positions, the effect of string position on mean pitch showed a positive trend but failed to achieve a customary level of statistical significance \( t = 1.89, p = 0.10 \).

![Mean Differences - Corrective Focus Minus Broad Focus](image)

Figure 5.4. Duration, mean intensity, and mean pitch values of the pre-focus position of tone 3 digits, separated by final vs. non-final positions. BF and CF are abbreviations for broad focus and corrective focus, respectively.

As for the post-focus position of tone 3 digits, Figure 5.5 describes the differences between initial vs. non-initial positions by aggregating the following parameters: duration,
mean intensity, and mean pitch. Each point of Figure 5.5 refers to the average value calculated by subtracting the values between broad focus and corrective focus. It is likely that the different post-focus positions reveal different kinds of post-focus effects. In initial positions, corrective-focus conditions clearly showed post-focus compression with reduced duration, mean intensity, and mean pitch values. However, non-initial positions did not show this post-focus compression. Rather, the duration, mean intensity, and mean pitch values showed a rebound effect immediately after the tone 3 focus; therefore, the values of these parameters were close to zero, meaning that the differences between the broad-focus and corrective-focus conditions were minimal in the non-initial post-focus conditions. The statistical analyses confirmed this observation for all the parameters [duration: \( t = 3.56, p < 0.001 \); mean intensity: \( t = 2.67, p < 0.05 \); mean pitch: \( t = 4.28, p < 0.01 \)]. The results clearly suggest that tone 3 focus had different post-focus effects depending on the position in a digit string: the initial post-focus positions showed the post-focus compression effect, whereas the non-initial post-focus positions displayed the post-low bouncing effect.
Figure 5.5. Duration, mean intensity, and mean pitch values of the post-focus position of tone 3 digits, separated by initial vs. non-initial positions. BF and CF are abbreviations for broad focus and corrective focus, respectively.

5.3. Perception

From the production experiment, we found that tone 3 focus was clearly marked by increased duration, intensity, and pitch expansion. At the same time, we found that pre-low raising and post-low bouncing effects appeared only within the same phrase. Therefore, this perception experiment was aimed to examine whether listeners can successfully identify tone 3 focus or whether the local dissimilatory effects hinder the recognition of tone 3 focus.

5.3.1. Data collection
The stimuli, design, and procedure were identical to the one that was described in Chapter 2. The perception data were classified in a confusion matrix to examine the extent to which the local dissimilatory effects hinder the recognition of tone 3 focus.

5.3.2 Results

Table 5.3 – repeated from 4.4 – shows the identification rate of corrected digits, where the columns correspond to positions in a string and the rows to tones 1-4. The overall identification rate was 90.8% for tone 1, 90.5% for tone 2, 77.1% for tone 3, and 92.5% for tone 4. Unlike other focused tones, tone 3 focus received a relatively low identification rate. The plausible reasons can be found in a confusion matrix (Table 5.4) that contains finer-grained information about the classifier performance of tone 3 focus.

<table>
<thead>
<tr>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td>95.0</td>
<td>82.5</td>
<td>88.8</td>
<td>97.5</td>
<td>98.8</td>
<td>88.8</td>
<td>98.8</td>
<td>95.0</td>
<td>98.8</td>
<td>63.8</td>
<td><strong>90.8</strong></td>
</tr>
<tr>
<td>Tone 2</td>
<td>85.0</td>
<td>95.0</td>
<td>90.0</td>
<td>90.0</td>
<td>75.0</td>
<td>100.0</td>
<td>80.0</td>
<td>100.0</td>
<td>95.0</td>
<td>95.0</td>
<td><strong>90.5</strong></td>
</tr>
<tr>
<td>Tone 3</td>
<td>70.5</td>
<td>76.5</td>
<td>75.5</td>
<td>76.0</td>
<td>78.0</td>
<td>79.0</td>
<td>88.0</td>
<td>75.5</td>
<td>88.5</td>
<td>63.0</td>
<td><strong>77.1</strong></td>
</tr>
<tr>
<td>Tone 4</td>
<td>91.7</td>
<td>91.7</td>
<td>96.7</td>
<td>91.7</td>
<td>96.7</td>
<td>95.0</td>
<td>96.7</td>
<td>91.7</td>
<td>98.3</td>
<td>75.0</td>
<td><strong>92.5</strong></td>
</tr>
</tbody>
</table>

As shown in Table 5.4, incorrect answers usually occurred within the same phrase, immediately preceding and/or following the focus position. Let us first illustrate the rate of incorrect answers in the pre-focus position. When position 3 was focused, listeners
chose position 2 15.0% of the time. When position 5 was focused, listeners chose position 4 17.5% of the time. When position 6 was focused, position 5 was chosen 11.0% of the time. When position 8 was focused, listeners chose position 7 at a rate of 19.5%. When position 10 was focused, listeners chose position 9 at a rate of 22.0%. Overall, the average rate of incorrect answers was 14.1% for the pre-focus position within the same phrase. The post-focus positions also showed a degree of confusion; however, this was to a lesser extent. The average rate of incorrect performance was 5.2% for the post-focus position within the same phrase. Even if we score by phrase rather than position, the identification rate of tone 3 focus would increase from 77.1% to 96.4%, suggesting that the confusion of tone 3 focus was indeed due to pre-focus raising and/or post-low bouncing effects. On the contrary, the confusion rate of tone 3 focus was very little or minimal across the phrase boundary. When position 3 was focused, listeners chose position 4 only 0.5% of the time. When position 6 was focused, listeners chose position 7 only 3.0% of the time. When position 4 was focused, listeners did not choose position 3 at all. When position 7 was focused, listeners chose position 6 only 0.5% of the time. Therefore, we can say that the pre-focus raising and/or post-low bouncing effects actually hindered the recognition of tone 3 focus within the same phrase but not across the phrase boundary.
Table 5.4. Confusion matrix of corrective focus perception for tone 3 digits (percentage values). Cells highlighted in black indicate focused positions. Cells highlighted in gray refer to pre- or post-focus positions across the phrase boundary. Vertical lines refer to phrase boundaries.

<table>
<thead>
<tr>
<th>Perceived</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70.5</td>
<td>5.5</td>
<td>6.0</td>
<td>6.0</td>
<td>1.5</td>
<td>3.0</td>
<td>5.0</td>
<td>0.0</td>
<td>1.5</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>6.0</td>
<td>76.5</td>
<td>9.5</td>
<td>4.0</td>
<td>1.0</td>
<td>0.5</td>
<td>0.0</td>
<td>1.5</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>15.0</td>
<td>75.5</td>
<td>0.5</td>
<td>1.5</td>
<td>2.0</td>
<td>2.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>1.5</td>
<td>0.5</td>
<td>0.0</td>
<td>76.0</td>
<td>9.0</td>
<td>8.0</td>
<td>4.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>17.5</td>
<td>78.0</td>
<td>1.5</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>1.0</td>
<td>2.0</td>
<td>2.0</td>
<td>1.5</td>
<td>11.0</td>
<td>79.0</td>
<td>3.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
<td>1.5</td>
<td>3.0</td>
<td>0.5</td>
<td>88.0</td>
<td>4.5</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>0.5</td>
<td>1.0</td>
<td>0.0</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
<td>19.5</td>
<td>75.5</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>9</td>
<td>1.0</td>
<td>0.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0</td>
<td>7.5</td>
<td>88.5</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>1.0</td>
<td>4.0</td>
<td>1.5</td>
<td>2.0</td>
<td>0.0</td>
<td>1.0</td>
<td>4.5</td>
<td>1.0</td>
<td>22.0</td>
<td>63.0</td>
</tr>
</tbody>
</table>

5.4. Discussion

The aims of this study were twofold: a) to investigate the prosodic nature of tone 3 focus and its local dissimilatory effects; and b) to examine whether listeners successfully identify tone 3 focus, or whether the local dissimilatory effects hinder the recognition of tone 3 focus. The method developed in this study allowed a systematic investigation of tone 3 focus and its local dissimilatory effects. We have observed that production and perception results for tone 3 focus were compatible with each other.
In production, tone 3 focus was realized with increased duration, intensity, and pitch expansion similar to other focused tones. Some interesting phenomena with tone 3 focus included local dissimilatory effects – pre-low raising and post-low bouncing were present within the same phrase but absent across the phrase boundary. In perception, tone 3 focus received relatively low identification rates compared to other focused tones – incorrect answers most likely occurred in the immediate pre- or post-focus position.

