THE REPRESENTATION AND COMPUTATION OF INTONATION

by

HYUNJUNG JOO

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Linguistics

Written under the direction of

Adam Jardine

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May 2026

ABSTRACT OF THE DISSERTATION

The Representation and Computation of Intonation

by HYUNJUNG JOO

Dissertation Director:

Adam Jardine

This dissertation investigates how *discrete* and *continuous* f0 information is encoded in the *representation and computation of intonation* using mathematical logic and speech perception data. Within intonational phonology, the autosegmental-metrical (AM) model treats discrete tonal targets as intonational primitives, whereas the configurational approach views holistic shapes as their primitives. However, they do not account for both discrete tonal targets and continuous f0 information in a restrictive way. Therefore, this dissertation defines intonation by connecting both discrete and continuous information using *model theory and logic*, a framework fruitfully used in computational phonology to represent and compute complex linguistic structures.

The first contribution of this dissertation is to provide an experimental finding that *continuous f0 shape information* plays a significant role in distinguishing the lexical pitch accents in South Kyungsang Korean. This supports existing findings that emphasize the importance of f0 shape information for phonological contrast.

The second contribution of this dissertation is to view intonation as *a quantifier-free (QF) logical interpretation of a metrical and prosodic structure*, focusing on different intona-

tional types: head-prominence (American English), edge-prominence (Seoul Korean), and head/edge-prominence (Tokyo Japanese). The discrete tonal targets in these languages were found to be *literal copies* of prosodic elements, such as accented syllables/moras or phrasal boundaries. Importantly, they were always linked *locally* to their tone-bearing units (TBUs), which were similar to other phonological patterns that fall in the regular upper bound of phonology. This is one step towards understanding the representational principles that shape similarities and differences across intonation patterns in languages.

The last contribution of this dissertation is to connect discrete and continuous f0 information of the pitch accents in American English, using two different types of logic, *Boolean* and *fuzzy logic*, within the model-theoretic framework. Based on First-Order and Monadic Second-Order logic, both discrete tonal targets and continuous f0 information are defined with perceptual primitives of intonation over the temporal domain, but only referencing f0 information. Then, the definitions were connected with the annotated real-world f0 data of American English by extending Boolean logic to fuzzy logic. The results showed an accuracy of 81.9% using those definitions, comparable to the performance of several machine learning models.

Taken together, this dissertation offers a novel computational perspective on the representation of intonation by showing the *interpretability* between continuous and discrete f0 information.

Acknowledgements

The first and foremost gratitude goes to my advisor, Professor Adam Jardine. It was a huge blessing for me to learn from you at Rutgers, and I would not have been able to write this dissertation without your guidance. Thank you for opening my eyes to the mathematics of language. I truly enjoyed every meeting and was glad to see myself growing as I learned to write logical formulas. You always provided immense support, making me feel relieved and encouraged whenever I faced challenges. You and your family have been incredibly kind to my family and me, especially to my baby boy, Kai. Thank you for the warm invitations and the food that made us feel at home.

I would like to express my sincere gratitude to my dissertation committee for your invaluable comments and feedback, which have deeply enriched this work. I will continue to reflect on your insightful guidance in my future research. Professor Adam McCollum, thank you for your kindness and for always prioritizing my well-being with such genuine care. I am also deeply grateful to you for inviting my family to Thanksgiving, making us feel so welcome and at home in a new environment. Your warm and encouraging words provided a great source of comfort and helped me persevere throughout this journey. Professor Bruce Tesar, thank you for your unwavering support and warm encouragement. I

truly enjoyed your seminars on formal methods and learnability, which were always fascinating and deeply inspiring to me. Professor Scott Nelson, I am grateful to you for your valuable comments and insights that helped me connect phonetics and phonology with model theory and logic.

Professor Dorothy Ahn, I am deeply grateful to you for the warm and kind support you have given me since my very first semester. I truly cherished our conversations and the wonderful time we spent together in Seoul. Your presence at Rutgers was a true blessing to me. I also thank you for consistently inviting me to lab meetings and providing opportunities to continue my experimental research.

Professors Taehong Cho and Sahyang Kim, I am profoundly thankful to you both for leading me into the world of prosody and intonation. I am especially grateful to you for building my academic foundation from the ground up. I wouldn't have been able to continue this journey without your unwavering support, which helped me overcome many difficulties, and I always felt at home whenever I visited the lab in Korea.

Professor Mariapaola D'Imperio, thank you for the opportunity to work with you and learn from you about intonation. Professor Jongho Jun, thank you for always being so kind and supportive of my academic endeavors. I also express my gratitude to Professor Jeremy Steffman for kindly providing the perceptual data used in this study.

To the Rutgers community, thank you for being my family in the US. To my cohort—Ying, Jiayuan, and Marjorie—and to Gerry, Erkan, Merlin, Ariela & Aidan, Beryl, Chenli, Vinny, Ziling, Chaoyi, Tatevik, and Deen: thank you all for your wonderful friendship and support.

My heartfelt gratitude goes to Sangyoung Bae and Nate Koser for being my sister and

brother in the US. I will never forget your thoughtful guidance and the genuine kindness you showed my family. I always felt deeply cared for through your support.

Special thanks go to Juhyun Oh, Hyeonjeong Kim, and Eunbi Cho: thank you for your constant support, understanding, and for being such wonderful friends to me throughout this journey.

Many thanks go to my friends who shared this path—Jungyun Seo, Jiyoung Jang, Jinyoung Jo, Sanghee Kim, and Sarang Jeong: thank you for your companionship, which made my years in the United States truly joyful and meaningful.

I would like to thank the Andrew W. Mellon Foundation for the 2025-2026 Andrew W. Mellon Dissertation Completion Fellowship. This fellowship provided me with the invaluable opportunity to focus entirely on my dissertation during my fifth year.

Lastly and most importantly, my deepest gratitude goes to my family, who always believed in me and supported me through everything along this long journey. I truly believe that nothing I have achieved in my life would have been possible without the unconditional love I received from all of you.

To my dad, Yongha Joo: It was you who first sparked my passion for language. By teaching me English through poems and pop songs since I was a child, you inspired me to pursue this academic path. Thank you for your guidance and for helping me finish this long journey safely.

To my mom, Youngsim Kim: Thank you for being my constant source of strength. During my most difficult moments, our long phone calls and your positive reminders that "I can do it" gave me the courage to persevere. Thank you for walking every step of this way with me.

To my grandmother, Imrye Seo: I am deeply grateful for your constant prayers. I know you stayed awake praying every time I boarded a flight, and it is thanks to your devotion that I have finished this journey safely. I promise to continue my studies with the courage and confidence you have always wished for me.

To my brother, Hyunwoo Joo: Thank you for all the support throughout my life. I am so grateful to have you by my side. My sincere thanks also go to Sumin Kim for her kindness and support.

To my sister, Seoyeon Joo: Thank you for being my companion during the final years of my PhD. Having you at the same university in the US was a true blessing, and your care and encouragement helped me survive the toughest times.

To Okyeop Kim and Hongseok Lee: I was able to successfully complete my doctoral program thanks to your support. Even though I have been away in the U.S. and could not visit often, thank you for always welcoming me with such warmth and for providing an environment that allowed me to focus on my studies.

To my husband, Gyeongtaek Lee: I honestly do not think I could have endured these challenging times without you. Thank you for standing by my side, supporting me, and believing in me at every single moment. You are the reason I was able to stay strong.

To my dearest son, Jihan Lee: Thank you for coming into my life. Giving birth to you is the best thing I have ever done. Meeting your father and having you is the greatest fortune of my life.

경택오빠와 지한이, 부모님, 할머니, 가족들 사랑합니다.

To my family

Table of Contents

Chapter 1

Introduction

1.1 Overview

Human language is brought to life with melodies, rhythms, and phrasing. *Intonation* delivers linguistic messages by modulating changes in Fundamental Frequency (f0). For example, in English, the utterance "*Amelia*" can be expressed as a statement with a rising-falling contour (H* L-L%), which can be an answer to a question "*What is her name?*". But the same utterance can be expressed as a question with a falling-rising contour (L* H-H%) equivalent to "*Did you say Amelia?*". Likewise, f0 is used to convey lexical and pragmatic meanings across languages.

The goal of this dissertation is to study how these contrastive f0 contours are *represented* and *computed* in the abstract phonological structure. Unlike consonants and vowels, perceived as discrete speech events, f0 contours are inherently continuous and dynamic. Therefore, in order to fully account for the continuous f0 contour, it is important to figure out what kind of f0 information should be encoded in the representation of intonation.

Existing models of Intonational Phonology are not sufficiently restrictive to capture the fundamental nature of intonation. The *configurational approach* (e.g., Bolinger 1951, Hart et al. 2003, Hirst & Di Cristo 1998, Kingdon 1958, Crystal 1969, OConnor & Arnold 2004, Wells 2006, Halliday & Greaves 2008, Xu 2005) claims that intonation is stored and processed as *holistic shapes* (e.g., rises, falls), while the *Autosegmental-Metrical (AM) model* of intonation (e.g., Beckman & Pierrehumbert 1986; Ladd 2008; Pierrehumbert 1980, Arvaniti 2022) represents intonation as sequences of discrete tonal targets, Hs (highs) and Ls (lows). For example, the melody of a question in American English "*Can you swim?*" can be illustrated as a single rising line in the configurational models, while in AM theory, it can be represented as one of the tonal sequences: L* L-H%, each mapping with a syllable. Both have their drawbacks: configurational models do not clearly predict possible or impossible intonational contours, while the AM model does not specify what an f0 contour looks like in between L and H targets.

To fill this theoretical gap, this dissertation investigates how *discrete* and *continuous* f0 information is encoded in the *representation and computation of intonation* using mathematical logic and speech perception data. Specifically, this dissertation connects both discrete and continuous f0 information by referring to the *perceptual primitives* of intonation, which is illustrated in Figure 1.1. To formalize these discrete and continuous properties of intonation, this dissertation is grounded on the *model theory and logic*, a framework fruitfully used in computational phonology to represent and compute complex linguistic structures (e.g., Jardine 2017, Chandlee & Jardine 2019a, Koser et al. 2018).

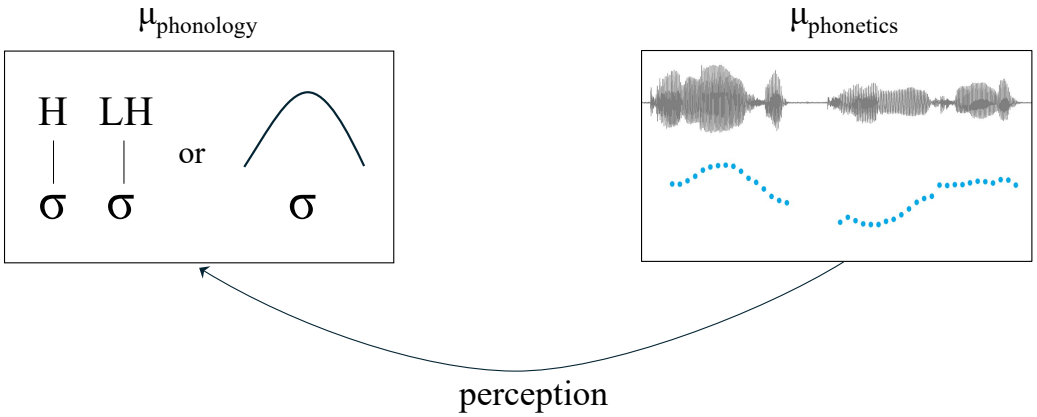


Figure 1.1: Overview of my dissertation.

Therefore, this dissertation first examines what kind of f0 information is crucial for defining intonation, by conducting a perception study on the lexical pitch accent in South Kyungsang Korean. This adds more evidence on the importance of f0 contour shapes to be encoded in the representation of intonation. With this finding, this dissertation views that not only the discrete tonal targets, but also continous f0 information should be encoded in the representation and computation of intonation.

Using model theory and logic, this dissertation defines intonation with the discrete tonal targets, by viewing intonation as a quantifier-free logical interpretation of a prosodic and metrical structure. By looking at different intonational patterns, head-prominence (American English), edge-prominence (Seoul Korean), and head/edge-prominence (Tokyo Japanese), this local logical interpretation explicitly shows similarities and differences across languages.

Lastly, this dissertation connects the discrete tonal units with the continuous information in the pitch accents in American English using Boolean and fuzzy logic. The fuzzy logical system defined for the pitch accent in American English is implemented using

annotated real-world f0 data, which is further evaluated by comparing with other data-driven machine learning methods.

The following section (1.2) lays out specific challenges in representing the discrete and continuous aspects of intonation, by contrasting two opposing frameworks: the AM model and the configurational approach.

1.2 Formal challenges in representing discrete and continuous information of intonation

This section specifies the formal challenges between these two representational approaches: the AM and the configurational approaches. A more detailed review of these approaches will be provided in Chapter 2. Although the AM model and configurational approach have their strengths, neither approach is restrictive enough to fully characterize the representation and computation of intonation.

First, the configurational approaches (e.g., Bolinger 1951, Hart et al. 2003, Hirst & Di Cristo 1998, Kingdon 1958, Crystal 1969, OConnor & Arnold 2004, Wells 2006, Halliday & Greaves 2008, Xu 2005) find it difficult to predict diverse f0 contours into a few configurations. For example, the configurationalists (e.g., Hart et al. 2003) attempt to reduce various f0 contours into a finite set of patterns, such as the pointed-hat, the hat pattern, and their combinations. In addition, by using the close-copy stylization method, they draw a minimal outline to capture various f0 configurations while maintaining perceptual equivalence. Moreover, other configurational frameworks such as Crystal (1969)

define intonational primitives as dynamic movements such as rises and falls, which inherently encode the continuous f0 information in intonation.

However, what is challenging for configurationalists is that there could be *infinitely many possible f0 configurations* if every perceptually relevant movement is taken into account. While studies in the configurational approaches propose a finite set of intonational primitives, the continous, gradient, and variable properties of f0 contour is difficult to be defined in a clear-cut way by just encoding the shapes and dynamics with the descriptive sets of primitives. Furthermore, we cannot constrain possible f0 patterns sufficiently to be encoded in the grammar within the computational capacity of human minds due to too many possible f0 patterns if we capture all the perceptually relevant variations.

On the other hand, while the AM model (e.g., Beckman & Pierrehumbert 1986; Ladd 2008; Pierrehumbert 1980, Arvaniti 2022) is more structured and generative in that it appears to account for all possible f0 patterns for intonation using discrete tonal targets, Hs and Ls. From the AM's viewpoint, however, it is difficult to account for the *shapes and dynamics* of an f0 contour, since f0 transitions between the discrete tonal targets are considered as byproducts of interpolating the targets.

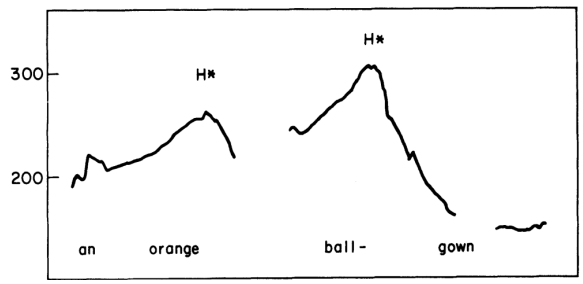
Several perceptual studies (e.g., D'Imperio & House 1997, Barnes et al. 2010, 2012, 2021, Kimball & Cole 2016) have shown that the detailed information such as f0 shapes plays a crucial role in distinguishing contrastive pitch accents in several languages (e.g., American English (Barnes et al. 2010, 2012, 2021, Kimball & Cole 2016), Neapolitan Italian (D'Imperio & House 1997), Bari Italian (Grice et al. 1995, French (Dorokhova & D'Imperio 2019)). Recently, Barnes et al. (2010, 2012, 2021) proposed a tonal center of gravity to account for both tonal target and f0 shape information. This various f0 information is re-

duced to a numeric value, but it does not explain how this continuous f0 information can be represented in the phonological representation of intonation. Therefore, it is important to figure out how continous f0 contours can be defined while incorporating the important information about the f0 contours—discrete tonal targets, shapes and dynamics.

In the literature of intonational phonology, there were no such restrictive models that explicitly formalize various f0 contours in intonation by both encoding the discrete tonal targets and the continuous f0 information in the representation of intonation. The goal of this dissertation is, therefore, to further examine what the perceptual primitives of intonation are and to provide an explicit and restrictive model that accounts for both the discrete and continuous f0 information in the representation and computation of intonation.

1.3 Research questions

A fundamental question of this dissertation is how intonation is *represented* and *computed*. More concretely, this dissertation deals with how *discrete* and *continuous* f0 information is encoded in the representation and computation of intonation. Figure 1.2 illustrates this overarching research question by showing that a *continuous* f0 contour can be represented as a sequence of *discrete* tonal targets, H* H* L- L% in the autosegmental structures, or vice versa.



INPUT: ⋊$_{\iota}$⋊$_{\varphi}$ ən ɔrɪnʤ bɔl gaʊn ⋉$_{\varphi}$⋉$_{\iota}$

OUTPUT: ⋊$_{\iota}$ ⋊$_{\varphi}$ ən ɔ* ɹɪnʤ bɔl* gaʊn ⋉$_{\varphi}$ ⋉$_{\iota}$

H* H* L- L%

Figure 1.2: An f_0 contour for a declarative in American English (Reprinted from Beckman & Pierrehumbert 1986.) and its autosegmental representation.

From a computational perspective of intonation, a simple string in the input can be mapped onto a hierarchically structured complex structure to the output, where starred tones are associated with starred syllables and phrasal and boundary tones are associated with the phrase-final syllable. Then, how do we compute intonation?

Concerning the representation and computation of intonation, this dissertation specifically asks the following questions:

RQ I: What kind of f0 information is crucial for defining intonation in the phonological representation?

RQ II: How can we mathematically define discrete tonal targets, Hs and Ls, in intonation?

RQ III: How can we connect the discrete and continuous f0 information in intonation?

Chapter 4 answers RQ I by empirically testing what kind of f0 information is the perceptual primitives of intonation. By conducting a perception experiment, this chapter examines how two main f0 cues, f0 peak alignment and rise shape, are perceived by South Kyungsang Korean listeners for H vs. LH distinction. The results show that *f0 rise shape information* plays a role in distinguishing lexical pitch accent in South Kyungsang Korean.

This chapter provides more evidence on the role of continuous f0 information for lexical distinction, further supporting the existing studies that highlighted the role of f0 shape information for sentential distinctions.

Based on this empirical evidence, the following chapters formalize the representation and computation of intonation using *both* discrete and continuous f0 information.

Chapter 5 answers RQ II by identifying what kind of discrete information is necessary to analyze intonational structure and its properties. This chapter defines intonation logically with reference to the hierarchical organization of stress, rhythm, and phrasing. The results show that the Hs and Ls patterns in intonation directly reflect these structures, similar to other phonological patterns that fall in the regular upper bound of phonology. This logical interpretation makes a strong prediction about the range of possible intonational patterns: intonational contours must reflect *local* information about stress and the beginning and ending of phrases. This is one step towards understanding the representational principles that shape similarities and differences across intonation patterns in languages.

Chapter 6 answers RQ III by connecting the *discrete* and *continuous* aspects of intonation using both Boolean and fuzzy logic within model theory. The definition of pitch accents in American English is formalized with a two-step process: first with Boolean logic (first-order and monadic second-order logic) and second with fuzzy logic. This shows how both the discrete and continuous representations of intonation can be formally accounted for within a logical framework. This combined approach offers a new perspective on bridging the two different kinds of intonational representation. The formal definition of the pitch accent in American English is tested with real-world f0 data by implementing the fuzzy logical system, and the results are compared with several machine learning

models.

1.4 Outline of this Dissertation

This dissertation is organized as follows: Chapter 2 provides a more detailed review of the theoretical background of two existing approaches: the AM (Section 2.1) and configurational models (Section 2.2). Intonational primitives of these approaches are mainly discussed.

Chapter 3 introduces the formal background used throughout this dissertation. It reviews previous studies on formalizing the representation and computation of intonation (Section 3.1). Then, the basics of model theory and logic are introuced (Section 3.2) by showing how autosegmental structures can be represented (Section 3.2.2). It also introduces two different types of logic, first-order and monadic second-order logic, with their syntax and semantics (Section 3.2.3 and Section 3.2.4).

Chapter 4 investigates perceptual primitives of intonation, focusing on those in a lexical pitch accent language, South Kyungsang Korean. A perceptual experiment was conducted to see what kind of f0 information, f0 peak timing or rise shape, is used to differentiate H versus LH pitch accents. The methods for the perceptual task are introduced (Section 4.3). Important empirical findings that f0 shape information is crucial to the phonological representation are provided and discussed (Sections 4.4 and 4.5).

Chapter 5 introduces the formal framework for analyzing intonation using discrete tonal targets. Specifically, logical transductions are introduced based on the study of other phonological structures, such as syllable structure (Section 5.1). Then, intonation

is viewed as a quantifier-free interpretation (Section 5.2). Within this framework, case studies are conducted to typologically examine different types of intonational patterns: head-prominence (American English; 5.3.1), edge-prominence (Seoul Korean; 5.3.2), and head/edge-prominence (Tokyo Japanese; 5.3.3) patterns.

Chapter 6 connects the discrete and continuous f0 information of intonation using both Boolean and fuzzy logic. Fuzzy logic is introduced and connected with a perception model (Section 6.1). Both the discrete and continuous information of intonation are defined with a two-step process: Boolean logic accounts for discrete aspects of intonation, followed by fuzzy logic to account for the gradient aspects of intonation (6.2). The definition of the pitch accents in American English is tested with real-world f0 data by implementing the fuzzy logical system, whose results are compared with several machine learning models (Section 6.3).

Lastly, Chapter 7 concludes this dissertation with discussion and conclusion.

Chapter 2

Theoretical Background on Intonation

This chapter reviews two theoretical approaches to account for intonation: the AM model and the configurational approach. The first line of theories is the *Autosegmental-metrical (AM)* model of intonational phonology (e.g., Beckman & Pierrehumbert 1986; Ladd 2008; Pierrehumbert 1980, Arvaniti 2022), which treats discrete tonal targets, Highs (Hs) and Lows (Ls), as phonological primitives of intonation. The global melody of intonation is said to be generated from sequences of the discrete tonal units. The f0 transitions between these tonal units are assumed to be computed via interpolation.

The AM's target-and-interpolation view contrasts with *configurational approach* (e.g., Bolinger 1951, Hart et al. 2003, Hirst & Di Cristo 1998, Xu 2005, Xu & Wang 2001), which views intonation as a gestalt or a holistic configuration. This view takes a functional perspective in which holistic shapes such as rises and falls directly reflects intonational categories (e.g., rising for questions, falling for statements, observed in diverse languages such as American English (Pierrehumbert & Hirschberg 2026), varieties of British English (Grabe et al. 2004), Spanish (Face 2007), and German (Niebuhr & Kohler 2004).

However, neither approach fully characterizes the fundamentals of the f0 contour in intonation. Thus, they are not sufficiently restrictive to deal with all possible f0 patterns in intonation. From the configurational perspective, it is difficult to predict diverse f0 contours into a few configurations. The AM model, while more structured, does not account for the shapes and dynamics of an f0 contour.

The following sections (2.1 and 2.2) provide a more detailed review of the two opposing theories of intonation, AM and configurational models, respectively. For a comprehensive comparison of these two theoretical frameworks, see Arvaniti (2009).

Portions of the literature review in this chapter were first published in Joo & D'Imperio (2025)[1].

2.1 Autosegmental-Metrical model of Intonational Phonology

The autosegmental-metrical (AM) model of intonational phonology has been the mainstream theory of intonational phonology since the early 1980s (e.g., Beckman & Pierrehumbert 1986; Ladd 2008; Pierrehumbert 1980, Arvaniti 2022). The notion of 'autosegmental' first came from the Autosegmental Phonology (e.g., Leben 1973, Williams 1976, Goldsmith 1976), which assumes that tones are autosegments, behaving independently from the segments.

Just like the autosegmental representation of lexical tones on an independent tier (e.g.,

[1]Joo, Hyunjung & D'Imperio, Mariapaola, The Perception of Lexical Pitch Accent in South Kyungsang Korean: The Relevance of Accent Shape, Language and Speech (OnlineFirst) pp. 1-33. Copyright © 2025 by Sage Publications. Reprinted by Permission of Sage Publications.

Leben 1973, Williams 1976, Goldsmith 1976), intonational tones can also be represented on an independent tonal tier, which was first introduced in the AM model (e.g., Beckman & Pierrehumbert 1986; Ladd 2008; Pierrehumbert 1980). Just like the segmental tier, the tonal tier in intonation consists of a sequence of phonological primitives such as Highs (Hs) and Lows (Ls), which are then connected by linear phonetic interpolation in most cases[2]. Importantly, these tonal units are associated with *metrically strong positions* (e.g., the head of a constituent) or phrasal boundaries (e.g., the right edges of intermediate phrase and Intonational Phrase), as shown in the prosodic structure of American English in Figure 2.1.

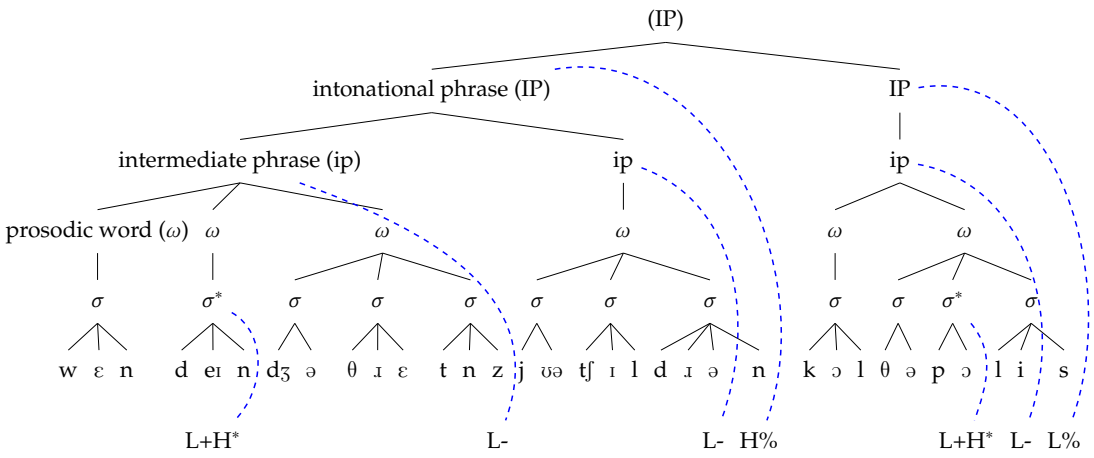


Figure 2.1: A prosodic structure of American English for an utterance, *When danger threatens your children, call the police*. Redrawn from Cho (2016).

In American English, a prosodic structure is hierarchically built from prosodic constituents. Syllables are put together to form a prosodic word, one or more of which are also put together to form an *intermediate intonational phrase* (ip). Then, one or more ips are grouped into an *Intonational Phrase* (IP). Within an ip, a *pitch accent* such as L*, H*, L*+H,

[2]Several studies (e.g., Pierrehumbert 1981, Ladd & Schepman 2003) reported a sagging transition between two H* pitch accents, which runs counter to the case of linear interpolation.

L+H* is associated with a *prominent syllable* (i.e., a starred syllable). A *phrasal tone* such as L-, H- is associated with an ip-final syllable, while a *boundary tone* such as L%, H% is associated with an IP-final syllable. The combination of these pitch accents, phrasal tones, and boundary tones becomes a global melody—*intonation*.

Intonation consists of at least one pitch accent, phrasal accent, and boundary tone. A finite-state acceptor in Figure 2.2 shows possible combinations of tones in intonation.

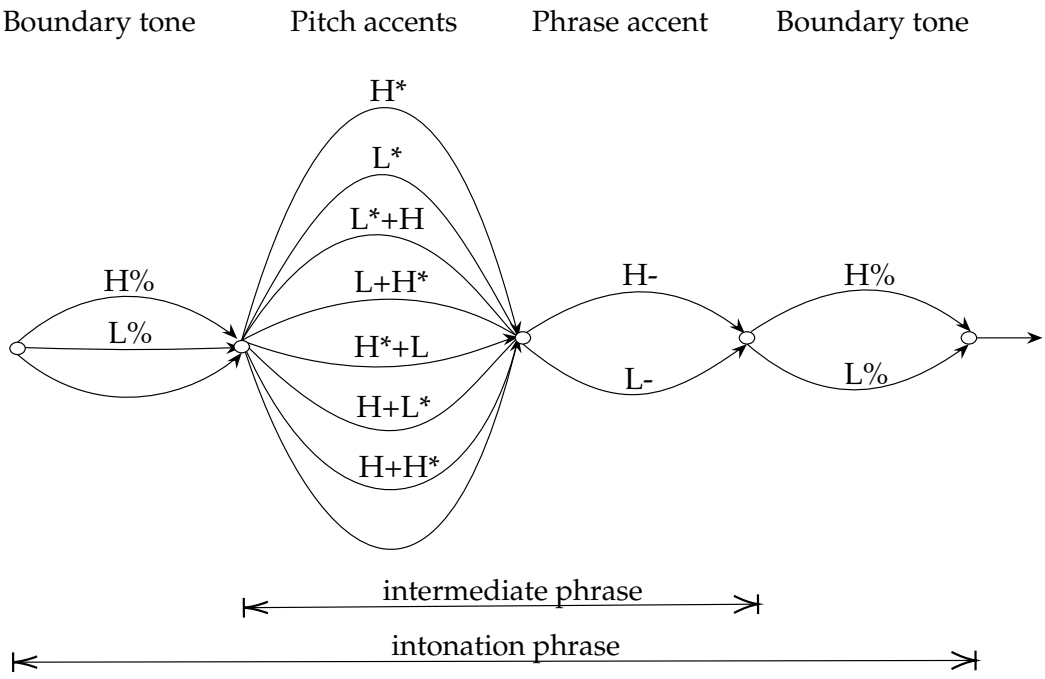


Figure 2.2: A finite-state acceptor for possible tonal combinations for American English intonation (Pierrehumbert 1980). For the later version, the bitonal pitch accents in American English have been reduced to L*+H, L+H*, H*+L, H+L*, excluding H+H*.

2.1.1 Intonational primitives

Within the AM model, intonational primitives are considered as *discrete tonal targets*, Hs and Ls (e.g., Beckman & Pierrehumbert 1986; Ladd 2008; Pierrehumbert 1980, Arvaniti 2022). These tonal targets are *associated* with *particular* tone-bearing units (TBUs) and are

realized as a melody. As shown in Figure 2.3, for instance, particular syllables, such as accent syllables or phrase-final syllables, are associated with tones, while some are not. Note that this sparse specification of intonational tones with TBUs is different from one-to-one mappings between lexical tones with TBUs in tonal languages.

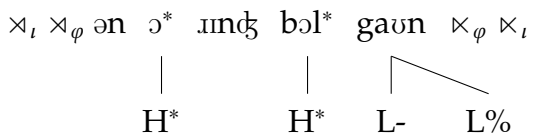


Figure 2.3: An autosegmental representation of a declarative intonational pattern in American English (H* H* L- L%).

According to Pierrehumbert (1980)'s AM description, pitch accents in American English are associated with *metrically strong positions*, such as the head of a constituent. The pitch accents are structured with either monotones (i.e., T*) such as L*, H*, or bitones (i.e., T*+T or T+T*) such as L*+H, L+H*, H*+L, or H+L*. A starred tone in the pitch accent (i.e., T*, T*+T, T+T*) is phonologically associated with a starred syllable. On the other hand, an *unstarred tone* in the pitch accent (i.e., T*+T, T+T*) is phonetically realized during the interval of the unstarred syllables before or after the starred tone. The unstarred tone before the starred tone is called a *leading tone*, while that after the starred tone is referred to as a *trailing tone*. In addition to pitch accents, phrasal tones such as H- or L- and boundary tones such as H% or L% combine to constitute a melody.

This AM framework has been successful due to its generativity and compositionality, demonstrating strong explanatory power by reducing continuous f0 contours into discrete tonal units (see Arvaniti (2022) for a review). Arvaniti (2022) illustrates this point by showing cross-linguistic f0 contours of a rise-fall-rise pitch pattern realized on mono-

syllabic or polysyllabic words. Similar rise-fall-rise melodies in monosyllabic words in English, Greek, and Polish showed varying f0 patterns across languages when stretched to polysyllabic words. Arvaniti (2022) notes that "neither can be said to be a stretched-out or squeezed version of the other".

The discrete tonal targets are phonetically realized in terms of *alignment*, as D'Imperio (2000) notes "alignment is a consequence of the notion of starredness and association." This tonal alignment in intonation has been substantiated by ample experimental evidence within the AM theory, showing that the tonal targets are consistently timed with segments in a principled way (e.g, Atterer & Ladd 2004; Arvaniti et al. 1998). Therefore, the tonal alignment with segments may be encoded in the intonational grammar within and across languages.

This timing relationship between tone and segment has been referred to as *segmental anchoring* in the AM theory (e.g., Arvaniti et al. 1998, see Ladd (2008) for a review). For instance, Arvaniti et al. (1998) found that the onset of the rise (i.e., f0 minimum) of the prenuclear pitch accent in Greek is aligned to the onset of the accented syllable, /ði/ in /roðitiko/ in (2.4a), and /rem/ in /paremvasi/ in (2.4b) in Figure 2.4. Also, the offset of the rise (i.e., the prenuclear accentual peak) is aligned to the onset of the following unstressed vowel, /ti/ in /roðitiko/ and /va/ in /paremvasi/. An important point here is that the alignment between f0 rising and a string of segments is relatively stable, especially in terms of L and H targets for the rise. That is, no matter what the intervening segments are (e.g., /ðit/ and /remv/), the L target is anchored to the onset of the accented syllable, while the H target is anchored to the onset of the following unstressed vowel.

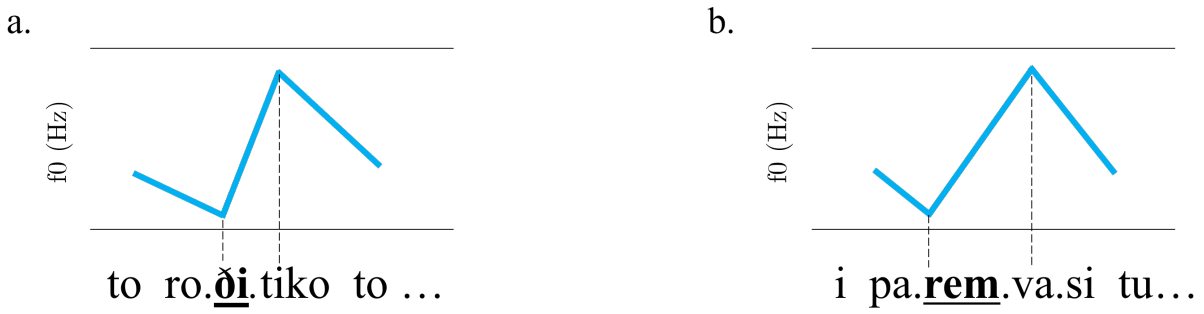


Figure 2.4: Schematized f0 contours of prenuclear pitch accent in Greek from Arvaniti et al. (1998). Redrawn based on Ladd (2008). The onset of the rise (i.e., f0 minimum) is aligned to the onset of the accented syllable, while the target of the rise (i.e., prenuclear accentual peak) is aligned to the onset of the following unstressed vowel.

This segmental anchoring also characterizes cross-linguistic differences in intonation. Atterer & Ladd (2004) show that rising pitch accents are associated with the same accented syllable in English, Greek, northern German, and Southern German. However, the specific timing of the tonal alignment within the segments differs across languages, as shown in Figure 2.5. For instance, the rising pitch accent is aligned slightly earlier with the stressed syllable for English, while it occurs slightly later for Greek. Even within the same language, the northern German dialect shows earlier alignment than the southern dialect.

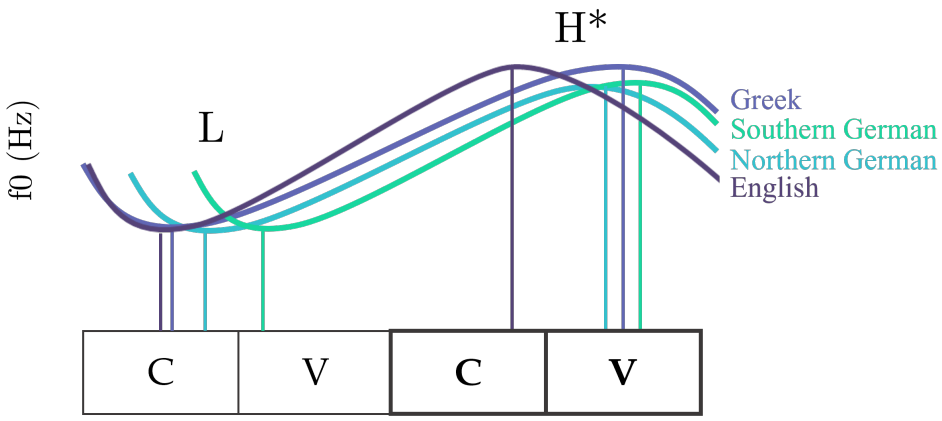


Figure 2.5: The alignment of a rising pitch accent in English, Greek, Southern and Northern German. CV in bold indicates the accented syllable. Redrawn based on Atterer & Ladd (2004).

According to the AM model, the only information related to the discrete tonal targets is encoded and decoded from the f0 contour, which serves as the primary perceptual reference for both speakers and listeners. Pierrehumbert & Steele (1989) found that the timing of f0 peak alignment is a crucial cue for distinguishing L*+H versus L+H* pitch accents in American English. The test sentence was *"only a millionaire"* with the pitch accent contrast, L*+H (meaning *uncertainty*) versus L+H* (meaning *assertion*). They conducted an imitation perception task, in which American English listeners were asked to imitate the f0 patterns, which were manipulated based on the f0 peak timing. The results showed earlier f0 peaks were produced as L+H*, while later f0 peaks as L*+H. This indicates that the f0 peaks are phonologically accessible information, such that they were able to deal with the continuum of the f0 peak timing into two discrete phonological categories.

However, many studies (e.g., D'Imperio 2000, Ladd & Schepman 2003, Pitrelli et al. 1994, Beckman & Hirschberg 1994) have pointed out the limitation of explaining intonational primitives solely in terms of the discrete tonal targets. For example, D'Imperio &

House (1997) examined the perceptual cues for a rise-fall pitch accent in distinguishing statement versus question in Neapolitan Italian. They manipulated two base sounds, a statement and a question, depending on the timing of f0 peak and rise/fall.

As for the f0 alignment, the entire rise-fall pitch pattern was shifted during the medial vowel in *dala̲la* based on the peak timing. Regardless of whether the base sound is a statement or a question, the overall results showed that earlier peaks were categorized as statements, while later peaks were perceived as questions.

As for the rise/fall timing, only the timing of the rise in the rise-fall pattern was manipulated for the question-based sounds, whereas only the timing of the fall in the rise-fall pattern was manipulated for the statement-based sounds.

Interestingly, the results varied depending on the base sounds. As for the different rise timing, the manipulated sounds were consistently perceived as questions. As for the different falling timing, the manipulated sounds with earlier falls showed question responses, which were gradually reduced with later falls.

D'Imperio & House (1997) argue that this is because the rise for the statement and the fall for the question that are located in the vowel are more perceptible than those located in the onset or coda of the syllable, such that the rise and the fall serve as a strong signal for the statement and the question, respectively. These results suggest that the perceptibility of the rising or falling is an important cue for pitch accent categorization in Neapolitan Italian, in addition to peak alignment. Further, this may offer evidence that information about the f0 turning points is not all that is needed to perceive rising or falling pitch accents.

In addition to the potential need for more than turning points in the intonational

phonology, many studies (e.g., D'Imperio 2000, Ladd & Schepman 2003, Pitrelli et al. 1994, Beckman & Hirschberg 1994) have pointed out that it is often very difficult to pinpoint the exact location of the f0 turning points. For instance, D'Imperio (2000) found that f0 contour on the stressed syllable *Lal-* in *Lalla* is perceived very clearly as a H+L* falling accent by Neapolitan Italian listeners. But in fact, the shape of the produced f0 contour is a plateau, which does not provide any discrete local points from continuous f0 contour.

This issue has also resulted in challenges for ToBI transcription, since f0 contours cannot be readily boiled down into f0 turning points (e.g., Ladd & Schepman 2003, Pitrelli et al. 1994, Beckman & Hirschberg 1994). For example, ToBI transcribers have difficulty distinguishing some pitch accent categories in American English, such as H* vs. L+H* or L+H* vs L*+H, because the differences between these categories does not appear to come from the tonal targets only but also from the contour shape. Despite these experimental results and transcription difficulties, AM views f0 contour shapes as epiphenomena, a byproduct interpolation between tonal targets.