In this study, the key issue at hand was to ascertain why tone 3 focus achieved low identification rates. From the perception data, we observed that the local dissimilatory effects of tone 3 focus (i.e., pre-low raising and post-low bouncing) clearly hindered the identification of tone 3 focus within the same phrase. In addition to these, there are, at least, two more reasons for the low identification rate: a) smaller opportunity for pitch expansion; and b) tone 3’s low intensity by nature. We discuss these one by one below.

First, lowering the pitch target results in a smaller opportunity for pitch expansion. Although other focused tones were expressed by pitch raising, tone 3 focus was expressed by pitch lowering. We assume that from a physiological point of view, it is more limited for one to lower the low pitch of tone 3 than to raise the high pitch of other tones given that human’s pitch range is within 100 Hz (Baken & Orlikoff, 2000; Keating & Kuo, 2012; Kuang, 2013), and a tone 3 syllable is produced at the floor of the pitch range. Our production data support that tone 3 focus showed just 1.56 st for pitch expansion, yet other focused tones showed the minimum of 2.37 st (tone 1: 2.37, tone 2: 2.52, tone 4: 3.32). Accordingly, the present study is in favor of Wang et al.’s (2002) finding that raising a high pitch target is a more important perceptual cue for identifying
focus and seems to support Shen’s (1992) theory that a top line of the pitch contour cues focus.

Another speculation is that tone 3 digits produce low intensity by nature. As shown in Table 5.5, tone 3 digits in the corrective-focus condition produced an average of 70.3 in decibels, which is even smaller than the intensity of other tones in the broad-focus condition. In a digit string, focused tone 3 digits were always surrounded by other tones. This is because in Mandarin, multiple tone 3 digits cannot appear in a row due to tone sandhi – tone 3 becomes tone 2 when followed by another tone 3. Therefore, we posit that the (seemingly) greater intensity of other adjacent tones may also affect the identification of tone 3 focus (at least to some extent).

Table 5.5. The mean intensity of tones 1-4. BF and CF are abbreviations for broad focus and corrective focus, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Tone 1</th>
<th>Tone 2</th>
<th>Tone 3</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>73.4</td>
<td>70.4</td>
<td>69.0</td>
<td>74.7</td>
</tr>
<tr>
<td>CF</td>
<td>74.8</td>
<td>72.1</td>
<td>70.3</td>
<td>75.9</td>
</tr>
</tbody>
</table>

In sum, this study enabled us to untangle the prosodic nature of tone 3 focus and its local dissimilatory effects. We found that although focused tone 3 digits were clearly marked by greater duration, intensity, and pitch expansion, the identification of tone 3 focus was not as high as other focused tones due to local dissimilatory effects, smaller opportunity for pitch expansion, and weak intensity by nature. It has been claimed that (purely) prosodic marking of focus is clear in Mandarin (e.g., Chen et al., 2009; Liu, 2009; Yuan, 2004). However, we found that even within a language where prosodic
marking of focus works very well, the location of prosodic focus can be difficult to identify in certain circumstances. Therefore, we claim that purely prosodic marking of focus is neither universal nor automatic, even within a given language. Instead, it behaves differently, conforming to the prosodic system of each language. Further examination will focus on languages with tonal patterns similar to tone 3 in Mandarin (e.g., Cantonese, Hakka) and languages where focus is characterized by a low pitch target (e.g. Turkish).
Chapter 6

An interaction effect between tonogenesis and prosodic focus in Seoul Korean

In Chapter 4, we have seen that prosodic marking of focus was neither clearly marked in production nor very well identified in perception in Seoul Korean. In this chapter, we reanalyze the production and perception data to test 1) whether prosodic marking of focus varies according to the tonal contrast (L vs. H) due to laryngeal articulations of onset initial consonants, and 2) whether prosodic marking of focus varies with different pitch-scaling conditions.

6.1. Introduction

We have learned that Seoul Korean has neither lexical stress nor lexical pitch accents (Jun, 1998, 2005; Song, 2005). Instead, the tonal pattern comes from a combination of phrasal and boundary tones.\(^1\) As Figure 6.1 illustrates, Seoul Korean includes three prosodic units: Intonation Phrase (IP), Intermediate Phrase (ip) and Accentual Phrase (AP)

\(^1\) It should be noted that we use “tone” not as lexical but post-lexical, as conventionally used in the K-ToBI system.
An IP is the highest prosodic unit marked by an IP-final boundary tone (%): when a sentence is declarative, the IP has a falling tone (L%); when a sentence is interrogative, the IP has a rising tone (H%). An ip is the domain of pitch reset marked by minor phrase-final lengthening. In default prosodic phrasing, each content word can form a small prosodic unit (AP) that is post-lexically marked. The basic melody of an AP is either LHLH or HHLH, depending on AP-initial onset consonants. The initial tone differs by the tonal contrast (L vs. H) derived from a process of tonogenesis in Seoul Korean (Jun, 2005; Kang, 2014; Kingston, 2011; Silva, 2006; Wright, 2007). The initial L of the AP’s basic melody (LHLH) is raised to H after an aspirated/tensed consonant. This indicates that the laryngeal articulations of phrase-initial consonants have split the existing phrasal tones in Seoul Korean – consonants with [+aspiration/tense] create a high pitch and those with [−aspiration/tense] create a low pitch.

![Figure 6.1. Intonation model of Seoul Korean (Adapted from Jun & Cha, 2011).](image-url)
In Seoul Korean, focus is expressed through prosodic phrasing spanning over the entire phrase.² Prosodic prominence takes place at the beginning of the focused phrase and its effect continues to the end of the phrase (Jun, 2011; Lee & Xu, 2010; Lee, 2012). Figure 6.2 – repeated from Figure 1.1 – depicts the prosodic focus effect of Seoul Korean, which demonstrates two noteworthy features.³ First, discourse-new focus produces a pitch range that is more expanded than broad focus. The pitch expansion clearly spans over the entire focused phrase. Second, and more importantly, the pitch expansion via focus is fairly small, with an increase of just 1.18 st. For this small scope of prosodic modulation, Seoul Korean’s prosodic marking of focus was neither clearly marked in production nor clearly identifiable in perception, as demonstrated in Chapter 4.

² The production and perception data of Seoul Korean in Chapter 4 clearly demonstrated this effect.
³ As previously mentioned, these focus conditions were produced in an experimental setting. In this setting, six native speakers of Seoul Korean read stimuli in isolation for broad focus and produced the same stimuli in a Q&A dialogue for discourse-new focus. The stimuli were repeated six times for both conditions.
Figure 6.2. Time-normalized pitch contours sampled at ten equidistant points. The shaded area indicates the focus position (Raw data from Lee & Xu, 2010). The sentence is *minsuga manduril məkninda* (Minsu is eating dumplings).

Prosodic focus in Seoul Korean has received considerable attention in the literature (e.g., Jun & Lee 1998; Lee & Xu 2010; Oh, 2008), yet our understanding of its exact nature is still incomplete. There has been no research examining how prosodic marking of focus interacts with the tonal contrast derived from different laryngeal gestures of an AP-initial onset consonant in Seoul Korean. In designing stimuli, previous studies have tended to exclude aspirated/tensed consonants in order to avoid pitch perturbation (e.g., Jo, Kang, & Yoon, 2006; Jun & Kim, 2007; Jun & Lee, 1998; Kim, Shin, & Kim, 2006; Lee & Xu, 2010; Lee, 2009, 2012; Oh, 2008). If these studies had included aspirated/tensed consonants, the potential differences caused by the tonal contrast might have been observed in marking prosodic focus. We assume that the pitch target of syllables with a high-tone-inducing onset becomes more increased in marking prosodic
focus due to the emergence of tonal contrast via tonogenesis in Seoul Korean, resulting in a better identification rate in perception.

The tonal contrast caused by the AP-initial onset consonant is an excellent case for testing whether prosodic marking of focus varies with different pitch-scaling conditions. As previously discussed in the Introduction, Liberman and Pierrehumbert (1984) suggest that the scaling of F0 values should be evaluated in terms of “baseline units above the baseline”, which can be formulated as \( F0 = Int \times B + B \) (see Chapter 1 for detailed information). Given that emphasis or focus has a multiplicative effect on the “Int” value, it is likely that higher pitches will demonstrate a greater impact than lower pitches, in terms of F0 ratios. This leads us to assume that higher pitches are more effective in marking prosodic focus than lower pitches, in both production and perception. With this point made, prosodic marking of focus, with a syllable beginning with aspirated/tensed consonants, will be more effective in production and also more identifiable in perception.

With these considerations in mind, we establish two research goals to increase our understanding of the nature of prosodic focus in Seoul Korean: a) to examine how prosodic modulation by focus differs by tonal contrast, and b) to test whether higher pitches are more effective in signaling prosodic focus. We approach these goals through production and perception experiments with 10-digit phone-number strings.

6.2. Production

6.2.1. Stimuli
We used the same production data described in Chapter 2. From the data set, we classified the numerical digits into two groups – High and Low – depending on the tonal contrast that they show. As shown in Table 6.1, the High tone group includes digits 3 [sam], 4 [sa], 7 [ṭeʰil], and 8 [pʰal], whose onset consonants are associated with aspiration/tenseness, as well as 1 [il], which is reported to be produced with a lexically specified H tone (Jun & Cha, 2011). The other digits (0, 2, 5, 6, 9) belong to the Low tone group. In the data set, each tone group was counterbalanced: there were 500 strings in the Low tone group and 500 strings in the High tone group.

Table 6.1. The onset consonant type of each digit and the tone group depending on the tonal contrast that each digit shows.

<table>
<thead>
<tr>
<th>Digit (IPA)</th>
<th>Onset Consonant Type</th>
<th>Tone Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (/goŋ/)</td>
<td>lenis</td>
<td>Low</td>
</tr>
<tr>
<td>1 (/il/)</td>
<td>vowel-initial</td>
<td>High</td>
</tr>
<tr>
<td>2 (/i/)</td>
<td>vowel-initial</td>
<td>Low</td>
</tr>
<tr>
<td>3 (/sam /)</td>
<td>aspirated</td>
<td>High</td>
</tr>
<tr>
<td>4 (/sa/)</td>
<td>aspirated</td>
<td>High</td>
</tr>
<tr>
<td>5 (/o/)</td>
<td>vowel-initial</td>
<td>Low</td>
</tr>
<tr>
<td>6 (/juk/)</td>
<td>glide</td>
<td>Low</td>
</tr>
<tr>
<td>7 (/ṭeʰil /)</td>
<td>aspirated</td>
<td>High</td>
</tr>
<tr>
<td>8 (/pʰal/)</td>
<td>aspirated</td>
<td>High</td>
</tr>
<tr>
<td>9 (/gu/)</td>
<td>lenis</td>
<td>Low</td>
</tr>
</tbody>
</table>

6.2.2. *A sketch of pitch contours*
Let us first describe some sample pitch contours in order to capture the prosodic differences between the broad focus and the corrective focus of each tone group. In this chapter, we obtained pitch contours sampled at ten equidistant points of each labeled digit, using ProsodyPro (Xu, 2013). Pitch contours in Hertz were then converted to semitones (st) by applying the following equation (Xu & Wang, 2009): $st = 12 \log_2 F0$.

Figure 6.3 displays the time-normalized pitch contours of broad focus and corrective focus, where the shaded area indicates a focus position, and the numerical digit in the shaded area refers to the target digit. For the sake of simplicity, Figure 6.3 only includes the first three digits (i.e., the first phrase) since the remaining digits are not necessary for the analysis. The digits 1 and 8 belong to the High tone group and the digits 5 and 6 to the Low tone group. Two noticeable features appear in Figure 6.3. First, prosodic marking of focus differs strikingly by tonal contrast: it seems that only the digits in the High tone group show a more expanded pitch range in the focus position, suggesting that prosodic marking of focus is more effective in the High tone group than in the Low tone group, and the High tone group would yield a higher accuracy rate in perception than the Low tone group. Second, it appears that focus does not modulate one single digit, but rather prosodic prominence spans over the entire phrase in both tone groups. All the plots of Figure 6.3 exhibit that the pitch level of positions 2 and/or 3 is also increased although only position 1 is focused. Therefore, prosodic marking of focus is still ambiguous even in the High tone group, which may result in a poor identification rate in perception.
Figure 6.3. Sample pitch contours for the Low and High tone groups in two focus conditions. BF and CF are abbreviations for broad focus and corrective focus, respectively.

6.2.3 Analyses and results

In evaluating how prosodic marking of focus differs by tonal contrast and testing whether or not higher pitches are more effective in marking prosodic focus, we compared directly the Low and High tone groups between the two focus conditions by the aggregate measures of mean pitch (st). For statistical analysis, we conducted a linear mixed model analysis based on the lmerTest package (Kuznetsova et al., 2013) in R (R Core Team,
Mean pitch was fitted into a model, where focus (broad vs. corrective) and tonal contrast (L vs. H tones) were used as fixed effects, and speakers (5 speakers), digit positions in a phone-number string (10 positions), and individual digits were used as random effects. In this model, broad focus in the Low tone group was the reference category. Furthermore, in order to assess whether or not both the Low and High tone groups produce an ambiguous prosodic marking of focus, our simple strategy was to calculate the mean pitch values (st) by subtracting corrective focus from broad focus in each string position, separated by the tone groups.

Figure 6.4a illustrates the mean pitch difference between broad focus and corrective focus separated by each tone group, and Figure 6.4b is the simplified version of Figure 6.4a. From Figure 6.4a, we observe that there was little or minimal difference between broad and corrective focus in the Low tone group. The difference between the two focus conditions, however, was highly discernable in the High tone group, and this difference was consistently observed in each of the focus positions.
Figure 6.4. Top panel: Points indicate mean values of each position. Bottom panel: Points indicate mean values and bars standard errors.

Table 6.2 shows the output of the linear mixed effects model of mean pitch (st). As shown in Table 6.2, the effect of focus on mean pitch was not significant ($p = 0.482$) in the Low tone group. The estimated mean pitch was 90.04 for broad focus and 90.27 for corrective focus, showing an increase of just 0.23 st extra pitch when marking prosodic focus in the Low tone group. The data indicate that speakers of Seoul Korean used a very minimal level of pitch in marking prosodic focus with the Low tone group. In addition, as expected, the effect of tonal contrast on mean pitch was significant ($p < 0.0001$). The
estimated pitch difference was 2.14 st between the Low tone and High tone groups in the broad-focus condition. That is, consonants beginning with [+aspiration/tense] induced higher pitches, whereas those beginning with [−aspiration/tense] induced lower pitches. An important point to recognize is that there was a significant interaction effect between focus and tonal contrast ($p < 0.0001$). As stated above, the estimated pitch difference between broad and corrective focus was 0.23 st in the Low tone group. That of the High tone group, however, was 1.62 st, meaning that the pitch expansion by focus was 1.39 st greater in the High tone group than in the Low tone group. The data clearly showed that prosodic marking of focus differed by tonal contrast; the High tone group showed clearer and more effective prosodic marking of focus than the Low tone group, as clearly demonstrated in Figure 6.4b. Therefore, the results of this study support our assumption that higher pitches are more effective than lower pitches for prosodic marking of focus.