2.1.2 Interpolation

Within the AM model, f0 transitions between two discrete tonal targets are generally assumed to be *linearly interpolated*. This is derived by phonetic rules for interpolation, rather than being encoded in the intonational phonology (Pierrehumbert 1980). Arvaniti (2022) notes that this f0 transition can be conceivable as an articulatory trajectory between constrictions in Articulatory Phonology (e.g., Browman & Goldstein 1992), where the trajectories are byproducts of target attainment. Several studies under the AM framework (e.g.,

Pierrehumbert 1980, 1981, Ladd & Schepman 2003) often point out that the f0 transitions are not necessarily linearly interpolated, as observed in sagging interpolations between the two H* pitch accents.

Importantly, from the AM's viewpoint, the interpolated f0 transitions are automatically generated by connecting the discrete tonal targets, so that they are not considered intonational primitives. Therefore, distinct f0 contour shapes for L*+H versus L+H* are viewed as arising from different tonal alignments (Arvaniti 2022). That is, what is the starred tone, L* versus H*, is the reason for having different contour shapes. However, the configurationalists (e.g., Barnes et al. 2010, 2012, 2021) may claim that distinct contour shapes themselves are what make the two bitonal pitch accents contrastive, which will be discussed in 2.2.1.

2.1.3 Intonational typology

Due to the AM's compositional nature of intonational structure, it is also possible to capture *similarities* and *differences* across languages within the typology of prosodic structure proposed by Jun (2006b, 2014, 2025). Jun (2006b) categorizes languages based on the properties of hierarchical intonational structure: *prominence* and *rhythmic/prosodic units*. Prominence refers to the type of prosodic constituents that are signaled with relative saliency compared to other constituents. It refers to whether the prominent constituents come from the lexical or post-lexical level, further specifying whether they are tone, stress, or lexical pitch accents (LPA) at the lexical level, or head or edge at the post-lexical level. As for the rhythmic/prosodic unit, the timing units for the rhythms are specified at the lexical

(e.g., mora, syllable, foot) or post-lexical levels (e.g., accentual phrase (AP), intermediate phrase (ip), and intonational phrase (IP)). Based on these properties, Table 2.1 provides the prosodic characterization of several languages from prosodic typology (Jun 2006b).

Language	**Prominence**					**Rhythmic/prosodic unit**					
	Lexical			Post-lexical		Lexical			Post-lexical		
	Tone	Stress	LPA	Head	Edge	Mora	Syll	Foot	AP	ip	IP
English		x		x				x		x	x
Spanish		x		x				x		(x)	x
German		x		x				x		(x)	x
Greek		x		x				x		x	x
Japanese			x	x	x	x			x	(x)	x
French				x	x		x		x	(x)	x
Basque			x	x	x		x		x	x	x
Korean					x		x		x		x

Table 2.1: Prosodic characterization of several languages from the prosodic typology (Jun 2006b). LPA refers to a lexical pitch accent, and AP indicates an accentual phrase. The parentheses indicate when the literature contains conflicting opinions.

In Table 2.1, for instance, English marks its prominence on the *heads* of constituents, which are predefined from the lexical stress. The rhythmic unit of English is hierarchically organized from the foot, the ip, to the IP. On the other hand, Korean marks its prominence only at the *edges* of constituents, and its rhythm is structured from the syllable to the AP, then to the IP. Japanese has both *head* and *edge* characteristics of prominence-marking at the post-lexical level, and its prominence is also determined by lexical pitch accents. The rhythms in Japanese prosody are defined in terms of AP and IP.

In Jun's recent studies (Jun 2014, 2025), these prosodic characterizations have been further refined by introducing three distinct intonational types: *head-prominence*, *edge-prominence*, and *head/edge-prominence*. For instance, as provided in Table 2.2, the *head-*

prominence languages include English, Spanish, and German, whereas the *edge-prominence* languages include Seoul Korean and West Greenlandic. *Head/edge-prominence* languages include lexical pitch accent languages such as Tokyo Japanese and Basque. In this way, languages can be compared with each other based on their intonational typology.

Table 2.2: Intonational types. Adapted from Jun (2025).

	Location of prominence	Example languages
Head-prominence	The *head* of constituents	English, German, Spanish
Edge-prominence	The *edge* of constituents	Korean, West Greenlandic
Head/Edge-prominence	Both the *edge* and *head* of constituents	Tokyo Japanese, Basque

Although Jun's prosodic typology (Jun 2006b, 2014, 2025) has *described* and *categorized* intonational patterns with crucial prosodic components, it is not *directly* and *explicitly* connected to the representation and computation of intonation. That is, this typological description has not been concretely materialized as a representation.

Therefore, this dissertation aims to show how this typological description can be computationally reflected in the representation of intonation using model theory and logic. Chapter 5 specifically defines what is being computed in terms of the input and output representations and their mappings. This will enable us to compare similarities and differences across different types of intonational systems.

2.2 Configurational Models

In configurational approaches, intonational primitives are viewed as a *gestalt* or *holistic configuration* such as rises and falls (e.g., Bolinger 1951, Hart et al. 2003, Hirst & Di Cristo

1998, Xu 2005, Xu & Wang 2001). This framework adopts a functional perspective in which distinct intonational patterns are directly mapped onto pragmatic meanings. This holistic view stems from Bolinger (1951)'s seminal work in the early 1950s, in which intonational primitives are considered as follows: "the basic entity of intonation is *a pattern*, not a pattern in the relatively abstract sense of grammatical recurrences, but in the fundamental, down-to-earth sense of *a continuous line that can be traced on a piece of paper*." In this perspective, intonation can be conceived as a sole and unique melodic line.

2.2.1 Intonational primitives

An influential work from Bolinger (1951) paved the way for the configurational view, by positing intonational primitives as *rises*, *falls*, and *sustentions*, which inherently encoded the f0 shape information. Bolinger conducted a series of perceptual experiments to see how listeners perceive similar configurations at different pitch levels and distinct configurations at the same pitch levels. They found out that similar configurations were identified as the same "pattern" regardless of pitch levels, supporting that f0 contours should be "geometrically" analyzed.

Following Bolinger (1951), configurationalists were largely grouped into two major schools in the 1990s: the Institute for Perception Research (IPO, Eindhoven) (e.g., Hart et al. 2003; de Pijper 1983), which treats *perceptually relevant pitch movements* as the primitives of intonation, and the British school (e.g., Cohen & Hart 1968, Halliday & Greaves 2008, see Nolan 2022 for a comprehensive review), which focuses on *distinct configurations* of nucleus in intonation.

A fundamental question in the IPO school is "what does the listener make of pitch in speech?" by mainly asking what kind of f0 information is encoded in the representation of intonation through perceptual studies (Hart et al. 2003). For this, they attempted to model different configurations of f0 contours varying in size, slope, duration, and position with respect to the syllable. One of their assumptions is that *only perceptually relevant pitch movements* must be taken into account based on four perceptual features of an f0 contour: directionality (rise or fall), f0 landmark timing with respect to syllable boundaries (early, late, or very late), change rate (fast or slow), and contour size (half or full) (Hart et al. 2003). For more details, see Hart et al. (2003).

Based on these perceptual features of an f0 contour, the IPO school describes f0 contours as being composed of a "pointed-hat" pattern (a rise-fall) or a "hat" pattern (a rise-plateau-fall) (Arvaniti 2009), as shown in (2.6a) and (2.6b) in Figure 2.6. The combination of these patterns makes a melody, as can be seen in (2.6c) in Figure 2.6. In the actual implementation, IPO researchers use a “close-copy stylization” method to capture the entire f0 movement pattern of an utterance. They draw the least number of straight lines that approximate a f0 contour, unless they are perceptually distinct from the original f0 contour.

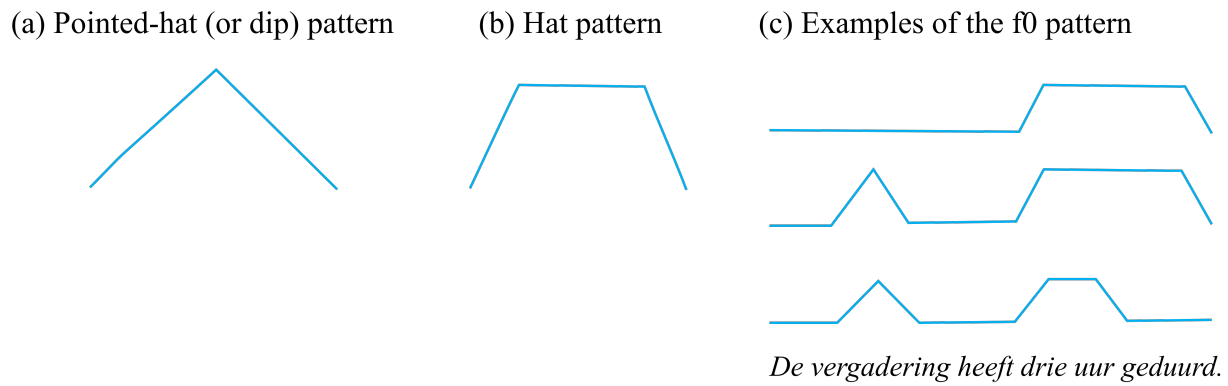


Figure 2.6: The intonational primitives in IPO school (Hart et al. 2003). Redrawn based on Hart et al. (2003).

Unlike the importance of segmental anchoring of f0 within the AM theory, the IPO school does not specify the timing relationship between the landmarks of intonation (e.g., turning points) and the segment in a principled way, due to the holistic view of f0 contour. They assume that the entire f0 configuration has an elastic property such that its length can be adjusted to the length of syllables.

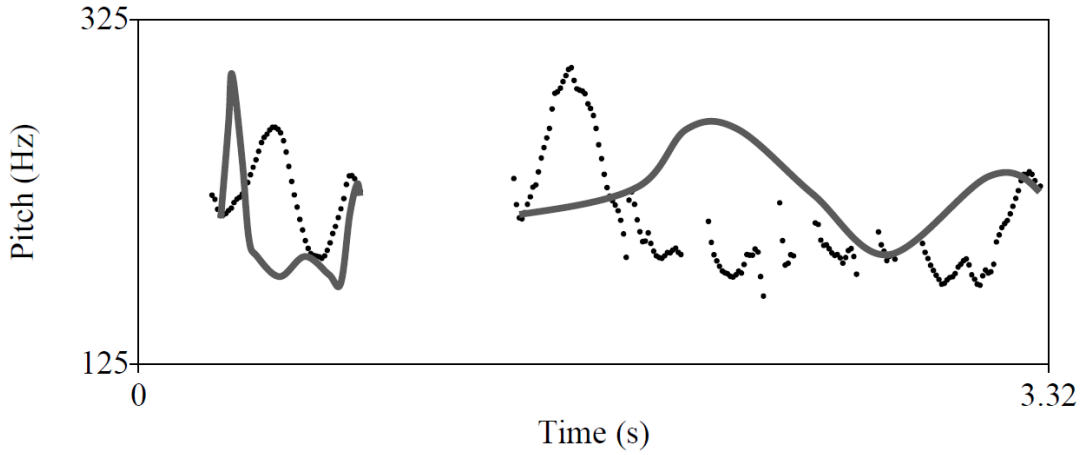


Figure 2.7: Examples of resized f0 contours depending on the number of syllables. From *The Representation of Intonation*, by Amalia Arvaniti, 2011. Copyright 2011 by John Wiley and Sons. Reprinted with permission.

In Figure 2.7, f0 contours are resized depending on the number of syllables. The dotted-

lined f0 contour on the left panel is expanded over the longer syllables, resulting in the gray-lined f0 contour on the right panel. In contrast, the dotted-lined f0 contour on the right panel is shrunk over the short syllables, resulting in the gray-lined f0 contour on the left panel. Note that the dotted- and gray-lined f0 contours in both shorter syllables and longer syllables do not show the same configurations. The f0 peaks and troughs are aligned very differently when the whole configuration is shrunk or expanded.

Under IPO, the mapping between specific landmarks of the f0 contour (e.g., peaks or troughs) and segments (e.g., vowel onset or offset) may allow for more variability than that assumed in AM. Nevertheless, the central goal of IPO work is to capture the various f0 configurations within an utterance, taking the overall shape of an utterance as a crucial element of intonation into account.

On the other hand, the British school (e.g., Kingdon 1958, Crystal 1969, OConnor & Arnold 2004, Wells 2006, Halliday & Greaves 2008) focuses on "dynamic pitch elements" in an f0 contour such as rises and fall-rise" (Nolan 2022). A main tenet of the British School is that "phonetic f0 trajectories are highly variable instantiations of abstract configurational sketches" as described by Iskarous & Cole (2026). Figure 2.8 shows an example representation of intonation in the British school, which encoded the shapes and dynamics in the representation of intonation.

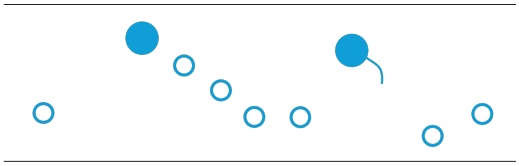


Figure 2.8: An example representation of intonation in the British school. Redrawn from Nolan 2022. The symbols in the text represent phrasal juncture, and the filled circles represent prominent syllables. A series of circles represents an f0 contour.

In one of the popular approaches within the British School (OConnor & Arnold 2004), intonation is structured with four elements, *prehead*, *head*, *nucleus*, and *tail*. Just like the nucleus in a syllable, a nucleus in intonation is considered an essential component that characterizes intonational categories such as questions or statements. The nucleus is realized with different types of f0 configurations over the prominent syllable, such as *low fall*, *high fall*, *rise-fall*, *low rise*, *high rise*, *fall-rise*, which directly reflect the overall f0 shape information.

In addition to the IPO and British school, another configurational view (e.g., Xu 2005, Xu & Wang 2001) captured a dynamic f0 contour in tone and intonation by positing four primitives that encode communicative meanings. The four primitives are local pitch target, pitch range, articulatory strength, and syllable duration. Based on these primitives, PENTA (Parallel Encoding and Pitch Target Approximation; e.g., Xu 2004, 2005) has been proposed to account for a "basic mechanism of tonal implementation", which is constrained by articulation (Xu 2005; p.227). Just like an articulator reaching its target, f0 asymptotes towards its tonal target, such as rise and low, by adjusting its velocity depending on the tonal target. This model also accounts for the tonal coarticulatory effect from the following f0 context. In line with other configurational approaches, the primi-

tives of the f0 contour in the PENTA model also encode its dynamics and movements for the target attainment.

Informed by configurational approaches, Barnes and colleagues (Barnes et al. 2010, 2012, 2021) propose a new way to account for intonational information that cannot be modeled in AM. Their Tonal Center of Gravity (TCoG) approach accounts for not only the alignment of tonal targets but also the transitions between the targets, such as rise/fall shapes, slopes, and duration. To account for variations in the f0 contour, TCoG calculates a weighted time point where the bulk of f0 rising (or falling) occurs, functioning as a perceptually important reference point.

As shown in 2.1, by quantifying the area under any portion of the f0 curve for rising (or falling) along the temporal dimension, the TCoG provides a unified measure point that captures variations of f0 configurations coming from durations, slopes, and shapes of f0 rising (or falling) contours. For instance, Figure 2.9 schematizes how TCoG works. (2.9a) and (2.9b) show different rising shapes, convex and concave shapes, respectively. By calculating the area under the overall f0 rising event averaged by time, we generate a left-shifted TCoG from the f0 peak for the concave shape and the right-shifted TCoG from the f0 peak for the convex shape.

$$Tonal\ Center\ of\ Gravity\ (TCoG) = \frac{\sum_i f0_i t_i}{\sum_i f0_i} \tag{2.1}$$

Recent studies (Barnes et al. 2010, 2012, 2021) have shown that f0 interpolation shape may influence the categorization of L+H* vs. L*+H in American English. Barnes et al. (2010) have examined the categorization of L+H* vs. L*+H pitch accent by manipulating

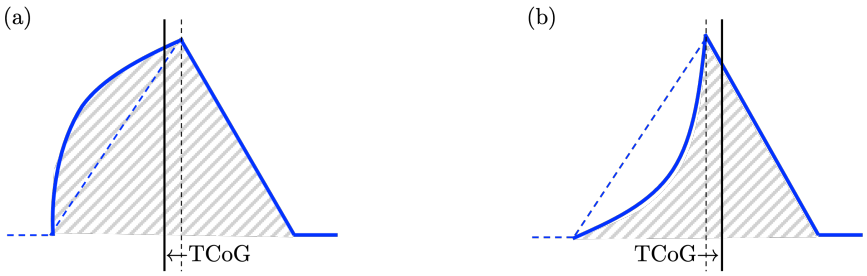


Figure 2.9: Schematized f0 contour of rising shape differences. (a) shows a domed rising shape, while (b) shows a scooped rising shape. The solid line indicates the TCoG. Redrawn based on Barnes et al. (2012).

both peak timing (earlier vs. later timed) and rise shape (scooped vs. domed). They have mainly looked at the interaction between peak timing and rise shape. Results showed that the earlier peaks and domed shapes were categorized as L+H*, while the later peaks and scooped shapes were categorized as L*+H, suggesting that both f0 configurations and tonal targets influence perception of pitch accents.

In Barnes et al. (2012), they specifically tested the performance of the TCoG and mainstream AM models. In the first experiment, they conducted a classification of L+H* and L*+H pitch accents produced by six American English speakers, and the results showed that the TCoG model significantly outperformed the AM model. In the second experiment, they added some noise to the data to test whether the TCoG and AM work when there are discontinuations in f0 contours and when it is difficult to pinpoint f0 peaks. Results showed that even in these tough situations to find f0 peaks, TCoG performed better than AM, showing a necessity of f0 interpolation shape as well as peak alignment to understand the overall f0 contours.

These results were confirmed in a follow-up study (Barnes et al. 2021). In this more recent study, they conducted a perception experiment to test whether not only f0 rising but

also falling shape (scooped or domed) matters in the categorization of L+H* and L*+H. Results showed that listeners responded to more scooped rise shapes as L*+H, while more domed rise shapes were categorized as L+H*. Moreover, a similar effect was observed by modifying the fall shapes within a rise-fall, suggesting that the entire f0 configurations including both f0 rising and falling need to be considered in pitch accent perception.

Further support for the significance of contour shape comes from Kimball & Cole (2016), arguing that shape information for the H* pitch accent for American English is actually stored in memory. They hypothesized that if f0 shape is a meaningful linguistic information, it would be retrieved from the memory much better in an AX discrimination task. Results showed that the plateau contours were more accurately identified than peaks in an immediate task, while this effect was even greater in delayed tasks. The greater discrimination effect in the delayed task shows that the plateau contours are remembered better than the peaks even after an interval, suggesting that fine-phonetic details about the plateau shapes may be encoded in the phonological representation of intonation. In sum, these studies support the significance of f0 contour shape in pitch accent categorization in American English.

In sum, these studies support the significance of f0 contour shape in pitch accent categorization in American English. Other (post-lexical) pitch accent languages have also shown the role of f0 shape in distinguishing sentential meaning. Grice et al. (1995) find that in Italian f0 contours within the pitch accent were categorically perceived depending on the size of the f0 dip within the L tone region. For the same sentence *lo mandi a Massimiliano* ('You send it to Massimiliano'), greater f0 dip in the pitch accent was identified as questions, while smaller f0 dip was related to commands. Recently, Dorokhova &

D'Imperio (2019) report similar effects in French. The rise shape of the LH* accent, more concave or more convex, influenced the perceptual shift of modality from continuation (*La beauté des canaux…* 'The beauty of the canals…') to question (*La beauté des canaux?* 'The beauty of the canals?'). Specifically, more concave shapes were identified as questions, while more convex shapes as continuations.

Chapter 3

Formal Background

This chapter introduces the formal framework used in formalizing the representation and computation of intonation in this dissertation. First, Section 3.1 reviews previous studies on formalizing the representation and computation of intonation. Then, Section 3.2 introduces the main mathematical framework, *model theory and logic*, which has been widely used in many representational studies in computational phonology. It also shows how the autosegmental structures are represented and computed explicitly within the model-theoretic framework in 3.2.2. Lastly, it introduces two different types of Boolean logic, *First-Order* and *Monadic Second-Order logic*, with their syntax and semantics in 3.2.3 and 3.2.4, respectively. These logical languages enable us to constrain the way the models are represented and computed in terms of their expressive power.

3.1 Previous studies on formalizing intonation

This section introduces how intonation has been represented and computed in terms of autosegmental representation. The autosegmental representation is grounded on the no-

tion of a metrical grid (e.g., Liberman 1975, Liberman & Prince 1977), providing a formal basis to build the autosegmental structures for the tone and TBU association in intonation (e.g., Liberman 1975, Liberman & Prince 1977, Goldsmith 1976, Pierrehumbert 1980).

The phonological association between tonal sequences and their corresponding TBUs is referred to as *text-tune association*. According to Liberman (1975), tunes (i.e., tonal sequences) are aligned with texts (i.e., TBUs) based on a metrical tree. Just like a syntactic tree, a metrical tree branches into two nodes that are labeled *s* (strong) and *w* (weak), respectively, as shown in Figure 3.1. An association rule first determines "the designated terminal node" (circled label), and matches each element in a tonal sequence to each node, as shown on the metrical tree on the left. Then, each syllable in the text is also matched to each node, as shown in the metrical tree on the right. With these matching processes, the mapping between tones and TBUs is computed.

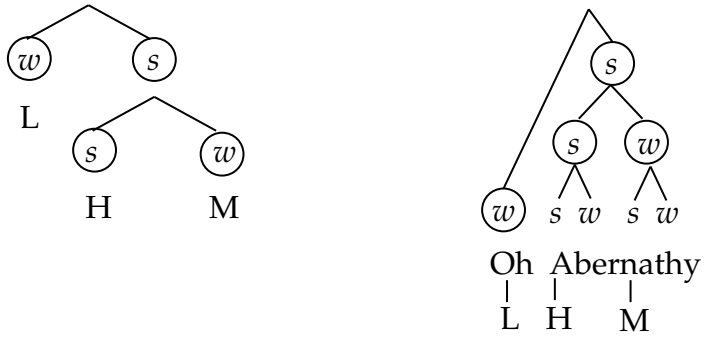


Figure 3.1: Text-tune association in Liberman (1975). Redrawn from Liberman (1975).

Based on the metrical structure in Liberman (1975) and Liberman & Prince (1977), Pierrehumbert (1980) also proposed a similar computation of tone-TBU mappings in intonation such that the texts are aligned with metrically defined positions, as shown in Figure 3.2. But Pierrehumbert (1980) used a more specific computation of tone-TBU map-

pings by defining the *s* and *w* metrical patterns with order pairs.

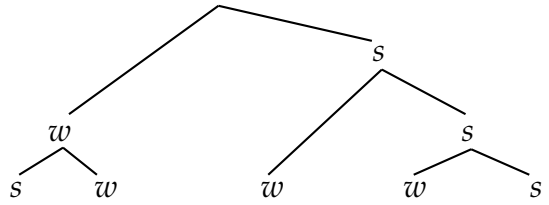


Figure 3.2: Text-tune association in Pierrehumbert (1980). Redrawn from Pierrehumbert (1980).

For the computation of tone-TBU mappings in Goldsmith (1976), association rules are based on the well-formedness condition (Goldsmith 1976, 1981). The well-formedness condition in (1) stipulates at least one-to-one mapping between tone and syllabic segment (1a), while banning line crossing (1b). This no-crossing constraint assumes the sequential ordering of the tone-TBU mappings.

(1) Well-formedness Condition (Goldsmith 1981):

a. Every syllabic segment should be associated with at least one tone.

b. Association lines should not be crossed.

Based on this well-formedness condition, the association rules in Goldsmith (1976) are computed with a total ordering. In Figure 3.3, when there are TBUs with five syllables and two tonal elements as shown in the left autosegmental representation, the tone-TBU association can be established according to the total ordering of the elements as shown in the right representation. The indices in the superscript indicate two independent tiers, TBUs for 1 and tones for 2, while those in the subscript indicate the ordering between the elements in each tier. Here, a_1^2 is associated with a_1^1, a_2^1, and a_3^1, sequentially. Then, a_2^2 is

associated with a_4^1 and a_5^1, in a sequential order, without any pairs crossing other pairs' association line.

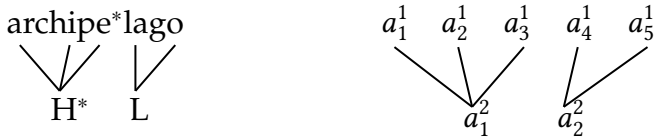


Figure 3.3: Text-tune association rules in Goldsmith (1976). Redrawn from Goldsmith (1976).

However, the partial or total orderings of the tone-TBU mappings in these studies (Liberman 1975, Goldsmith 1976, 1981, Pierrehumbert 1980) are not restrictive enough, as they focus only on the mapping between tones and TBUs. They do not explicitly explain the mechanism of the tone-TBU association with reference to the computational property of intonation. Also, their mechanism is difficult to compare with other phonological structures. However, by using model theory and logic, this dissertation explicitly defines the representation of intonational structures and their computations. This logical framework also enables us to compare other phonological structures with respect to computational complexity, which will be discussed in Chapter 5.

3.2 Model theory and logic

This dissertation defines intonation mathematically mainly using model theory and logic. From a linguistic perspective, model theory can deal with *any* sorts of linguistic representation, since it allows us to create *any structures* of both the underlying and the surface forms while explicitly specifying the mechanisms or derivations of their relations as well. Heinz (forthcoming) views this model-theoretic framework as an "ontology" by handling

the way we represent and compute the entities in the structures. Therefore, many studies have used this framework due to its versatility in representing structures, such as autosegmental structures (Jardine 2017, Chandlee & Jardine 2019a, Koser et al. 2018), syllable structures (Strother-Garcia 2019, Strother-Garcia & Heinz 2017), articulatory gestures (Nelson 2022a, 2023), phonological features (Nelson 2022b), and phonology-syntax interface (Czarnecki 2025).

Importantly, this model-theoretic representation is closely related to representing phonological primitives stored in the abstract mental structures. Heinz (forthcoming) writes:

> "From a psychological perspective, the primitive set of facts a model-theoretic representation encodes about a word can be thought of as *primitive psychological units*. In its strongest form, the model-theoretic representation of words as embodied in its signature makes a concrete claim about the psychological reality of the ways are represented mentally." (Heinz forthcoming; p.32)

This quote also sheds light on the relationship between model theory and speech perception. Model theory provides a very concise, yet fully explanatory vocabulary of the primitives, which we can figure out what is encoded in our mental representation through perceptual studies.

Therefore, this dissertation uses a model-theoretic approach to develop a theory of representing and computing intonation by providing the definitions of intonation, which enables us to make some connections with other phonological patterns. This dissertation especially deals with phonological primitives of intonation that were found to be crucial perceptual cues in encoding intonation through perceptual experiments.

The following subsection (3.2.1) introduces the basics and details about strings and models. After that, the subsection (3.2.2) shows how autosegmental structures can be represented in model theory, by first looking at those defined for the string of lexical tones, H́HL̀L̀, and then extending those to one of the declarative intonational patterns, H*H*L-L%. The subsection (3.2.3) introduces the syntax and semantics of first-order logic, which is mainly used throughout this dissertation.

3.2.1 Strings and models

Strings Let Σ be an alphabet, a finite set of symbols, and let Σ^* be a set of all strings over Σ. For example, the string CVCV can be specified with a set of symbols $\Sigma = \{C, V\}$. To indicate boundaries, we can use special boundary symbols, $\rtimes, \ltimes \notin \Sigma$. $\rtimes$ and $\ltimes$ indicate boundary symbols at the left and right edges, respectively, while C and V indicate consonants and vowels, respectively. Thus, we can get a string like $\rtimes$CVCV$\ltimes$.

Model Using model theory, a *model* can be created for precisely describing properties or relations of elements and functions between the elements in various structures (e.g., Enderton 2001, Libkin 2004). The focus of this dissertation is building a model for a linguistic structure, *a hierarchical intonational structure*. A simpler version of this model for intonation is briefly explained using a graphic representation in the next subsection (3.2.2). A full version of the intonational model is dealt in Chapter 5.

But first, let's take a look at a model with a simple string, $\mathcal{M}_{\rtimes\text{CVCV}\ltimes}$. The model in (2) specifies the domain, properties, and relations for $\rtimes$CVCV$\ltimes$, and it is visualized with a graphic representation in Figure 3.4. Throughout this subsection (3.2.1), I explain the def-

initions of properties and relations of the model by talking about this ⋊CVCV⋉ example.

$$\mathcal{M}_{\rtimes CVCV\ltimes} = \left\langle \begin{array}{l} \mathcal{D} = \{0, 1, 2, 3, 4, 5\}; \\ P_C = \{1, 3\}, \quad P_V = \{2, 4\}, \\ P_{\rtimes} = \{0\}, \quad P_{\ltimes} = \{5\}, \\ p = \{(1, 0), (2, 1), (3, 2), (4, 3), (5, 4)\}, \\ s = \{(0, 1), (1, 2), (2, 3), (3, 4), (4, 5)\} \end{array} \right\rangle \tag{2}$$

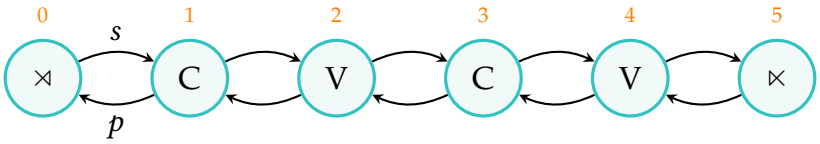


Figure 3.4: An illustration of a model $\mathcal{M}_{\rtimes CVCV\ltimes}$ constructed for the string ⋊CVCV⋉.

Domain Domain $\mathcal{D}$ is a set of indices {0, 1, 2, ⋯, n}, where each index indicates the position of a node that is labeled with a symbol in a string. For example, the domain $\mathcal{D}$ in (2) is {0, 1, 2, 3, 4, 5}, which are marked with indices above the nodes (circles) in Figure 3.4. With these domain indices, we can find where the nodes labeled with certain symbols are located.

Within the domain $\mathcal{D}$, sets of *relations* $\mathcal{R}$ and *functions* f in the model can be defined as relations and functions on $\mathcal{D}$ in (3):

$$\langle \mathcal{D}; \mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_k, f_1, \ldots, f_l \rangle. \tag{3}$$

Relations As provided in (3), $\mathcal{R}$ is a finite set of k relations over $\mathcal{D}$. Two types of relations are used in this dissertation: *unary* relations are defined with $\mathcal{R}_i \subseteq \mathcal{D}$ and *binary*

relations are defined with $\mathcal{R}_i \subseteq \mathcal{D} \times \mathcal{D}$. Unary relations simply refer to an individual domain element in $\mathcal{D}$, while binary relations refer to a pair of domain elements in $\mathcal{D} \times \mathcal{D}$. Here, a predicate P_σ is specified with unary relations, whose symbols $\sigma \in \Sigma$ are defined within the signature $\mathcal{S}$.

In (2), for example, P_C and P_V hold true for the sets of positions, $\{1, 3\}$ and $\{2, 4\}$ over $\mathcal{D}$, respectively. $P_{\ltimes}$ and $P_{\rtimes}$ are true for the positions, $\{0\}$ and $\{5\}$, respectively. In Figure 3.4, the labels in each node indicate these unary predicates.

Binary predicates hold true for pairs of positions (x, y) in $\mathcal{D} \times \mathcal{D}$. This binary predicate is used for defining the association relation between tones and TBUs in the next section (3.2.2).

Functions f is a finite set of l functions over $\mathcal{D}$. Functions return an output element in the domain as a result of an input element in the domain. Two functions are mainly used in this study: the predecessor p and the successor s functions. They return the input element's preceding and following elements in the domain, respectively.

The p function, indicated with the right arrows in Figure 3.4, is an immediate predecessor function with two variables, $p(x) \approx y$, such that y precedes x. Thus, the element in the first position in the model immediately precedes the element in the second position ($p(2) \approx 1$), which immediately precedes the third element ($p(3) \approx 2$), which in turn immediately precedes the fourth element ($p(4) \approx 3$)), which in turn immediately precedes the final element ($p(5) \approx 4$). The function s works in the opposite direction, as indicated by the left arrows in Figure 3.4.

Signature A *signature* $\mathcal{S}$ is a set of all *all symbols* used to formalize the relations and functions in the model, $\{P_{\sigma \in \Sigma}, p, s\}$. A unary predicate $P_\sigma \in \mathcal{R}$ with a variable x holds true for a position x in $\mathcal{D}$ if and only if x is labeled with the symbol $\sigma \in \Sigma$. The signature of (2) is provided in (4), where C and V stand for consonants and vowels, respectively, while $\rtimes$ and $\ltimes$ refer to left and right boundaries, respectively.

(4) $\{P_C, P_V, P_{\rtimes}, P_{\ltimes}, p, s\}$

Importantly, the signature is made up of the *primitives* in the representation. That is, what we store in the signature for intonation is considered the primitives of intonation. Therefore, the *discrete* and *continuous* properties of intonation are reflected in the signatures defined in Chapter 5 and Chapter 6, respectively.

3.2.2 Representing autosegmental structures in model theory

Within the model-theoretic framework, as shown in the previous subsection (3.2), a string can be represented as a model, where each node represents each symbol in a string, and the relations between the nodes are defined with the predecessor p and successor s functions.

Using a model, it is also possible to build more complex linguistic structures such as autosegmental structures (Jardine 2016) and syllable structures (Strother-Garcia 2019). Crucially, Jardine (2016) presented a graphic representation of the autosegmental structures. For instance, a toned string in (5a) can be autosegmentally represented as in (5b).

(5) a. félàmà 'junction'

b. fe la ma

H L

Within model theory, the autosegmental structure in (5b) can be represented with a *graph* composed of nodes and their relations, based on the model $\mathcal{M}_{\acute{H}\grave{L}\grave{L}}$ in (6). In Figure 3.5, each node consists of the symbols defined in the signature $\mathcal{S} = \{\sigma, H, L, p, s, \mathcal{A}\}$, where σ here is a syllable, H and L are tones, p and s are predecessor and sucessor functions, and $\mathcal{A}$ is an association relation. We can think of the sequence of σ nodes representing a TBU tier, while the sequence of H and L nodes represents a tonal tier.

In (6), a set of indices for the positions of the symbol in $\mathcal{S}$ are specified in the domain $\mathcal{D} = \{0, 1, 2, 3, 4\}$. P_σ holds true for $\{0, 1, 2\}$, while P_H and P_L hold true for $\{3\}$ and $\{4\}$, respectively.

$$\text{(6)} \qquad \mathcal{M}_{\acute{H}\grave{L}\grave{L}} = \left\langle \begin{array}{l} \mathcal{D} = \{0, 1, 2, 3, 4\}; \\ P_\sigma = \{0, 1, 2\}, \quad P_{\text{H}} = \{3\}, \quad P_{\text{L}} = \{4\}, \\ p = \{(1, 0), (2, 1), (4, 3)\}, \\ s = \{(0, 1), (1, 2), (3, 4)\}, \\ \mathcal{A} = \{(0, 3), (1, 4), (2, 4)\} \end{array} \right\rangle$$

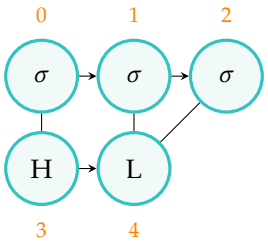


Figure 3.5: A graphic representation of the model $\mathcal{M}_{\acute{H}\grave{L}\grave{L}}$ that describes the autosegmental structure in Mende. Adapted from Jardine (2016).

The order of the nodes in each tier is specified with the predecessor p and successor s functions. That is, as for the TBU tier, the σ node at the 0th position precedes that at the 1st position, which in turn precedes that at the 2nd position in the domain. In the tonal tier, the H tone node at the 3rd position precedes the L tone node at the 4th position. The association between the TBU and tonal tiers is specified with the association relation $\mathcal{A}$. For the case of Mende, the first syllable at the 0th position is associated with the H tone at the 3rd position, while the rest of the following syllables (1st and 2nd positions) are associated with the L tone at the 4th position.

Following Jardine (2016), this dissertation represents intonation based on the autosegmental structure within the model-theoretic framework. The autosegmental structure of intonation in Figure 3.6 can be represented with a *graph*, based on the ingredients defined in the model $\mathcal{M}_{H^*H^*L-L\%}$ in (7).

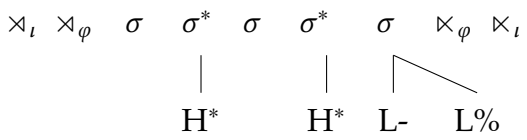


Figure 3.6: The autosegmental structure of American English intonation.

For intonation, a signature $\mathcal{S}$ for (7) consists of $\{\ltimes_\iota, \rtimes_\iota, \ltimes_\varphi, \rtimes_\varphi, \sigma, \sigma^*$, H*, L-, L%, $\mathcal{A}\}$. $\ltimes_\iota$ and $\rtimes_\iota$ symbols stand for the left and right intonational phrase (IP) boundaries, respectively, while $\ltimes_\varphi$ and $\rtimes_\varphi$ indicate the left and right intermediate phrase (ip) boundaries, respectively. σ and σ^* stand for unstarred and starred (prominent) syllables, respectively. H* indicates an H pitch accent, while L- and L% denote L phrasal and boundary tones, respectively. $\mathcal{A}$ indicates an association relation.

In (7), a set of indices for the positions of the symbol is specified in the domain $\mathcal{D}$ =

$\{0, 1, 2, 3, \cdots, 10, 11, 12\}$. Unary relations for boundaries and TBUs are defined: $P_{\rtimes_\iota}$ and $P_{\ltimes_\iota}$ hold true for $\{0\}$ and $\{8\}$; $P_{\rtimes_\varphi}$ and $P_{\ltimes_\varphi}$ for $\{1\}$ and $\{7\}$; P_σ and P_{σ^*} for $\{2, 4, 6\}$ and $\{3, 5\}$. Unary relations, $P_{\text{H*}}$, $P_{\text{L-}}$, $P_{\text{L\%}}$, are defined for the sets of positions $\{9, 10\}$, $\{11\}$, and $\{12\}$, respectively. The symbols in both the string tier (including TBUs and boundaries) and the tonal tier are visualized as a graph in Figure 3.7.

(7)
$$\left\langle \begin{array}{l} \mathcal{D} = \{0, 1, 2, 3, \cdots, 10, 11, 12\}; \\ P_{\rtimes_\iota} = \{0\}, \quad P_{\ltimes_\iota} = \{8\}, \quad \mathrm{P}_{\rtimes_\varphi} = \{1\}, \quad P_{\ltimes_\varphi} = \{7\}, \quad \mathrm{P}_\sigma = \{2, 4, 6\}, \quad P_{\sigma^*} = \{3, 5\}, \\ P_{\text{H*}} = \{9, 10\}, \quad P_{\text{L-}} = \{11\}, \quad P_{\text{L\%}} = \{12\}, \\ p = \{(1, 0), (2, 1), (3, 2), (4, 3), (5, 4), (6, 5), (7, 6), (8, 7), (10, 9), (11, 10), (12, 11)\}, \\ s = \{(0, 1), (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 7), (7, 8), (9, 10), (10, 11), (11, 12)\}, \\ \mathcal{A} = \{(3, 9), (5, 10), (6, 11), (6, 12)\} \end{array} \right\rangle$$

The relations between the nodes within the same tier are sequentially ordered as specified by the p and s functions, while the relations of the nodes between the tiers are *associated* with the association relation $\mathcal{A}$. Within the TBU tier, the $\rtimes_\iota$ node at the 0th position precedes the $\rtimes_\varphi$ node at the 1st position, which in turn precedes the σ node at the 2nd position. The same goes for the predecessor relation of other nodes from the 3rd to 8th positions in this tier. As for the tonal tier, the H* node at the 9th position precedes the H* node at the 10th position, which in turn precedes the L- node at the 11th position, which finally precedes the L% node at the 12th position.

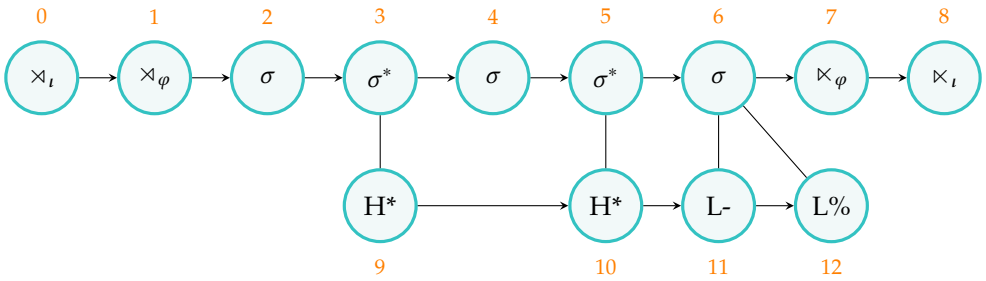


Figure 3.7: A graph representing the autosegmental structure of American English intonation.