Table 6.2. The output of the linear mixed effects model. Broad focus in the Low tone group is the reference category.

|                      | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | 90.04    | 2.02       | 44.55   | 0.0001 *** |
| Focus                | 0.23     | 0.32       | 0.73    | 0.482    |
| Tonal contrast       | 2.14     | 0.15       | 14.10   | 0.0001 *** |
| Focus*Tonal contrast | 1.39     | 0.21       | 6.46    | 0.0001 *** |

*Note: Significance code: “***” 0.001

Although prosodic marking of focus was clearer and more effective in the High tone group than in the Low tone group, it should be noted that both the Low and High tone groups showed no clear indication of focus positions, as demonstrated in Figure 6.5.
In the Low tone group, the pitch cues to focus positions were unclear and ambiguous: when position 1 was focused, position 2 was even higher; when position 4 was focused, position 5 was even higher; and when position 7 was focused, position 8 was even higher. The High tone group also showed similar levels of ambiguity, but to a lesser extent: when position 1 was focused, position 2 was even higher; when position 9 was focused, position A was even higher; and when positions 4 and 7 were focused, the following positions showed a similar (but smaller) level of pitch.

![Figure 6.5. Means by corrective focus minus broad focus. Characters 1-A indicate focus position from 1 to 10.](image)

6.3. Perception

6.3.1. Data collection
The stimuli, design, and procedure were the same as described in Chapter 2. The perception data were classified in a confusion matrix to see how accurately each focus position is identified for each tone group.

6.3.2. Results

Table 6.3 exhibits a confusion matrix for the identification of corrected digits in each tone group. The overall identification rate was 50.8% for the High group, but it was just 23.8% for the Low group. The results indicate that tonal contrast and prosodic focus interacted asymmetrically in Seoul Korean, improving the identification of focus positions for the digits in the High tone group, but not for the digits in the Low group, confirmed by a binary logistic regression analysis ($\chi^2 = 262.4, df = 1, p < 0.0001$). The results of the perception experiment clearly demonstrated that higher pitches were more identifiable than lower pitches in perception.

---

4 One may wonder about the identification performance rate for each digit. As Table 6.i shows, there is a clear distinction between the tone groups; the High tone group received at least 41.5% for the identification rates, while the Low tone group received identification rates of no greater than 33.5%.

Table 6.i. The mean identification rate (%) for each digit. The areas shaded in gray refer to the High tone group.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10.4</td>
<td>55.8</td>
<td>19.6</td>
<td>41.5</td>
<td>42.7</td>
<td>33.5</td>
<td>21.9</td>
<td>52.3</td>
<td>58.1</td>
<td>28.8</td>
</tr>
</tbody>
</table>
Table 6.3. Confusion matrix of corrective focus perception (percentage values). Numbers highlighted in gray indicate correct identification rates. Dotted lines indicate a phrase boundary in a string. (Top panel: the Low tone group, Bottom panel: the High tone group)

<table>
<thead>
<tr>
<th>Perceived</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>19</td>
<td>31</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>18</td>
<td>27</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>4</td>
<td>63</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>8</td>
<td>15</td>
<td>27</td>
<td>7</td>
<td>18</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>14</td>
<td>6</td>
<td>17</td>
<td>28</td>
<td>16</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>5</td>
<td>12</td>
<td>17</td>
<td>4</td>
<td>25</td>
<td>13</td>
<td>10</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>16</td>
<td>6</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>15</td>
<td>15</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>3</td>
<td>10</td>
<td>23</td>
<td>4</td>
<td>4</td>
<td>23</td>
<td>9</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>8</td>
<td>18</td>
<td>11</td>
<td>3</td>
<td>8</td>
<td>19</td>
<td>4</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38</td>
<td>15</td>
<td>15</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>55</td>
<td>17</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>5</td>
<td>41</td>
<td>10</td>
<td>2</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>2</td>
<td>11</td>
<td>45</td>
<td>3</td>
<td>7</td>
<td>16</td>
<td>1</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>19</td>
<td>49</td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>66</td>
<td>16</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>66</td>
<td>10</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>6</td>
<td>25</td>
<td>21</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>59</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>68</td>
</tr>
</tbody>
</table>

From Table 6.3, we can observe that incorrect answers usually appeared within the same phrase, immediately before or after the focus position for both tone groups. For example, in the Low tone group, when position 1 was focused, listeners chose position 2
at a rate of 19% and position 3 at a rate of 31%. When position 2 was focused, listeners chose position 1 at a rate of 22% and position 3 at a rate of 27%. A similar trend was also observed in the second and third phrases. The High tone group also exhibited a similar confusion rate (but to a lesser extent). When position 1 was focused, listeners chose positions 2 and 3 at a rate of 15%. When position 2 was focused, listeners chose position 1 at a rate of 11% and position 3 at a rate of 17%. Again, the second and third phrases of the High tone group displayed a similar degree of confusion. The identification results of Table 6.3 clearly indicate that the ambiguity of prosodic modulation hindered the identification of corrected digits in both the Low and High tone groups.

Even if we calculated the identification rate by phrase, as illustrated in Table 6.4, the overall identification rate of the Low tone group would become 58.3% and that of the High tone group would become 70.0%. The identification rate of the Low tone group is still lower than that of the High tone group. This is because, as expected, the Low tone group yielded a smaller increase in pitch expansion when marking prosodic focus. Nonetheless, even though we score by phrase, the identification rate is still not high in both tone groups, suggesting that focus marking by prosodic modulation was actually weak in Seoul Korean.
Table 6.4. The phrase-by-phrase confusion matrix (Top panel: Low tone group, Bottom panel: High tone group).

<table>
<thead>
<tr>
<th>Target</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; phrase</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; phrase</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; phrase</td>
<td>73.1</td>
<td>11.4</td>
<td>15.5</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; phrase</td>
<td>30.6</td>
<td>53.3</td>
<td>16.1</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; phrase</td>
<td>29.1</td>
<td>22.2</td>
<td>48.5</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; phrase</td>
<td>68.8</td>
<td>19.6</td>
<td>11.6</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; phrase</td>
<td>12.7</td>
<td>69.2</td>
<td>18.0</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; phrase</td>
<td>17.8</td>
<td>10.1</td>
<td>71.9</td>
</tr>
</tbody>
</table>

6.4. Discussion

The goals of this study were 1) to examine how prosodic marking of focus interacts with (post-lexical) tonal contrast (H vs. L tones) found in Seoul Korean, and 2) to determine whether prosodic marking of focus varies with different pitch-scaling conditions. The method developed here led to a systematic and effective approach for evaluating the interaction between tonogenesis and prosodic focus and for identifying the effectiveness of higher pitches in marking prosodic focus. In our study, production and perception experiments produced congruent results.

We found an interesting interaction effect between tonal contrast and prosodic focus in Seoul Korean. The digits in the High tone group were realized with an even higher pitch under corrective focus than broad focus, whereas those in the Low tone group did not show such a difference. This effect was also shown in perception – the identification rate of the High tone group was twice as high as that of the Low tone group.
We posit that the pitch target of digits with a high tone becomes more increased in marking prosodic focus due to the tonal contrast via tonogenesis in Seoul Korean (Jun, 2005; Kang, 2014; Kingston, 2011; Silva, 2006; Wright, 2007).

The interplay between tonal contrast and prosodic focus provides a new insight into the prosodic structure of Seoul Korean. We have learned that the basic melody of an AP in Seoul Korean is patterned either LHLH or HHLH, where the initial tone differs according to the AP-initial onset consonants. It has long been believed that tonal contrast (L vs. H) is limited to the AP-initial position. Conversely, given that the difference between broad focus and corrective focus was clearly seen regardless of AP position for the high tone group, we speculate that tonal contrast is not only limited to the AP-initial position, but can appear in any AP position. This result suggests that the prosodic structure of Seoul Korean is now undergoing a change: the tonal contrast (L vs. H) is spreading to other AP positions, meaning that it is now being phonologized in Seoul Korean.

The findings of this study clearly confirmed that prosodic marking of focus actually varied with different pitch-scaling conditions. In this study, the High tone group induced higher pitches, whereas the Low tone group induced lower pitches. The higher pitches actually made an increase of 1.62 st in marking prosodic focus, whereas the lower pitches made just a mere increase of 0.23 st. The greater increase in pitch enabled listeners to identify the High tone group in a clearer and more effective way. Additional research is certainly necessary to support our findings with other languages, such as Cantonese, where there are four level tones, and Yoruba, where there are three level tones. It would be very interesting to see if prosodic marking of focus actually varies depending
on the pitch height of each level tone in these languages. It may be physiologically very demanding for the speakers of Cantonese and Yoruba to distinguish all different level tones in marking prosodic focus, given that the most comfortable pitch range is fairly limited for human speech: 100-180 Hz for male speakers; and, 170-250 Hz for female speakers (Kuang, 2013). If pitch is not sufficient to distinguish all the level tones, then which acoustic parameters will play a primary role in signaling focus? Future research with this question will provide a better platform to broaden our understanding of prosodic focus.

This study leaves us with one important question: why is prosodic marking of focus so ineffective in Seoul Korean? Given that prosodic focus requires simple exertion by a mere increasing of vocal effort, our production and perception data are striking. Future research needs to be done by collecting more prosodically similar and different languages in verifying the relationships of our findings with typological dimensions, such as stress-timing vs. syllable-timing (Abercrombie, 1967; Pike, 1945), stress (dynamic) vs. non-stress (melodic) accents (Beckman, 1986) and/or head-prominence vs. edge-prominence marking (Jun, 2005, 2012, 2014). If the existing prosodic typologies do not work, a wider range of data may provide a more useful basis for refining and/or developing a new prosodic typology, especially for prosodic focus.

In sum, this study revealed several interesting findings. First, tonogenesis and prosodic focus interacted in Seoul Korean so that digits with a (post-lexical) H tone were

---

5 We will re-address this issue in more detail in Chapter 7.
6 We will attempt to verify the relationships of our findings with previously-described typological dimensions in Chapter 7.
produced with a higher pitch under contrastive focus. Second, digits with [+aspiration/tense] consonants were more effective in marking prosodic focus when compared to those with [−aspiration/tense] consonants, which supports that higher pitches are more effective in production and more identifiable in perception. In this chapter, we found that even within a language where purely prosodic marking of focus is unclear and ambiguous, the identification rate of focus can be higher under certain circumstances. In line with the previous chapter, we continue to support the claim that purely prosodic marking of focus is neither universal nor automatic, even within the same language. Rather, we continue to show that prosodic marking of focus varies according to the specific prosodic system of a language.
Chapter 7

General discussion, future directions, and conclusion

This chapter is divided into four sections. In Section 7.1, we discuss the main findings of this dissertation by stating and answering each of the research questions that this dissertation addressed. Section 7.2 addresses the potential limitations that previously-described prosodic typologies had with respect to prosodic marking of focus between languages. In Section 7.3, we propose a series of open questions that will further expand and extend our working knowledge of prosodic focus. Finally, in the last section, we conclude this dissertation.

7.1. Main findings

The first question we asked was whether and how prosodic marking of focus varies across languages. We have learned that focus marking by prosodic modulation indeed differed across languages. On the basis of the findings, we have classified the languages into three groups, as shown in (1).
(I) Group 1: American English, Mandarin Chinese, and Standard French
Group 2: Seoul Korean, South Kyungsang Korean, and Tokyo Japanese
Group 3: Suzhou Wu

First, American English, Mandarin Chinese (except for tone 3), and Standard French showed clear prosodic marking of focus. Increased prosodic changes in duration, pitch, and intensity clearly indicated each focus position, and those cues were not found in adjacent positions. Therefore, each focus position received a high identification rate of over 90%. We believed that the “clear” prosodic modulation was due to the higher degree of freedom; these languages had enough room for variability in the distribution of prosodic focus, enabling any position to be prosodically prominent in a string.

Second, Seoul Korean, South Kyungsang Korean, and Tokyo Japanese exhibited weak and ambiguous prosodic marking of focus. Not only did focus positions display a minimal level of increased duration, pitch, and intensity, but increased prosodic cues were also often observed in surrounding positions within the same phrase. As a consequence, each focus position was not well identified in perception. We speculated that these languages had a lesser degree of freedom with little room for variability, leading to a “messy” distribution of prosodic focus.

Third, unlike in other languages, we have learned that Suzhou Wu’s prosodic marking of focus was position-dependent in both production and perception experiments. Some focus positions showed relatively clear prosodic marking of focus (but to a lesser extent), whereas other focus positions showed weak and ambiguous prosodic marking of focus. The perception results also displayed a position-dependent difference. Positions 1, 3, and 4 received an identification rate of over 60%, whereas positions 5 and 10 received
an identification rate of about 45% and 24%, respectively. The production and perception data suggest that although Suzhou is still a tone language, its prosodic patterns show a trend of changing from a tone language to a pitch-accent language.Phrase-medial and phrase-final syllables no longer retain their citation tones; instead, Suzhou Wu includes fixed tonal melodies determined by the initial tone of a phrase-initial syllable. As a result, the identification performance of positions 5 and 10 turned out to be poorer than other focus positions.

Taken as a whole, this dissertation demonstrates that prosodic marking of focus indeed varies across languages. Given our findings that prosodic modulation by focus was weak and ambiguous in some languages, our research challenges the view that focus (as a semantic/pragmatic notion) is expressed through a universal phonetic symbolism. This symbolism is thought to involve prosodic changes of increased duration, pitch, and intensity, whereby a focused element is more prominent than its adjacent elements within the focus domain (Büring, 2010; Samek-Lodovici, 2005; Truckenbrodt, 1995). Instead, we claim that purely prosodic marking of focus is neither completely universal nor automatic; instead, it varies according to the prosodic system of each language.

The second question of this dissertation concerned whether and how prosodic marking of focus differs within the same language. To pursue this question, we conducted two case studies by reanalyzing the production and perception data of Mandarin Chinese and Seoul Korean. We discuss the main finding of each case study one by one.

In Mandarin Chinese, focus positions were clearly marked with increased duration, intensity, and pitch range: listeners identified those focus positions correctly more than 90% of the time. But since tone 3 focus was expressed by lowering its pitch target, it offered a
smaller opportunity for pitch range expansion, and also yielded less intensity increase; in addition, local dissimilation increased the duration, intensity, and pitch range of adjacent digits around the focused tone 3 digit within the same phrase. As a result, tone 3 focus was less well identified by listeners (77.1%), which was actually at least 13% smaller than the identification rate of other focused tones. We suggested that the relatively poor identification of tone 3 focus was due to the smaller opportunity for pitch expansion, the confusion from local dissimilatory effects, and the relatively weak intensity. This case study supports the view, established in other work (Cao, 2012; Genzel & Kügler, 2010), that prosodic marking of focus is not always associated with pitch raising. Instead, in a tone language, prosodic marking of focus is expressed differently according to the pitch target of each lexical tone.

In Seoul Korean, prosodic marking of focus was neither clearly marked in production nor well identified in perception. However, we revealed a novel finding that prosodic marking of focus behaved differently depending on the tonal contrast (low vs. high tones) derived from a tonogenetic sound change recently introduced in Seoul Korean. Prosodic marking of focus with low-tone digits was neither effective nor well identified. In contrast, high-tone digits produced a more effective prosodic marking of focus in production and received an identification performance that was twice as high as that of low-tone digits.

The findings drawn from these two case studies underline the importance that 1) purely prosodic marking of focus varies even within a single language, in which the location of prosodic focus can be more difficult to identify in certain circumstances, and
that 2) prosodic marking of focus is neither universal nor automatic, but is expressed through the different prosodic conditions of the same language.

The third question this dissertation addressed was the phonological unit carrying various prosodic focus effects, such as prosodic focus marking and post-focus compression across languages. Based on the findings, we have divided the languages into three groups with the same clustering as previously shown in (1).