The association relation $\mathcal{A}$ associates the nodes between tones and particular syllables. Only starred syllables are associated with the starred tones sequentially, such that the σ^* nodes at the 3rd and 5th position are associated with the H* nodes at the 9th and 10th position, respectively. The final syllable is associated simultaneously with both the phrasal and boundary tones, such that the σ at the 6th position is associated with both the L- and L% nodes at the 11th and 12 position, respectively.

This graph representation of autosegmental structures is further refined by specifying the input and the larger output structures via *logical transductions* (e.g., Courcelle 1994, Engelfriet & Hoogeboom 2001, Filiot & Reynier 2016), which are introduced in Section 5.1 of Chapter 5. By inputting a string and outputting a larger autosegmental structure, logical transductions clearly and explicitly allow for explaining the computational mapping of intonation.

3.2.3 First-order logic

First-order (FO) logic is a mathematical language by which we can describe the properties and relations in the model. It is constrained by *syntax* to build grammatical expressions,

and then it is interpreted in terms of the concrete model by *semantics*. Throughout this dissertation, FO logic is mainly used to formalize the relations for computing the properties of intonation. FO logic is more restrictive than *monadic second-order* (MSO) logic, since FO logic only defines the properties of *individual* elements, while MSO logic is used to characterize the *set* of elements (Enderton 2001, McNaughton & Papert 1971). Except for the definitions of intervals to select the prosodic domain for the realization of intonation in Chapter 6, which are expressed with MSO logic, FO logic is mostly used in Chapter 5 and Chapter 6.

3.2.3.1 Syntax of FO logic

FO variables are variables x, y, z, $\cdots$ (notated with lowercase letters). We use a *term* to refer to the element over the domain $\mathcal{D}$. The term t includes variables x, y, z and a function of terms $f(t)$, where $f \leftarrow \mathcal{S}$. The building blocks of FO logic are *atomic predicates* or *atomic formula* defined as $\mathcal{R}(t_1, \cdots, t_n)$, where the terms specify positions of elements in the domain. For example, $P_C(x)$ and $P_V(x)$ are unary predicates with one variable, while $\mathcal{A}(x, y)$ is a binary predicate with two variables.

These building blocks are glued together to form a *formula* by using logical connectives, $\neg$ (not), $\vee$ (or), $\wedge$ (and), $\rightarrow$ (if...then), $\approx$ (equal). Formulas are considered to be *well-formed* if they meet the following rules:

(8) Well-formed formulas (WFFs):

a. Atomic formula itself is a WFF.

b. If ϕ is a WFF, then $\neg\phi$ is a WFF.

c. If ϕ and ψ are WFFs, then $\phi \vee \psi$ is a WFF.

d. If ϕ and ψ are WFFs, then $\phi \wedge \psi$ is a WFF.

e. If ϕ and ψ are WFFs, then $\phi \rightarrow \psi$ is a WFF.

f. If ϕ and ψ are WFFs, then $\phi \approx \psi$ is a WFF.

g. If x is a variable and ϕ is a WFF where x is free, then $\exists x[\phi]$ is a WFF.

h. If x is a variable and ϕ is a WFF where x is free, then $\forall x[\phi]$ is a WFF.

The formulas from (8a) to (8e) have *free variables*, which can only be interpreted in terms of the model $\mathcal{M}$. In contrast, the variables in the formulas (8g) and (8h) are *bound* by quantifiers.

An important notion in FO logic is that if formulas do not contain any quantifiers, it is called *quantifier-free* (QF). This QF characteristic is considered to be very restrictive in expressing properties and relations. The following chapter (Chapter 5) claims that intonation can be characterized as a QF logical interpretation of metrical and prosodic structure. This local property of logic enables us to begin examining intonational structures by imposing very restrictive computational power on the representation and computation of intonation. The ingredients for the syntax of FO logic are provided in Table 3.1.

Table 3.1: Ingredients for the syntax of FO logic.

Ingredients	**Symbols**
Variables	$x, y, z, \cdots$
Predicates	$P(x), \mathcal{A}(x, y), \cdots$
Connectives	$\neg, \wedge, \vee, \rightarrow, \approx$
Quantifiers	$\exists, \forall$

Lastly, the predicates can be defined using WFFs, with the symbol $\stackrel{\text{def}}{=}$ denoting *logical*

equivalence. For example, the predicate $\mathcal{A}(x, y)$ is said to be logically equivalent to WFFs, $(H^*(x) \wedge \sigma^*(y)) \wedge (L(x) \wedge \sigma(y) \wedge \ltimes_\varphi(p(y)))$, with the following formula: $\mathcal{A}(x, y) \stackrel{\text{def}}{=} (H^*(x) \wedge \sigma^*(y)) \wedge (L(x) \wedge \sigma(y) \wedge \ltimes_\varphi(p(y)))$.

3.2.3.2 Semantics of FO logic

WFFs are *interpreted* via a concrete model $\mathcal{M}$ defined from a signature $\mathcal{S}$. We can get the truth value of WFFs ϕ, whether ϕ is *true* or *false*, based on $\mathcal{M}$.

Now the variables $t_1, \cdots, t_n$ can be interpreted as *individual* elements over the domain $\mathcal{D}$. Let a subset of n domain element of $\mathcal{M}$ be $(i_1, \cdots, i_n)$. For example, in (10), $(i_1, \cdots, i_n)$ can be interpreted as $(2, 4, 6)$ for P_σ. We match each element of $(i_1, \cdots, i_n)$ in $\mathcal{M}$ with each element in a WFF with n variables $\phi(t_1, \cdots, t_n)$. If there exist some matched pairs of $(i_n, \phi(t_n))$ when $i_n \in \mathcal{R}_\sigma$ and $\sigma \in \Sigma$, we say ϕ is satisfied with $\mathcal{M}$. This satisfaction can be notated as $i_n \models \sigma(t_n)$. WFFs in (8) can be translated into the actual elements within $\mathcal{M}$ using the semantic rules in (9).

(9) Interpretation of WFFs:

a. $i \models \sigma(x)$ if and only if $i \in \mathcal{R}_\sigma$ and $\sigma \in \Sigma$.

b. $(i_1, \cdots, i_n) \models \neg\phi(t_1, \cdots, t_n)$ iff $(i_1, \cdots, i_n) \not\models \phi(t_1, \cdots, t_n)$.

c. $(i_1, \cdots, i_n, j_1, \cdots, j_n) \models \phi(t_1, \cdots, t_n) \wedge \psi(y_1, \cdots, y_n)$ iff $((i_1, \cdots, i_n) \models \phi(t_1, \cdots, t_n)) \wedge ((j_1, \cdots, j_n) \models \psi(t'_1, \cdots, t'_n))$.

d. $(i_1, \cdots, i_n, j_1, \cdots, j_n) \models \phi(t_1, \cdots, t_n) \vee \psi(t'_1, \cdots, t'_n)$ iff $((i_1, \cdots, i_n) \models \phi(t_1, \cdots, t_n)) \vee ((j_1, \cdots, j_n) \models \psi(t_1, \cdots, t_n))$.

e. $(i_1, \cdots, i_n, j_1, \cdots, j_n) \models \phi(t_1, \cdots, t_n) \approx \psi(t'_1, \cdots, t'_n)$ iff $((i_1, \cdots, i_n) \models \phi(t_1, \cdots, t_n)) \approx$

$((j_1, \cdots, j_n) \models \psi(t_1, \cdots, t_n))$.

f. $(i_1, \cdots, i_n) \models \forall x[\phi(x, t_1, \cdots, t_n)]$ iff for all i, $(i_1, \cdots, i_n) \models \phi(x, t_1, \cdots, t_n)$.

g. $(i_1, \cdots, i_n) \models \exists x[\phi(x, t_1, \cdots, t_n)]$ iff for some i, $(i_1, \cdots, i_n) \models \phi(x, t_1, \cdots, t_n)$.

Therefore, using the rules in (9), we can evaluate the truth value of WFFs with reference to the model $\mathcal{M}$. For instance, a WFF $P_\sigma(x)$ is true for certain positions $\{2, 4, 5\}$, otherwise false within the model (10). $P_\sigma(x)$ holds true for the positions $\{3, 5\}$, otherwise false in the model.

$$
(10) \quad \left\langle \begin{array}{l}
\mathcal{D} = \{0, 1, 2, 3, \cdots, 10, 11, 12\}; \\
P_{\rtimes_\iota} = \{0\}, \quad P_{\ltimes_\iota} = \{8\}, \quad \mathrm{P}_{\rtimes_\varphi} = \{1\}, \quad P_{\ltimes_\varphi} = \{7\}, \quad \mathrm{P}_\sigma = \{2, 4, 6\}, \quad P_{\sigma^*} = \{3, 5\}, \\
P_{\mathrm{H}^*} = \{9, 10\}, \quad P_{\mathrm{L}\text{-}} = \{11\}, \quad P_{\mathrm{L}\%} = \{12\}, \\
p = \{(1, 0), (2, 1), (3, 2), (4, 3), (5, 4), (6, 5), (7, 6), (8, 7), (10, 9), (11, 10), (12, 11)\}, \\
s = \{(0, 1), (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 7), (7, 8), (9, 10), (10, 11), (11, 12)\}, \\
\mathcal{A} = (3, 9), (5, 10), (6, 11), (6, 12)
\end{array} \right\rangle
$$

When a WFF is the following formula in (3.1), it is true for the set of pairs $\{(3, 9),(5, 10),(6, 11),(6, 12)\}$ in the model (10), otherwise false.

$$
\mathcal{A}(x, y) \stackrel{\text{def}}{=} (H^*(x) \wedge \sigma^*(y)) \wedge (L(x) \wedge \sigma(y) \wedge \ltimes_\varphi(p(y))) \tag{3.1}
$$

"Associate positions x and y iff x is labeled H* and y is labeled σ^*,

and x is labeled L and y's preceding element is labeled σ."

For such positions in the model, the WFF then can be said to be satisfied.

3.2.4 Monadic second-order logic

Monadic second-order (MSO) logic is another mathematical language by which we can describe the properties and relations in the model. Since MSO logic deals not only with individual elements but also with *sets* of the individual elements, it is less restrictive than FO logic. In this dissertation, MSO logic is used to define an interval for the pitch accent realization in Chapter 6.

3.2.4.1 Syntax of MSO logic

MSO variables are variables X, Y, Z, $\cdots$ (notated with upper case letters), in addition to variables x, y, z. Therefore, everything except for adding the sets X, Y, Z is the same for the syntax of MSO logic.

Atomic formulas are recursively combined together using logical connectives, $\neg$, $\vee$, $\wedge$, $\rightarrow$, just like those in FO logic. On top of the WFF rules in (8) with set variables X, Y, Z, we can add the following two rules for MSO logic:

(11) WFFs for MSO logic:

a. If ϕ is a WFF and X is a variable, then $\exists X[\phi]$ is a WFF.

b. If ϕ is a WFF and X is a variable, then $\forall X[\phi]$ is a WFF.

The set variables in the formulas (11a) and (11b) are *bound* by quantifiers. The ingredients for the syntax of MSO logic are provided in Table 3.2.

Table 3.2: Ingredients for the syntax of MSO logic

Ingredients	**Symbols**
Variables	$x, y, z, X, Y, Z, \cdots$
Variables in a Set	$X(x), X(y), X(z), Y(x), Y(y), \cdots$
Predicates	$P(x), \mathcal{A}(x, y), P(X), R(X, Y), \cdots$
Connectives	$\neg, \wedge, \vee, \rightarrow$
Quantifiers	$\exists, \forall$

For instance, in Chapter 6, an interval is defined with the set variables as follows:

$$interval(X) = \forall x, y, z[X(x) \wedge X(z) \wedge x \prec y \wedge y \prec z \rightarrow X(y)]$$

where for all x, y, and z, if x and z belong to a set X and y exists between x and z, y should belong to the set X. In this way, contiguous elements can be expressed by using X, Y, Z.

3.2.4.2 Semantics of MSO logic

WFFs are *interpreted* via a concrete model $\mathcal{M}$ defined from a signature $\mathcal{S}$. We can get the truth value of WFFs ϕ, whether ϕ is *true* or *false*, based on $\mathcal{M}$.

In addition to FO variables, the MSO variables $X_1, \cdots, X_n$ can be defined over *sets* of individual elements over the domain $\mathcal{D}$. We add a subset of n sets of domain elements of $\mathcal{M}$ as $(S_1, \cdots, S_n)$. We match each set of $(S_1, \cdots, S_n)$ in $\mathcal{M}$ with each set in the WFFs with n variables $\phi(X_1, \cdots, X_n)$. If there exist some matched pairs of $(S_n, \phi(X_n))$ when $S_n \in \mathcal{R}_\sigma$ and $\sigma \in \Sigma$, we say ϕ is satisfied with $\mathcal{M}$. This satisfaction can be notated as $S_n \models \sigma(X_n)$. WFFs in (11) can be interpreted via $\mathcal{M}$ using the semantic rules in (12). The following rules are just plugging the set variables X, Y, Z into (9).

(12) Interpretation of WFFs:

a. $S \models \sigma(X)$ if and only if $S \in \mathcal{R}_\sigma$ and $\sigma \in \Sigma$.

b. $(S_1, \cdots, S_n) \models \neg\phi(X_1, \cdots, X_n)$ iff $(S_1, \cdots, S_n) \not\models \phi(X_1, \cdots, X_n)$.

c. $(S_1, \cdots, S_n, T_1, \cdots, T_n) \models \phi(X_1, \cdots, X_n) \wedge \psi(Y_1, \cdots, Y_n)$ iff $((S_1, \cdots, S_n) \models \phi(X_1, \cdots, X_n)) \wedge$ $((T_1, \cdots, T_n) \models \psi(Y_1, \cdots, Y_n))$.

d. $(S_1, \cdots, S_n, T_1, \cdots, T_n) \models \phi(X_1, \cdots, X_n) \vee \psi(Y_1, \cdots Y_n)$ iff $((S_1, \cdots, S_n) \models \phi(X_1, \cdots, X_n)) \vee$ $((T_1, \cdots, T_n) \models \psi(X_1, \cdots, X_n))$.

e. $(S_1, \cdots, S_n) \models \forall X[\phi(X_1, \cdots, X_n)]$ iff for all S, $(S_1, \cdots, S_n) \models \phi(X_1, \cdots, X_n)$.

f. $(S_1, \cdots, S_n) \models \exists X[\phi(X_1, \cdots, X_n)]$ iff for some S, $(S_1, \cdots, S_n) \models \phi(X_1, \cdots, X_n)$.

For MSO logic, we can obtain the truth value of WFFs using both individual variables x, y, z and set variables X, Y, Z.

3.3 Summary and Conclusion

This chapter reviewed the mathematical framework used in formalizing the representation and computation of intonation throughout this dissertation. First, it reviewed previous studies on formalizing the representation and computation of intonation. Then, the basics of model theory and logic have been introduced so that we can relate them to the primitives of intonation in the following chapters. Lastly, the mathematical languages for analyzing the discrete and continuous information of f0 were restricted to FO and MSO logic. For the discrete tonal targets, FO logic will be used to describe the individual elements in the domain. In contrast, for the continuous f0 information, MSO logic will be used to express the sets of elements in the domain. These logical languages will enable

us to constrain how models are represented and computed in terms of their expressive power.

Chapter 4

Perceptual primitives of intonation

This chapter investigates what kind of f0 information is crucial in distinguishing phonological contrasts in intonation. In the AM theory, phonological primitives of intonation are considered discrete tonal targets (Hs and Ls) (e.g., Beckman & Pierrehumbert 1986; Ladd 2008; Pierrehumbert 1980, Arvaniti 2022). As opposed to the AM theory, configurational approaches emphasize the importance of the f0 contour as a whole (e.g., Bolinger 1951, Hart et al. 2003, Hirst & Di Cristo 1998, Xu 2005, Xu & Wang 2001). Importantly, this chapter assumes that *both* discrete and continuous f0 information of intonation should be considered to fully capture the phonological nature of intonation. Hence, this chapter adds more evidence that continuous f0 information such as f0 contour shape plays a crucial role in the phonological contrast of intonation.

Building on recent findings in the post-lexical pitch accent languages, American English, where not only discrete tonal targets (e.g., Beckman & Pierrehumbert 1986, Pierrehumbert 1981) but also continuous f0 information (e.g., Barnes et al. 2010, 2012, 2021) play an essential role in distinguishing sentential meaning, this chapter provides evidence

that such f0 shape information in intonation is also crucial for lexical distinctions. Since f0 information at both lexical and post-lexical levels is realized in intonation, it is possible to look at the f0 contour with respect to lexical contrast. Also, one of the experimental methods to identify the intonational primitives of a language is to examine what kind of phonological information is encoded in our mental representation.

Hence, a perception study was conducted for a lexical pitch accent system to figure out whether both tonal target and f0 shape information work as intonational primitives. Thus, this chapter focuses on South Kyungsang Korean for an empirical test case of a lexical pitch accent language (e.g., Chang 2007, Do & Kenstowicz 2010, Kim & Jun 2009, Lee & Zhang 2014).

Much of the material in this chapter was first published in Joo & D'Imperio (2025)[3]. However, it has been significantly rewritten for this dissertation, including a different experimental design, new datasets, and updated results.

4.1 Lexical pitch accents in South Kyungsang Korean

South Kyungsang Korean is a dialect spoken in the southern region of Korea. Different from an intonational language, Seoul Korean (Jun 2006a), South Kyungsang Korean is considered to be categorized as a lexical pitch accent system (e.g., Chang 2007, Do & Kenstowicz 2010, Kim & Jun 2009, Lee & Zhang 2014, Cho et al. 2019), similar to Tokyo Japanese (Beckman & Pierrehumbert 1986, Venditti 2005). In this system, f0 is used to

[3]Joo, Hyunjung & D'Imperio, Mariapaola, The Perception of Lexical Pitch Accent in South Kyungsang Korean: The Relevance of Accent Shape, Language and Speech (OnlineFirst) pp. 1-33. Copyright © 2025 by Sage Publications. Reprinted by Permission of Sage Publications.

distinguish not only lexical contrast from pitch accents but also grammatical contrast determined by phrase-level prosody.

For instance, as shown in Figure 4.1 below, f0 patterns differ depending on the lexical meaning: [pal] with H tone means *foot*, while [pal] with LH tone means *shade*. More specifically, f0 reaches its H tone target much earlier for the H-toned word than for the LH-toned word. Also, note that the f0 contour shape in the H-toned word appears to be convex, while that in the LH-toned word is more likely to have a concave shape. On the other hand, their preceding and following f0 contexts exhibits similar tonal patterns[4], constrained by both lexical and phrasal prosody.

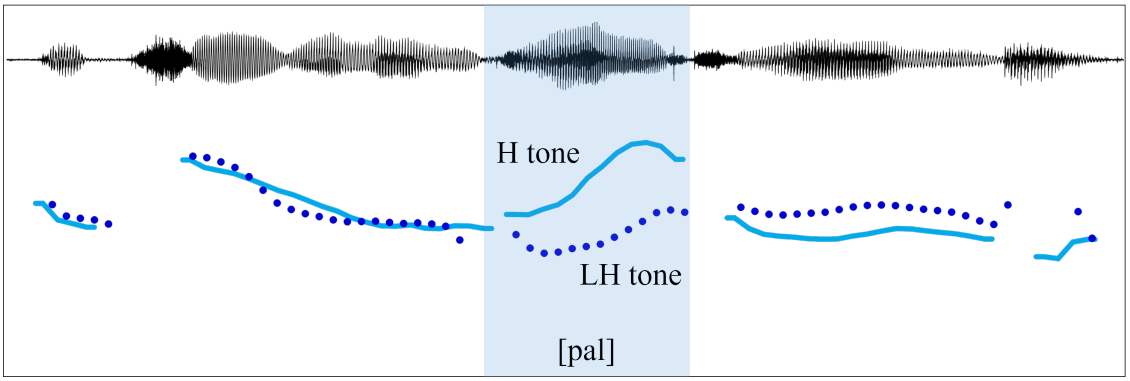


Figure 4.1: f0 contours of High (H) versus Rising (LH) contrast for [pal] at a phrase-medial position.

Importantly, this chapter concerns f0 patterns of monosyllabic H- and LH-tone minimal pairs embedded in a phrase-medial position. Studies in South Kyungsang Korean (e.g., Chang 2007, Do & Kenstowicz 2010, Kim & Jun 2009, Lee & Zhang 2014) basically characterized monosyllabic words in South Kyungsang Korean with two contrastive

[4]The following f0 context may appear to differ, with the f0 context after the LH item showing slightly higher f0 values than that after the H item. But meaningwise, there is neither lexical or sentential contrast for this f0 difference. However, it should be further investigated how lexical prosody affects its preceding or following context. Another possibility could be due to the sparse tonal specification of lexical pitch accent, such that the pitch accents from the lexical item can possibly spread to the following f0 contour.

tonal categories, H and LH. Specifically, H tone was described as having a quite higher initial f0, followed by a high f0 peak, with a shorter duration. LH tone (or a low rising contour) was described as having a slightly lower initial f0 and a later f0 peak alignment, with a longer duration.

Building on these observations, a perceptual study by Chang (2013) examined the perceptual primitives in distinguishing the lexical contrast in South Kyungsang Korean: timing of peak alignment, onset f0 values, and syllable duration. This study found that H tone perception was associated with earlier peaks, higher f0 values on the word onset, and longer syllable duration, while LH tone identification was related to later peaks, lower onset f0 values, and shorter syllable duration. However, within a broader view of intonational phonology, their investigations were largely confined to viewing discrete tonal targets as primitives, as proposed by the AM model. Furthermore, Chang (2013) did not account for different lexical pitch accents in relation to intonation, as their test stimuli were isolated tokens without any preceding or following phrasal contexts.

Cruically, according to Barnes and their colleagues (e.g., Barnes et al. 2010, 2012, 2021), f0 shape information plays a crucial role in distinguishing sentential meaning in American English. They found that more convex shapes were categorized as L+H* (meaning *incredulity*), whereas more concave shapes were identified as L*+H (meaning *uncertainty*). These findings shed light on the possible perceptual relevance of f0 shape information for lexical distinction, which can be extended to South Kyungsang Korean. Therefore, this chapter addresses the theoretical question of the nature of intonational primitives by testing two opposing frameworks: discrete tonal targets within the AM theory or continuous f0 information within configurational approach. For this, a perceptual study is conducted

to figure out what kind of f0 information is stored in our phonological representation.

4.2 Hypotheses and predictions

This chapter aims to explore the phonological primitives of intonation encoded in our phonological representation by looking at the perception of lexical pitch accents, H versus LH, in South Kyungsang Korean. Two factors, the timing of f0 peak alignment and f0 contour shape, are used to determine H versus LH pitch accents. The timing of f0 peak alignment is used to test the AM's discrete target model, while the f0 shape is used to examine the holistic configurational approach. Therefore, two independent hypotheses can be made as the following:

- **Hypothesis 1**: Different timing of discrete tonal targets influences pitch accent distinction.
  - Prediction: Earlier targets lead to more H responses, whereas later targets have more LH responses.
- **Hypothesis 2**: Different f0 shapes influence pitch accent distinction.
  - Prediction: More convex shapes lead to more H tone responses, while more concave shapes as LH tone responses.

Hypothesis 1 states that the timing of the discrete tonal target influences pitch accent distinction. Since H tone has one tonal target—H tone, while LH tone has two tonal targets—LH tone, the f0 peak is aligned much later for LH tone than H due to the preced-

ing L target. Therefore, the prediction would be earlier targets inducing more H responses and later targets leading to more LH responses.

Hypothesis 2 concerns the impact of f0 shape on the pitch accent categorization. As observed in Figure 4.1, if the f0 shape information is closely related to the tonal contrast, convex shapes are more likely to be categorized as H tones, while concave shapes as LH tones.

From these two hypotheses, we can test what kind of f0 information is available for the lexical pitch accent distinction in South Kyungsang Korean.

4.3 Methods

In order to test the two hypotheses above, a perception experiment was conducted to investigate whether f0 peak alignment (discrete tonal targets) or f0 rise shape (f0 configurations) is an intonational primitive of South Kyungsang Korean. The experimental task was a two-alternative forced choice (2AFC), in which upon hearing the manipulated sounds, South Kyungsang Korean listeners were asked to choose one of the two pictures that matched the sounds. The auditory stimuli were monosyllabic pitch accent minimal pairs embedded in a carrier sentence, manipulated depending on two factors: the timing of f0 peak alignment (earlier vs. later peak) and the f0 rise shape (scooped vs. domed). The results were then analyzed with mixed-effects logistic regression models to see which factors, the f0 peak alignment or the rise shape, or both, are statistically significant to be considered as perceptual primitives of lexical pitch accent in South Kyungsang Korean.

4.3.1 Reference speech material

The reference speech material consists of three target word pairs and twelve distractor words. The target word pairs were contrastive for pitch accents: /kan/ (H tone: *taste*, LH tone: *liver*), /pam/ (H tone: *night*, LH tone: *chestnut*), and /pal/ (H tone: *foot*, LH tone: *shade*). The distractor words were contrastive for segments, differing in either the onset (e.g., /tan/ vs. /kan/) or the coda (e.g., /kaŋ/ vs. /kan/). Both the target and the distractor words are listed in Table 4.1.

Table 4.1: Stimuli consisting of target and distractor words.

	Word	**Meaning**	**Contrastive feature**
Target words	/kan/	*taste*	**Tone** (H vs. LH)
	/kan/	*liver*	**Tone** (LH vs. H)
	/pam/	*night*	**Tone** (H vs. LH)
	/pam/	*chestnut*	**Tone** (LH vs. H)
	/pal/	*foot*	**Tone** (H vs. LH)
	/pal/	*shade*	**Tone** (LH vs. H)
Distractor words	/tan/, /nan/	*podium, orchid*	**Onset** (/t, n/ vs. /k/)
	/kam/, /kaŋ/	*persimmon, river*	**Coda** (/m, ŋ/ vs. /n/)
	/nam/, /tam/	*south, fence*	**Onset** (/n, t/ vs. /p/)
	/pan/, /paŋ/	*class, room*	**Coda** (/n, ŋ/ vs. /m/)
	/kʰal/, /tal/	*knife, moon*	**Onset** (/kʰ, t/ vs. /p/)
	/pan/, /paŋ/	*class, room*	**Coda** (/n, ŋ/ vs. /l/)

Both the target and the distractor words were embedded in a phrase-medial position of an Accentual Phrase (AP): [jobʌnænɯn maʊsɯɾo k'amansæk _____ k'ɯlɾikhæa] (‘*Click the black _____ with the mouse this time.*’) In order to induce a clear rise for both the H-toned and LH-toned target words, the preceding word before the target words was a LH-toned word, [k'amansæk] (‘*black*’). Also, to avoid any tonal alternation of the monosyllabic target

words, no affix was attached since the underlying tonal pattern of monosyllabic words in South Kyungsang Korean surfaces differently with some affixes. Importantly, the target words were assigned a contrastive focus, to make sure that target were is clearly accented since the deaccentuation occurs in South Kyungsang Korean (Kim & Jun 2009).

15 repetitions of these sentences were recorded by a female South Kyungsang Korean speaker in a sound-proof booth in Hanyang Institute of Phonetics and Cognitive Sciences Of Languages (HIPCS) in Seoul, South Korea, using a SHURE KSM44A microphone. The prosodic realization of each recorded utterance was checked by a trained K-ToBI (Korean Tones and Break Indices, Jun 2000) transcriber.

The produced tokens were used for reference values for the sound manipulation, as shown in Figure 4.2. f0 values at specific points in Figure 4.2 were calculated by averaging across 15 repetitions of the produced tokens. For both H- and LH-toned target words, f0 minimum (4.2a) and f0 maximum (4.2b) values were extracted and averaged. f0 minimum values for the preceding word (4.2c) were also obtained and averaged. The target word duration from the start of consonant release to the end of the rime (4.2d) was also averaged to avoid the effect of the duration.

To manipulate peak alignment and rising shape of the f0 contour in the target word, a pitch stylizing function was used in Praat (Boersma 2009). The pitch stylizing function simplifies a continuous f0 contour into several discretized points, which enables us to modify the pitch values manually.

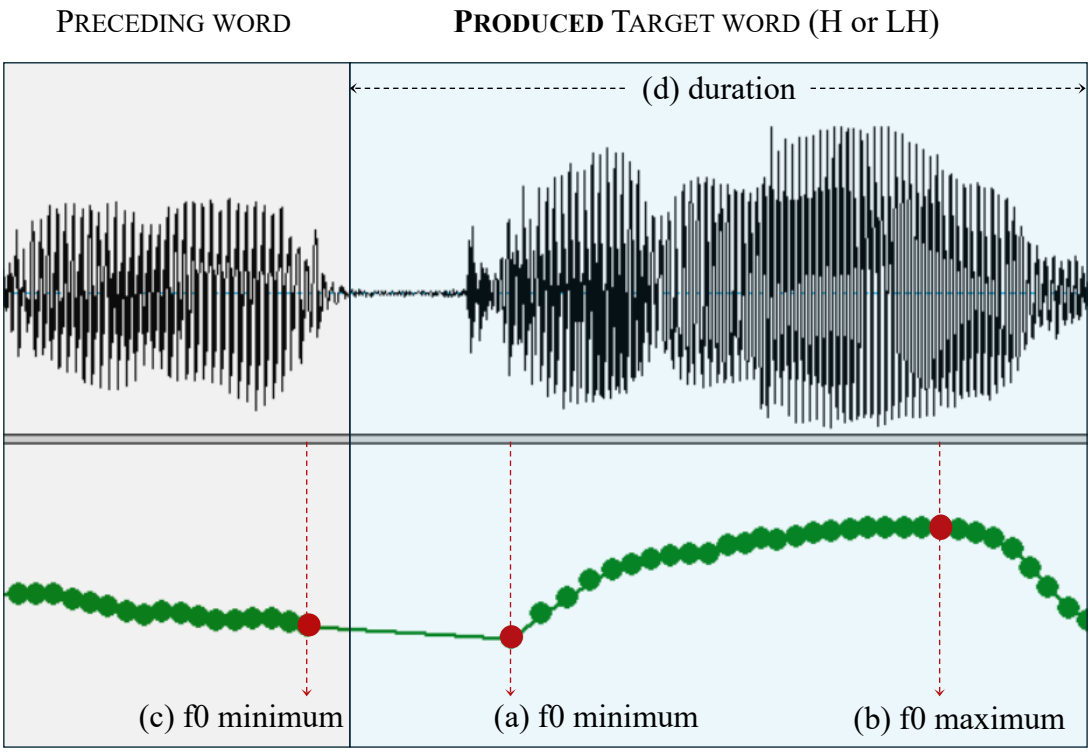


Figure 4.2: Reference values for sound manipulation. (a) and (b) refer to the f0 minimum and maximum values of each rising contour for H and LH, respectively. (c) indicates the f0 minimum value of the preceding value. (d) indicates the duration of the target words.

4.3.2 Stimulus resynthesis

4.3.2.1 Manipulation of peak alignment

First, to test how f0 peak alignment plays a role in South Kyungsang Korean listeners' perception, peak alignment was adjusted as illustrated in Figure 4.3. Segmental duration was set to be ambiguous, with the mean duration of all H and LH words (290 ms for /kan/, 252 ms for /pal/, and 280 ms for /pam/). However, the timing of peak alignment was varied from 20%, 40%, 60%, 80% to 100% of the rime duration, resulting in 5 steps. Each alignment step was increased with 20% of the rime duration (50 ms for /kan/, 46

ms for /pal/, 48 ms for /pam/). As for the f0 value of the target word, it started with 210 Hz and increased up to 270 Hz to reach its H target (f0 peak). After reaching its target, the f0 value ended with 210 Hz. The f0 rising contour shape remained the same but only the timing of the f0 contour differed along the continuum. The rise-fall portion of the contour lasted 40% of the total target word duration in all cases. Additionally, the f0 value of the preceding HL tone ended with the same f0 value (210 Hz) as at the vowel onset of the target words.

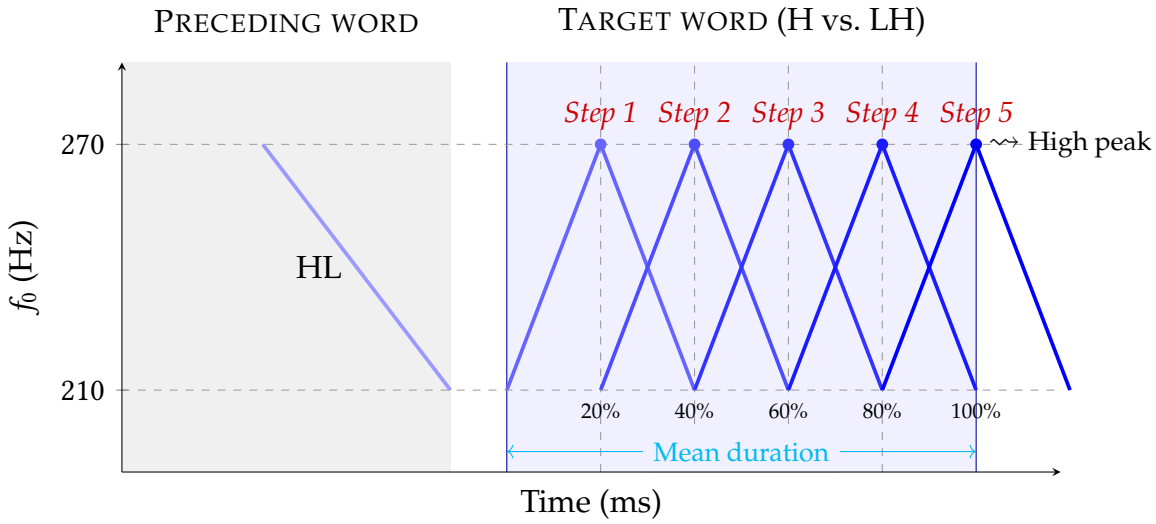


Figure 4.3: Manipulation of peak alignment. Step 1 indicates the earliest peak, while Step 5 indicates the latest peak.

4.3.2.2 Manipulation of rise shape

Next, to test whether f0 rise shape matters in the pitch accent categorization of SKK, f0 rise shape was manipulated, as shown in Figure 4.4. The f0 of the target word started from vowel onset (210 Hz) and ended at the offset of the sonorant coda (270 Hz). In order to create convex and concave rise shapes, the f0 value at the midpoint of the rime was increased by 10 Hz steps from 215 Hz to 265 Hz, resulting in 5 steps. Thus, for the most

convex shape, the f0 value started with 210 Hz, increased by 5 Hz at the vowel midpoint, and then reached its H target with 270 Hz. For the most concave shape, the f0 value started 210 Hz, but increased greater by 65 Hz at the vowel midpoint, and then reached its H target with 270 Hz. Adjusting the f0 value at the midpoint made it possible to make either concave or convex rise shapes. The f0 value of the preceding HL tone ended at the same level (210 Hz) as the vowel onset of the target word. Since longer duration is typically associated with LH-tone words (Chang 2013), the duration of each target word was set to a value ambiguous between the shorter H-tone words and the longer HL-tone words: 290 ms for /kan/, 252 ms for /pal/, and 280 ms for /pam/.

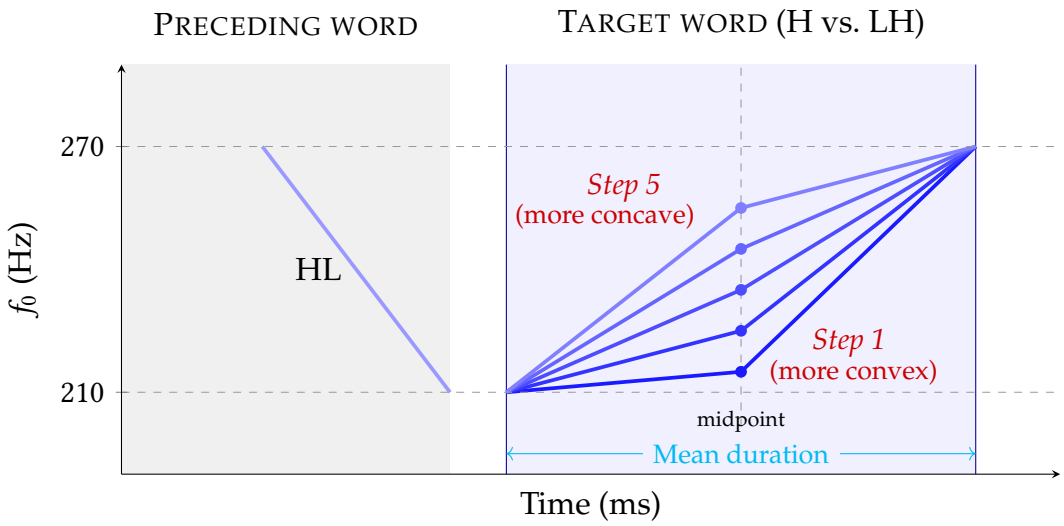


Figure 4.4: Manipulation of rise shape. Step 1 indicates the most convex f0 shape, while Step 5 indicates the most concave shape.

4.3.3 Participants

Participants were 22 native South Kyungsang Korean listeners (11 females, 11 males), whose age ranged from 20 to 28 (mean = 23.33). To control for dialectal difference, the participants and their parents were life-long residents of South Kyungsang region (Busan

and surrounding areas). All participants had no auditory impairments. They received compensation for their time and effort.

4.3.4 Procedure

Participants saw visual stimuli on a laptop computer and heard sound stimuli using a headphone (Sony MDR-7509) in a sound-proof booth in Ulsan, South Korea. A two-alternative forced choice (2AFC) experiment was carried out using PsychoPy (Peirce 2009). Before the experimental session, the participants were familiarized with the experimental setting during a short training session. In the main experiment, in each trial participants heard an auditory stimulus and were then asked to select the visual stimulus that matched the audio stimulus. Visual stimuli were presented side by side for each minimal pair (e.g., for the word /pal/, *foot* on the left and *shade* on the right), as shown in Figure 4.5.

In each sub-experiment (peak alignment, and rise shape), 120 target were used to see the effect of peak alignment and rise shape: 2 factors (5 f0 peak alignments or 5 f0 rise shapes) x 3 items (/kan/ for taste and liver, /pam/ for night and chestnut, /pal/ for foot and shade) x 4 repetitions. In order to mask the purpose of this study, 180 filler stimuli were added, which consisted of 10 combinations of pairs differing in either onset or coda from the target words (/kam/, /kang/, /pan/, /pang/, /tan/, /nan/, /nam/, /tam/, / khal/, /tal/). A total of 300 stimuli were presented in the study. The four blocks (repetitions) were presented in a randomized order, with an interval between the blocks (1st interval: 30 sec, 2nd interval: 5 mins, and 3rd interval: 30 sec). Among the 22 subjects, one participant was excluded because they responded to all stimuli as H tone, regardless of

f0 peak alignment and rise shape.

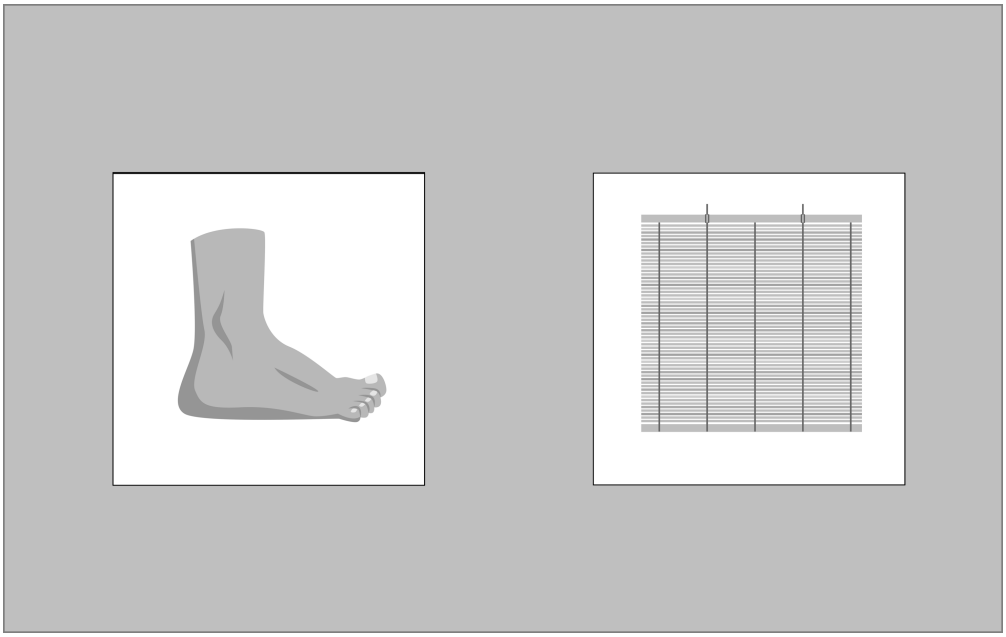


Figure 4.5: Visual stimuli for the 2AFC task.