(2) Group 1: American English, Mandarin Chinese, and Standard French
    Group 2: Seoul Korean, South Kyungsang Korean, and Tokyo Japanese
    Group 3: Suzhou Wu

In the first group, the phonological unit carrying prosodic focus was a word. Each focused word (or each focused digit) was clearly marked by increased prosodic changes. Post-focus compression was consistently seen immediately after the focused word. In the second group, the phonological unit carrying prosodic focus was a phrase. Although each digit (or word) was focused, prosodic focus effects spanned over the entire phrase as a focused one. In this group, post-focus compression appeared across the phrase boundary. This means that although the first digit (or word) was focused in a string, post-focus compression did not occur until the second phrase was spelled out. In Suzhou Wu, increased prosodic changes aligned with each focused digit (or word) in some positions, while those changes were also observed in surrounding digits in other positions. This is why additional research is necessary to make a firm statement.

The fourth question dealt with whether higher pitches are more effective in marking prosodic focus than lower pitches in both production and perception. The
findings generated from the interaction effect between tonal contrast and prosodic focus in Seoul Korean successfully answered the question. In Seoul Korean, as previously discussed, syllables beginning with [+aspiration/tense] consonants induce a higher pitch, whereas those beginning with [−aspiration/tense] induce a lower pitch. The results showed that when the digits inducing a higher pitch were contrastively focused, prosodic modulation by focus was relatively clear with an increase of 1.61 st extra pitch. In comparison, when the digits inducing a lower pitch were focused, prosodic modulation by focus was little or very minimal with an increase of just 0.23 st. The difference was also shown in perception. The digits inducing a higher pitch demonstrated an identification rate that is twice higher than those inducing a lower pitch (23.8% vs. 50.8%). The findings from both production and perception experiments support the view that the scaling of pitch values should be expressed in terms of “baseline units above the baseline” (Liberman & Pierrehumbert, 1984). To our knowledge, this finding is the first indication that prosodic marking of focus varies with different pitch-scaling conditions, and it underlines the need for additional research to test for cross-linguistic validity.

7.2. Limitations of existing prosodic typologies

The findings of this dissertation present us with a mystery. It is not clear why speakers of American English, Mandarin Chinese, and Standard French successfully produced clear prosodic marking of focus and why speakers of Seoul Korean, South Kyungsang Korean, and Tokyo Japanese produced weak and ambiguous marking of focus. In Chapter 4, we speculated that the languages with clear prosodic marking of focus had a higher degree of
freedom in the distribution of prosodic marking of focus. Therefore, each focus position had clear prominence by enhancing syntagmatic contrast with increased prosodic modulation between adjacent positions. In addition, we speculated that the languages with weak and ambiguous prosodic marking of focus had a lower degree of freedom in the distribution of prosodic marking of focus.

The question now concerns the mechanism by which prosodic modulation by focus becomes clear and prominent. To identify the driving force behind this mechanism, let us compare these seven languages based on rhythmic units as well as types of word-level and phrase-level prominence marking, as Table 7.1 illustrates. Please note that the classification of languages in Table 7.1 was based on observation of the production data of a phone number string, so Tokyo Japanese is categorized as edge-prominent (see Chapter 3 for more information). First, with respect to the rhythmic unit, American English is stress-timed, Tokyo Japanese is mora-timed, and the rest are syllable-timed. Therefore, the rhythmic unit cannot be the main factor accounting for clear prosodic marking of focus since those languages with clear prosodic marking of focus showed mixed rhythmic units. Second, regarding word-level prominence marking, American English has lexical stress; Mandarin Chinese and Suzhou Wu have lexical tones; South Kyungsang Korean has lexical pitch accent; Standard French, Seoul Korean, and Tokyo Japanese have none of these features. Again, we cannot say that a clear prosodic marking of focus was derived from a certain word-level prosodic feature. This is because languages with lexical tones showed different prosodic behaviors in prosodic focus, and

---

1 Table 7.1 does not include accent-based typology since it cannot cover all the seven languages, as demonstrated in Chapter 3.
Standard French with no word-level prosodic feature produced clear prosodic marking of focus. Third, based on phrase-level prominence marking, American English, Mandarin Chinese, and Suzhou Wu are head-prominent. Standard French and South Kyungsang Korean are head/edge-prominent. Seoul Korean and Tokyo Japanese are edge-prominent. Once again, phrase-level prominence marking is unable to provide a clear reason behind why prosodic marking of focus was clear and prominent in some languages but weak in others.

Table 7.1. The classification of languages based on rhythmic unit and type of prominence marking at the word level and at the phrase level. Languages above the dotted line produced clear prosodic marking of focus, whereas languages below that line did not produce such a clear prosodic marking of focus (AE: American English, MC: Mandarin Chinese, SF: Standard French, LPA: lexical pitch accent, Head: Head-prominent, Head/edge: Head/edge-prominent, Edge: Edge-prominent).

<table>
<thead>
<tr>
<th>Language</th>
<th>Rhythmic unit</th>
<th>Word level</th>
<th>Phrase level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stress</td>
<td>Tone</td>
</tr>
<tr>
<td>AE</td>
<td>stress-timed</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>syllable-timed</td>
<td>(∧)</td>
<td>√</td>
</tr>
<tr>
<td>SF</td>
<td>syllable-timed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>syllable-timed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKK</td>
<td>syllable-timed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TJ</td>
<td>mora-timed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW</td>
<td>syllable-timed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This classification was based on the production data of digit strings.
However, given the finding that prosodic focus is expressed through the prosodic system of each language, we know there must be a similar prosodic force by which these languages demonstrate similar prosodic effects for focus marking. But the different prosodic patterns that these languages showed for focus marking do not seem to line up with any existing prosodic typologies. So for now, this must be regarded as an indicator of a new typological dimension, or as a function of a new typological space. In order to discover the mechanism of this mystery, we certainly need more data from various languages. Once we have a very large range of data, an opportunity to make a contribution to the study of prosodic typology will be available.

7.3. Implications

In this dissertation we have found that prosodic marking of focus is neither universal nor automatic, the reason for this being that prosodic marking of focus is expressed while being intimately connected with each language’s prosodic system. Nevertheless, we still believe that the universal phonetic symbolism of prosodic focus is to attract prominence and thereby enable a focused element to stand out from its adjacent elements. One may wonder why a language like Seoul Korean produced “weak” prosodic marking of focus. The reason is that this language actually had a smaller capacity for achieving sufficient prominence when compared to other languages, such as American English. However, when Seoul Korean had the opportunity to access a larger capacity for prominence through a tonogenesis-like sound change, this language began to produce clearer prosodic marking of focus. Therefore, the resulting prosodic marking of focus differed according
to tonal contrast (low vs. high tones) when marking prosodic focus. This difference suggests that low tones have a smaller capacity for prominence, whereas high tones have a larger capacity for prominence.

A similar line of evidence can be found by examining Suzhou Wu. In this language prosodic marking of focus was position-dependent. Some positions showed relatively clear prosodic marking of focus, while, in other cases, prosodic marking of focus turned out to be weak. Again, we believe that those positions with clearer prosodic marking of focus are associated with a larger capacity for prominence than other positions. Mandarin Chinese also demonstrated a similar trend. As opposed to other focused tones, tone 3 focus actually showed a smaller capacity for pitch range expansion.

From these three lines of evidence, we speculate that the universal nature of prosodic focus is expressed through prominence to clearly deliver important information to a listener. The different phenomena of prosodic focus across languages are solely attributed to the different prosodic system of each language; in particular, the difference appears to be dependent upon whether or not languages have sufficient room for triggering prominence.