4.3.5 Measurement and statistical analysis

Response data was analyzed with a generalized linear mixed effects model with a binomial linking function within the lme4 package in R (R Core Team 2020). Two models were created, one for data from the peak alignment sub-experiment and one for data from the rise shape sub-experiment.[5] To examine the effects of these factors on the pitch accent categorization, response (H vs. LH) was used as the dependent variable. H responses were coded as 1, and LH responses were coded as 0. The peak alignment model included one fixed effect, *continuum step* (5 steps from earlier to later peaks). The rise shape model

[5]R code for mixed-effects logistic regression models

- Peak alignment model: glmer (response ~ steps-peak + (1 + steps-peak | subject) + (1 + steps-peak | item), data = peak-alignment)
- Rise shape model: glmer (response ~ steps-shape + (1 + steps-shape | subject) + (1 | item), data = rise-shape)

included one fixed effect, *rise shape* (5 steps from convex to concave shapes). Both fixed effects were coded as continuous variables. The peak alignment model included random slopes for subject and item. The rise shape model, on the other hand, included only one random slope, for subject, with item as a random intercept. Maximal models were fitted, but due to convergence failure, the rise shape model was simplified, excluding the random slope for item (Barr et al. 2013). Due to convergence failure, random intercepts and slopes were included in order to keep maximal random structures. p-values less than .05 are reported as significant.

4.4 Results

4.4.1 Peak alignment

Figure 4.6 shows the percentage of H responses for each step of the continuum from earlier to later peak alignments. Results showed that South Kyungsang Korean listeners did not differentiate H vs. LH depending on f0 peak alignment, regardless of whether the peak is aligned earlier or later. The H response percentage for all continuum steps for peak alignment was below chance level (mean = 0.16, sd = 0.37). In the peak alignment model, differences in peak alignment did not exert a significant effect on response data (β = -0.09, z = -0.89, p <.001), as shown in Table 4.2. That is, regardless of the timing of f0 peak, the listeners responded to all steps in the continuum mainly as LH.

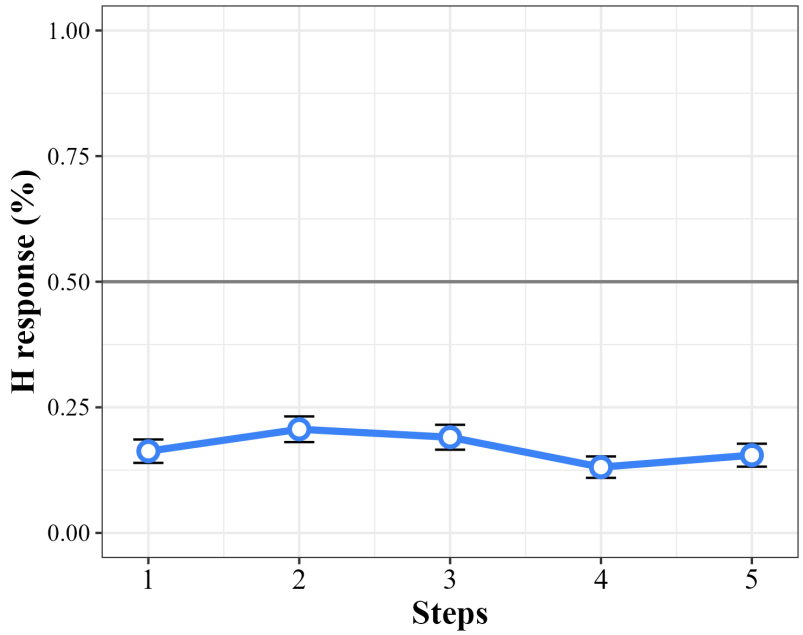


Figure 4.6: H response percentage for peak alignment. Step 1 represents the earliest peak, while Step 5 represents the latest peak. The horizontal black line indicates the 50% crossover point; error bars indicate standard errors.

Table 4.2: Statistical result of peak alignment

	Estimate	Std. Error	*z* value	*p* value
(Intercept)	-1.643	0.364	-4.509	< .001
steps-peak	-0.090	0.102	-0.890	0.373

The insignificance of varying peak alignment on tonal categorization was further confirmed by examining the by-item and by-subject data. Figure 4.7 displays the H response percentage for peak alignment by-item (/kan/, /pal/, /pam/). The H response percentage for all continuum steps for was below chance level for all three lexical items (/kan/: mean = 0.17, sd = 0.38; /pal/: mean = 0.17, sd = 0.38; /pam/: mean = 0.17, sd = 0.37). In sum, regardless of the timing of the f0 peak, the listeners responded as LH for all the continuum steps for the different lexical items (/kan/, /pal/, and /pam/).

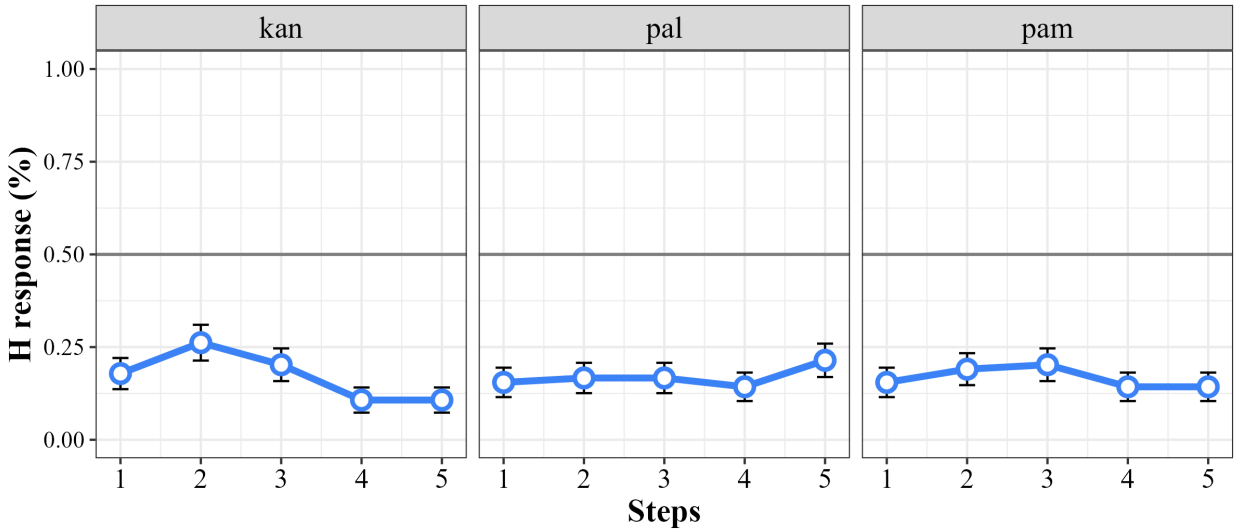


Figure 4.7: H response percentage for peak alignment by-item (/kan/, /pal/, /pam/). Step 1 represents the earliest peak, while Step 5 represents the latest peak. The horizontal black line indicates the 50% crossover point and error bars indicate standard error.

Figure 4.8 shows the percentage of H responses by-subject. The plots in Figure 4.8 indicate that for all subjects, the H response percentage was below chance level, indicating that they did not distinguish H vs. LH based on the earlier or later f0 peaks. Four subjects (F07, F08, F10, M09) exhibited a tendency for later peaks to be more likely identified as LH tone, but all the steps were still below chance, undermining any clear by-subject effect. Two subjects (M04, M05) showed H response percentages above chance at Steps 4 and 5, though this ran counter to predictions, since these steps on the continuum were predicted to favor LH responses. That is, for these two participants, their H response percentages increased from earlier to later peaks, indicating that the later peaks were identified as H tone, not LH tone. Thus, there were no individual differences in terms of the peak alignment effect on tonal categorization. Overall, the results of by-item and by-subject responses suggest that South Kyungsang Korean listeners do not use peak alignment cues to distinguish H vs. LH pitch accents.

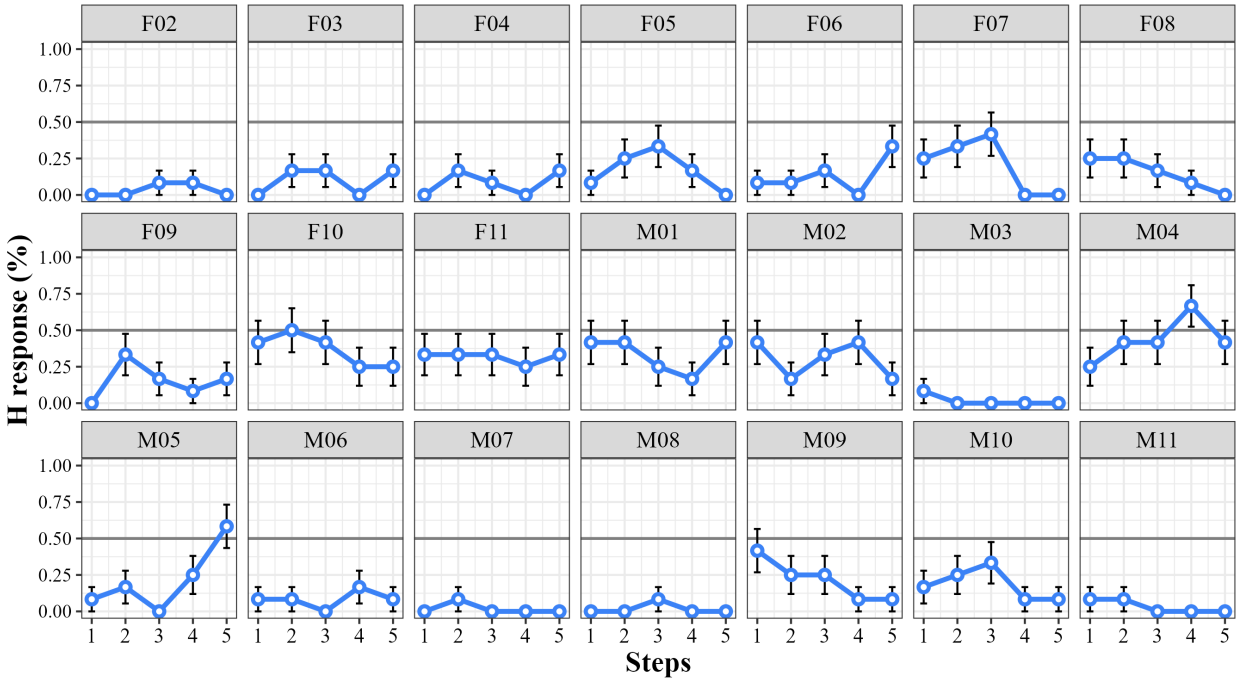


Figure 4.8: H response percentage for peak alignment by subject. Towards Step 1 represents the earliest peak; Step 5 represents the latest peak. The black horizontal line indicates the 50% crossover point; error bars indicate standard error.

4.4.2 Discussion of Peak alignment

Results from Section 4.4.1 show that South Kyungsang Korean listeners did not perceive H vs. LH differently depending on f0 peak alignment. They responded to all continuum steps as LH, regardless of whether the peak is aligned earlier or later. Thus, no effect of f0 peak alignment was observed for South Kyungsang Korean's tonal categorization. The absence of peak alignment effect was further confirmed across items (/kan/, /pal/, /pam/) and subjects (21 subjects), as observed in the absence of by-item and by-subject effects.

Interestingly, these results were very different from the previous findings on the perception of the lexical pitch accent in South Kyungsang Korean, where earlier peaks were

identified as H, while later peaks were identified as LH (Chang 2013). The only difference between this chapter and Chang (2013) is the design of the stimuli for the peak alignment, specifically the differences in slopes for the f0 rise and fall.

In Chang's study, the slope of rise and fall changed in every step, while in the present study, the slope of rise and fall remained the same for all steps. One might expect that if South Kyungsang Korean listeners were to be influenced by the slope-related confounds, the listeners in Chang's study would have had a hard time being sensitive to the peak alignment, while those in this chapter would have been able to focus on the peak alignment effect.

In fact, the results from these two studies were the opposite of what we had expected. That is, as for the present study, despite the fully-controlled slopes, South Kyungsang Korean listeners did not perceive any difference for earlier or later peaks, while as for Chang's study, despite the slope difference, South Kyungsang Korean listeners were able to distinguish H vs. LH depending on the peak alignment.

One explanation for this surprising result may be the existence of low f0 values before the rise towards the f0 peak. In this chapter, f0 was sustained at a low value (210 Hz) for a while and then initiated rising from Step 2 to Step 4 in the continuum. However, in Chang (2013)'s study, f0 rising started at the vowel onset towards the f0 peak for all continuum steps. That being said, it would be difficult to explain why Step 1 in this chapter, in which f0 rises directly towards peak from vowel onset, was still consistently treated as LH.

Two stimuli-related possibilities may have affected participant responses here. First, H target words in Figure 4.1 begin with a higher f0 than f0 offset of the preceding word. In contrast, LH target words begin with a lower f0, roughly comparable to the f0 at the

end of the preceding HL word. Forcing all target words to begin with a lower f0 may have triggered a much higher rate of LH responses since this is one cue of the H vs. LH tonal contrast in the language.

Furthermore, Chang (2013) found evidence to support the relevance of initial f0 for tonal perception. The amount of the target rhyme that involved raised f0 was 40% across all continuum steps. When this is compared to pitch tracks in Figure 4.1, it is clear that H-toned syllables are produced with raised f0 for the entire syllable while LH syllable show less rasied f0 due to the initial L tone.

Another evidence comes from my initial pilot study Joo & D'Imperio (2025). In Joo & D'Imperio (2025), which implemented the same design for peak alignment as Chang (2013), results indicated that speakers were using the peak alignment to differentiate tonal contrasts. Crucially, in both Chang (2013) and Joo & D'Imperio (2025), the amount of the target syllable that exhibited raised f0 was much larger than in this chapter.

Future studies should use stimuli that involve longer rise-fall excursions, as these are both more consistent with pitch tracks in Figure 4.1 and accord with results from Chang (2013) and Joo & D'Imperio (2025). As a result of this methodological shortcoming, it is difficult to interpret results from this section.

4.4.3 Rise shape

The second sub-experiment tested whether or not South Kyungsang Korean listeners attend to rise shape for the categorization of H vs LH in monosyllables. Results, which are shown in Figure 4.9, indicate that the rate of H responses increases as the f0 rise becomes

more concave.

Specifically, the percentage of H responses gradually increased from Step 1 to Step 5; steps-shape was found to be significant in the statistical model (Rise shape model, β = 0.68, z = -7.61, p <.001), as shown in Table 4.3. The parameter estimates of the model indicated that a one unit increase in steps-shape yielded a change in the log odds of selecting H tone by 0.69, with a positive difference of 33% in the probability of selecting H tone.

The category boundary (50% crossover point) was at Step 2, as shown in the vertical line of Figure 4.9. Note that below Step 2, when the rise is more convex, listeners were more likely to respond as LH, while above Step 2, when the rise is more concave, they were more likely to respond as H. At the two extremes of the continuum, listeners treated stimuli at Step 1 as LH-toned (mean = 0.37, sd = 0.48), whereas they treated stimuli at Step 5 as H-toned (mean = 0.85, sd = 0.35). In short, the shape of the rise conditions South Kyungsang Korean listeners' perception of tonal contours.

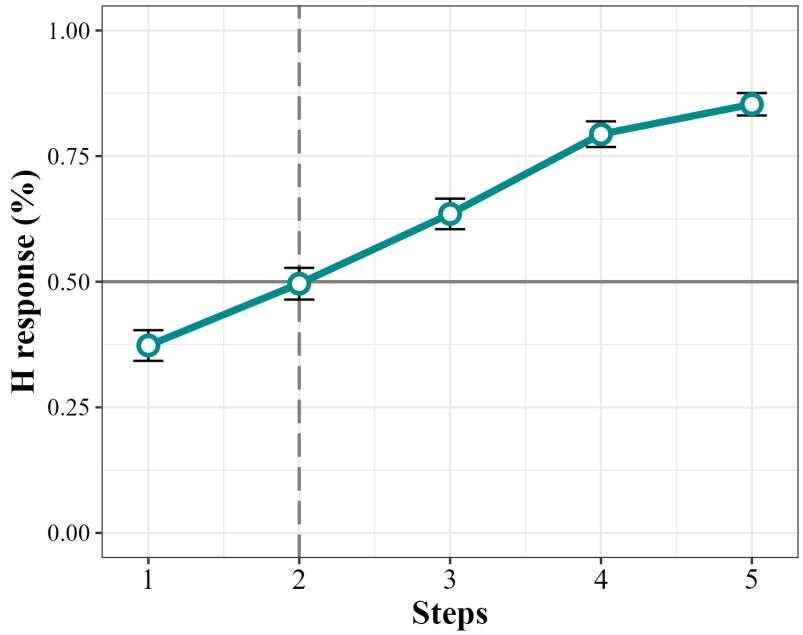


Figure 4.9: H response percentage for rise shape model. Step 1 represents the most convex shape, while Step 5 represents the most concave shape. The horizontal black line indicates the 50% crossover point; error bars indicate standard error.

Table 4.3: Statistical result of rise shape.

	Estimate	Std. Error	*z* value	*p* value
(Intercept)	-1.362	0.293	-4.649	< .001
steps_shape	0.689	0.091	7.612	< .001

Figure 4.10 shows the percentage of H response for by item (/kan/, /pal/, /pam/). The percentage of H responses shows the same pattern across all three lexical pairs. The category boundary for all the items was at Step 2. The response percentage increased from Step 1 to Step 5, showing that the more convex shapes were perceived as LH, the more concave shapes were perceived as H. At the two extremes in the continuum, stimuli at Step 1 were typically perceived as LH (/kan/: mean = 0.35, sd = 0.48; /pal/: mean = 0.37, sd = 0.49; /pam/: mean = 0.41, sd = 0.49), while stimuli at Step 5 were consistently perceived as H (/kan/: mean = 0.86, sd = 0.35; /pal/: mean = 0.81, sd = 0.40; /pam/:

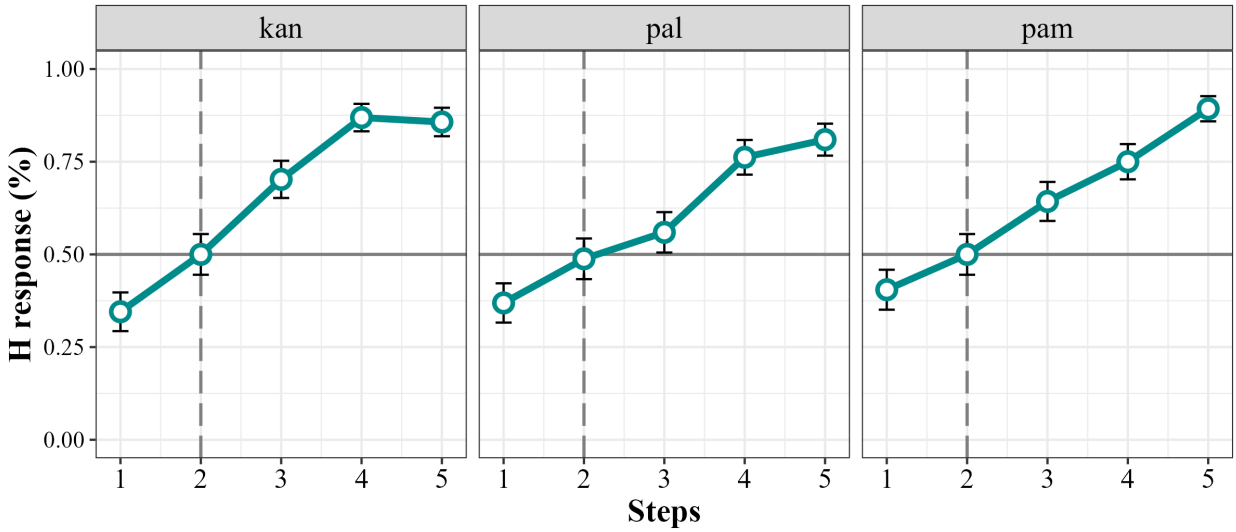


Figure 4.10: Percentage for rise shape by-item (/kan/, /pal/, /pam/). Step 1 represents the most convex shape, while Step 5 represents the most concave shape. The horizontal black line indicates the 50% crossover point; error bars indicate standard error.

mean = 0.89, sd = 0.31). The uniformity of the effect across items further supports the effect of rise shape on tonal categorization in South Kyungsang Korean.

Figure 4.11 presents by-subject percentage of H responses for rise shape. Among the 21 subjects, 13 subjects with orange borders in Figure 23 (F02, F03, F06, F07, F09, F01, M01, M03, M05, M06, M08, M09, M10) showed a categorical boundary around Step 2 or 3. Below the categorical boundary, when the rise shape is convex, they categorized stimuli as LH, while above it, when the rise shape is more concave, they categorized stimuli as H.

The subjects without orange borders in Figure 4.11 (F04, F05, F08, F10, M02, M04, M07, M11) showed a similar tendency to the overall response pattern, such that the convex rise shapes were more likely to be categorized as LH, while the concave rise shapes were categorized as H. Despite the similar tendency, all the response percentage was above chance level, showing that the listeners perceived all continuum steps as H tone. To summarize,

by-item and by-subject results support the perceptual salience of rise shape for the categorization of H vs. LH tones for South Kyungsang Korean listeners.

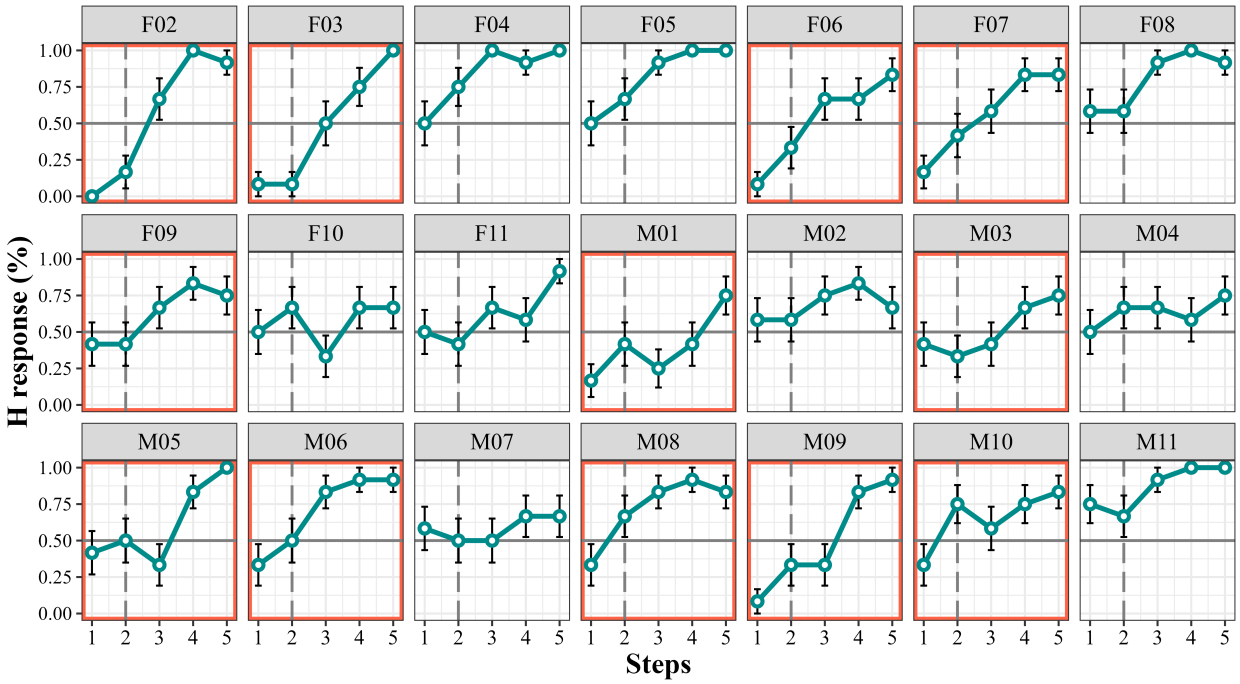


Figure 4.11: Percentage of H responses for rise shape by subject. Step 1 represents the most convex shape, while Step 5 represents the most concave shape. The horizontal black line indicates the 50% crossover point; error bars indicate standard error.

4.4.4 Discussion of Rise shape

Results showed an effect of f0 rise shape on tonal categorization. The percentage of H responses significantly increased when participants were exposed to more concave rise shapes with a category boundary between H and LH at Step 2. In short, South Kyungsang Korean listeners treated more convex shapes as the LH tone, while more concave shapes were treated as the H tone.

This pattern was further confirmed across items (/kan/, /pal/, /pam/), finding no meaningful by-item differences for H response percentage. Moreover, by-subject results

confirm this interpretation of the data. Most subjects (13 out of 21) showed a categorical shift for H responses from more convex to concave shapes at Step 2 or 3, while the rest of the subjects showed similar tendency though their overall percentage of H responses was generally above chance. Overall, all subjects showed a similar effect of f0 rise shape on H response percentage – more convex shapes were perceived as LH; more concave shapes were perceived as H.

The effect of the f0 rise shape was also consistent with the results in Joo & D'Imperio (2025), where 11 continuum steps from convex to concave shapes were used for the rise shape, but with a greater f0 range (190-290 Hz) than that (210-270 Hz) in the present study. Note that H response percentage for f0 rise shape in Joo & D'Imperio (2025) exhibited a sigmoidal curve. The likelihood of H response gradually increased from convex to concave shapes, with a category boundary at Step 7 among the 11 continuum steps. Taken together, the results in Joo & D'Imperio (2025) and the present study strongly suggest that the rise shape is an important cue for distinguishing H vs. LH pitch accents.

4.5 Discussion

The aim of this chapter was to understand what kind of f0 information is represented in our intonational phonological grammar, and to further understand which phonological primitives of intonation South Kyungsang Korean listeners attend to in a perception task.

First, the results of the peak alignment showed that South Kyungsang Korean listeners did not differentiate between H and LH tones based on the f0 peak alignment. They perceived all continuum steps as LH, regardless of whether the peak was aligned earlier

or later. The absence of a peak alignment effect was consistent across items (/kan/, /pal/, /pam/) and subjects (21 subjects).

Interestingly, these results differed from both the findings from Chang (2013) and Joo & D'Imperio (2025). Recall that Chang (2013) showed that South Kyungsang Korean listeners responded to earlier peaks as H while later peaks as LH, showing the existence of the peak alignment effect on the lexical pitch accent perception in South Kyungsang Korean.

One of the possible explanations was due to the slope-related confounds in the design of the stimuli in the present study vs. Chang (2013). That is, the slopes remained the same for the present study, whereas the slopes changed every step for Chang (2013). In fact, the same slopes across the steps induced the LH responses, while different slopes for every step induced earlier peaks as the H responses and later peaks as the LH responses. These results make it hard to explain why South Kyungsang Korean listeners sometimes perceived the differences in peak alignment but sometimes did not perceive at all.

Moreover, this chapter showed different results from Joo & D'Imperio (2025), where South Kyungsang Korean listeners responded to all continuum steps as H, regardless of the peak alignment timing, which was exactly opposite from our main experiment. The slope-related confounds in Joo & D'Imperio (2025) were eliminated in the experiment of this chapter such that we expected to see some results that are at least the H responses across the steps as shown in Joo & D'Imperio (2025) or the peak alignment effect for earlier vs. later peaks. However, the responses did not meet expectations, and further investigation of this likely requires revising the stimuli to increase the percentage of the rhyme in which a rise-fall excursion occurs, reflecting pitch tracks in Figure 4.1 and Chang's results.

Another possible explanation for the surprising rate of LH responses could be that the particular rise-fall excursions manipulated resulted in unnatural stimuli. Since the slope and duration of the f0 did not resemble those in the production, it is possible that listeners would have responded to the unnatural sounds as less familiar lexical items. That is, lexical items for H tone may have a higher lexical frequency than those for LH tone (/kan/: H, taste vs. LH, liver; /pam/: H, night vs. LH, chestnut; /pal/: H, foot vs. LH, shade). If lexical frequency came into play, listeners may have categorized the peak alignment stimuli based on the frequency of those competing minimal pairs. We do not have corpus data at present to thoroughly test this possibility, though it should inform future work.

Second, the results of the rise shape sub-experiment showed that South Kyungsang Korean listeners use rise shape as a perceptual cue to distinguish H vs. LH lexical pitch accents. Overall, all subjects showed a similar effect of f0 rise shape on H response percentage, such that more convex shapes were perceived as LH, while more concave shapes as H. This finding provides significant support for the need for more than simply tonal targets in the intonational grammar. In particular, AM claims that tonal targets, a collection of L and H tones, is sufficient to model the intonational grammar. This entails that the shape of the curve from one turning point to another is entirely epiphenomenal for the phonological grammar. However, these results suggest just the opposite, that the shape of the rising f0 curve in South Kyungsang Korean is highly perceptually salient to listeners. Therefore, in order to fully account for the perception of lexical pitch accent in South Kyungsang Korean, theoretical frameworks must integrate the role of continuous f0 information as well as discrete tonal targets.

South Kyungsang Korean listeners' sensitivity to the rise shape is consistent with several post-lexical pitch accent (intonational) languages in the literature, where f0 perception differed depending on the shape of the pitch accents in distinguishing sentential meanings (Barnes et al. 2010, 2012, 2021, Dorokhova & D'Imperio 2019).

This may suggest that the perception of pitch accents is closely related to the perception of rise shape, no matter whether the pitch accent is specified lexically as in the lexical pitch accent languages such as South Kyungsang Korean, or post-lexically as in the post-lexical pitch accent languages such as English and French. This is consistent with the TCoG hypothesis, claiming that the rise (and fall) shape is perceptually crucial to determine pitch accent category in American English (Barnes et al. 2010, 2012, 2021). This also suggests that the fine-phonetic details about f0 shape may be stored in the phonological representation of the pitch accents (Barnes et al. 2010, 2012, 2021, Kimball & Cole 2016), in line with the configurational approach (e.g., Bolinger 1951, Hart et al. 2003, Hirst & Di Cristo 1998).

The multiplicity of cues, vowel duration, f0 scaling (i.e., the relative height of the f0 maximum for each contour), initial f0, in addition to rise shape and any potential effect of peak alignment, renders it difficult to conclude which feature(s) are primary in tonal perception in South Kyungsang Korean. This calls for further studies on how South Kyungsang Korean listeners exploit perceptual cues for distinguishing H vs. LH pitch accents — whether they use f0 rise shape as a primary cue or as a secondary cue in addition to another primary perceptual cue, and how this interacts with the other cues involved in the contrast.

There could be several specific ways that we can further examine South Kyungsang

Korean's perception of lexical pitch accents. First, it is important to understand the contradicting results of peak alignments from the previous findings (Chang 2013, Joo & D'Imperio 2025). Although the design of the main experiment in the present study controlled the slope-related confounds, the results of the main experiment were different from the results of Chang (2013) and Joo & D'Imperio (2025) where the slope-related confounds were not fully controlled. We need to see whether these different results are derived solely from the slope difference.

Alternatively, as discussed above, the overall LH responses of the experiment could be due to the sustained low f0 values before the rise onset, which may indicate the existence of the L tone target. This was a by-product of the experimental design for the peak alignment and we did not explicitly include the L tone targets as a factor in the study. Note that initial f0 was an important factor for the pitch accent categorization of South Kyungsang Korean, as found in Chang (2013). Thus, it would be necessary to take into account the initial f0 values as another factor.

Moreover, it is noteworthy that f0 height for the overall contour differs between H and LH pitch accents in the production of f0 contour, as shown in Figure 4.1 above. This may also indicate that the overall tonal scaling can be an important factor to consider. Then, we may have to tease apart whether the perceptual difference comes from the initial f0 or the overall f0 scaling.

Second, we also need to account for segmental duration since f0 slope and duration may co-vary. That is, H tone may be accompanied by gradual slope and shorter duration, while LH tone may be related to steeper slope and longer duration. Note that in Joo & D'Imperio (2025), segmental duration was one of the cues to differentiate H vs. LH pitch

accent, although it was not as robust as the effect of the rise shape. Thus, it would be crucial to see how the slope and the duration interact with each other, so that we can explain which cues can be accounted for in the perception of the pitch accents.

It is also possible that the rise shape effect may not be the primary factor in distinguishing H vs. LH pitch accents, but rather it could be a result of phonetic enhancement on some other phonological feature (e.g., Keyser & Stevens 2006, Stevens & Keyser 2010). Enhancement effects refers to when distinctive phonological units are phonetically realized with feature-related acoustic/articulatory correlates and feature-enhancing correlates. For example, when distinguishing two contrastive phonemes in English /ʃ/ vs. /s/, the phonetic, non-phonological lip rounding of /ʃ/ may enhance the perceptual saliency of this contrast, such that the perceptual categorization of these two sibilants /ʃ/ and /s/ improves a lot. In this case, the distinctive phonological feature is [anterior], which, among sibilants, is phonetically realized by differences in center of gravity. However, lip rounding also modulates center of gravity, with lip rounding for /ʃ/ further differentiating the spectra of these sounds. One of the phonological feature-related acoustic correlates is the place of articulation, but feature-enhancing feature, the lip rounding, facilitates the distinction between the two phonemes. Likewise, it is also possible that one of the phonologically relevant correlates for f0 contour is peak alignment, but the rising shape enhances the perceptual distinction of H vs. LH pitch accents. More concave shapes approach an f0 maxima sooner than more convex rise shapes. Given the convexity of LH in South Kyungsang Korean, a concave H may serve to simply enhance the distinction between an earlier peak for H-toned syllables and a later peak for LH-toned syllables. Then, we may be able to account for the absence of the peak alignment effect but the presence

of the rising shape. But this requires further investigation by examining the interaction of the peak alignment and the rising shape. Thus, we need to include the interaction between the peak alignment and the rising shape, such that we may be able to see which cue is more salient than the other in the context of where both cues are given.

Lastly, we have only looked at monosyllabic words embedded in a phrase-medial position. It would be also interesting to extend to longer words such as bisyllabic or trisyllabic and see how the peak alignment and rise shape factors extend to these different phrasal contexts. We can also add suffixes to monomorphemic noun stems (e.g., /kan/, /pal/, /mal/). In South Kyungsang Korean, some studies (Do et al. 2014, Lee & Zhang 2014) suggest that when a vowel-initial suffix is attached to the monosyllabic words, H tone surfaces as HH or HL, while LH tone surfaces as LH. This is because the underlying tonal pattern of H tone is actually not H, but H with a floating H or L. If we test the monosyllabic items + a vowel-initial suffix, we may be able to understand the underlying tonal patterns that are actually stored in our phonological representation.

To sum up, future studies are required to test whether we can generalize findings from the present study to other types of words or with other additional factors, which will provide a deeper understanding of the perception of the lexical pitch accent in South Kyungsang Korean, and more generally, how intonational contours are represented in the grammar.

4.6 Summary and conclusion

The chapter investigated how South Kyungsang Korean listeners perceive lexical pitch accents, H vs. LH, using different theoretical approaches to the phonological representation of f0 contour.

The result showed that South Kyungsang Korean listeners did not perceive any difference in H vs. LH pitch accents depending on f0 peak alignment. However, f0 rise shape showed a significant effect on H response percentage, showing that South Kyungsang Korean listeners perceive more concave shapes as LH and more convex shapes as H. These findings suggest that continuous f0 information is crucial for accounting for H vs. LH pitch accent perception in South Kyungsang Korean.

This study adds evidence that f0 rise shape could be a perceptual primitive for distingiushing lexical pitch accent in South Kyungsang Korean, extending previous studies of post-lexical pitch accent languages that highlighted the role of f0 shape for sentential contrast. A potential impact of additional factors such as target f0 height and its interactions with rise shape and segmental duration need to be examined in further studies.

Chapter 5

Defining intonation mathematically with discrete tonal targets

Intonation has been productively studied from theoretical and experimental perspectives, but the representation and computation of intonation has yet to be explored *mathematically* using *model theory and logic*. This model-theoretic framework defines linguistic information more precisely and explicitly, offering a new perspective on representing input and output structures and computing the relations between them. Specifically, *logical transductions* have been employed in building linguistic structures in the field of computational phonology (e.g., Strother-Garcia 2019; Jardine 2017; Jardine et al. 2021).

The goal of this chapter is to investigate how discrete tonal targets in intonation are mathematically analyzed, specifically using *logical transduction*. For this, it starts with a hypothesis that *intonation is a Quantifier-Free (QF) logical interpretation of a metrical and prosodic structure*. It extends Jardine (2017) and Strother-Garcia (2019) by viewing the autosegmental representations as additional structures imposed on an input string. This QF property of intonation shows the restrictiveness of intonational structure in terms of com-

putational complexity and typologically compares differences and similarities across languages in terms of logical interpretation. Therefore, to understand the computational representation and mechanisms of tonal mappings in intonation, this chapter examines three distinct types of intonational patterns: a *head-prominence* language, American English; an *edge-prominence* language, Seoul Korean; a *lexical pitch accent* (head/edge-prominence) language, Tokyo Japanese.

Before moving onto the case studies of these languages (section 5.2), the following section 5.1 briefly overviews the basic structure of logical transduction. Based on that, the intonational structure is viewed as a QF logical interpretation of a metrical and prosodic structure. Different intonational patterns are analyzed with two logical transductions step-by-step. This approach provides a way to compare the intonational typology of languages by measuring computational complexity. Note that the focus of this chapter is on the basic "phonological" patterns of intonation, although the intonational patterns may include a lot of phonetic variations in the actual realization.

Much of this chapter was first presented at the Society for Computation in Linguistics (SCiL 2025) and published in the peer-reviewed proceedings (Joo & Jardine 2025). It has been considerably rewritten for the general audience.

5.1 Logical transduction

Logical transductions are a mathematical tool to build a new structure in the output from an input structure (Courcelle 1994, Engelfriet & Hoogeboom 2001, Filiot & Reynier 2016). In this framework, an input structure is said to be *interpreted* into the output structure,

which consists of a finite number of *copies*, using logical formulas. To define the discrete tonal targets, this chapter uses transductions defined in first-order (FO) logic. An example graph in Figure 5.1 illustrates a logical transduction of syllable structure building for ⋊CVCV⋉.

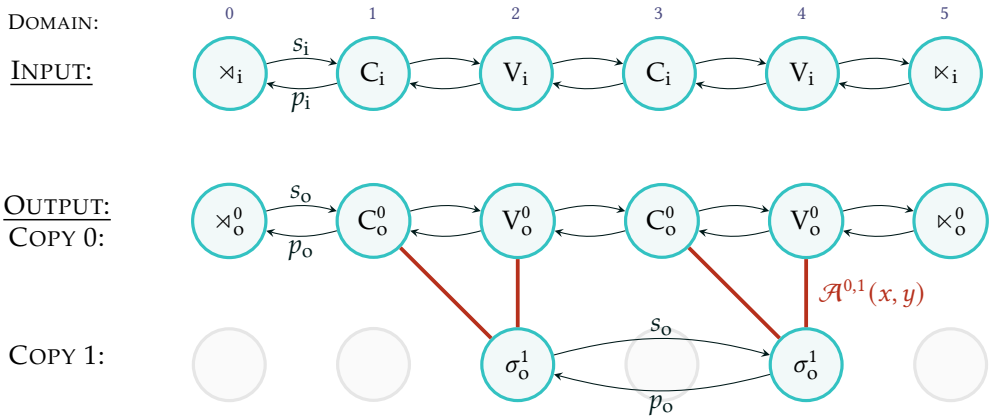


Figure 5.1: A graph illustrating a logical transduction for ⋊CVCV⋉.

Let a logical transduction be $\tau : \Sigma \rightarrow \Gamma$. Σ is an input structure, $\langle \mathcal{D}; \mathcal{R}_1, \mathcal{R}_2, \cdots, \mathcal{R}_n \rangle$, while Γ is an output structure, $\langle \mathcal{D}'; \mathcal{R}'_1, \mathcal{R}'_2, \cdots, \mathcal{R}'_n \rangle$. Via a logical transduction τ, the output structure Γ is interpreted as a finite set of copies C of the input structure Σ, using the domain formulas ϕ_{dom}. The input structure Σ is specified with its signature $\mathcal{S}_i$, while the output structure Γ is specified with its signature $\mathcal{S}_o$.

Let $C = \{0, 1, 2, \cdots, k\}$ be the indices of copy sets. The domain of the output structure $\mathcal{D}'$ is therefore $\mathcal{D} \times (k + 1)$ copies of the domain elements in the input structure. For example, the input structure consists of a set of nodes whose positions are marked with a set of domain indices {0, 1, 2, 3, 4, 5}, while the output structure is two copy sets[6] of the

[6]Throughout this dissertation, the term *Copy Set* is used, which is different from the *copies* defined in Enderton (2001). Each *Copy Set* here corresponds to each of the TBU or tonal tier in the hierarchical structure of intonation. For instance, in American English, COPY SET 0 is the TBU tier, while COPY SET 1-3 are the pitch accent, phrasal accent, and boundary tone tiers, respectively. A set of Copy Sets as a whole represents a single hierarchical prosodic structure.

input structure, as illustrated Figure 5.1. In this way, the output can be a larger structure based on the input structure.

Unary relations in COPY SET m are denoted as $\mathcal{R}'^{m}$, while binary relations between COPY SETS m and n are denoted as $\mathcal{R}'^{m,n}$, where $m, n \in C$. These unary and binary relations are specified with node formulas, $\phi^{m}_{\sigma}(x)$ and $\phi^{m,n}_{\sigma'}(x, y)$ respectively, where $\sigma, \sigma' \in \mathcal{S}_o$, $m, n \in C$, with free variables x and y. The unary relations $\mathcal{R}'^{m}$ is defined as $\{u^m |\ u \in \Sigma, m \in C$, there is exactly one $\sigma \in \mathcal{S}_o$ such that $\Sigma \models \phi^{m}_{\sigma}(x)\}$. That is, $\mathcal{R}'^{m}$ is a set of the elements u in COPY SET m that is labeled as σ, which is satisfied by the output structure Γ. The binary relations $\mathcal{R}'^{m,n}$ is defined as $\{u^m, v^n |\ u, v \in \Sigma, m, n \in C, \sigma' \in \mathcal{S}_o$ such that $\Sigma \models \phi^{m,n}_{\sigma'}(u, v)\}$. That is, $\mathcal{R}'^{m,n}$ is a set of the paired elements, u in COPY SET m and v in COPY SET n, whose formula is labeled as σ', which is satisfied by the output structure Γ. For the node formulas, $\phi^{m}_{\sigma}(x) \stackrel{\text{def}}{=} \sigma^{m}_{o}(x)$ and $\phi^{m,n}_{\sigma'}(x, y) \stackrel{\text{def}}{=} \sigma'^{m,n}_{o}(x, y)$. These unary and binary relations can be defined from the formulas in the input. For the notation, the output elements are indicated by a subscript $_o$, and the input elements, by a subscript $_i$.

For instance, in Figure 5.1, a unary relation, $V^{0}_{o}(x) \stackrel{\text{def}}{=} V_i(x)$ holds true if position x is labeled V in COPY SET 0 if and only if x is labeled V in the input. A binary relation, $\mathcal{A}^{0,1}_{o}(x, y) \stackrel{\text{def}}{=} V(x) \wedge \sigma(x) \wedge x \approx y$, holds true if and only if position x labeled V in COPY SET 0 and position y labeled σ in COPY SET 1 are associated, and they are at the same position in the input.

As for the relations between the output domain elements, we follow Chandlee & Jardine (2019b) such that the order of the elements in the copy sets (i.e., p_o, s_o) is assumed to be preserved from the order of the input elements (i.e., p_i, s_i). But for the intonational transductions, the order preservation occurs separately for TBU and tonal copy

sets, which will be discussed in detail in the next section (Section 5.2).

Importantly, Strother-Garcia (2019) used logical transductions to define *syllable structures* from an input string. This approach provides a valuable insight into this chapter by mathematically formulating the mapping from a simple string in the input to a more complex phonological structure in the output. In this way, the representation of a linguistic entity can be "enriched" with hierarchical linguistic information (Strother-Garcia 2019). Inspired by this approach, this chapter defines the hierarchical structure of intonation in the output from an input string. First, let's look at how Strother-Garcia (2019) builds syllable structures from an input string, $\rtimes$CVCV$\ltimes$.

As defined in the formulas (5.1 - 5.4) and graphically represented in Figure 5.2, every element in the input string is copied in COPY SET 0. Any output nodes that are a consonant or a vowel, $C_o^0(x)$ or $V_o^0(x)$, are defined from those elements in the input, $C_i(x)$ or $V_i(x)$. Also, boundaries in the output, $\rtimes_o^0(x)$ or $\ltimes_o^0(x)$, come from the input, $\rtimes_i(x)$ or $\ltimes_i(x)$.

Note that this chapter follows the format of interpreting each formula as in Strother-Garcia (2019).

$$C_o^0(x) \stackrel{\text{def}}{=} C_i(x) \tag{5.1}$$

"Position x in COPY SET 0 is labeled C iff x is labeled C in the input."

$$V_o^0(x) \stackrel{\text{def}}{=} V_i(x) \tag{5.2}$$

"Position x in COPY SET 0 is labeled V iff x is labeled V in the input."

$$\ltimes_o^0(x) \stackrel{\text{def}}{=} \ltimes_{\varphi_i}(x) \tag{5.3}$$

"Position x in COPY SET 0 is labeled $\ltimes$ iff x is labeled $\ltimes$ in the input."

$$\rtimes_o^0(x) \stackrel{\text{def}}{=} \rtimes_i(x) \tag{5.4}$$

"Position x in COPY SET 0 is labeled $\rtimes$ iff x is labeled $\rtimes$ in the input."

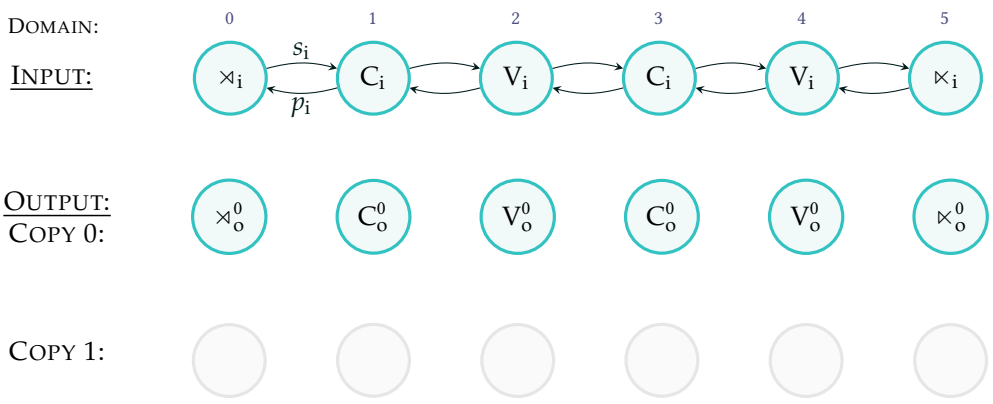


Figure 5.2: COPY SET 0 for syllable structural building from the input string $\rtimes$CVCV$\ltimes$.

Importantly, for the second copy (COPY SET 1), syllables in the output, $\sigma_o^1(x)$, is defined from a vowel in the input, $V_i(x)$, as provided in the formula (5.5) and in Figure 5.3. This shows that every syllable is a *reflection* or an *interpretation* of the nucleus. If there are no formulas defined for some nodes, it means the formulas are false for those nodes indicated

by the gray empty nodes.

$$\sigma_o^1(x) \stackrel{\text{def}}{=} V_i(x) \tag{5.5}$$

"Position x in COPY SET 1 is labeled σ iff x is labeled V in the input."

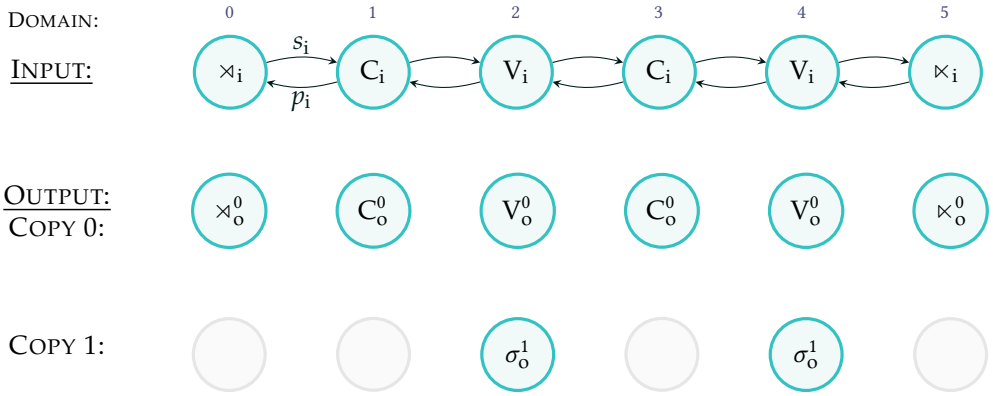


Figure 5.3: COPY SET 1 for syllable structural building from the input string $\rtimes$CVCV$\ltimes$.

Based on the definition of the copy sets in the output, we can establish some relations between them to build syllable structures. ${\mathcal{A}_o}^{m,n}(x, y)$ defines an association relationship between the output positions x and y, where $m, n \in C$, by referring to the input positions. As provided in the formulas (5.6), ${\mathcal{A}_o}^{0,1}(x, y)$ associates each C and V in COPY SET 0 with each syllable in COPY SET 1, respectively, with two conditions: when the input position of the C nodes in COPY SET 0 (i.e., the C nodes in the input) precedes the input position of the syllable nodes in COPY SET 1 (i.e., the V nodes in the input), or when the input position of the V nodes in COPY SET 0 (i.e., the V nodes in the input) is at the same position as the input position of the syllable nodes COPY SET 1 (i.e., the V nodes in the

input). This association relation is indicated with the red lines in Figure 5.4. Lastly, the order of the elements in the input is preserved in the output. In this way, phonological structure building can be seen as an *interpretation* of a more basic structure.

$$\mathcal{A}_o^{0,1}(x,y) \stackrel{\text{def}}{=} (C_i(x) \wedge V_i(y) \wedge y \approx s(x)) \vee (V_i(x) \wedge V_i(y) \wedge y \approx x) \tag{5.6}$$

"Associate positions x in COPY SET 0 and y in COPY SET 1

iff x is labeled as C, y is labeled as V, and x precedes y, or

iff x is labeled as V, y is labeled as V, and they are at the same position in the input."

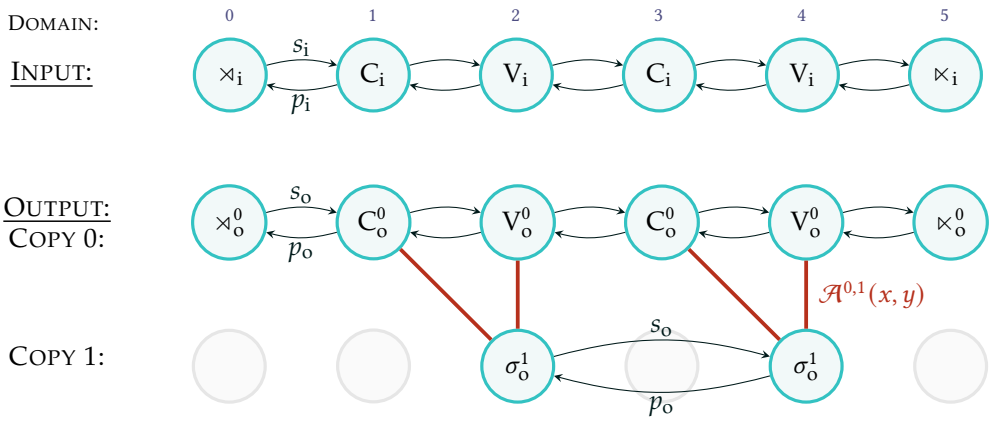


Figure 5.4: Association for syllable structural building from the input string ⋊CVCV⋉.

By using logical transduction, we can impose restrictions on the logic, which in turn impose restrictions on intonational typology. This also connects to computational complexity, which we can measure the intonational structure within the regular upper bound of phonology. Studies (e.g., Chandlee 2014, Chandlee & Heinz 2018) have proved that local phonological process is QF-definable, which is very restrictive in logic. Chandlee & Jardine (2019b) proved that even long-distance phonological processes are QF definable

with limited recursion. From this *local* property, we can establish a strong hypothesis for phonology and structure building.

In this chapter, we start with a hypothesis that *tone-TBU mappings in intonation are QF-definable*. By extending Jardine (2017) and Strother-Garcia & Heinz (2017), this chapter views intonational representations as additional structures imposed on an input string. For this, we will do some case studies on three different intonational types: American English for *head-prominence*, Seoul Korean for *edge-prominence*, and Tokyo Japanese for *head/edge-prominence* (and lexical pitch accent) patterns.

5.2 Intonation as a quantifier-free interpretation

This section shows how an intonational structure can be viewed as a QF logical interpretation of a metrical and prosodic structure. It first introduces basic ingredients of logical transductions to define a hierarchical intonational structure from an input string. Then, it shows how an intonation is analyzed using logical transduction with an example of a declarative intonation in American English.

5.2.1 Preliminaries

Signature Recall from 3.2.1 that a *signature* is a vocabulary that consists of symbols specified for properties and relations. For intonation, an input signature $\mathcal{S}_i$ is $\{\sigma, \sigma^*, \ltimes_\varphi, \rtimes_\varphi, \ltimes_\iota, \rtimes_\iota, p, s, p^*, s^*, \prec\}$. An output signature is $\mathcal{S}_o$ is $\{\sigma, \sigma^*, \ltimes_\varphi, \rtimes_\varphi, \ltimes_\iota, \rtimes_\iota, p, s, p^*, s^*, T, T^*, \mathcal{A}\}$.

Unary relations Unary relations are indicated as $symbol(x)$, where position x is labeled with one of the following symbols from the input and output signatures:

- σ and σ^* stand for (non-prominent) syllables and prominent syllables, respectively;
- $\rtimes_\varphi$ and $\ltimes_\varphi$ for the left and right edge of an ip, respectively;
- $\rtimes_\iota$ and $\ltimes_\iota$ for the left and right edge of IP, respectively;
- T for non-starred tones;
- T^* for starred tones.

Binary relations Binary relations are represented as $symbol(x, y)$, where the relation between the positions x and y is specified with this symbol:

- $\mathcal{A}$ stands for an association relation between a tone and a TBU;
- $\prec$ stands for a precedence relation to determine the order between the elements. This is only used to preserve the input order in the output.

Functions Two types of ordering functions are used in this chapter:

- p and s stand for the *immediate* predecessor and successor functions, respectively, which operate over all positions in the domain;
- p^* and s^* stand for the *starred* predecessor and successor functions, respectively, which only operate over the set of positions labeled as σ^* in the domain.

The immediate predecessor p and successor s functions are used to define the ordering relations between *any* elements in the domain. Notably, special starred predecessor p^* and

successor s^* functions are used to define the ordering relations between *particular prosodic elements* such as metrically strong syllables (e.g., σ^*) and phrasal boundaries (e.g., $\ltimes_\iota$, $\rtimes_\iota$, $\ltimes_\varphi$, $\rtimes_\varphi$). For this, a tier-based representation using a metrical grid is assumed to refer to the ordering of a particular tier from a set of multiple tiers. This assumption draws on the metrical grid representation for stress in Liberman (1975) and Liberman & Prince (1977).

Table 5.1 shows that the p function operates not only on the syllables but also on the prominent syllables and phrasal boundaries. For example, for 'o' in 'ɔrəndʒ', $\sigma(p(x))$ is true since x's preceding element is a non-starred syllable on the first tier, which is 'ən'. In contrast, the p^* function operates only on the prominent syllables and phrasal boundaries. For instance, for 'o' in 'ɔrəndʒ', $\ltimes_\varphi(p^*(x))$ is true since x's preceding element is an ip-boundary on the second tier, which is $\ltimes_\varphi$.

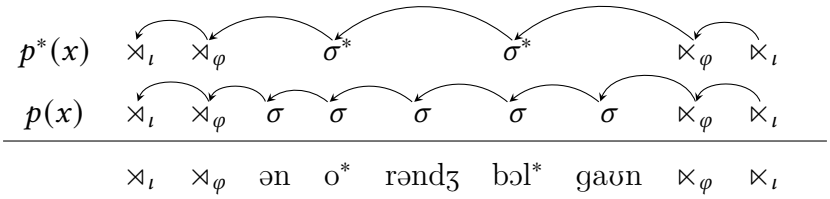


Table 5.1: The p and p^* functions based on a tier-based representation.

Order preservation When determining the order of elements within a copy set and across copy sets, this chapter follows Chandlee & Jardine (2019b) to preserve the input order to the output. However, this order works independently for TBU and tonal copy sets. For the TBU copy set (COPY SET 0), the order of the elements is exactly the same as the input. But for the tonal copy set (COPY SET 1-k), the order of the elements *within a copy set* basically follows the input order, whereas the order of the elements *across copy sets*

follows the hierarchical ordering between the copy sets, if the elements are at the same position. A full version of these orderings is illustrated in Figure 5.5.

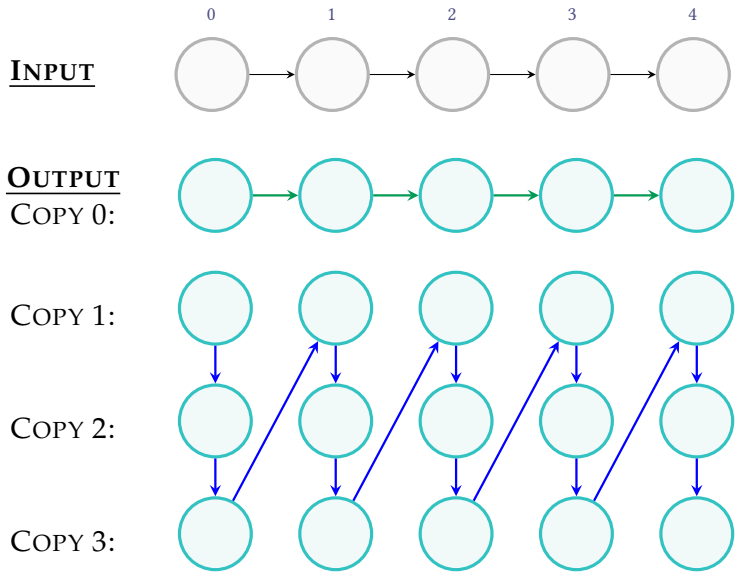


Figure 5.5: The order between the elements independently for TBU and tonal copy sets following Chandlee & Jardine (2019b).

For COPY SET 0, the order between the elements is identical to that in the input, as defined in $p_o^0(x) \stackrel{\text{def}}{=} p_i(x)$. As shown in Figure 5.5, the green arrows in COPY SET 0 preserve the input ordering.

For COPY SET 1 to k, if two elements x and y are in different positions *with the same copy set*, the input order ($m < n$) is preserved in the output ($x^m < y^n$). If the elements x and y are at the same position ($q \approx r$) *across the copy sets*, their order is determined by the order of the copy sets ($1 < 2 < \cdots < k$), where the position in COPY SET q precedes the same position in COPY SET r if $q < r$. The blue arrows in Figure 5.5 indicate the ordering of elements across COPY SET 1-3. Figure 5.6 illustrates the order preservation with an American English example.

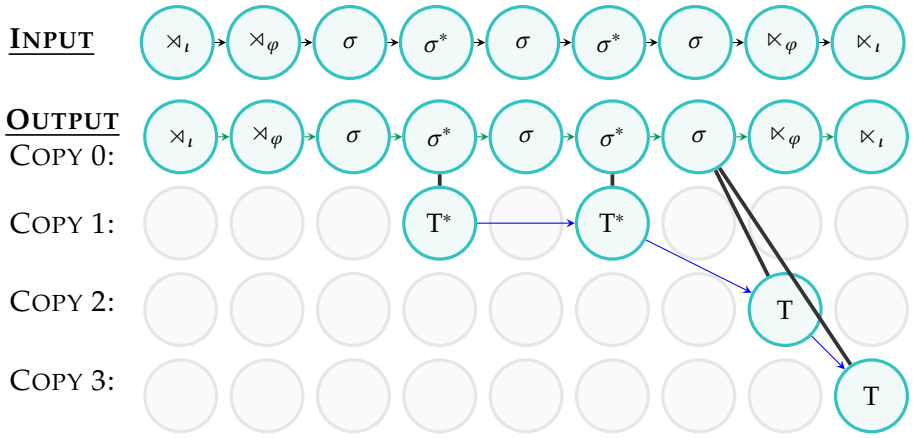


Figure 5.6: A graph illustrating the order preservation with an American English example.

The independent ordering of the TBU tier versus the tonal tier is theoretically meaningful, given that both are independent but autonomous tiers within AM theory.

5.2.2 Intonational tranductions

One of the main ideas in defining discrete tonal targets of intonation using logical transductions is that melodies in intonation are viewed as *literal copies* of prosodic and metrical elements. That is, the discrete tonal targets are *direct interpretation* of starred syllables and phrasal boundaries. These tonal targets are realized as a melody by being associated with their adjacent TBUs. The transductions from an input string to an output hierarchical intonational structure offer a theoretically informed and computationally restrictive perspective on the analysis of intonation.

Intonational transductions consist of two transductions: *melodic transduction* and *meaning transduction*. In the *melodic transduction*, the discrete tonal targets of intonation are computed by structurally defining the positions of prosodic elements with unspecified tones (i.e. Ts). Then, they are associated locally with their TBUs.

In the *meaning transduction*, the unspecified Ts can be filled with any combination of tones based on the intonational meanings (e.g., declarative, interrogative). The final output is a single complete intonational melody, since the QF transductions are *closed* under composition as long as they do not bring in arbitrary deletions or insertions (Chandlee & Lindell 2021).[7] The whole process of computing intonation using logical transductions is provided in Figure 5.7.

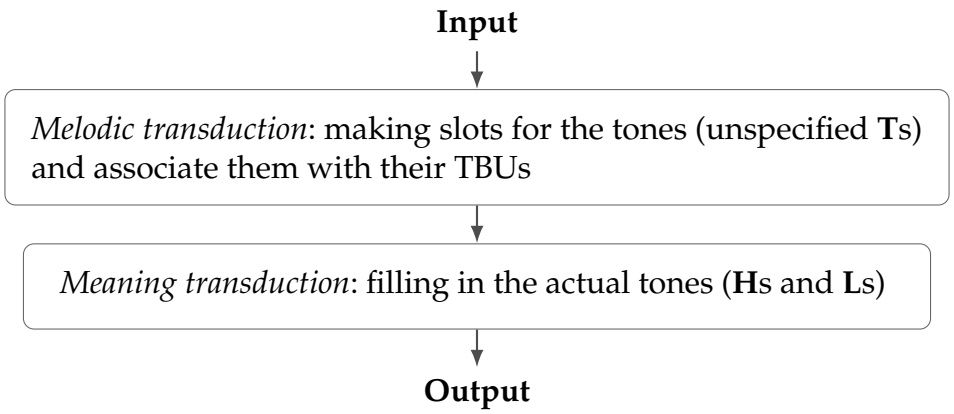


Figure 5.7: Intonational transductions for defining intonation: *melodic* and *meaning* transductions

5.2.3 An example analysis

This section briefly shows how logical transductions work with an example from American English provided in Figure 5.8. The detailed steps of transductions and their formal definitions of each language are discussed in the following sections.

[7]The detailed discussion is found in Chandlee & Lindell (2021). They claim the closure property of QF transductions, since the QF transductions work only when the input structure is maintained in the output structure, without deleting or adding any elements in the structure, but only replacing the properties of the elements in the structures. Therefore, this transduction will be closed under composition.

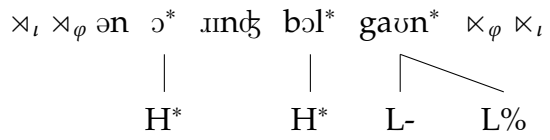


Figure 5.8: Autosegmental representation of a declarative intonational pattern in American English (H* H* L- L%).

In Figure 5.8, the string tier consists of TBUs as well as phrasal boundaries ($\rtimes_\iota/\ltimes_\iota$, $\rtimes_\varphi/\ltimes_\varphi$), while the tonal tier consists of the tonal sequences, H* H* L- L%. Two pitch accents, H*s, are linked to the accented syllables, ɔ and *bɔl*. L phrasal tone is associated with the last syllable of an ip, *gaʊn*. L boundary tone is linked to the last syllable of an IP, *gaʊn*. Importantly, not all TBUs are associated with tones, but only the *heads* of a constituent or phrasal *edges* in the utterance are associated with pitch accents or phrasal and boundary tones.

By using logical transductions, the output intonational tones are defined by the properties and relations in the input structure, based on logical formulas provided in Table 5.2. The transductions are graphically presented in Figure 5.9.

a. COPY 0	$\sigma_{\mathrm{o}}^{0}(x) \stackrel{\mathrm{def}}{=} \sigma_{\mathrm{i}}(x)$ $\quad \sigma^{*}{}_{\mathrm{o}}^{0}(x) \stackrel{\mathrm{def}}{=} \sigma^{*}{}_{\mathrm{i}}(x)$ $\ltimes_{\varphi}{}_{\mathrm{o}}^{0}(x) \stackrel{\mathrm{def}}{=} \ltimes_{\varphi_{\mathrm{i}}}(x)$ $\quad \ltimes_{\iota}{}_{\mathrm{o}}^{0}(x) \stackrel{\mathrm{def}}{=} \ltimes_{\iota\mathrm{i}}(x)$ $\rtimes_{\varphi}{}_{\mathrm{o}}^{0}(x) \stackrel{\mathrm{def}}{=} \rtimes_{\varphi_{\mathrm{i}}}(x)$ $\quad \rtimes_{\iota}{}_{\mathrm{o}}^{0}(x) \stackrel{\mathrm{def}}{=} \rtimes_{\iota\mathrm{i}}(x)$
b. COPY 1	$\mathrm{T}^{*}{}_{\mathrm{o}}^{1}(x) \stackrel{\mathrm{def}}{=} \sigma^{*}{}_{\mathrm{i}}(x)$
c. COPY 2	$\mathrm{T}_{\mathrm{o}}^{2}(x) \stackrel{\mathrm{def}}{=} \ltimes_{\varphi_{\mathrm{i}}}(x)$
d. COPY 3	$\mathrm{T}_{\mathrm{o}}^{3}(x) \stackrel{\mathrm{def}}{=} \ltimes_{\iota\mathrm{i}}(x) \vee \rtimes_{\iota\mathrm{i}}(x)$
e. ASSOCIATION	$\mathcal{A}_{\mathrm{o}}{}^{0,1}(x,y) \stackrel{\mathrm{def}}{=} x \approx y$ $\mathcal{A}_{\mathrm{o}}{}^{0,2}(x,y) \stackrel{\mathrm{def}}{=} \sigma_{\mathrm{i}}(x) \wedge \ltimes_{\varphi_{\mathrm{i}}}(y) \wedge y \approx s(x)$ $\mathcal{A}_{\mathrm{o}}{}^{0,3}(x,y) \stackrel{\mathrm{def}}{=} (\sigma_{\mathrm{i}}(x) \wedge \rtimes_{\iota\mathrm{i}}(y) \wedge y \approx p(p(x))) \vee (\sigma_{\mathrm{i}}(x) \wedge \ltimes_{\iota\mathrm{i}}(y) \wedge y \approx s(s(x)))$

Table 5.2: The definitions of intonation in American English.

First, in the *melodic transduction*, we build a hierarchical intonational structure with TBUs, unspecified tones (*T*s), and their associations, as shown in Figure 5.9a. In the first copy (COPY SET 0), all the elements in the input (i.e., σ, σ^*, $\rtimes_\varphi$, $\ltimes_\varphi$, $\rtimes_\iota$, $\rtimes_\iota$) are first copied to make TBUs, as defined in Table 5.2a. The rest of the copies (COPY SET 1-3) are defined with *T*s to make melodic slots. Specifically, only starred tones (i.e., pitch accents) in COPY SET 1 are directly defined from the input starred syllables (i.e., σ^*), as defined in Table 5.2b. The phrasal and boundary tones in COPY SET 2 and COPY SET 3 are directly defined based on ip boundaries (i.e., $\rtimes_\varphi$, $\ltimes_\varphi$) and IP boundaries (i.e., $\rtimes_\iota$, $\ltimes_\iota$), respectively, as defined in Table 5.2b-c.

To generate a complete melody, the tonal elements in COPY SET 0 are associated with the TBU elements in COPY SET 0, as defined in Table 5.2e. Their associations are defined *locally* only using *p* and *s* functions, demonstrating the QF nature of intonation transduction. The order of the elements in COPY SET 0 and COPY SET 0 is preserved respectively from the order of the input elements.

Then, in the *meaning transduction*, a melodic pattern (i.e., a sequence of *T*s) is specified with *H*s and *L*s for the declarative meaning (H*H*L-L%), as shown in Figure 5.9a.

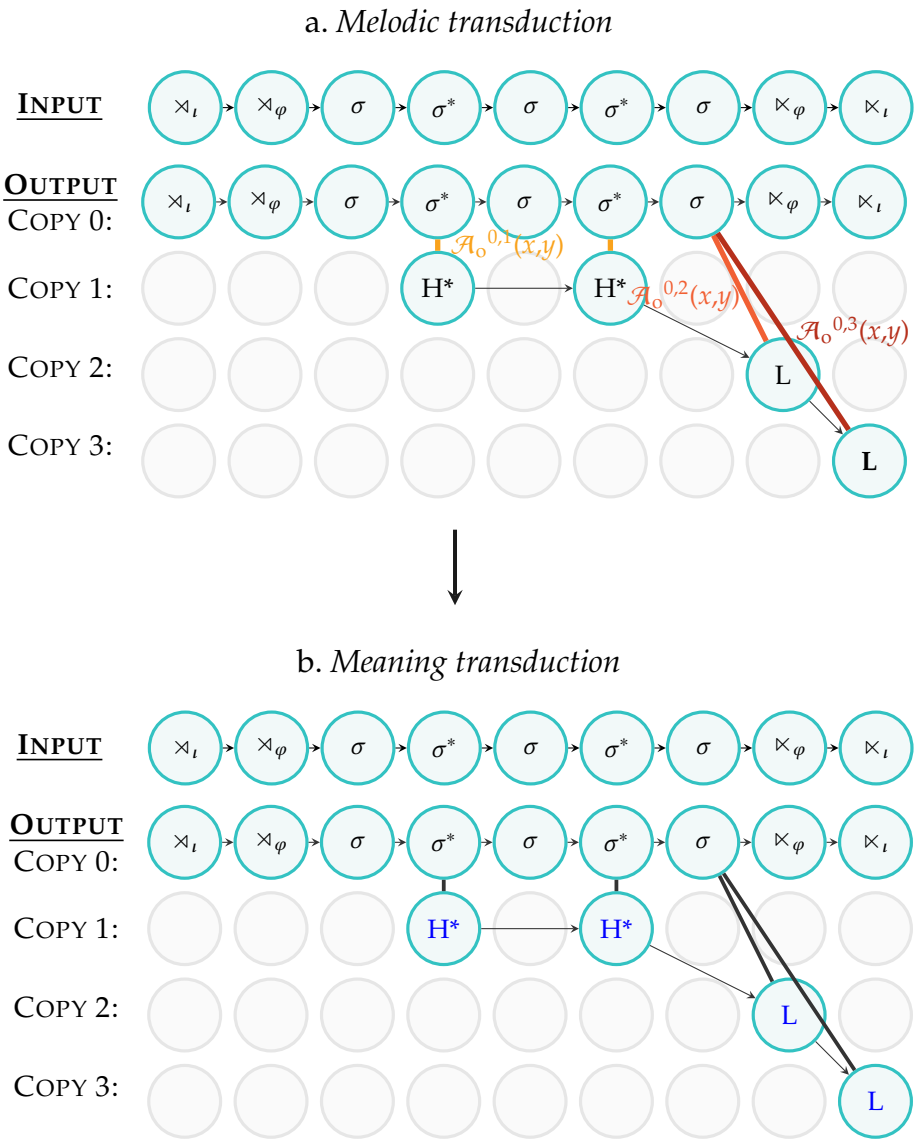


Figure 5.9: An overview of two transductions for intonation in American English.

5.3 Case studies

This section analyzes three distinct intonational systems, *head-prominence* (post-lexical), *edge-prominence* (post-lexical), and *lexical pitch accent* (head/edge-prominence), by looking at American English, Seoul Korean, and Tokyo Japanese as their test cases. Basic in-

tonational patterns for each type are first described, and then analyzed using *melodic* and *declarative transductions*.

5.3.1 American English

5.3.1.1 Basic intonational pattern

The first intonational type analyzed in this chapter is a *head-prominence* intonational language. In this system, pitch accents are realized at the *head* of a prosodic constituent, specifically at the *metrically strong positions* in a phrase (Jun 2006b, 2014, 2025). American English is one of the languages that employs the head-prominence intonational system (Beckman & Pierrehumbert 1986, Pierrehumbert 1980).

In American English, pitch accents (e.g., H^*, L^*, $L+H^*$, L^*+H, $H+L^*$, H^*+L) are realized on accented syllables (σ^*) within an ip (φ). Phrasal tones (e.g., H-, L-) occur at the right edge (i.e., the final syllable) of an ip, while boundary tones (e.g., H%, L%) are realized at the right edge (i.e., the final syllable) of an IP (ι). One or more ips are hierarchically organized into an IP, one or more of which recursively form an IP. The prosodic structure of American English is provided in Figure 5.10.

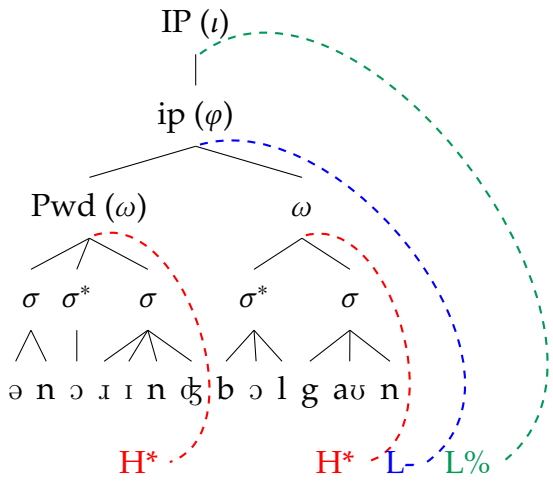


Figure 5.10: The prosodic structure of American English (Beckman & Pierrehumbert 1986, Pierrehumbert 1980).

One of the declarative intonational patterns, H* H* L- L%, in American English is analyzed mathematically using logical transductions. An example test case is provided in Figure 5.11. This section only includes an IP that consists of an ip, but the logical transductions of an IP with multiple ips are provided in Appendix A.

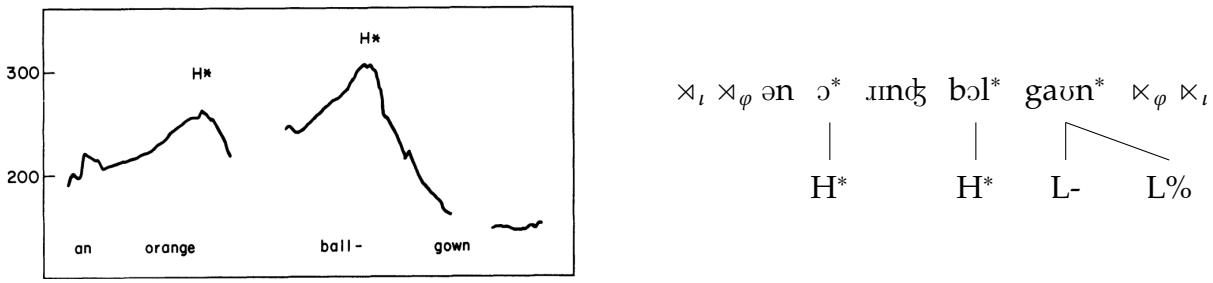


Figure 5.11: A declarative intonation in American English (Reprinted from Beckman & Pierrehumbert 1986.) and its autosegmental representation.

This basic intonational pattern of a declarative in American English first undergoes *melodic transduction* by creating tonal slots with unspecified Ts. Then, an actual melodic pattern with Hs and Ls is inserted into those tonal slots via *declarative transduction*.

5.3.1.2 Melodic transduction

Step 1: Copying In the first copy set (COPY SET 0), all of the elements in the input string are copied in the output, as provided in the formulas (5.7 - 5.12). This step is just like an identity mapping between the input and the output to make TBUs. The formulas are visually presented in the graph in Figure 5.12.

$$\sigma_o^0(x) \stackrel{\text{def}}{=} \sigma_i(x) \tag{5.7}$$

"Position x in COPY SET 0 is labeled σ iff x is labeled σ in the input."

$$\sigma_o^{*0}(x) \stackrel{\text{def}}{=} \sigma_i^*(x) \tag{5.8}$$

"Position x in COPY SET 0 is labeled σ^* iff x is labeled σ^* in the input."

$$\ltimes_{\varphi_o}{}^0(x) \stackrel{\text{def}}{=} \ltimes_{\varphi_i}(x) \tag{5.9}$$

"Position x in COPY SET 0 is labeled $\ltimes_\varphi$ iff x is labeled $\ltimes_\varphi$ in the input."

$$\rtimes_{\varphi_o}{}^0(x) \stackrel{\text{def}}{=} \rtimes_{\varphi_i}(x) \tag{5.10}$$

"Position x in COPY SET 0 is labeled $\rtimes_\varphi$ iff x is labeled $\rtimes_\varphi$ in the input."

$$\ltimes_{\iota_o}{}^0(x) \stackrel{\text{def}}{=} \ltimes_{\iota_i}(x) \tag{5.11}$$

"Position x in COPY SET 0 is labeled $\ltimes_\iota$ iff x is labeled $\ltimes_\iota$ in the input."

$$\rtimes_{\iota_o}{}^0(x) \stackrel{\text{def}}{=} \rtimes_{\iota_i}(x) \tag{5.12}$$

"Position x in COPY SET 0 is labeled $\rtimes_\iota$ iff x is labeled $\rtimes_\iota$ in the input."

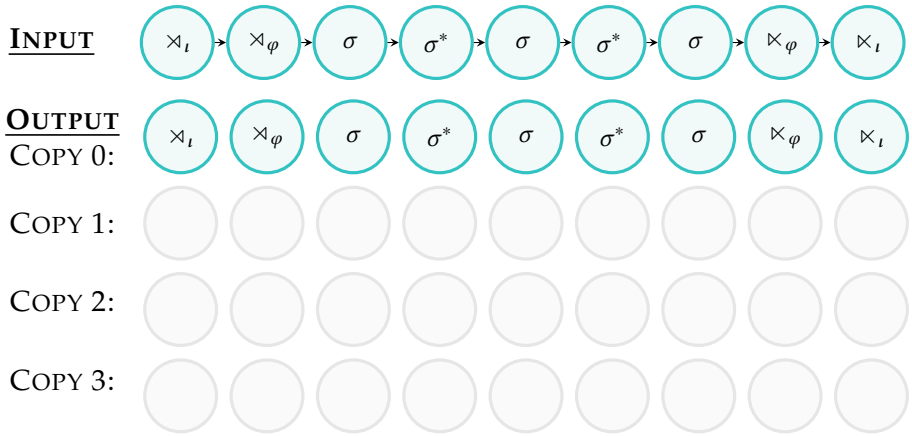


Figure 5.12: A graph illustrating COPY SET 0 for American English intonation based on the formulas (5.7 - 5.12).

As for the rest of the copy sets (COPY SET 1-3), only starred syllables and boundaries in the input are copied and *interpreted* as tones.

In the second copy set (COPY SET 1), the starred tones (i.e., pitch accents) are defined from the starred syllables (i.e., accented syllables) in the input string, as provided in the formula (5.13) and illustrated in Figure 5.13. Importantly, this COPY SET 1 shows the *head-prominence* characteristics of American English, where the starred syllables in the input are directly *interpreted* as the starred tones in the output.

$$T^{*1}_{o}(x) \stackrel{\text{def}}{=} \sigma^{*}_{i}(x) \tag{5.13}$$

"Position x in COPY SET 1 is labeled T^* iff x is labeled σ^* in the input."

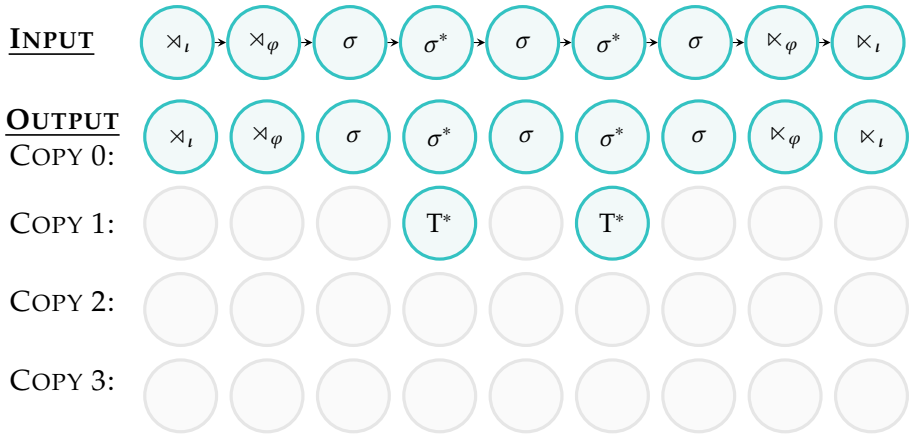


Figure 5.13: A graph illustrating COPY SET 1 for American English intonation based on the formula (5.13).

In the third copy set (COPY SET 2), the phrasal tones are defined based on the right edge of an ip boundary in the input, as provided in the formula (5.14) and illustrated in Figure 5.14.

$$T^2_o(x) \stackrel{\text{def}}{=} \ltimes_{\varphi_i}(x) \tag{5.14}$$

"Position x in COPY SET 2 is labeled T iff x is labeled $\ltimes_\varphi$ in the input."