Another implication is methodological. The technique used in the dissertation helped us determine the exact phonological unit carrying prosodic focus and post-focus compression in the languages of Seoul Korean, South Kyungsang Korean, and Tokyo Japanese. Previous studies of these languages have tended to include natural sentences as test stimuli, which did not allow identification of the exact phonological unit carrying prosodic focus. Since each word or phrase normally forms an accentual phrase (AP) in these languages, we only had the opportunity to observe prosodic focus effects at the
phrase level. However, by using a phone-number string, each string position (i.e., each word/digit) was effectively compared between broad focus and corrective focus, and this design offered an opportunity to identify the exact phonological unit carrying prosodic focus in these languages. The two panels of Figure 7.1 illustrate the similarities of prosodic focus effects between the natural sentence and the phone number string in Seoul Korean. While the first phrase in Figure 7.1a was focused, and the first word (i.e., digit) in Figure 7.1b was focused, the two panels clearly displayed similar prosodic focus effects. A small degree of prominence by focus spanned the entire phrase, and a minimal level of post-focus compression occurred towards the end of the second phrase. As a result, we can say that the exact phonological unit carrying prosodic focus is a phrase, and post-focus compression appears across the phrase boundary, regardless of whether a single word or a whole phrase is focused.
Figure 7.1. Time-normalized pitch contours, where the area in gray indicates a focus position. Figure 1a includes a natural sentence *Minsuga mandwil makinnda* (‘Minswu is eating dumplings’); Figure 1b consists of a digit string 637-686-7664. Abbreviations are BF (broad focus), DF (discourse-new focus), and CF (corrective focus).

7.4. Future directions

Although this dissertation has revealed several interesting findings about the area of prosodic focus, many open questions still remain (like many other dissertations). We hope that these questions will stimulate future research that may ultimately help improve our understanding of prosodic focus.
First, it is not evident whether clear prosodic marking of focus, without post-focus compression, is sufficient for the recognition of prosodic focus, or if a combination of clear prosodic marking and post-focus compression serves as a sufficient cue for the recognition of prosodic focus. We found that American English received an identification rate of 98.7% for position 10 without post-focus compression, while Mandarin Chinese received an identification rate of 74.2% for the same position. This result, therefore, challenges the view that post-focus compression is a highly effective cue for identifying focus (Liu & Xu, 2005; Xu et al., 2004; Xu et al., 2012). Alternatively, it may be possible that Mandarin Chinese relies on post-focus compression more than American English does when it comes to the recognition of prosodic focus. Whichever the case may be, additional research is required to explore the detailed relationship between the presence or absence of post-focus compression and perceptual accuracy. Since it is almost impossible for one to control his or her vocal folds to create certain pitch patterns, we assume that synthesized/resynthesized stimuli are preferred over natural sentences for this kind of research. Consider Figure 7.2, which illustrates some samples of resynthesized pitch patterns. We obtained the outputs in Figure 7.2 by manipulating the pitch contour of a digit string 233-653-8838 produced by five American English speakers and averaged. For simplicity’s sake, Figure 7.2 only includes the first two digits.
Figure 7.2. Time-normalized pitch contours, where the first area includes a focused digit and the dotted lines refer to a digit boundary. BF and CF abbreviate broad focus and corrective focus, respectively.

Figure 7.2a features clear prosodic marking of focus, followed by a clear indication of post-focus compression, labeled as Model 1. Figure 7.2b includes clear prosodic marking of focus without any indication of post-focus compression, labeled as Model 2. Figure 7.2c shows weak prosodic marking of focus, but followed by a clear indication of post-focus compression, labeled as Model 3. We know that the model in Figure 7.2a will be very well identified with an identification of over 90%. What about Figure 7.2b?
Although it shows clear prosodic marking of focus, it lacks post-focus compression in the post-focus area. If this model receives an identification rate of over 90%, the result will suggest that prosodic focus can be well identified even without post-focus compression. Conversely, if it is relatively poorly identified, this will be the evidence that post-focus compression serves as a highly effective cue to focus. In addition, if the model in Figure 7.2c is well identified, this will indicate that post-focus compression is a useful prosodic cue that contributes positively to the recognition of prosodic focus. This line of research will be significant in determining whether a close correlation between the presence/absence of post-focus compression and perceptual accuracy exists.

The next question that requires a thorough investigation is whether multiple level tones, which are already crowded in the tonal space, can effectively contrast with each other when marking prosodic focus. We found that prosodic marking of focus varies with different pitch-scaling conditions. Focus scaling was expressed as multiplicative in terms of F0 ratios. As a consequence, higher pitches or higher tones had a greater impact on prosodic focus than lower pitches or lower tones. Our interest here is whether focus scaling also works well for tonal languages with multiple level tones such as Cantonese and Black Miao. Cantonese consists of four level tones (Chao, 1948): tone 4 [55], tone 3 [33], tone 6 [22], and tone 4 [11], where the numbers in square brackets refer to the pitch height of each tone, where 1 is lowest and 5 is highest. Black Miao includes five level tones (Kuang, 2013): tone 8 [11], tone 4 [22], tone 6 [44], and tone 3 [55]. If we now apply focus scaling to these level tones, although there will be a certain amount of increase in pitch for each focused level tone, higher tones will produce higher pitch values. What remains uncertain is whether these focused tones are able to maintain their
level-pitch contrasts in the tonal space. Given that our pitch range is fairly limited, within 100 Hz (Baken & Orlikoff, 2000; Keating & Kuo, 2012; Kuang, 2013), it will be very difficult to effectively disperse all the focused level tones in the tonal space. If pitch is not sufficient to distinguish all the focused level tones, then it makes sense to identify which acoustic parameters, such as duration, intensity, or voice quality, play key roles in the multiple level-tone contrasts in terms of prosodic marking of focus.

Fourth, the findings of this dissertation raise questions for second-language acquisition. We found that some languages like American English produced a clear prosodic marking of focus, while other languages like Seoul Korean produced a weak and ambiguous prosodic marking of focus. It will be interesting to see whether native speakers of a language like Seoul Korean can accurately perceive the intended position of corrective focus in a language like American English and can themselves convey this information as L2 speakers. If they are easily able to do so, this raises the question of why analogous kinds of modulation are not used in speaking their own language. If their L2 production and perception abilities are limited in this respect, it raises questions for second-language instruction.

Another important question that needs to be investigated further is that of modeling the detailed relationship between prosodic modulation and perceptual accuracy. In this dissertation, the languages of American English, Mandarin Chinese, and Standard French received an identification rate of over 90%. However, the location of corrective focus was not clearly identified (showing a rate of far less than 90%) in the languages Seoul Korean, South Kyungsang Korean, Tokyo Japanese, and Suzhou Wu. If we say that an identification rate of > 90% is considered a “good” identification rate, it will be
interesting to identify what kinds of prosodic cues are involved in reaching “good” identification performance. In other words, what are the minimum threshold values in duration, intensity, and pitch necessary to produce a clear indication of prosodic focus? In addition, it will also be interesting to examine what features listeners pay attention to and to see whether we will be able to use machine-learning techniques to match their successes (or failures).

Next, the study of prosodic focus will be useful as a means of resolving the long-standing debate that surrounds the prosodic structure of certain languages like Yoruba. It is well known that Yoruba has three contrastive level tones: high (H), mid (M), and low (L). Among these, the exact nature of the mid tone is still open to debate. It has been claimed that a mid tone is phonologically unspecified (Ajibóyè, Déchaine, Gick, & Pulleyblank, 2011; Akinlabi & Liberman, 2000), reflecting only the intermediate status between H and L. Such a phonologically unspecified tone gets its pitch values by interpolation (Pierrehumbert & Beckman, 1988). Given the finding that prosodic marking of focus is expressed through the pitch target of each lexical tone in a language like Mandarin Chinese, we assume that if a mid tone has its own pitch target, its prosodic marking of focus will be in accord with that pitch target, thus behaving differently from other tones. On the other hand, if a mid tone is unspecified as previously suggested, then its prosodic marking of focus will vary depending on the surrounding tones. We believe that this approach to understanding will help future scholars find a new paradigm for clarifying the issues associated with the prosodic structure of certain languages, especially where a certain tone is considered unspecified.
7.5. Conclusion