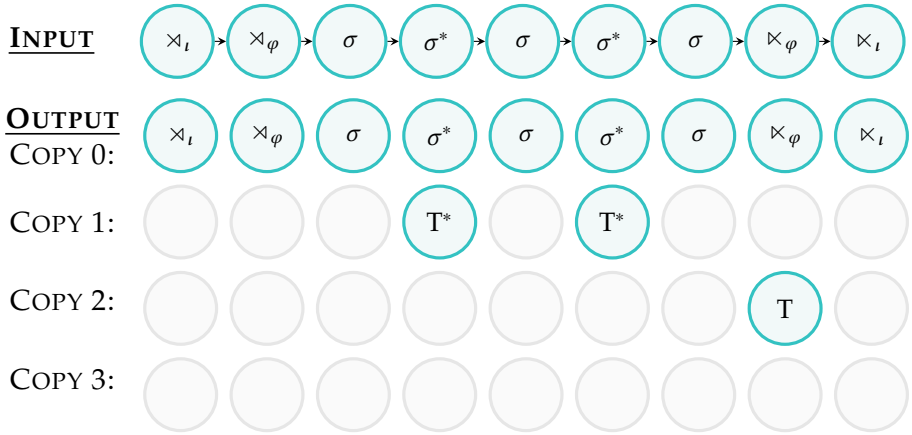


Figure 5.14: A graph illustrating Copy Set 2 for American English intonation based on the formula (5.14).

In the fourth copy set (COPY SET 3), the boundary tones are defined from both edges of an IP boundary, as provided in the formula (5.15). This formula includes the cases where there is an initial boundary tone. However, our example (H* H* L-L%) only has a final boundary tone, so the boundary only at the right edge is copied and interpreted as the boundary tone, as shown in Figure 5.15.

$$T_o^3(x) \stackrel{\text{def}}{=} \rtimes_{\iota i}(x) \vee \ltimes_{\iota i}(x) \tag{5.15}$$

"Position x in COPY SET 3 is labeled T iff x is labeled $\rtimes_\iota$ or $\ltimes_\iota$ in the input."

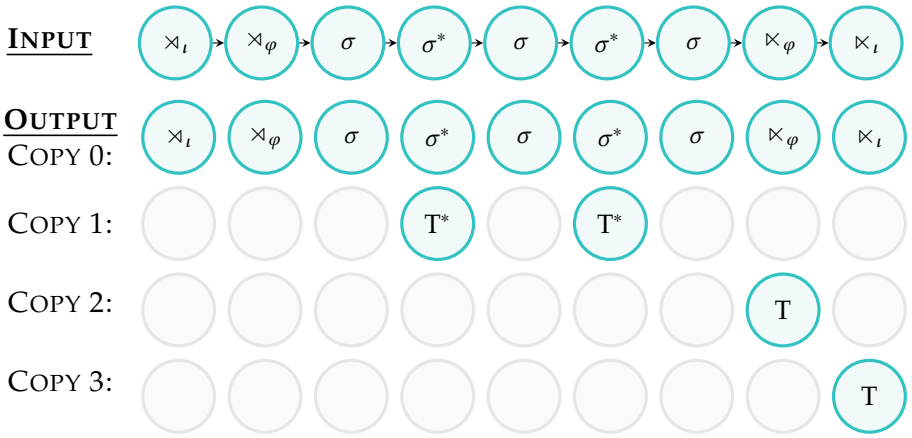


Figure 5.15: A graph illustrating COPY SET 3 for American English intonation based on the formula (5.15).

Step 2: Tone-TBU association After creating four copy sets, the tones can be connected with the tones with the TBUs, as shown in Figure 5.16, using the formulas in (5.16 - 5.18). Importantly, the association relations in the output $\mathcal{A}_o$ link the tones in each copy set with the TBUs in the first copy set, if and only if the inputs of all the copy sets are in *local relations* using only the p or s function. This local property demonstrates a QF characteristic of tone-TBU association in American English intonation.

$\mathcal{A}_o{}^{0,1}(x, y)$ associates the pitch accents in COPY SET 1 with their TBUs in COPY SET 0 if and only if the input position of the pitch accents is the same as the input position of the TBUs, as defined in the formula (5.16).

But for the phrasal and boundary tones, the tones are linked *locally* with the TBUs using *p* or *s* function. That is, in the formula (5.17), $\mathcal{A}_o{}^{0,2}(x, y)$ associates the ip-final tones in COPY SET 2 with the phrase-final syllables on the right edge in COPY SET 0, if and only if the input position of the ip-final tone (i.e., the ip boundary in the input) *immediately follows* the input position of the syllable (i.e., the syllable in the input).

In the formula (5.18), $\mathcal{A}_o{}^{0,3}(x, y)$ associates the boundary tones in COPY SET 3 with the syllables on both edges in COPY SET 0, with two conditions: first, if and only if the input position of the tone at the left edge (i.e., the left IP boundary) is two nodes before the input position of the phrase-initial syllable (i.e., the syllable in the input), or second, if and only if the input position of the tone at the right edge (i.e., the right IP boundary) is two nodes after the input position of the phrase-final syllable (i.e., the syllable in the input).

$$\mathcal{A}_o^{0,1}(x) \stackrel{\text{def}}{=} x \approx y \tag{5.16}$$

"Position x in COPY SET 0 and position y in COPY SET 1 are associated iff they are at the same position in the input."

$$\mathcal{A}_o^{0,2}(x) \stackrel{\text{def}}{=} \sigma_i(x) \wedge \ltimes_{\varphi i}(y) \wedge y \approx s(x) \tag{5.17}$$

"Position x in COPY SET 0 and position y in COPY SET 2 are associated iff x is labeled σ, y is labeled $\ltimes_\varphi$, and y follows x in the input."

$$\mathcal{A}_o^{0,3}(x) \stackrel{\text{def}}{=} (\sigma_i(x) \wedge \rtimes_{\iota i}(y) \wedge y \approx p(p(x))) \vee (\sigma_i(x) \wedge \ltimes_{\iota i}(y) \wedge y \approx s(s(x))) \tag{5.18}$$

"Position x in COPY SET 0 and position y in Copy Set 3 are associated iff x is labeled σ, y is labeled $\rtimes_\iota$, and y is two nodes before x in the input, or x is labeled σ, y is labeled $\ltimes_\iota$, and y is two nodes after x in the input."

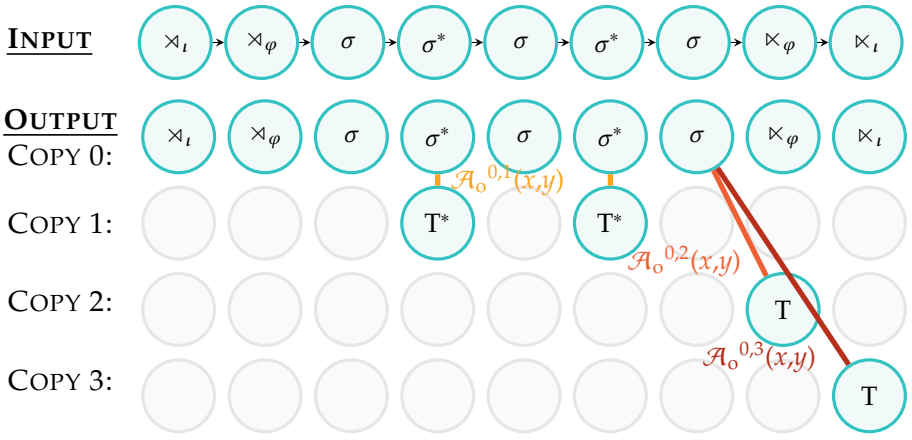


Figure 5.16: A graph illustrating the binary association relations $\mathcal{A}_o$ between tones and TBUs for American English intonation based on the formulas (5.16 - 5.18).

Step 3: Defining the order in melodies The last step of the *melodic transduction* is determining the ordering relations between the elements. The ordering relations are defined separately for the TBU and tonal tiers, while both are still perserving the input order, following Chandlee & Jardine (2019b).

As for the TBU tier (COPY SET 0), the green arrows in Figure 5.17 indicate the ordering relation between the elements, which are the same as the input ordering relation.

Independent of the TBU tier, the order of the elements across the tonal tiers (COPY SET 1 - 3) is indicated by the blue arrows in Figure 5.17. That is, the two pitch accents are ordered sequentially, and then the ip-final and IP-final tones are ordered, such that all the tones are combined to be realized as a melody, T* T* T- T%.

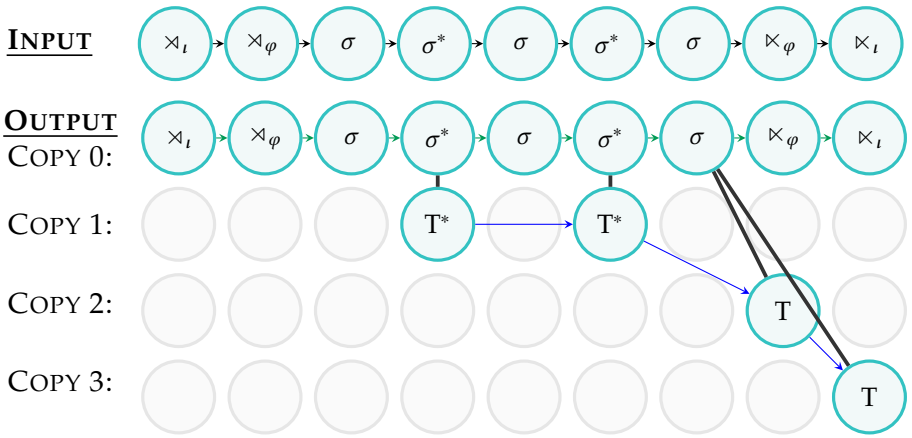


Figure 5.17: A graph illustrating the ordering relations of the TBU and tonal tiers for American English intonation following Chandlee & Jardine (2019b).

5.3.1.3 Declarative transduction

Now that the hierarchical structure of intonation is constructed using the copy sets via a melodic transduction, the unspecified tones (i.e., T*s and Ts) can be specified with actual H and L tones via a *declarative transduction*. Recall that the declarative intonation used for

the test case is a H* H* L- L% melody.

An input signature $\mathcal{S}_{\text{i}}$ for the declarative transduction is $\{\sigma, \sigma^*, \ltimes_\varphi, \ltimes_\iota, \rtimes_\varphi, \rtimes_\iota, T, T^*\}$, while the output signature $\mathcal{S}_{\text{o}}$ for the declarative transduction is $\{\sigma, \sigma^*, \ltimes_\varphi, \ltimes_\iota, \rtimes_\varphi, \rtimes_\iota, H, L, H^*, L^*, L + H^*, L^* + H, H^* + L, H + L^*\}$. Using the formulas (5.19 - 5.20), the T*s and Ts are interpreted into Hs and Ls. This declarative transduction is visually presented in Figure 5.18.

$$H^*_o(x) \stackrel{\text{def}}{=} T^*_i(x) \tag{5.19}$$

"Position x in the output is labeled H^* iff x is labeled T^* in the input."

$$L_o(x) \stackrel{\text{def}}{=} T_i(x) \tag{5.20}$$

"Position x in the output is labeled L iff x is labeled T in the input."

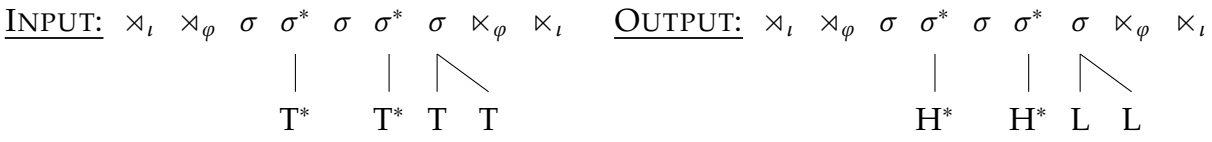


Figure 5.18: Declarative transduction for the H* H* L-L% melody in American English.

Via both the *melodic* and *declarative transductions*, a complete structure of a declarative intonation in American English can be constructed from the input string, as shown in Figure 5.19.

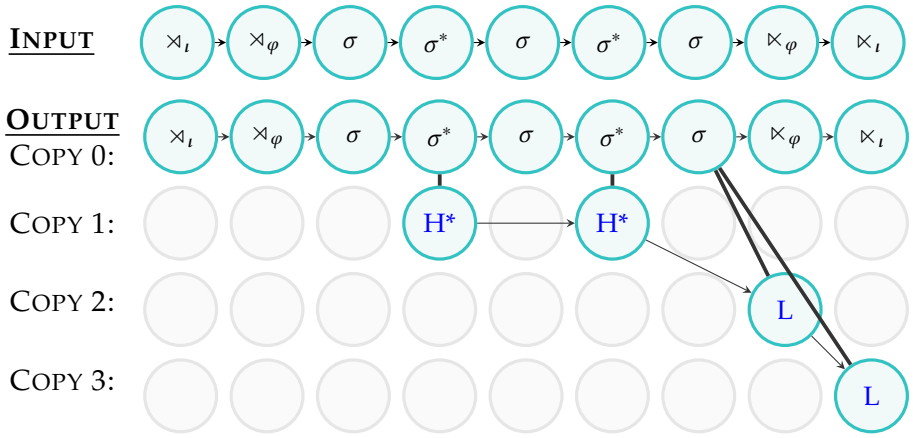


Figure 5.19: The final output from the two transductions, *melodic* and *declarative transductions*, in American English.

5.3.1.4 Summary

Results showed that the intonational pattern in American English can be viewed as a QF logical interpretation of a metrical and prosodic structure. The melodies in the output were *literal copies* of starred syllables and boundaries in the input structure. Crucially, copying the starred syllables was able to capture the *head-prominence* characteristic of American English intonation, showing that the pitch accents were the *melodic interpretations* of the heads of the prosodic unit – starred syllables. Also, the tone-TBU associations were defined *locally* from the input structure without using any quantifiers. That is, the predecessor and successor functions were able to compute the association relations. Therefore, we can conclude that the head-prominence intonational pattern in American English is a QF logical interpretation of a metrical and prosodic structure.

5.3.2 Seoul Korean

5.3.2.1 Basic intonational pattern

The second intonational type is an *edge-prominence* intonational system, in which specific tonal patterns signal the edges (boundaries) of a prosodic constituent (Jun 2006b, 2014, 2025). Seoul Korean—one of the edge-prominence languages—marks an Accentual Phrase (AP; α) with either LH...LH or HH...LH tonal patterns (Jun 2006a). When the AP-initial consonant is [+aspirated] or [+tense], the tonal sequence of an AP is HH...LH, otherwise LH...LH. That is, the first LH or HH edge tones are realized at the left edge of an AP, while the other LH edge tones are aligned at the right edge. One or more APs are hierarchically organized into an IP (ι), and one or more IPs, in turn, are organized into an IP. At the right edge of an IP, a boundary tone (e.g., L%, H%, LHL%, HLH%, LHLH%, HLHL%, LHLHL%) is realized on the phrase-final syllable, instead of the AP-final H tone. The prosodic structure of Seoul Korean is provided in Figure 5.20.

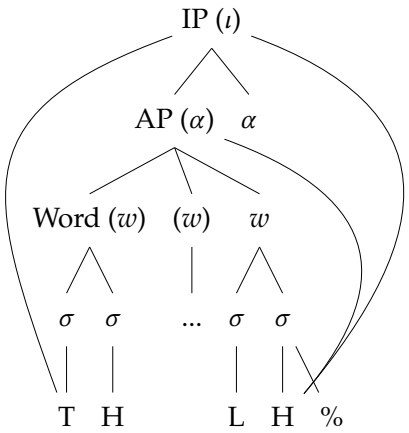


Figure 5.20: The prosodic structure of Seoul Korean. Redrawn from Jun (2006a).

One of the declarative intonational patterns, LHLHa LLHa LHLL%, in Seoul Korean

is provided on the left panel in Figure 5.21. But as shown in the right panel in Figure 5.21, only the final AP that is nested in an IP, LHLL%, is used as an example for analyzing mathematically using logical transductions. The transduction of multiple APs within an IP, LHLHa LLHa LHLL%, is provided in Appendix A.

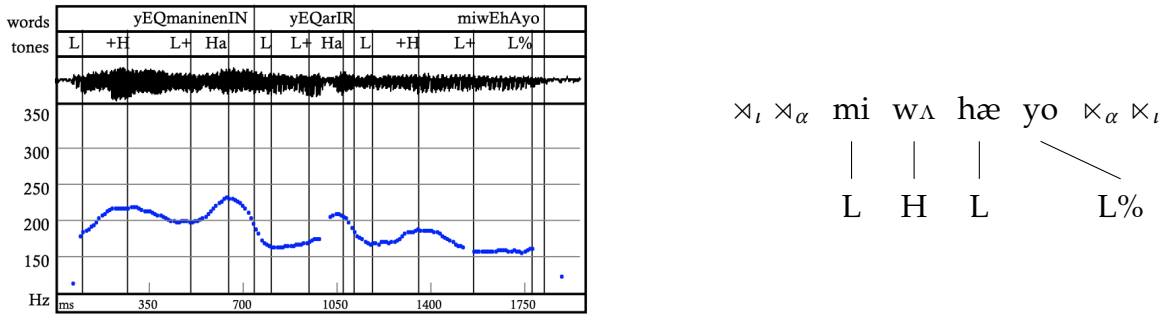


Figure 5.21: A declarative intonation in Seoul Korean (Reprinted from Jun 2006a.) and its autosegmental representation.

This basic intonational pattern of a declarative in Seoul Korean first undergoes *melodic transduction* by creating tonal slots with unspecified Ts. Then, an actual melodic pattern with Hs and Ls is inserted into those tonal slots via *declarative transduction*.

5.3.2.2 Melodic transduction

Step 1: Copying For the first copy (COPY SET 0), every input element is copied such that syllables and boundaries in the output is *interpreted* the same as those in the input, as provided in the formulas (5.21 - 5.26) and as illustrated in Figure 5.22.

Note that for Seoul Korean, the stiffness feature [+stiff] is used with syllables to represent aspirated or tense consonants in those syllables. In this way, we can refer to this featural information in the cases where HH...LH should be computed due to the aspi-

rated or tense consonants, as provided in the formula (5.22).

$$\sigma_o^0(x) \stackrel{\text{def}}{=} \sigma_i(x) \tag{5.21}$$

"Position x in COPY SET 0 is labeled σ iff x is labeled σ in the input."

$$\sigma_{F\,o}^{\ 0}(x) \stackrel{\text{def}}{=} \sigma_i(x) \tag{5.22}$$

"Position x in COPY SET 0 is labeled σ_F (F = [+stiff]) iff x is labeled σ in the input."

$$\ltimes_{\alpha o}^{\ 0}(x) \stackrel{\text{def}}{=} \ltimes_{\alpha i}(x) \tag{5.23}$$

"Position x in COPY SET 0 is labeled $\ltimes_\alpha$ iff x is labeled $\ltimes_\alpha$ in the input."

$$\rtimes_{\alpha o}^{\ 0}(x) \stackrel{\text{def}}{=} \rtimes_{\alpha i}(x) \tag{5.24}$$

"Position x in COPY SET 0 is labeled $\rtimes_\alpha$ iff x is labeled $\rtimes_\alpha$ in the input."

$$\ltimes_{\iota o}^{\ 0}(x) \stackrel{\text{def}}{=} \ltimes_{\iota i}(x) \tag{5.25}$$

"Position x in COPY SET 0 is labeled $\ltimes_\iota$ iff x is labeled $\ltimes_\iota$ in the input."

$$\rtimes_{\iota o}^{\ 0}(x) \stackrel{\text{def}}{=} \rtimes_{\iota i}(x) \tag{5.26}$$

"Position x in COPY SET 0 is labeled $\rtimes_\iota$ iff x is labeled $\rtimes_\iota$ in the input."

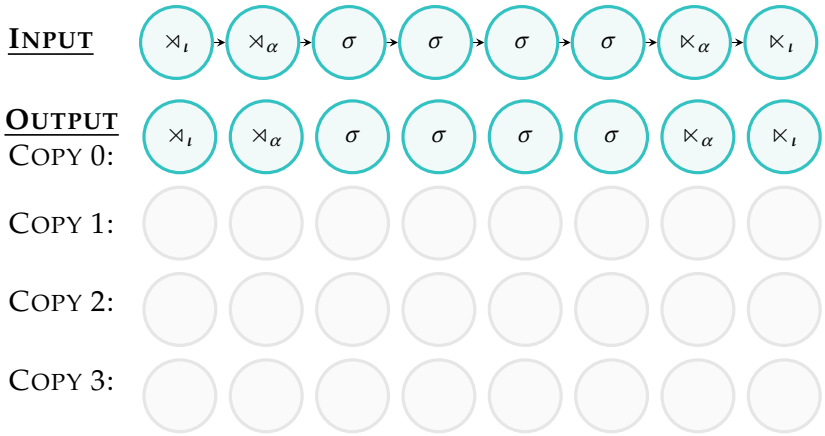


Figure 5.22: A graph illustrating COPY SET 0 for Seoul Korean intonation based on the formulas (5.21 - 5.26).

As for the rest of the copies (COPY SET 1-3), only *boundaries* in the input are copied and interpreted as tones, showing a crucial characteristic for the *edge-prominence* intonational system.

As defined in the formulas (5.27) and (5.28), the tones in the second and third copy sets (COPY SET 1 and 2) are defined from the AP boundaries at the left or right edge in the input. In the last copy set (COPY SET 3), the boundary tones are defined based on the right edge of an IP, as defined in the formula (5.29). The definition of both the phrasal and

boundary tones is visually presented in Figure 5.23.

$$T_o^1(x) \stackrel{\text{def}}{=} \rtimes_{\alpha i}(x) \vee \ltimes_{\alpha i}(x) \tag{5.27}$$

"Position x in COPY SET 1 is labeled T iff x is labeled $\rtimes_\alpha$ or $\ltimes_\alpha$ in the input."

$$T_o^2(x) \stackrel{\text{def}}{=} \rtimes_{\alpha i}(x) \vee \ltimes_{\alpha i}(x) \tag{5.28}$$

"Position x in COPY SET 2 is labeled T iff x is labeled $\rtimes_\alpha$ or $\ltimes_\alpha$ in the input."

$$T_o^3(x) \stackrel{\text{def}}{=} \ltimes_{\iota i}(x) \tag{5.29}$$

"Position x in COPY SET 3 is labeled T iff x is labeled $\ltimes_\iota$ in the input."

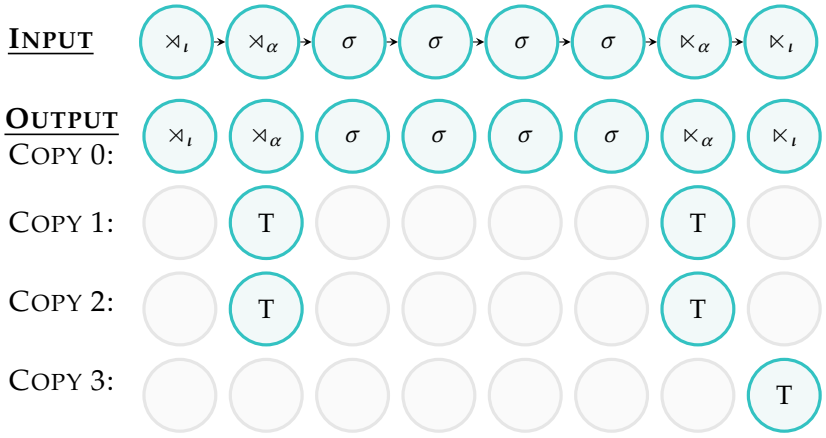


Figure 5.23: A graph illustrating COPY SET 1-3 for Seoul Korean intonation based on the formulas (5.27 - 5.29).

Step 2: Tone-TBU association Now, each tonal tier (COPY SET 1-3) is associated *locally* with the TBU tier (COPY SET 0), by only using the p or s functions. These assocations are based on the formulas (5.30 - 5.32) and visually presented in Figure 5.24. First, $\mathcal{A}_o{}^{0,1}(x, y)$ associates the phrasal tones in COPY SET 1 with their TBUs in COPY SET 0, if and only if the input position of a phrasal tone on the left edge (i.e., the left AP boundary in the

input) immediately precedes the input position of a syllable (i.e., the syllable in the input), or the input position of a phrasal tone on the right edge (i.e., the right AP boundary in the input) is two nodes after the input position of a syllable (i.e., the syllable in the input). That is, in COPY SET 1, the phrasal tone at the right edge is linked with the phrase-initial syllable, while that at the left edge is linked with the second-to-last syllable.

Similarly, $\mathcal{A}_o{}^{0,2}(x, y)$ associates the phrasal tones in COPY SET 2 with their TBUs in COPY SET 0, with two conditions: first, if and only if the input position of a left AP boundary (i.e., the left edge of an AP boundary in the input) is two nodes before the input position of a syllable (i.e., a syllable in the input); second, if and only if the input position of a right AP boundary (i.e., the right edge of an AP boundary in the input) is right after the input position of a syllable (i.e., a syllable in the input), but only when it is not followed by the input position of a right IP boundary (i.e., an IP boundary in the input). That is, the right-edged phrasal tone in COPY SET 2 is linked to the second syllable in an AP, while the left-edged phrasal tone in COPY SET 2 is linked to the last syllable in an AP, but only when it does not accompany the right-edged IP boundary.

Lastly, $\mathcal{A}_o{}^{0,3}(x, y)$ associates an boundary tone in COPY SET 3 with its TBU in COPY SET 0, if and only if the input position of the boundary tone (i.e., an IP boundary in the input) is two nodes after the input position of a syllable (i.e., the syllable in the input). That is, it links an IP boundary tone with the last syllable before an IP boundary. Not linking the AP-final phrasal tone in COPY SET 2 with the final syllable in COPY SET 0, but linking the final boundary tone in COPY SET 3 with the final syllable in COPY SET 0, expresses the boundary tone's overriding of the AP-final phrasal tone.

$$\mathcal{A}_o^{0,1}(x,y) \stackrel{\text{def}}{=} \sigma_i(x) \wedge (\rtimes_{\alpha i}(y) \wedge y \approx p(x)) \vee (\ltimes_{\alpha i}(y) \wedge y \approx s(s(x))) \tag{5.30}$$

"Positions x in COPY SET 0 and y in COPY SET 1 are associated
iff x is labeled as σ, with y that is labeled as $\rtimes_\alpha$, and y immediately precedes x, or
with y that is labeled as $\ltimes_\alpha$, and y is two nodes after x."

$$\mathcal{A}_o^{0,2}(x,y) \stackrel{\text{def}}{=} \sigma_i(x) \wedge (\rtimes_{\alpha i}(y) \wedge y \approx p(p(x))) \vee (\ltimes_{\alpha i}(y) \wedge y \approx s(x) \wedge \neg(\ltimes_{\iota i}(y) \wedge y \approx s(s(x))) \tag{5.31}$$

"Positions x in COPY SET 0 and y in COPY SET 2 are associated
iff x is labeled as σ, with y that is labeled as $\rtimes_\alpha$ and is two nodes before x,
or with y that is labeled as $\ltimes_\alpha$ and immediately precedes x,
but not with y that is labeled as $\ltimes_\iota$ and is two nodes after x in the input."

$$\mathcal{A}_o^{0,3}(x,y) \stackrel{\text{def}}{=} \sigma_i(x) \wedge \ltimes_{\iota i}(y) \wedge y \approx s(s(x)) \tag{5.32}$$

"Positions x in COPY SET 0 and y in COPY SET 3 are associated
iff x is labeled as σ, and y is labeled as $\ltimes_\iota$,
and y is two nodes after x in the input."

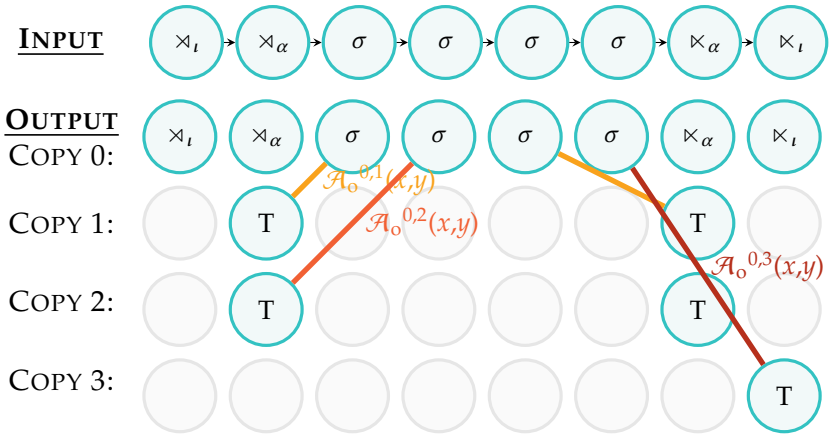


Figure 5.24: A graph illustrating the binary association relations $\mathcal{A}_o$ between tones and TBUs for Seoul Korean based on the formulas (5.30 - 5.32).

Step 3: Defining the order in melodies Due to the inherited precedence relations separately for the TBUs and the tonal tiers from the input, the ordering relations in the output are automatically defined, as shown in Figure 5.25. That is, the order of the elements in COPY SET 0 is the same as that in the input. As for the tonal tiers, two tones at the left edge of an AP are ordered sequentially, since COPY SET 1 precedes COPY SET 2 in terms of the ordering relation between the tonal tiers if the elements are at the same position. Similarly, two tones at the right edge of an AP are also ordered sequentially. Lastly, a boundary tone ends the melody.

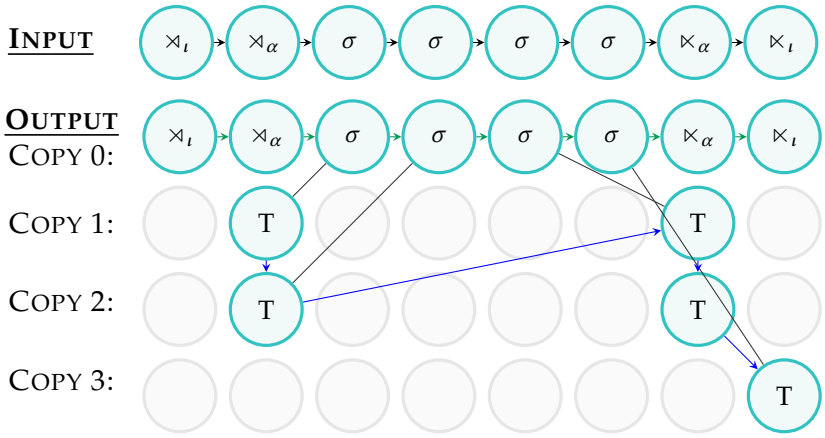


Figure 5.25: A graph illustrating the ordering relations of the TBU and tonal tiers for Seoul Korean intonation following Chandlee & Jardine (2019b).

5.3.2.3 Declarative transduction

After the melodic transduction, the unspecified tones (Ts) are filled with Hs and Ls via *declarative transduction* in Seoul Korean. Recall that the example case for Seoul Korean is the LHLL% tonal sequence. An input signature $\mathcal{S}_i$ for Seoul Korean declarative transduction is $\{\sigma, \ltimes_\alpha, \ltimes_\iota, \rtimes_\alpha, \rtimes_\iota, T\}$, while an output signature $\mathcal{S}_i$ is $\{\sigma, \ltimes_\alpha, \ltimes_\iota, \rtimes_\alpha, \rtimes_\iota, H, L\}$. Using the formulas (5.33) and (5.34), the Ts are specified with Ls and Hs, as shown in Figure 5.26.

$$H_o(x) \stackrel{\text{def}}{=} T_i(x) \wedge \rtimes_{\alpha i}(p(p(x))) \tag{5.33}$$

"Position x in the output is labeled H iff x is labeled T,

and an element that is two nodes before x is labeled $\rtimes_\alpha$ in the input."

$$L_o(x) \stackrel{\text{def}}{=} T_i(x) \wedge (\rtimes_{\alpha i}(p(x)) \vee \ltimes_{\alpha i}(s(s(x)))) \tag{5.34}$$

"Position x in the output is labeled L iff x is labeled T, x's preceding element is labeled $\rtimes_\alpha$,

and an element that is two nodes after x is labeled $\ltimes_\alpha$ in the input."

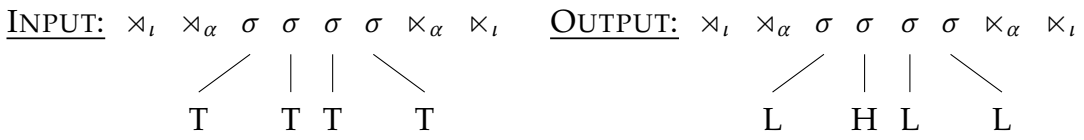


Figure 5.26: Declarative transduction for the LHLL% melody in Seoul Korean.

Via both the *melodic* and *declarative transductions*, a full structure of a declarative intonation in Seoul Korean can be constructed from the input string, as shown in Figure 5.27.

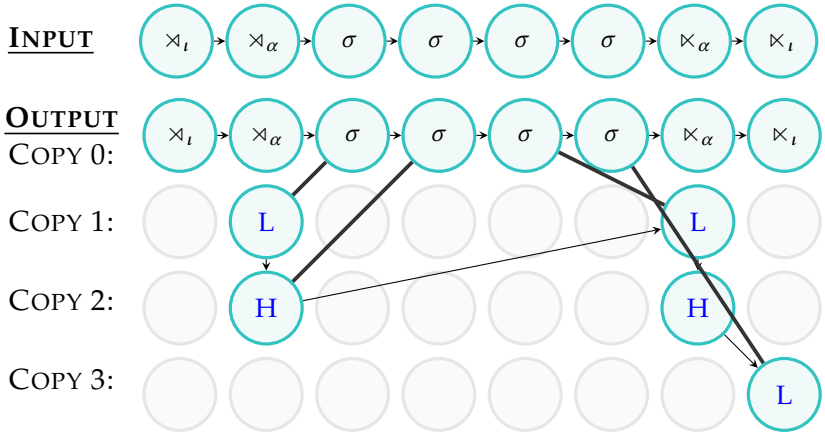


Figure 5.27: The final output of the *melodic* and *declarative transductions* in Seoul Korean.

5.3.2.4 Summary

Results showed that the intonational pattern in Seoul Korean can be defined using a logical interpretation of a metrical and prosodic structure. The melodies in the output were literal copies of *only boundaries* in the input structure. Unlike the head-prominence intonational pattern in American English, copying only boundaries was able to capture the *edge-prominence* characteristic of Seoul Korean intonation, showing that the edge tones were the *melodic interpretations* of the phrasal edges. Just like the local association in American English, the tone-TBU associations were defined *locally* from the input structure without using any quantifiers. Thus, we can conclude that the edge-prominence intonational pat-

tern in Seoul Korean is a QF logical interpretation of a metrical and prosodic structure.

5.3.3 Tokyo Japanese

5.3.3.1 Basic intonational pattern

The last intonational type is a *head/edge prominence* system, where tonal patterns are specified not only from the head of a prosodic constituent but also from the edge of a prosodic constituent (Jun 2006b, 2014, 2025). Tokyo Japanese—one of the head/edge prominence languages—is a lexical pitch accent language (Beckman 1988, Beckman & Pierrehumbert 1986, Venditti 2005), in which some tones are lexically specified, while the rest are post-lexical tones. It is known for a fixed pitch pattern that rises at the beginning of an AP and gradually declines towards the end of the AP. This pattern is closely related to the accentedness of lexical items in an AP. If the AP-initial lexical item is unaccented, an initial L boundary tone and a phrasal H tone are realized. But if the lexical item is accented, an initial L boundary tone and a lexically specified H tone in HL are realized.

A basic prosodic domain in Tokyo Japanese is an Accentual Phrase (AP), where at most one pitch accent is realized and boundaries are marked with a AP-initial phrasal H tone, an utterance-initial and an AP-final L% boundary tone[8]. The prosodic structure for Tokyo Japanese is provided in Figure 5.28.

[8]In, Beckman & Pierrehumbert (1986), a L% boundary tone, except for the utterance-initial boundary tone, is analyzed to belong to the right edge of an AP but realized at the first syllable of the following AP. This may be because the f0 value of the L% boundary tone was comparable to the reset f0 value, such that the boundary tone is considered not Lα, but L%.

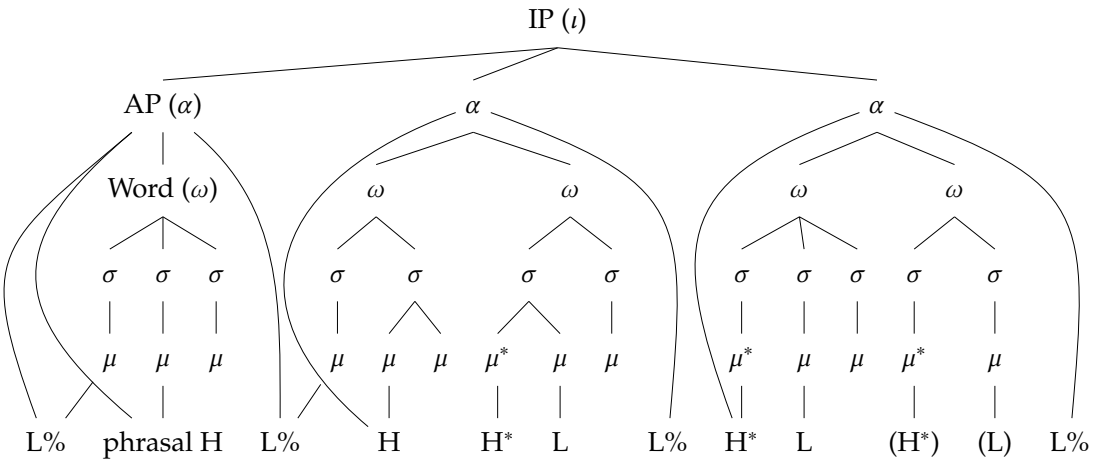


Figure 5.28: The prosodic structure in Tokyo Japanese. Redrawn from Beckman (1988). Non-starred H tones are phrasal H tones, and the parentheses indicate the deaccentuation.

The basic intonational patterns in Tokyo Japanese depend on the presence or absence of lexical pitch accent. When the first syllable of the first lexical item in an AP is *accented* (e.g., kágeboosi), H*L is associated to the first mora of the accented syllable. Usually, the H* tone is realized on the first mora, and the following L tone is realized on the second mora. Due to the realization of the lexical pitch accent on the first and second moras, a link between a phrasal H tone and the second mora is blocked, and an L% boundary tone in the preceding AP is associated with the final mora of the preceding AP, instead of being associated with the first mora of the AP.

When the first syllable of the first lexical item in an AP *unaccented* (e.g., toomórokoski), a phrasal H tone is usually linked to the second sonorant mora and L% boundary tone of the preceding AP is associated to the first mora of the following AP.

Lastly, an initial L% boundary tone is always realized at the utterance-initial position. A post-lexical rule deletes any accents after the first lexical accent in an AP, which is known as deaccentuation.

It is important to highlight that in lexical pitch accent languages, the lexical pitch accents specified from the input are maintained in the output, whereas other post-lexical melodies are only realized in the output.

One of the declarative intonational patterns, L%Ha L%HLL%, in Tokyo Japanese is provided on the left panel in Figure 5.29. For the purpose of explanation, the structure on the left panel in Figure 5.29 is analyzed using logical transductions. Therefore, only some APs are analyzed in this section, but the transductions of multiple phrases are provided in Appendix A.

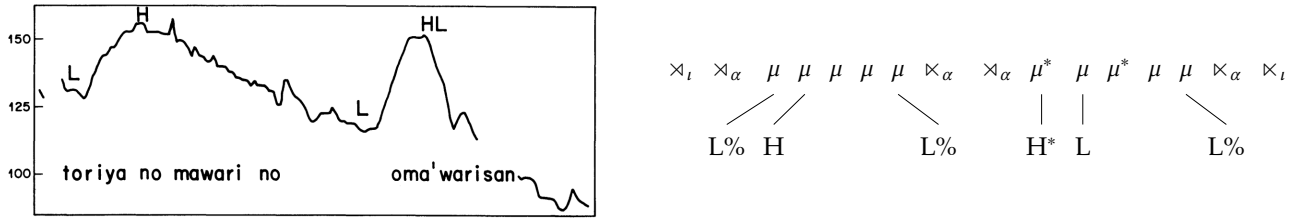


Figure 5.29: An f0 contour for a declarative in Tokyo Japanese (Reprinted from Beckman & Pierrehumbert 1986.) and its autosegmental representation.

This basic intonational pattern of a declarative in Tokyo Japanese first undergoes *melodic transduction* by creating tonal slots with unspecified Ts. Then, an actual melodic pattern with Hs and Ls is filled into those tonal slots via *declarative transduction*.

5.3.3.2 Melodic transduction

Step 1: Copying For the first copy (COPY SET 0), every element in the input is copied such that moras and boundaries in the output are interpreted the same as those in the input, as provided in the formulas (5.35-5.40) and illustrated in Figure 5.36.

$$\mu_o^0(x) \stackrel{\text{def}}{=} \mu_i(x) \tag{5.35}$$

"Position x in COPY SET 0 is labeled μ iff x is labeled μ in the input."

$$\mu_o^{*0}(x) \stackrel{\text{def}}{=} \mu_i^*(x) \tag{5.36}$$

"Position x in COPY SET 0 is labeled μ^* iff x is labeled μ^* in the input."

$$\ltimes_{\varphi_o}{}^0(x) \stackrel{\text{def}}{=} \ltimes_{\varphi_i}(x) \tag{5.37}$$

"Position x in COPY SET 0 is labeled $\ltimes_\varphi$ iff x is labeled $\ltimes_\varphi$ in the input."

$$\rtimes_{\varphi_o}{}^0(x) \stackrel{\text{def}}{=} \rtimes_{\varphi_i}(x) \tag{5.38}$$

"Position x in COPY SET 0 is labeled $\rtimes_\varphi$ iff x is labeled $\rtimes_\varphi$ in the input."

$$\ltimes_{\iota_o}{}^0(x) \stackrel{\text{def}}{=} \ltimes_{\iota i}(x) \tag{5.39}$$

"Position x in COPY SET 0 is labeled $\ltimes_\iota$ iff x is labeled $\ltimes_\iota$ in the input."

$$\rtimes_{\iota_o}{}^0(x) \stackrel{\text{def}}{=} \rtimes_{\iota i}(x) \tag{5.40}$$

"Position x in COPY SET 0 is labeled $\rtimes_\iota$ iff x is labeled $\rtimes_\iota$ in the input."

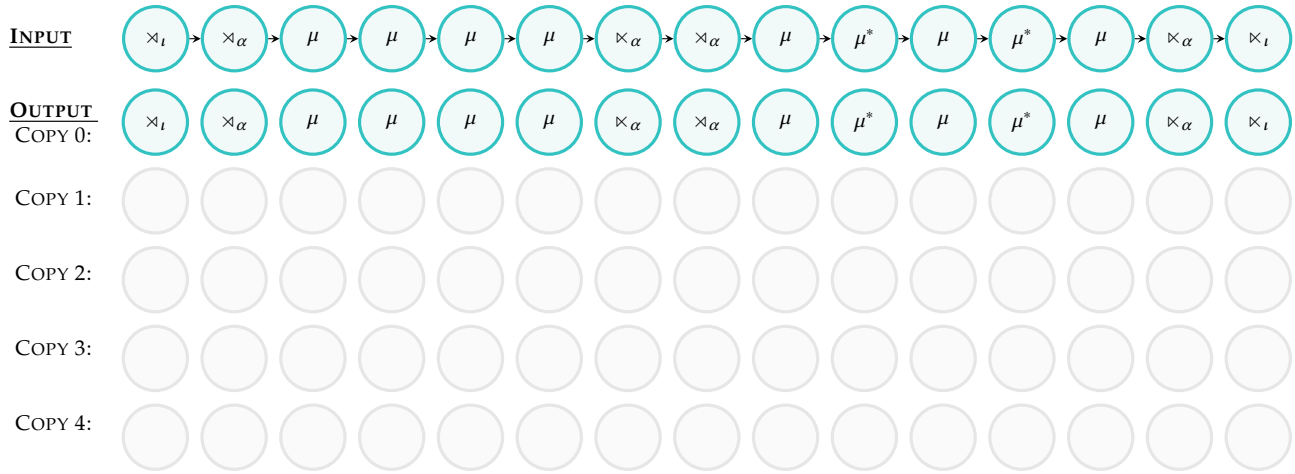


Figure 5.30: A graph illustrating COPY SET 0 for Tokyo Japanese intonation based on the formulas (5.35 - 5.40).

As for the second and third copy sets (COPY SET 1-2), the HL lexical pitch accents in the output are defined from the starred moras in the input, as defined in the formulas (5.41) and (5.42). Since these pitch accents are underlyingly determined from the lexicon, the starred moras are *directly interpreted* as the actual HL tones in the output, which is different from the post-lexical tones that are left unspecified with Ts until they undergo the meaning transductions. These two copy sets are illustrated in Figure 5.31.

$$H^{*1}_{o}(x) \stackrel{\text{def}}{=} \mu^{*}_{i}(x) \tag{5.41}$$

"Position x in COPY SET 1 is labeled H^* iff x is labeled μ^* in the input."

$$L^{1}_{o}(x) \stackrel{\text{def}}{=} \mu^{*}_{i}(x) \tag{5.42}$$

"Position x in COPY SET 1 is labeled L iff x is labeled μ^* in the input."

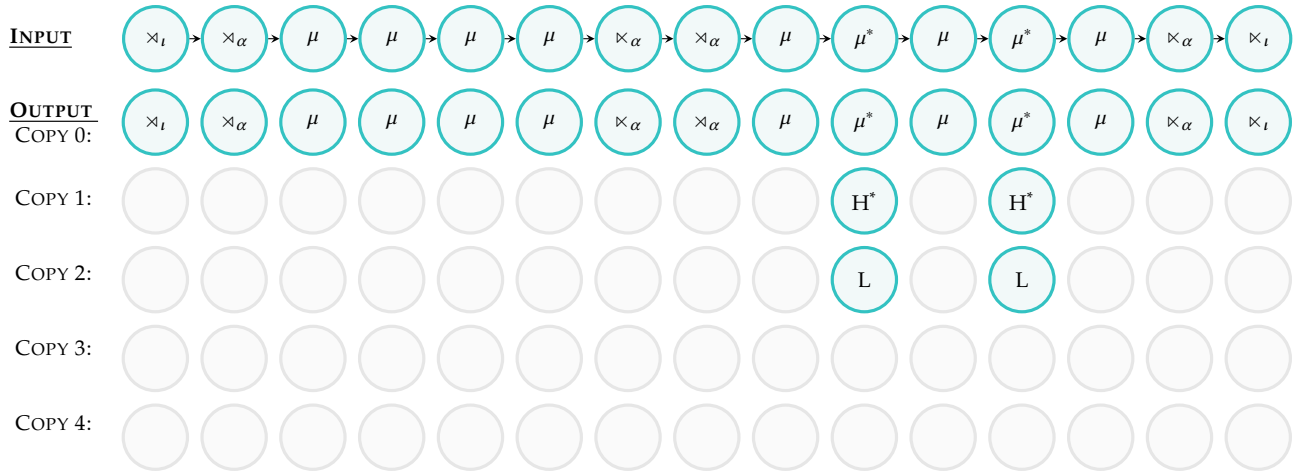


Figure 5.31: A graph illustrating COPY SET 1-2 for Tokyo Japanese intonation based on the formulas (5.41 - 5.42).

As for the last two copies (COPY SET 3-4), the post-lexical tones in the output (i.e., unspecified Ts) are defined from the left edge of an AP boundary in the input (5.43), while the boundary tones are defined from the left edge of an IP boundary or the right edge of an AP boundary (5.44). These two copies are visually presented in Figure 5.32.

$$T^3_o(x) \stackrel{\text{def}}{=} \rtimes_{\alpha i}(x) \tag{5.43}$$

"Position x in COPY SET 3 is labeled T iff x is labeled $\rtimes_\alpha$ in the input."

$$T^4_o(x) \stackrel{\text{def}}{=} \rtimes_{\iota i}(x) \vee \ltimes_{\alpha i}(x) \tag{5.44}$$

"Position x in COPY SET 4 is labeled T iff x is labeled $\rtimes_\iota$ or $\ltimes_\alpha$ in the input."

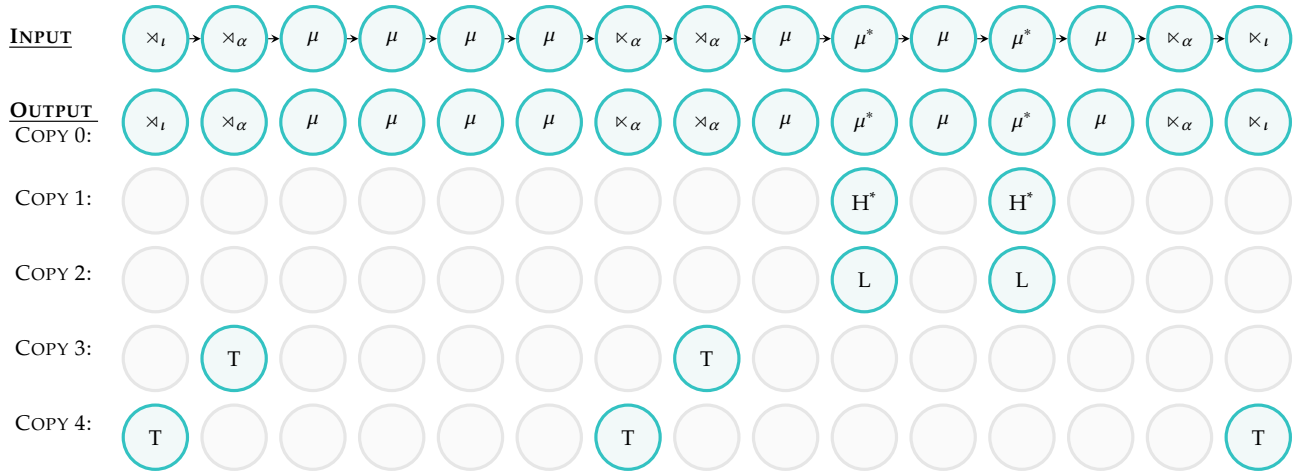


Figure 5.32: A graph illustrating COPY SET 3-4 for Tokyo Japanese intonation based on the formulas (5.43 - 5.44).

Crucially, the *direct interpretation* from the starred moras to the lexical pitch accents and the *indirect interpretation* from phrasal boundaries to the unspecified post-lexical tones exhibit the head/edge-prominence properties in a lexical pitch accent language, Tokyo Japanese.

Step 2: Tone-TBU association The tonal tiers (COPY SET 1-4) are linked with the TBU tier (COPY SET 0), as defined in the formulas (5.45-5.48) and shown in Figure 5.33. As for the lexical pitch accents, only the first pitch accent sequence (i.e., H* in COPY SET 1 and L in COPY SET 2) is realized and others are deaccented. Therefore, $\mathcal{A}_{\mathrm{o}}{}^{0,1}(x,y)$ associates the H* lexical pitch accents in COPY SET 1 and their starred TBUs in COPY SET 0, if and only if the input position of the H* lexical pitch accent (i.e., a starred mora in the input) is the same as the input position of a starred mora (i.e., a starred mora in the input), and the starred mora's preceding starred element is an AP boundary at the left edge, using the p^* function.[9] This is for only selecting the very first lexical pitch accent within an

[9]Recall the special function $p^*(x)$ that only projects starred syllables and boundaries using a tier-based representation, but for Tokyo Japanese, the starred moras are used to define p^*, instead of the starred sylla-

AP, after which the subsequent lexical pitch accents within an AP are not realized due to deaccentuation in Tokyo Japanese. ${\mathcal{A}_{\mathrm{o}}}^{0,2}(x,y)$ associates the L lexical pitch accents in COPY SET 2 and their immediately following non-starred TBUs in COPY SET 0, if and only if the input position of the L lexical pitch accent (i.e., the starred mora in the input) immediately precedes the input position of a non-starred mora (i.e., a non-starred mora in the input).

${\mathcal{A}_{\mathrm{o}}}^{0,3}(x,y)$ associates the phrasal tones in COPY SET 3 and the second mora in an AP in COPY SET 0, if and only if the input position of the phrasal tone (T) (i.e., the right AP boundary in the input) is two nodes before the non-starred mora (i.e., a non-starred mora in the input), but only when not followed by a lexical pitch accent. Therefore, if the following elements are the lexical pitch accents, the phrasal tones cannot be realized. As for the boundary tones, ${\mathcal{A}_{\mathrm{o}}}^{0,4}(x,y)$ associates the boundary tones in COPY SET 4 and a mora in COPY SET 0 with three conditions: first, if and only if the input position of the IP-initial boundary tone (i.e., the right IP boundary in the input) is two nodes before the input position of the non-starred mora (i.e., a non-starred mora in the input); second, if and only if the input position of the AP-final boundary tone (i.e., the left AP boundary in the input) is two nodes before the input position of the non-starred mora (i.e., a non-starred mora in the input); third, if and only if the input position of the IP-final boundary tone (i.e., the left AP boundary in the input) is two nodes after the input position of the non-starred mora (i.e., a non-starred mora in the input). This association enables us to associate the boundary tones with the moras not only at the edges of an IP but also at the left edge of an AP.

---

bles. Therefore, p^* for Tokyo Japanese is defined as follows: $p^* \stackrel{\text{def}}{=} {\mu^*}_i(x) \vee \ltimes_{\alpha i}(x) \vee \rtimes_{\alpha i}(x) \vee \ltimes_{\iota i}(x) \vee \rtimes_{\iota i}(x)$.

$$\mathcal{A}_o^{0,1}(x,y) \stackrel{\text{def}}{=} \mu_i^*(x) \wedge \mu_i^*(y) \wedge x \approx y \wedge \rtimes_{\alpha i}(p^*(x)) \tag{5.45}$$

"Positions x in COPY SET 0 and y in COPY SET 1 are associated

iff x is labeled as μ^*, y is labeled as μ^*, x's preceding starred element is $\rtimes_\alpha$,

and x and y are at the same position in the input."

$$\mathcal{A}_o^{0,2}(x,y) \stackrel{\text{def}}{=} \mu_i(x) \wedge \mu_i^*(y) \wedge y \approx p(x) \tag{5.46}$$

"Associate positions x in Copy set 0 and y in Copy set 1

iff x is labeled as μ, y is labeled as $\rtimes_\alpha$, and x precedes y in the input."

$$\mathcal{A}_o^{0,3}(x,y) \stackrel{\text{def}}{=} \mu_i(x) \wedge \rtimes_{\alpha i}(y) \wedge y \approx p(p(x)) \wedge \neg\mu^*(s(x)) \tag{5.47}$$

"Associate positions x in Copy set 0 and y in Copy set 3

iff x is labeled as μ, y is labeled as $\rtimes_\alpha$, and y is two nodes away from y,

but only when x's following element is not μ^* in the input."

$$\mathcal{A}_o^{0,4}(x,y) \stackrel{\text{def}}{=} \mu_i(x) \wedge (\rtimes_{\iota i}(y) \vee \ltimes_{\alpha i}(y) \wedge y \approx p(p(x))) \vee (\ltimes_{\iota i}(y) \wedge y \approx s(s(x))) \tag{5.48}$$

"Associate positions x in Copy set 0 and y in Copy set 3

iff x is labeled as μ, y is labeled as $\rtimes_\iota$ or $\ltimes_\alpha$, and y is two nodes before x,

or iff y is $\ltimes_\alpha$ and y follows x in the input."

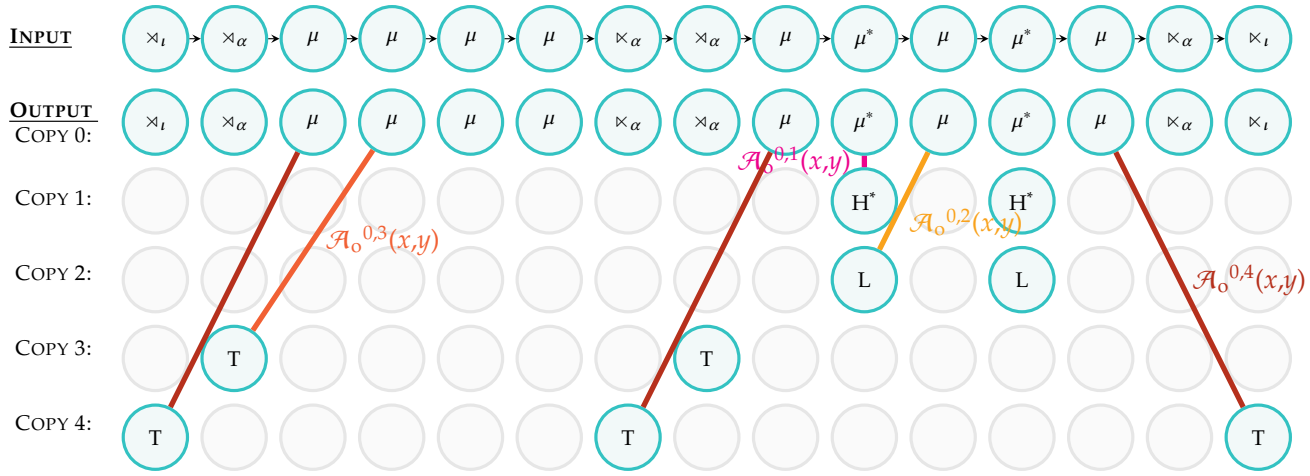


Figure 5.33: A graph illustrating COPY SET 1-2 for Tokyo Japanese intonation based on the formulas (5.41 - 5.42).

Step 3: Defining the order in melodies The ordering of the elements in the output is fixed based on that in the input, but separately for the TBUs and the tones, as shown in Figure 5.34. Specifically, the TBUs have the same order as the inputs. As for the melodies, a boundary tone at the left edge of an utterance is linked with the first phrasal tone of the first AP. Then, since there is no lexical pitch accent in the first AP, another boundary tone at the right edge of the first AP and another phrasal tone at the left edge of the second AP come next. Then, the two lexical pitch accent seqeunces, H*Ls, are connected. But here an important assumption is that only the association between the tones and the TBUs leads to the tonal realization. Even if the tones are in the copy sets but are not associated with the TBUs, they are not realized in the output melodies. Therefore, only the first H*L sequence is realized due to its association with the TBUS, while the second sequence is not. Lastly, a boundary tone at the end of the last AP ends the melody.

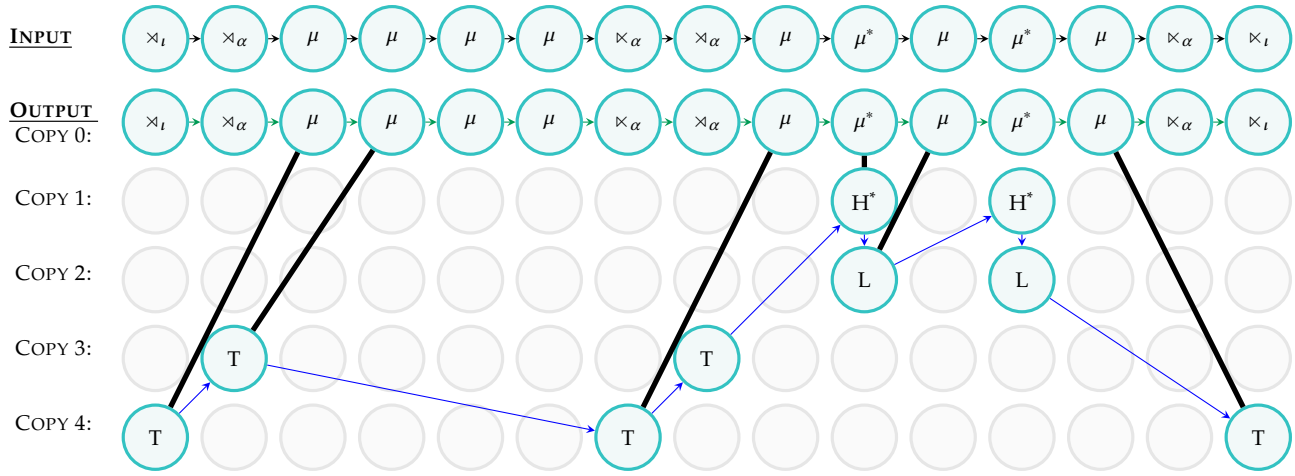


Figure 5.34: A graph illustrating COPY SET 1-2 for Tokyo Japanese intonation based on the formulas (5.41 - 5.42).

5.3.3.3 Declarative transduction

Via a *declarative transduction*, the unspecified post-lexical tones are then filled with Hs and Ls for the declarative in Tokyo Japanese, as shown in Figure 5.35. Note that the lexical pitch accents are already specified with H* and L. An input signatures $\mathcal{S}_i$ is $\{\mu, \mu^*, \ltimes_\alpha, \ltimes_\iota, \rtimes_\alpha, \rtimes_\iota, T, H^*, L\}$ and an output signature $\mathcal{S}_o$ is $\{\mu, \mu^*, \ltimes_\alpha, \ltimes_\iota, \rtimes_\alpha, \rtimes_\iota, H^*, H, L\}$. The following formulas specify Ts with Hs and Ls:

$$L_o(x) \stackrel{\text{def}}{=} T_i(x) \wedge \rtimes_{\alpha i}(p(x)) \vee \ltimes_{\alpha i}(s(x)) \tag{5.49}$$

"Position x in the output is labeled L iff x is labeled T,

and x's preceding element is labeled $\rtimes_\alpha$, or x's following element is labeled $\ltimes_\alpha$ in the input."

$$H_o(x) \stackrel{\text{def}}{=} T_i(x) \wedge \rtimes_{\alpha i}(p(p(x))) \tag{5.50}$$

"Position x in the output is labeled H iff x is labeled T,

and the element that is two nodes before x is labeled $\rtimes_\alpha$ in the input."

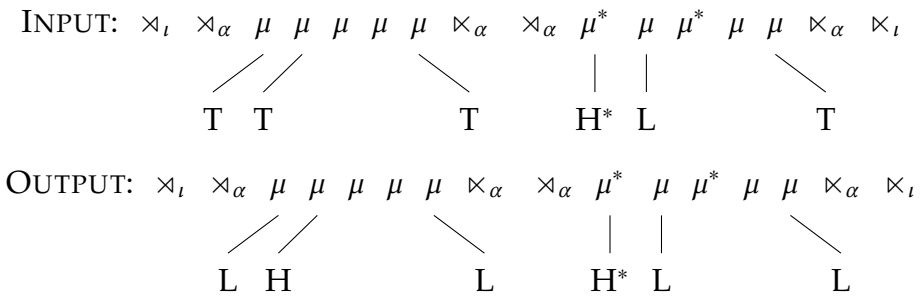


Figure 5.35: Declarative transduction for Tokyo Japanese.

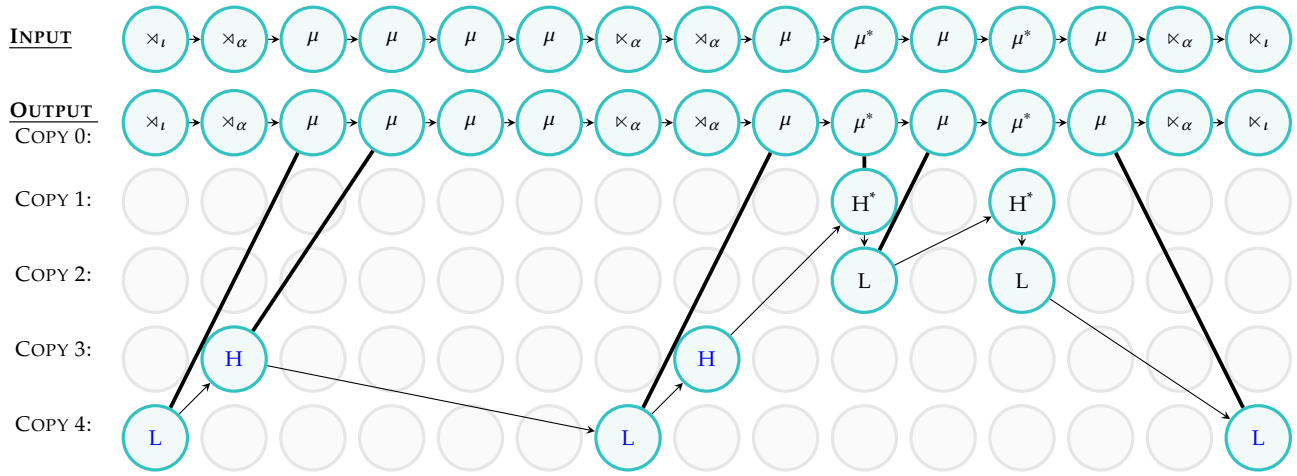


Figure 5.36: A graph illustrating COPY SET 1-2 for Tokyo Japanese intonation based on the formulas (5.41 - 5.42).

5.3.3.4 Summary

Results showed that the intonational pattern in Tokyo Japanese can be defined using a logical interpretation of a prosodic and metrical structure. The melodies in the output were *literal copies of starred moras and boundaries* in the input structure, reflecting the *head/edge-prominence* characteristic of Tokyo Japanese. Unlike the post-lexical (head-prominence and edge-prominence) intonational patterns in American English and Seoul Korean, the *direct interpretation* of the starred moras to H* and L showed the lexically specified pitch accents in Tokyo Japanese. Also, the *indirect interpretation* of boundaries to the unspecified

Ts showed the realization of post-lexical tones. Both the lexical and the post-lexical pitch accent patterns captured the typical initial rising pitch accent in an AP in Tokyo Japanese. Moreover, even considering the deaccentuation pattern that only selects the first lexical pitch accent in an AP, the tone-TBU associations were defined *locally* from the input structure without using any quantifiers. Thus, we can also conclude that the lexical pitch accent pattern in Tokyo Japanese is a QF logical interpretation of a metrical and prosodic structure.

5.4 Discussion

The chapter aimed to explore how we can capture the tone-TBU association patterns in intonation using a logical interpretation of a metrical and prosodic structure. Specifically, we have examined three different intonation patterns: a *head-prominence* language, American English; an *edge-prominence* language, Seoul Korean; a *lexical pitch accent* language, Tokyo Japanese. Importantly, we were able to see that tones in intonational melodies are literal copies of elements, such as starred TBUs or different types of phrasal boundaries, in the metrical and prosodic structure, which made a typological distinction between the intonational types. Also, the intonational tone-TBU associations were able to be defined as a local process. Therefore, the results overall showed that intonation is a QF logical interpretation of a metrical and prosodic structure, which can be defined *locally* from the input structure.

First, interpreting prosodic elements from the input to the discrete tonal targets in the output explicitly showed different intonational patterns. Specifically, the head-prominence

characteristic in American English intonation was able to be captured by copying the starred syllables, showing that the pitch accents are the direct interpretations of the heads — starred syllables (more prominent elements). In contrast, the edge-prominence characteristic in Seoul Korean was captured by copying the phrasal boundaries, showing that the edge tones are the direct interpretations of the phrasal edges. This indicates that the prosodic elements in the input strings are not realized the same way, but the way they are logically interpreted leads to the characterization of different metrical and prosodic realizations in intonation.

Moreover, the lexical pitch accent pattern in Tokyo Japanese intonation was captured by copying the starred moras, which were directly interpreted into specified tones (H^*L). However, the phrasal pitch accent patterns were captured by copying boundaries without specifying tonal sequences. While both the lexical and phrasal pitch accents showed the typical initial rising pitch pattern in Tokyo Japanese, the distinction between the H^*L lexical pitch accents and the unspecified Ts showed different tonal realizations of lexical vs. post-lexical pitch accents, allowing us to account for the lexical status of the tones.

Crucially, the computational nature of the intonational tone-TBU association patterns turned out to be local. Referring only to particular elements, such as starred TBUs and boundaries, does not seem to be local. But if we assume the order preserving from the input to the output structure as proposed in Chandlee & Jardine (2019b), the intonational patterns can be analyzed locally without using any quantifiers. The tone-TBU associations in American English and Seoul Korean were defined locally with the predecessor and successor functions, while those in Tokyo Japanese showed a more complex pattern. The deaccentuation pattern in Tokyo Japanese only chooses the first one of the lexical pitch

accents and does not choose the rest of them in an AP. By using a tier-based predecessor function p^*, the deaccentuation pattern was able to be computed without any quantifiers. Therefore, we can see that the intonational patterns in these languages are found to be locally QF-interpretable from the input structure.

By defining the intonational structure as a QF logical interpretation of a metrical and prosodic structure that are input strictly local, we were able to create an *intonational theory* that is restrictive enough to characterize different intonational patterns of the languages. From the typological view of intonation, the head-prominence intonational pattern in American English was defined with the copies of both starred syllables (i.e., heads) and boundaries, whereas the edge-prominence pattern in Seoul Korean was defined with the copies of only boundaries (i.e., edges). The lexical pitch accent pattern in Tokyo Japanese was defined with the copies of the starred moras for the lexical pitch accent and the copies of the boundaries for the post-lexical tones.

Moreover, we were able to make the typological distinctions of the intonational patterns. Differences and similarities of the intonations across languages were specified in terms of what kind of prosodic component is being computed and how they were associated with their TBUs – locally or not. Jun (2006a, 2014, 2025) provided the prosodic typology depending on the prominent and rhythmic/prosodic elements, focusing on *describing* the intonational patterns. However, the logical interpretation of intonational patterns made this typological information more *explicit*, such that we were actually able to see what kind of prosodic element is being computed in the output structure from the input structure and how the tone-TBU associations are established, using the QF formulas.

In terms of possible intonational patterns, we began with the hypothesis that intona-

tional patterns must be at least QF logically interpretable, since the tonal sequences like Midpoint Pathology do not exist. In the Declarative Transduction for American English, Seoul Korean, and Tokyo Japanese, we were also able to define Hs and Ls sequences using FO logic without any need of quantifiers or more powerful logic. At least for those languages, there was no such case that the even number of starred syllables must be H tones, which may also work for Question Transduction with H boundary tone at the end of an IP. At least for certain meaning transductions in several languages, more powerful logic, such as Monadic Second Order logic, does not appear to be restrictive enough to exclude the even-number case (e.g., Graf 2010), while less powerful FO logic may possibly be sufficiently expressive to account for the possible intonational patterns. However, further research with more data is needed to logically define the computational nature of these intonational patterns.

Based on these results, we may able to ask several questions to predict the intonational patterns: 1) what kind of prosodic elements are being copied in the output? Is it a head of a constituent? Is it a phrasal boundary? Or are they both?; 2) when are the tones specified during the derivation from the input to the output? Is it directly specified from the input to the output in a melodic transduction? Or is it specified during the meaning transduction? These questions can provide valuable predictions of possible intonational patterns in the typology of the languages.

However, further research is needed to generalize these results by examining other languages that fall into the same intonational type: *head-prominence*, *edge-prominence*, and *head-edge prominence*. For instance, Spanish can be another example of the head-prominence type (Beckman et al. 2002). In Spanish, pitch accents such as H^*, L^*+H, L+H^*, H+L^* are

realized in the stressed syllables in an ip and boundary tones such as L%, H% are realized at the end of an IP. The computation of these prosodic patterns in Spanish may show similarities to that in American English, since the pitch accent realization may be closely related to the head of a constituent in both languages.

For another edge-prominence pattern, West Greenlandic can be an interesting example, since HLH tonal patterns are typically realized at the end of a phonological word, with a potential L tone at the beginning of a phonological word (Arnhold 2014). In this language, moras are grouped into phonological words, which in turn form an IP. West Greenlandic may show similarities with Seoul Korean in the sense that the edges of the prosodic units are marked by typicall tonal patterns. However, the basic prosodic unit for the edge tones is a phonological word, which may differ from Seoul Korean.

Lastly, for the lexical pitch accent pattern, Lekeitio Basque may exhibit similar patterns as in Tokyo Japanese such that a H_*+L lexical pitch accent is realized in an AP and %L boundary tone on the first syllable of an AP (Elordieta 1998). And there are initial and final L% and H% boundary tones. However, there is no deaccentuation pattern in Lekeitio Basque, such that this may differ from that in Tokyo Japanese.

Likewise, we need to provide further analyses on the intonational pattern of other languages to generalize the results that intonation is a QF logical interpretation of a metrical and prosodic structure that are defined locally. But in this way, we can provide a theory of intonation that makes restrictive predictions about the typology of intonation and measure the complexity of intonational structures.

5.5 Summary and conclusion

The chapter explored how the tone-TBU association patterns in intonation can be defined using a logical interpretation of a metrical and prosodic structure. Importantly, in this framework, tones in intonational melodies were viewed as literal copies of elements in the metrical structure, such as starred syllables or boundaries. Also, the intonational tone-TBU associations were defined locally without any quantifiers. Typologically, head-prominence and edge-prominence intonational patterns were QF metrical grids, whereas lexical pitch accent patterns were more complex. By defining intonation as a logical interpretation, we were able to figure out the computational nature of intonation and predict the typology of intonation. This chapter provided a better understanding of intonation in the theory of intonational and computational phonology.

Chapter 6

Connecting discrete and continuous f0 information in intonation

This chapter shows how the discrete and continuous f0 information in intonation can be mathematically and computationally connected. Section 6.1 first introduces fuzzy logic as a framework to deal with the continuous values of f0 and then introduces a fuzzy logical model that connects fuzzy logic with perception. Section 6.2 defines how discrete and continuous representations of intonation can be connected using model theory and logic, and provides formal definitions of the pitch accents in American English. Section **??** implements the fuzzy logical system of the pitch accent in American English with real-world f0 data and compares its performance with other data-driven machine learning models.

6.1 Fuzzy Logic

Fuzzy logic is a type of logic that is computed with *degrees* of truth values or membership with the values from 0 to 1 (e.g., Zadeh 1965, 1975, Zadeh et al. 1996, Zadeh 2008, 2023). This logic is used for statements that describe uncertainty or ambiguity, which cannot always return true (valued 1) or false (valued 0). It differs from *boolean logic*, which returns only 1 or 0 as the truth value. For example, as shown in Table 6.1, for a statement "*Is it hot?*", Boolean logic returns only either 1 or 0, by answering whether the statement is true ("*Yes, it is hot.*") or false ("*No, it isn't hot.*"). However, fuzzy logic returns the values between 0 and 1, by expressing whether the statement is partially true ("*It is a bit hot.*") or partially false ("*It is a bit cold.*").

Fuzzy logic was first introduced by Zadeh (1965) to import the concept of vagueness and gradiency into the logical world. The core concept of fuzzy logic is to capture how humans make decisions, which cannot always be reducible to true or false, but are *partially* true or false. Zadeh (2008) views fuzzy logic as "a precise logic of imprecision and approximate reasoning", which is closely related to decision-making processes whose tasks do not require accurate calculation or operation, and whose situations are not crystal-clear to be accompanied by categorical decisions.

Boolean logic		**Fuzzy logic**	
Is it hot?	Truth value	**Is it hot?**	Truth value
Yes	1	*hot*	1
No	0	*warm*	0.75
		cool	0.5
		cold	0.25
		very cold	0

Table 6.1: Comparison of Boolean logic and fuzzy logic.

6.1.1 Preliminaries

The domain of a fuzzy set is the universe of discourse $\mathcal{U}$. Within the domain, a linguistic variable X is defined with a 5-tuple:

$$\langle \mathcal{U}, X, t(X), \mathcal{P}, \mathcal{Q} \rangle,$$

where the linguistic variable X consists of a term set of X, $t(X)$, which is defined by a syntactic rule $\mathcal{P}$. Then, these terms are interpreted with a semantic rule $\mathcal{Q}$.

Suppose we are defining a linguistic variable *temperature*, as shown in Figure 6.1. Here a linguistic variable X is *temperature*, and its term sets are valued with {*very cold*, *cold*, *cool*, *warm*, *hot*} via the syntactic rule $\mathcal{P}$. With this set of terms, the gradiency of temperature can be expressed from *very cold* to *hot*. Via the semantic rule $\mathcal{Q}$, each term can then be semantically interpreted as a fuzzy set that is represented with actual values (i.e., from -30°C to 30°C) within the domain $\mathcal{U}$. The value of the fuzzy set is defined with the degree of membership ranging over a closed interval [0,1].

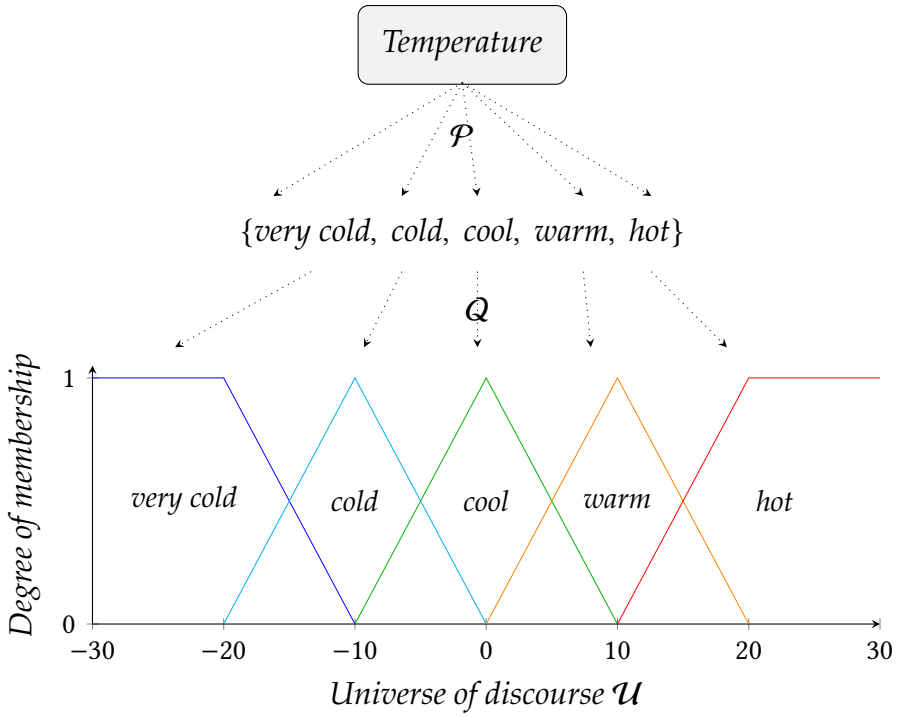


Figure 6.1: An example of a linguistic variable *temperature* defined with fuzzy sets. Redrawn from Kimia Lab (2019).

6.1.2 Syntax of fuzzy logic

A *fuzzy set* A is defined with the ordered pairs $(x, \mu_A(x))$ as the following:

$$A = \{(x, \mu_A(x)) \mid x \in A,\ \mu_A(x) \in [0, 1]\},$$

where each element x in the fuzzy set A is defined with its membership function $\mu_A(x)$ such that the belongingness of x in the set A is defined within a closed interval [0,1]. If x fully belongs to A, the degree of membership is 1, whereas if x does not belong to A at all, the degree of membership is 0. For instance, as for the fuzzy set *hot* for the linguistic variable *temperature* defined in Figure 6.3, if x is 17°C then $\mu_{hot}(17)$ has the membership degree of 0.7 for the fuzzy set *hot*. This is shown with the red dashed vertical and horizontal lines.

In this way, the degree to which the temperature is "*a bit hot*" can be expressed.

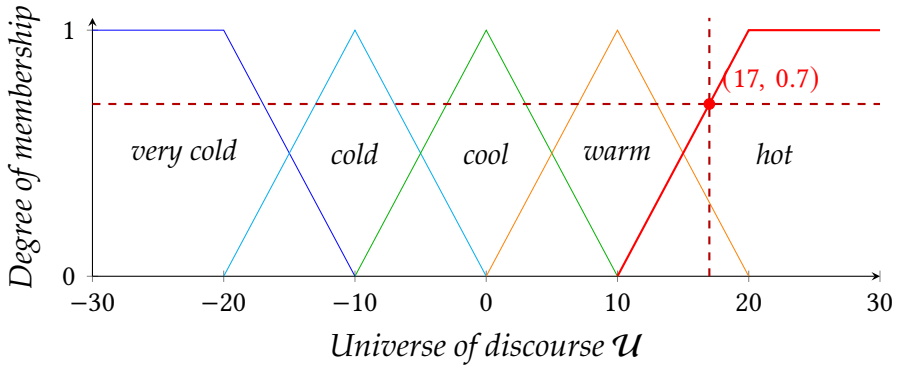


Figure 6.2: A fuzzy membership function $\mu_{hot}(x)$ for *temperature*. If x is 17°C (indicated by the vertical red dashed line), the degree of membership $\mu_{hot}(17)$ is 0.7 for the fuzzy set *hot* (indicated by the horizontal red dashed line). Redrawn from Kimia Lab (2019).

Fuzzy sets are translated into the membership value between 0 and 1 using membership functions. Similar to Boolean logic, they also use $\wedge$, $\vee$, and $\neg$. However, they are calculated using the numeric values of the membership degree. The conjunction $\wedge$ of the membership functions, $\mu_A(x)$ and $\mu_B(x)$, computes the minimum value between them, while the disjunction $\vee$ of them computes the maximum value between them. The negation of the membership function $\mu_A(x)$ is $1-\mu_A(x)$. The ingredients for the syntax of fuzzy logic are listed Table 6.2.