This project was aimed at increasing our understanding of prosodic focus, both cross-linguistically and intra-linguistically. We conducted both production and perception experiments using the comparable data collection procedure with a phone-number string in the style of American English. This technique enabled us to examine the various prosodic focus effects of seven languages in a comparable way. This dissertation has demonstrated that prosodic marking of focus varies across languages. Some languages showed clear prosodic marking of focus in production, resulting in a good identification rate in perception. Other languages showed weak and ambiguous marking of focus, leading to poor identification performance. This dissertation also revealed that prosodic marking of focus differed even within a single language. In a language (Mandarin Chinese) where prosodic marking of focus was clear, the location of corrective focus was more difficult to identify under certain circumstances. Furthermore, in a language (Seoul Korean) where prosodic marking of focus was weak and ambiguous, the location of corrective focus was easier to identify under certain circumstances. This dissertation, therefore, underscores the claim that prosodic marking of focus is neither universal nor automatic; instead, it is expressed through the prosodic structure of each language.
import random
Results = range(0,10)
Results[0] = range(0,10)*10
random.shuffle(Results[0])
for position in range(0,9):
    Results[position+1] = range(0,100)
NextValues = [x[:] for x in [range(0,10)]*10]
for n in range(0,10):
    random.shuffle(NextValues[n])
for count in range(0,100):
    Results[position+1][count] = NextValues[Results[position][count]].pop()
for line in range(0,100):
    numStr = ""
    for item in range(0,10):
        if item == 3 or item == 6:
            numStr += "-"
        numStr += str(Results[item][line])
    print numStr
Appendix B - Strings for broad focus

(Boldfaced and underlined digits are target digits)

1. 787-412-4699
2. 105-601-2318
3. 734-592-8426
4. 011-427-7050
5. 366-802-0565
6. 446-187-6163
7. 149-665-0810
8. 654-645-3455
9. 759-166-6017
10. 499-715-5960
11. 996-209-6296
12. 518-634-9587
13. 861-333-0905
14. 395-840-1475
15. 389-343-4492
16. 040-776-9911
17. 158-317-5785
18. 076-790-6956
19. 579-498-0139
20. 581-994-8704
21. 233-653-8838
22. 875-246-5197
23. 845-397-4866
24. 344-142-6806
25. 320-404-7259
26. 257-175-4972
27. 410-616-4790
28. 791-873-6416
29. 770-120-5328
30. 129-883-2724
31. 648-472-2851
32. 947-956-8349
33. 182-449-3176
34. 471-252-1382
35. 662-018-3333
36. 609-554-4273
37. 485-568-2071
38. 036-678-1547
39. 460-852-0379
40. 298-832-3889
41. 065-977-0461
42. 543-270-8629
43. 026-988-4358
44. 335-084-5027
45. 174-214-6615
46. 988-103-7578
47. 007-079-7934
48. 052-547-3522
49. 839-221-3208
50. 450-962-7842
51. 280-506-1114
52. 213-385-7369
53. 304-323-1895
54. 112-813-3921

155
Appendix C - Strings for corrective focus

(Boldfaced and underlined digits are corrected digits)

2. Q: 105-611-2318  A: 105-601-2318
26. Q: 258-175-4972  A: 257-175-4972
29. Q: 770-130-5328  A: 770-120-5328
32. Q: 047-956-8349  A: 947-956-8349
34. Q: 471-262-1382  A: 471-252-1382
36. Q: 619-554-4273  A: 609-554-4273
38. Q: 036-678-1548  A: 036-678-1547
40. Q: 298-832-3880  A: 298-832-3889
41. Q: 065-977-0462  A: 065-977-0461
42. Q: 543-270-8729  A: 543-270-8629
43. Q: 026-988-4359  A: 026-988-4358
44. Q: 335-084-6027  A: 335-084-5027
46. Q: 988-203-7578  A: 988-103-7578
47. Q: 007-079-7935  A: 007-079-7934
48. Q: 152-547-3522  A: 052-547-3522
49. Q: 939-221-3208  A: 839-221-3208
50. Q: 460-962-7842  A: 450-962-7842
52. Q: 223-385-7369  A: 213-385-7369
54. Q: 112-913-3921  A: 112-813-3921
Appendix D – Strings for Suzhou Wu

There are 10 strings in each row and column as shown in the following matrix. We randomized the digit combination for each speaker and provided each speaker with a different set of 10 10-digit strings.

```
0  8  2  4  5  3  6  1  9  7
2  5  3  7  1  8  0  4  6  9
9  3  8  0  4  7  1  5  2  6
4  1  6  5  2  0  9  3  7  8
6  9  5  3  8  2  4  7  1  0
3  7  0  9  6  4  8  2  5  1
1  0  4  8  7  6  5  9  3  2
8  4  1  2  3  9  7  6  0  5
7  6  9  1  0  5  2  8  4  3
5  2  7  6  9  1  3  0  8  4
```

Digit strings for each speaker:

<table>
<thead>
<tr>
<th>Speaker 1</th>
<th>Speaker 2</th>
<th>Speaker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(527) 691-3084</td>
<td>(980) 673-2514</td>
<td>(906) 851-2473</td>
</tr>
<tr>
<td>(082) 453-6197</td>
<td>(013) 248-6957</td>
<td>(638) 129-4750</td>
</tr>
<tr>
<td>(253) 718-0469</td>
<td>(128) 354-9760</td>
<td>(123) 548-7609</td>
</tr>
<tr>
<td>(938) 047-1526</td>
<td>(234) 169-5078</td>
<td>(514) 267-3098</td>
</tr>
<tr>
<td>(416) 520-9378</td>
<td>(341) 526-7809</td>
<td>(792) 083-6541</td>
</tr>
<tr>
<td>(695) 382-4710</td>
<td>(467) 835-0291</td>
<td>(857) 610-9234</td>
</tr>
<tr>
<td>(370) 964-8251</td>
<td>(572) 490-1386</td>
<td>(480) 372-1965</td>
</tr>
<tr>
<td>(104) 876-5932</td>
<td>(695) 702-8143</td>
<td>(249) 735-0816</td>
</tr>
<tr>
<td>(841) 239-7605</td>
<td>(706) 981-3425</td>
<td>(071) 496-5382</td>
</tr>
<tr>
<td>(769) 105-2843</td>
<td>(859) 017-4632</td>
<td>(365) 904-8127</td>
</tr>
<tr>
<td>Speaker 4</td>
<td>Speaker 5</td>
<td>Speaker 6</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>(702) 583-6941</td>
<td>(842) 793-6051</td>
<td>(902) 573-6814</td>
</tr>
<tr>
<td>(438) 219-7605</td>
<td>(017) 629-3548</td>
<td>(218) 649-3057</td>
</tr>
<tr>
<td>(243) 168-7509</td>
<td>(230) 148-5769</td>
<td>(023) 168-5749</td>
</tr>
<tr>
<td>(514) 627-3098</td>
<td>(324) 516-7890</td>
<td>(374) 216-8590</td>
</tr>
<tr>
<td>(857) 941-0236</td>
<td>(153) 264-8907</td>
<td>(157) 824-9306</td>
</tr>
<tr>
<td>(129) 035-4687</td>
<td>(478) 935-0216</td>
<td>(489) 735-0261</td>
</tr>
<tr>
<td>(071) 496-8325</td>
<td>(681) 470-9325</td>
<td>(531) 490-7628</td>
</tr>
<tr>
<td>(365) 804-9172</td>
<td>(965) 807-4132</td>
<td>(865) 907-4132</td>
</tr>
<tr>
<td>(986) 750-2413</td>
<td>(796) 051-2483</td>
<td>(796) 051-2483</td>
</tr>
<tr>
<td>(690) 372-1854</td>
<td>(509) 382-1674</td>
<td>(640) 382-1975</td>
</tr>
</tbody>
</table>
Bibliography


doi:10.1016/j.specom.2013.07.010


Kim, J. (2009). Focus realization: Not by focus-to-accent but prosodic structure. In M. Abdurrahman, A. Schardl, & M. Walkow (Eds.), *NELS 38, volume 2* (pp. 1–14). Amherst: GLSA.