Table 6.2: Ingredients for the syntax of fuzzy logic (Zadeh 1965, 1975)

Ingredients	**Symbols**	**Definition**
Linguistic variables	$X, Y, Z, \cdots$	
Fuzzy sets (Linguistic values)	$A, B, C, \cdots$	
Variables in a fuzzy set	$x, y, z, \cdots$	
Membership functions	$\mu_A(x),\ \mu_B(x),\ \mu_C(x), \cdots$	
Conjunction	$\wedge$	$\min(\mu_A(x),\ \mu_B(x))$
Disjunction	$\vee$	$\max(\mu_A(x),\ \mu_B(x))$
Negation	$\neg$	$1-\mu_A(x)$

6.1.3 Semantics of fuzzy logic

The fuzzy set A can be interpreted as individual elements $u_1, \cdots, u_n$ over the domain $\mathcal{U}$, where $U = u_1 + \cdots + u_n$. The fuzzy set A can be defined with the membership degree μ_i for the element u_i, such that $A = \mu_1/u_1 + \mu_2/u_2 + \cdots + \mu_n/u_n$. Note that the symbols here are not the arithmetic operators. '+' indicates the union of elements, not addition, and '/' is used as a separator between the membership degree and the element, not division. For instance, suppose the fuzzy set *hot* over the universe of discourse $\mathcal{U} = \{-20, -10, 0, 10, 20\}$. If the membership degree is $\{0, 0.25, 0.5, 0.75, 1\}$ for each element, respectively, the fuzzy set *hot* can be specified with the membership degree for each element: $0/-20 + 0.25/-10 + 0.5/0 + 0.75/10 + 1/20$.

6.1.4 Extension Principle

Boolean logic can be extended to fuzzy logic using the Extension Principle (Zadeh 1965, 1975). This principle is used to extend the definition of the pitch accent in American English defined within Boolean logic to fuzzy logic in order to calculate the gradiency.

We first define the function f over the Boolean domain $\mathcal{D}$ such that $f : X_1 \times \cdots \times X_n \rightarrow Y$. Then, using the Extension Principle, this f function can be extended to the function f' over the fuzzy domain $\mathcal{U}$, where fuzzy sets $A_1, \cdots, A_n$ are defined. The function f' maps $A_1 \times \cdots \times A_n$ to B such that $f' : A_1 \times \cdots \times A_n \rightarrow B$. Now B can be defined with the membership function as the following:

$$\mu_B(x) = \sup_{x:f(x_1,\cdots,x_n)=x} \Big(\min\big(\mu_{A_1}(x_1), \cdots, \mu_{A_n}(x_n)\big)\Big),$$

where the membership values of n number of fuzzy sets $A_1, \cdots, A_n$ are combined to a minimum value, whose value is the supremum over all possible combinations of $x_1, \cdots, x_n$, where $f(x_1, \cdots, x_n) = x$. As a result, the output of $\mu_B(x)$ is a numeric value of the membership degree.

For example, suppose we have two linguistic variables: *door status* and *temperature*. *Door status* is defined over the Boolean domain where its fuzzy set $\mu_{doorOpen}$ is computed with either 0 or 1. In contrast, *temperature* is defined over the continuous domain where its fuzzy set μ_{hot} is computed with the gradient values from 0 to 1. These two fuzzy sets can be combined by calculating $\min\big(\mu_{doorOpen}(x_1), \mu_{hot}(x_2)\big)$, where $x_1 \in \mathcal{U}_1$ and $x_2 \in \mathcal{U}_2$. If $\mu_{doorOpen}(x_1)$ is 1 and $\mu_{hot}(x_2)$ is 0.25, the combined membership degree $\mu_B(y)$ is the minimum value between the two, which is 0.25. In this way, we can connect the Boolean and fuzzy definitions of the pitch accent predicates in American English.

Fuzzy logic undergoes a specific inference procedure, a *fuzzy logic system*, which is introduced in the following section.

6.1.5 Fuzzy logic system

Fuzzy logic is widely used in various control systems and applications, whose internal factors are defined with gradient or continuous properties. These systems and applications use a fuzzy system, in which various factors reinterpreted as fuzzy values are combined to produce a numeric output.

Now, let us look at how the fuzzy logic system works in an air conditioning system, which is one of the popular examples (e.g., Sobhy & Khedr 2015). Suppose we need to

control an air conditioning system that decides the *cooling power* based on two factors (linguistic variables): *temperature* and *fan speed*. Note that the inputs (*temperature* and *fan speed*) and the outputs are defined with the gradient terms. Figure 6.3 shows the whole structure of the fuzzy logic system.

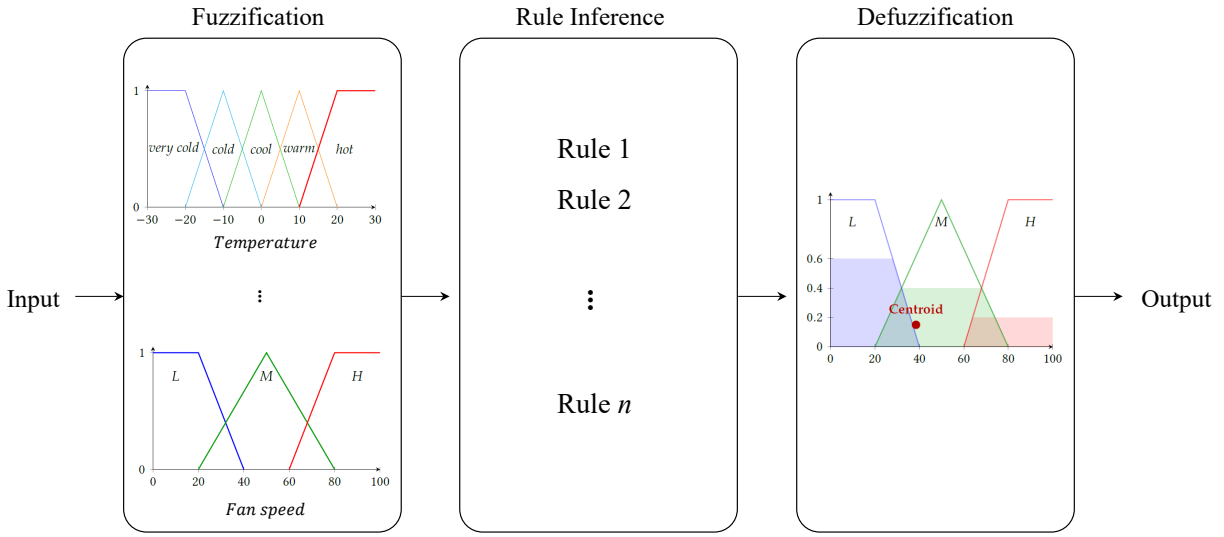


Figure 6.3: A fuzzy logic system.

First, for the *inputs*, let's say the inputs are 17°C for *temperature* and 60 out of 100 for *fan speed*, which are numeric values. In the *fuzzification* stage, each numeric input is fuzzified into the membership degree ranging from 0 to 1. For an input *temperature*, 17°C is interpreted as 0.7 membership degree for the fuzzy set μ_{hot} and 0.3 membership degree for the fuzzy set μ_{warm}, as shown in Figure 6.4. For another input *fan speed*, 60% of the fan speed is interpreted as 0.67 membership degree of the fuzzy set μ_M, as shown in Figure 6.5.

After the fuzzification stage, the fuzzified values are calculated based on the linguistic rules at the *rule inference* stage. In this stage, *if-then* rules evaluate the fuzzified values of the inputs.

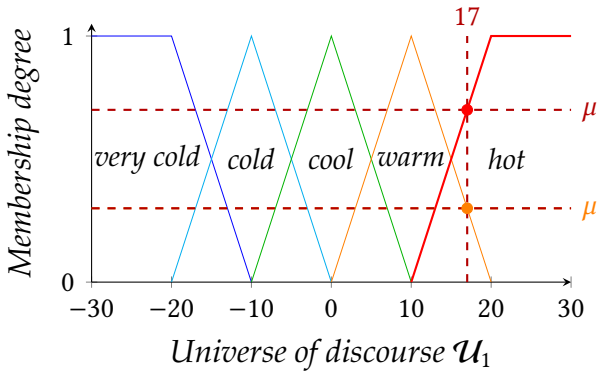


Figure 6.4: Temperature

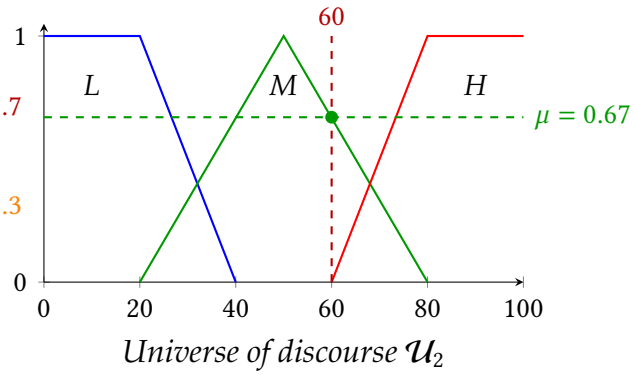


Figure 6.5: Fan speed

For instance, if a rule says: "If the temperature is *warm* AND the fan speed is *M*, then the cooling system is *M*." This rule considers the membership degree only for μ_{warm}, which is 0.3, and membership degree only for μ_M, which is 0.67. Then, due to the logical connective AND, $min(0.3, 0.67)$ results in 0.3 for *M* in the cooling system. Likewise, other rules are evaluated based on the fuzzy values of the inputs.

The last step is the *defuzzification* stage, where the output values of all the applicable rules are combined into a single numeric value. As shown in Figure 6.6, the output fuzzy set is {*L*, *M*, *H*} defined over the percentage of the cooling power (0-100%). If each of the rules has its output value (e.g., Rule 1: 0.6 for μ_L, Rule 2: 0.4 for μ_M, Rule 3: 0.2 for μ_H), the areas below the output value for each rule are combined and defuzzified into a numeric output (0.38). This final output is calculated based on centroid (center of gravity) of the combined shaded region on the right in Figure 6.6. The final output thus is 38% of the cooling power. In this way, the two gradient factors can be combined to obtain a numeric value.

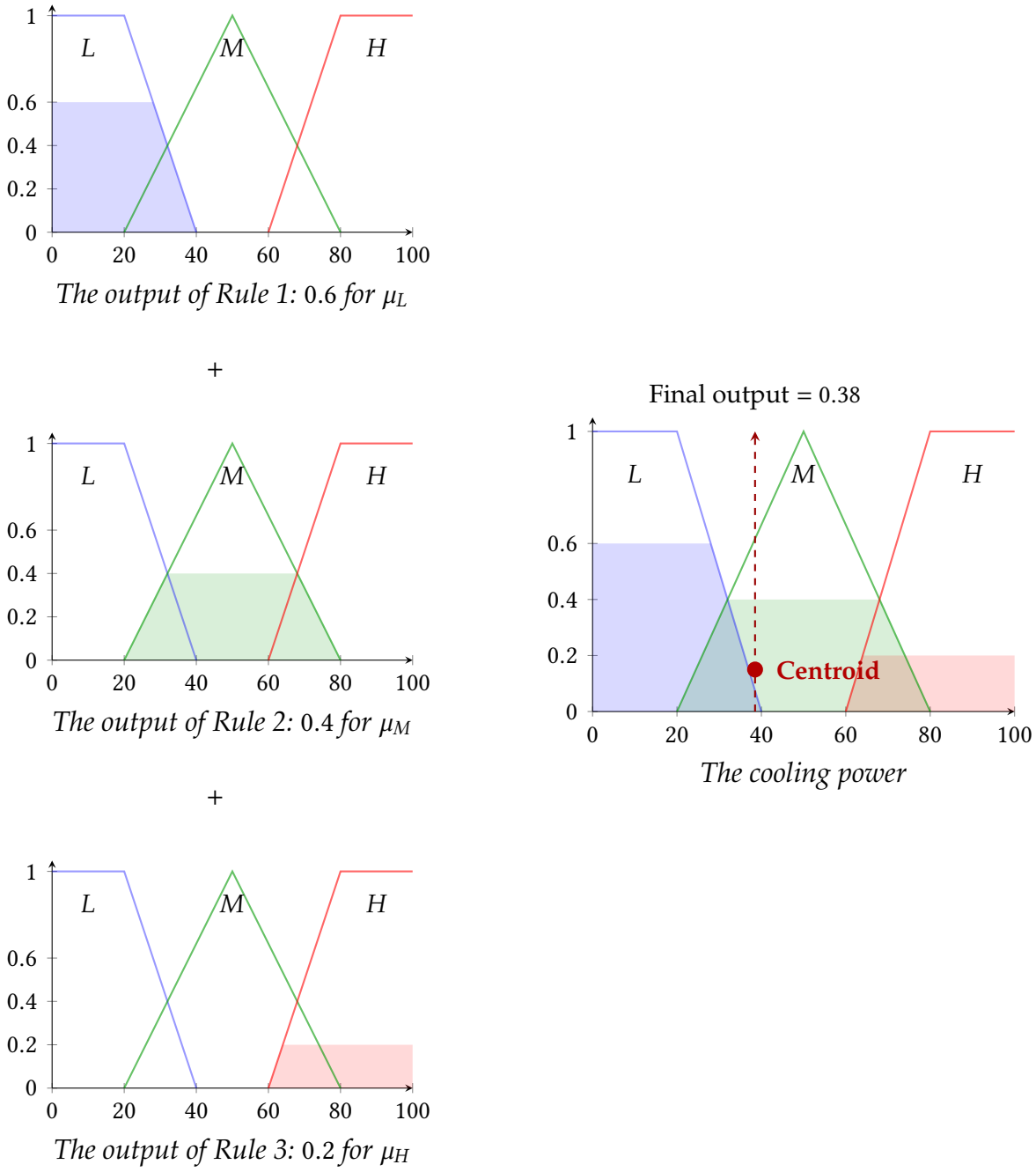


Figure 6.6: The combined output value for the cooling power

Now, we know how the fuzzy logical system works. The following section discusses how this fuzzy logical system relates to accounting for speech perception.

6.1.6 A fuzzy logical model of speech perception

Massaro (1989) uses fuzzy logic to account for speech perception and proposes a perceptual model that explains a logical mechanism of how various perceptual cues for a certain phonological category can be combined to be perceived as that category in the phonological representation. This perceptual model is called a *fuzzy logical model of perception (FLMP)*. This model explains how continuums of several perceptual cues are categorically perceived so as to be represented as one category or the other in our phonological representation by presenting three stages of the logical mechanism: *evaluation*, *integration*, and *classification*. That is, "continuously valued features are evaluated, matched, and matched against prototype descriptions in memory and an identification decision is made on the basis of the relative goodness of match." (Massaro 1989; p.400). This mechanism is a simplified version of the fuzzification-defuzzification system introduced in Section 6.1.

The main idea of FLMP is based on the assumption that the category of a sound is represented in our memory with a set of its properties. Massaro (1989) assumes that the perceptual primitives are characterized by "summary description", which is called a *prototype* (category), and this prototype is composed of a set of properties, which is called a *feature*. A system of FLMP is illustrated in Figure 6.7. At the evaluation stage, a set of features (visual feature, V_i; auditory feature, A_i) is evaluated with respect to prototypes in memory. In the next stage, integration, these two visual and auditory features are combined and matched to one of those prototypes. At this stage, the degree to which the combined features match the prototype P_{ij} is determined. In the classification stage, the accuracy rate, R_{ij}, is measured and determined which of the prototypes is the best match

among the prototypes.

Figure 6.7: A fuzzy logical perceptual model proposed by Massaro (1989)

Massaro (1989) conducted a categorical perception experiment in which perceptual cues American English listeners use to categorize /ba/ versus /da/. This study assumed that the relevant features for the prototype /ba/ are a lip closing gesture and a rising F1 and F2, while those for the prototype /da/ are a lip opening gesture and a falling F1 and F2. As shown in Figure 6.8, the sound stimuli were manipulated depending on these two features, a lip gesture (visual feature, V_i) and the formant transitions (auditory features, A_i). The lip gesture was manipulated in a categorical way by testing whether it is /ba/, /da/, or neither. In contrast, the formant transitions were manipulated in a continuous way with 10 steps from /ba/ to /da/. The results of the categorical perception were then evaluated using FLMP, showing that FLMP outperformed other perceptual models.

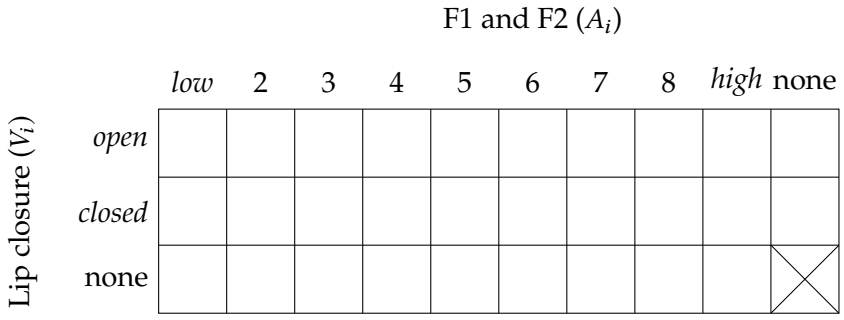


Figure 6.8: Stimuli design of the categorical perception /ba/ versus /da/ in Massaro (1989). Redrawn based on Massaro (1989).

In FLMP, fuzzy logic is used to account for speech perception since it can deal with various sources of information such as visual and auditory features, whose dimensions

are characterized either categorically or continuously, into a single system. Also, Massaro (1989; p.334) describes fuzzy logic as a "natural development of a quantitative description of the phenomenon of interest." Fuzzy logic, therefore, enables us to account for various perceptual cues that are represented in various forms in speech perception.

Drawing on Massaro's FLMP, the following section uses fuzzy logic to encode the continuous property of the pitch accents in American English. By handling gradient information using membership degrees between 0 and 1, fuzzy logic helps categorize various f0 contours as either H* or L+H*.

6.2 Connecting discrete and continuous representation of intonation

This chapter *mathematically* defines the pitch accents in American English by connecting discrete and continuous information of the f0 contour within a logical framework. Specifically, the discrete and continuous information of the pitch accents is first defined with *Boolean logic—first-order* (FO) and *monadic second-order* (MSO) logic, and then these definitions are converted into *fuzzy logic*.

According to the AM model, continuous f0 contours are abstracted into distinctive tonal categories, using combinations of Ls and Hs (e.g., Pierrehumbert 1980, Beckman & Pierrehumbert 1986, Ladd 2008, Arvaniti 2022). In American English, there are two types of pitch accents: monotonal pitch accents (e.g., H*, L*) and bitonal pitch accents (e.g., L+H*, L*+H, H+L*, H*+L) (Pierrehumbert 1980, Beckman & Pierrehumbert 1986).

In these pitch accents, starred tones (e.g., H*, L+H*, L*+H) are only realized on a starred syllable (i.e., an accented syllable), while non-starred tones before or after the starred tone (e.g., L+H*, L*+H) are realized on unstarred syllables. These non-starred tones are called *leading* or *trailing tones*, which are realized over an interval that spans one or more syllables before the starred tone.

There has been considerable attention to a three-way contrast of the rising pitch accents, H*, L+H*, and L*+H, in American English (e.g., Arvaniti et al. 2007, Beckman & Pierrehumbert 1986, Dainora 2002, Ladd 2008, Ladd & Schepman 2003, Pierrehumbert 1980, Steffman et al. 2024, Welby 2003). As shown in the schematized f0 contours of these pitch accents in Figure 6.9, they both appear to have a rising contour whose f0 value starts from the lower f0 and reaches the H tone target; however, the degree of f0 slope and the size of f0 excursion varies depending on the pitch accents. That is, the f0 contour for H* exhibits a gradual slope and narrow excursion, compared to that for L+H* and L*+H. As for the f0 contours between L+H* and L*+H, the f0 peak shows later alignment for L*+H than for L+H*.

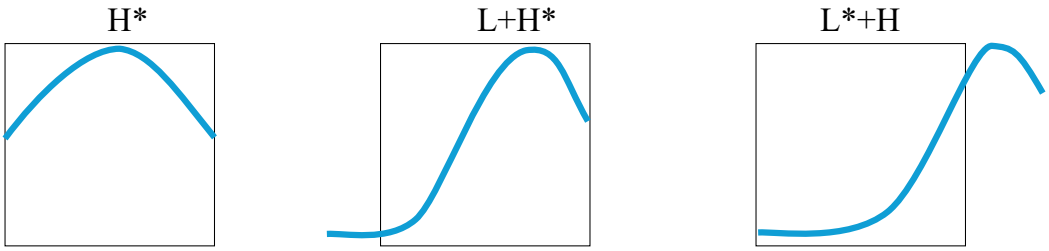


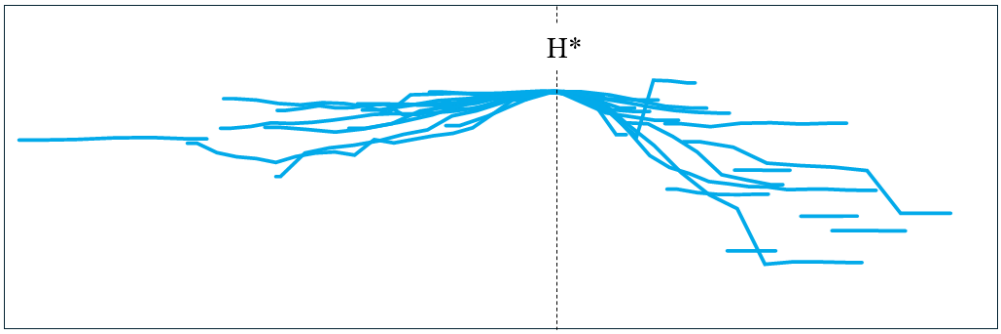
Figure 6.9: Schematized f0 contours of H*, L+H*, and L*+H in American English. Redrawn from Arvaniti et al. (2007).

Studies (e.g., Beckman & Pierrehumbert 1986, Ladd 2008) have reported that L*+H

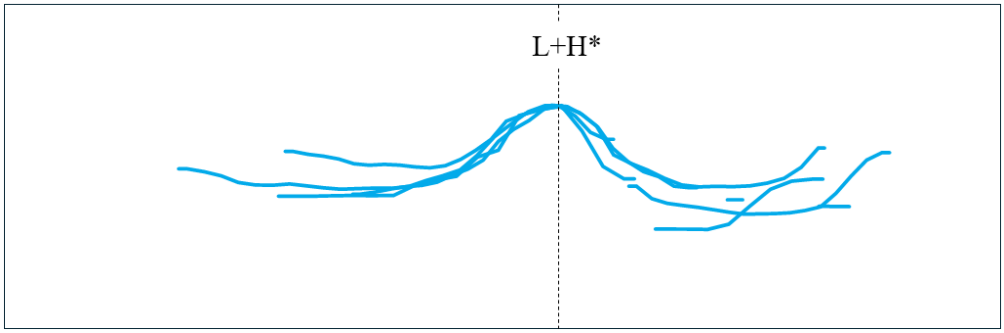
is distinguishable from H* due to the timing of peak alignment and the existence of an L tone. Also, Pierrehumbert & Steele (1989) found that L*+H is aligned much later than L+H* in terms of peak timing. However, studies (e.g., Arvaniti et al. 2007, Steffman et al. 2024, Ladd & Schepman 2003, Ladd 2008 have pointed out that it is quite difficult to distinguish L+H* versus H*[10]. Ladd & Schepman (2003) views this pitch accent contrast as "a matter of gradient variation between two ends of a continuum". Although both pitch accents show the rising trajectories, one of them for H* is considered "incidental" due to the interpolation, while the rising for L+H* is phonologicially meaningful due to the existence of the L leading tone (Arvaniti et al. (2007), Pierrehumbert (1980), Ladd & Schepman (2003). L+H* also shows a greater pitch excursion than H*. However, the f0 variations or gradiency does not give a clear-cut way of contrasting these pitch accents.

More evidence can be found in the superimposed f0 contours of H* and L+H* from ToBI lecture notes (Veilleux et al. 2006) in Figure 6.10. The f0 contours for L+H* clearly show the leading tone and wider excursion, compared to those for H*.

[10]This L+H* vs. H* contrast is also closely related to pragmatics, in which L+H* is considered to be an "emphatic" version of H* (Calhoun 2004, Ladd & Morton 1997).



(a) Superimposed f0 contours of H*



(b) Superimposed f0 contours of L+H*

Figure 6.10: Superimposed f0 contours of H* and L+H* from ToBI lecture notes (Veilleux et al. 2006). The contours were overlapped with respect to the f0 peak to see the f0 variability before the f0 peak, so their f0 values may be different from the actual f0 values.

In addition to the f0 peak timing, importantly, recent studies (Barnes et al. 2010, 2012, 2021) found that rise shape contributes to the pitch accent distinction in American English. One of their main findings is that the scooped rise shapes were perceived as L*+H, while the domed rise shapes were perceived as L+H* by American English listeners. These findings together shed light on the nature of the intonational primitives, which should encode both the discrete tonal targets and continuous f0 information.

Then, a fundamental question arises as to how both the discrete and continuous f0 information is represented and computed in intonation. Specifically speaking, how do we define a *finite* set of contrastive pitch accent categories (e.g., H* vs. L+H*) from infinitely possible f0 contours? How do we account for the shape information? To illustrate these

questions, Figure 6.11 shows that some of the rising contours are categorized as H*, while others are classified as L+H* in American English.

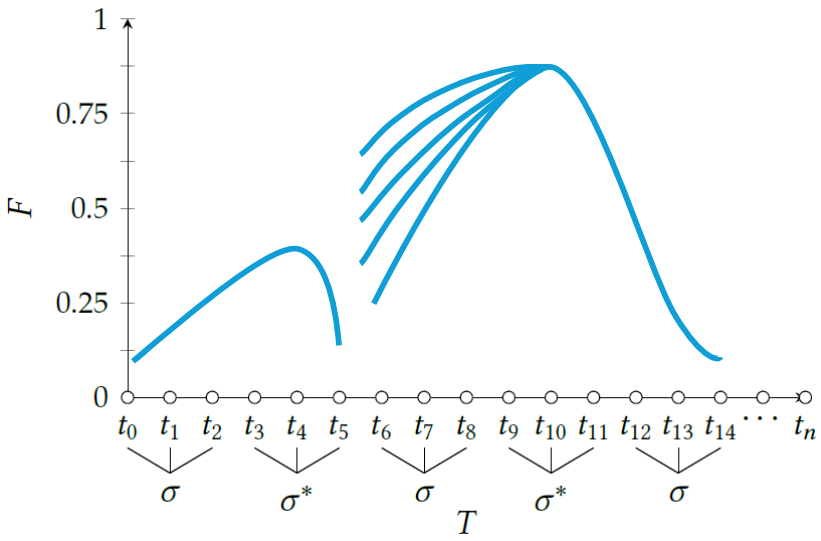


Figure 6.11: Possible rising contours for H* and L+H*.

Therefore, this chapter aims to connect the discrete and continuous f0 information by proposing a mathematical definition of the pitch accents in American English using Boolean (FO and MSO) and fuzzy logic. It uses a decision-making process to determine the pitch accent categories in American English: first, by defining discrete tonal targets (starred tones) and a search space for potential leading or trailing tones (non-starred tones) of the pitch accents with FO and MSO logic, and then by extending the definitions defined in Boolean logic to those in fuzzy logic, which allows for taking the continuous and gradient properties of intonation. The whole process is provided in Figure 6.12.

An important assumption here is that the process starts from the unspecified bitonal form, so either T+T* or T*+T. Then, we decide what the starred tone is. There could be an H* group (i.e., T+H* or H*+T) or an L* group (i.e., T+L* or H*+L). Then, we decide

what the preceding or following f0 context before the starred tone is. For the leading tone, if there isn't any f0 context that is phonological meaning, then we just decide they are monotones, H* or L*. But if there is a leading tone, then we decide whether it is L+H* or H+L*, depending on the phonetic details of the preceding f0 context. The flow works similarly for the trailing tones. The main focus of this chapter is the pitch accent categorization of H* versus L+H* in American English.

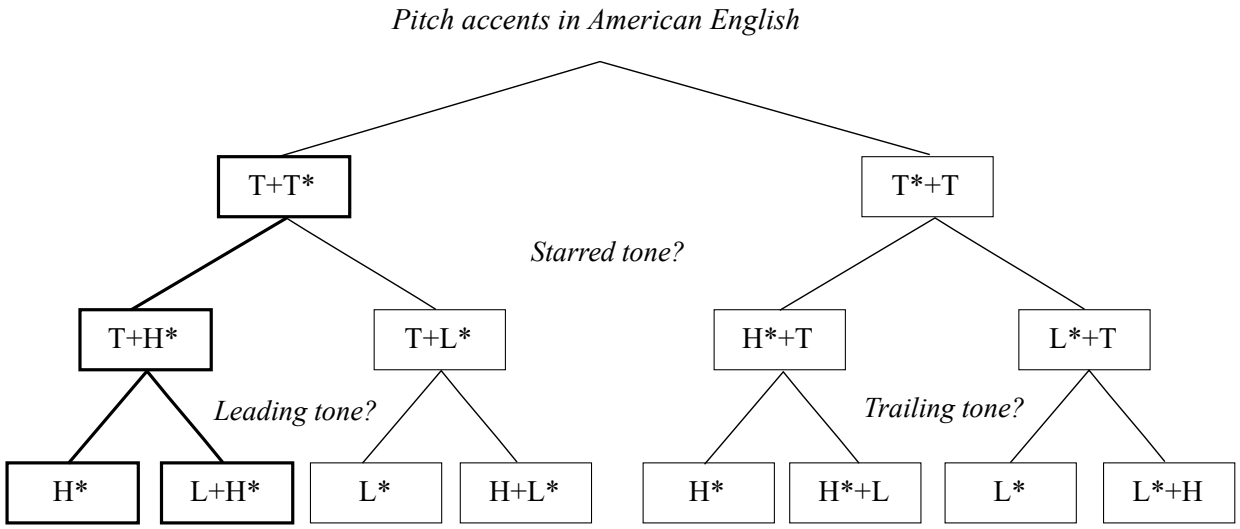


Figure 6.12: A decision-making process of this chapter to determine the pitch accent categories. The main focus of this chapter is the pitch accent categorization of H* versus L+H*, indicated by a thick black line.

The following section introduces the preliminaries to mathematically define American English pitch accents using FO, MSO, and fuzzy logic.

6.2.1 Preliminaries

6.2.1.1 Properties and relations

Recall that a *signature* S defines functions and relations for the input and the output. An input signature $\mathcal{S}_i$ is

$$\langle \mathcal{D};\ \prec_T, \prec_F, P_\sigma, P_{\sigma^*}, P_{[_\sigma}, P_{]_\sigma}, P_{\text{steep}}, P_{\text{gradual}} \rangle,$$

where functions and relations are defined over a domain

$$\mathcal{D} = \{t_0, t_1, \ldots, t_n\}$$

which is structured with discrete timepoints, $\{t_0, t_1, t_2, ..., t_n\}$ in a temporal domain. A schematic representation of the input and output domain is illustrated in Figure 6.13.

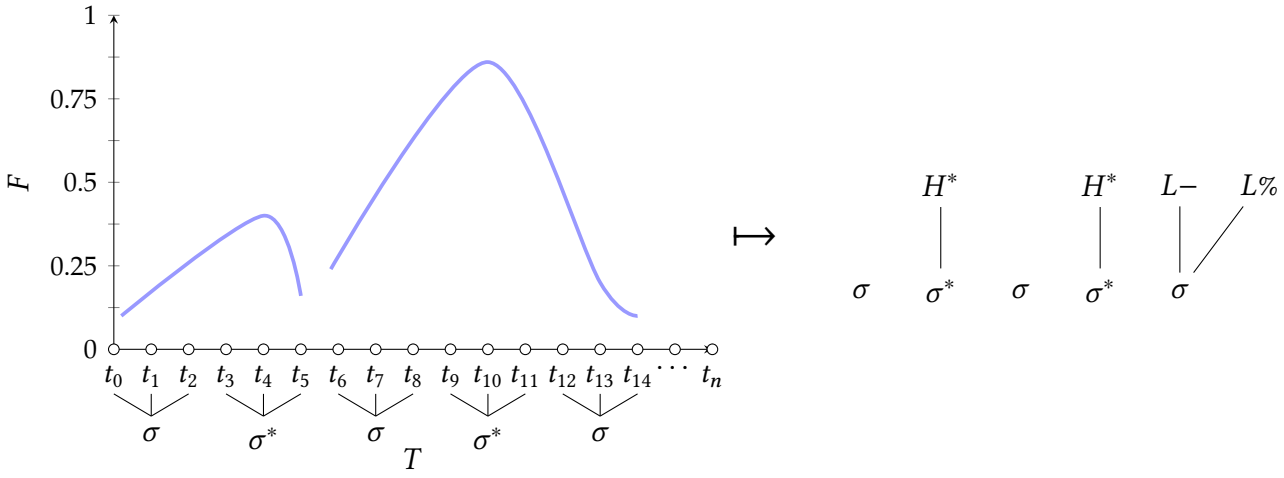


Figure 6.13: A schematic representation of the input and output domain.

$\prec_T$ defines a precedence relation over the set of timepoints ($T \times T$), while $\prec_F$ defines an

ordering relation, with respect to the f0 values in the domain F, over the T-domain ($T \times T$). Crucially, this $\prec_F$ relation is a *core concept* of this chapter in accounting for the continuous f0 information. This predicate is defined based on the timepoints in the T-domain, but it refers to the ordering of f0 values at each timepoint. It does not require us to use the entire continuous f0 space or to model all possible curves, since it costs a lot of computational complexity. Instead, by effectively using the relationship of the temporal ordering with respect to the f0 ordering, the relationship between actual f0 values can be boiled down into the *relative ordering* of f0 values.

Within the domain D, several unary predicates are defined in order to encode the *properties* of individual timepoints or sets of timepoints: P_σ and P_{σ^*} denote the properties of (non-starred) syllables and prominent (starred) syllables, respectively; $P_{[_\sigma}$ or $P_{]_\sigma}$ denote the properties of a syllable boundary at the left or right edge, respectively; and, P_{steep} and $P_{gradual}$ denotes the properties of f0 contour shape information, steep or gradual, respectively. It is noteworthy that these shape predicates encode the *properties* of an interval (i.e., sets of timepoints), since such shapes are perceived as characteristics of an entire contour rather than individual timepoints.

An output signature $\mathcal{S}_o$ consists of

$$\{\sigma, \sigma^*, L, H, L^*, H^*, L + H^*, L^* + H, H^* + L, p, s, \mathcal{A}\},$$

where each property and relation is defined as follows: σ and σ^* denote syllables and prominent syllables, respectively; L, H, L^*, H^*, $L + H^*$, $L^* + H$, and $H^* + L$ denote pitch accent categories; and p and s specify predecessor and successor relations, respectively,

while $\mathcal{A}$ denotes an association relation.

To define both the discrete and continuous f0 information, two types of Boolean logic are used: FO and MSO logic. Recall that in Chapter 3, FO variables x, y, and z are defined over the *individual* elements over the domain, while MSO variables, X, Y, and Z, are defined over the *set* of individual elements over the domain. FO variables allow for characterizing the property of a discrete tonal target at each timepoint over the domain. For example, $P_\sigma(x)$ refers to a timepoint x that belongs to a syllable. MSO variables, however, are used to describe a continuous configuration over a set of time points. For instance, *steep*(X) refers to a set of timepoints, X, that belongs to a property of steep. Therefore, the power of MSO logic enables us to capture f0 shapes and dynamics that span over an interval of the domain, which cannot be reduced to an individual timepoint.

6.2.1.2 Basic predicates

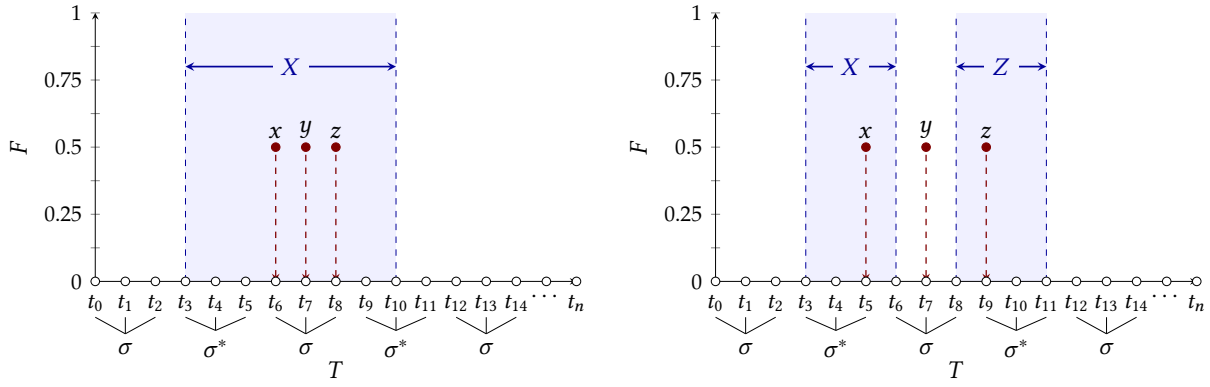
Basic predicates are introduced to characterize both the discrete and continuous information of f0 contours of the pitch accents in American English. First, we need to confine the continuous temporal domain into a specific interval, so that this *search space* allows for locating any Hs and Ls of certain pitch accents. This search space is defined based on the specification of the syllable properties. For example, for a T+T* pitch accent, an interval begins with a non-starred syllable for the realization of the leading tone and ends with a starred syllable for the realization of the starred tone. As for a T*+T pitch accent, an interval begins with a starred syllable for the starred tone and ends with a non-starred syllable for the trailing tone. Second, based on the definition of an interval, *discrete tonal targets* can be defined by finding a single timepoint with the maximum f0 value for an H

tonal target and a single timepoint with the minimum f0 value for an L tonal target within that interval. Lastly, *continuous information* of f0 contour can be defined by encoding the properties of shapes and dynamics to an interval.

Search space An *interval* is defined to restrict the search space for the pitch accents over the temporal domain T. The definition of an interval is as follows:

$$interval(X) \stackrel{\text{def}}{=} \forall x, y, z[X(x) \wedge X(z) \wedge x \prec_T y \wedge y \prec_T z \rightarrow X(y)]$$

where $interval(X)$ is true for all the timepoints x, y, and z, iff both x and z belong to a set X, and y exists between x and z in the T-domain and should belong to the set X. That is, $interval(X)$ is true for a contiguous set of timepoints in the T domain, as visualized in Figure 6.14a. Note that $interval(X)$ is false for three contiguous timepoints x, y, z, if y is not in a set, but x and z belong to distinct sets, X and Z, as illustrated in Figure 6.14b. Therefore, to satisfy the definition of $interval(X)$, all the elements between x and z should also be in the same set as x and z.



(a) A contiguous set of timepoints in an interval.

(b) A contiguous set of timepoints *not* in an interval.

Figure 6.14: A schematic representation of $interval(X)$.

In addition to the definition of an interval, the start and end points for an interval can be specified by introducing the non-starred or starred syllable positional predicates. The non-starred syllables at the left or right edges are represented with $leftEdge_{\sigma}(x)$ and $rightEdge_{\sigma}(x)$, respectively. The starred syllables at the left or right edges are represented with $leftEdge_{\sigma^*}(x)$ and $rightEdge_{\sigma^*}(x)$, respectively. These syllable positional predicates only consider a particular (non-starred or starred) syllable, which does not allow any other *immediately* preceding or following syllables. For this, the immediate precedence relation $\lhd_T$ is defined using the precedence relation $\prec_T$[11]. Therefore, the starred and non-starred syllables at both edges are defined as follows:

$$leftEdge_{\sigma}(x) \stackrel{\text{def}}{=} P_{\lfloor_{\sigma}}(x) \wedge \neg\exists y[P_{\sigma}(y) \wedge y \lhd_T x],$$

$$rightEdge_{\sigma}(x) \stackrel{\text{def}}{=} P_{\rfloor_{\sigma}}(x) \wedge \neg\exists y[P_{\sigma}(y) \wedge y \rhd_T x],$$

$$leftEdge_{\sigma^*}(x) \stackrel{\text{def}}{=} P_{\lfloor^*_{\sigma}}(x) \wedge \neg\exists y[P_{\sigma^*}(y) \wedge y \lhd_T x],$$

$$rightEdge_{\sigma^*}(x) \stackrel{\text{def}}{=} P_{\rfloor^*_{\sigma}}(x) \wedge \neg\exists y[P_{\sigma^*}(x) \wedge y \rhd_T x],$$

where $leftEdge_{\sigma}(x)$ is true if x is a syllable on the left edge and there isn't any preceding syllable y before x, whereas $rightEdge_{\sigma}(x)$ is true if x is a syllable on the right edge and there isn't any following syllable y after x. That is, $leftEdge_{\sigma}(x)$ refers to the first non-starred syllable and $rightEdge_{\sigma}(x)$ refers to the last non-starred syllable. Similarly, $leftEdge_{\sigma^*}(x)$ and $rightEdge_{\sigma^*}(x)$ identify the first starred syllable on the left edge and the last starred syllable on the right edge, respectively.

Now, using the definition of the interval and the (non-starred or starred) positional

[11] $\lhd_T(x, y) \stackrel{\text{def}}{=} x \prec_T y \wedge \neg\exists z[x \prec_T z \wedge z \prec_T y]$. The immediate successor relation $\rhd_T$ is defined symmetrically by reversing the order.

predicates, a *specific* interval for pitch accent can be defined with an interval that spans leftward or rightward with respect to the starred syllable. For the pitch accents T+T*, a leftward interval spans from a starred syllable to its first preceding non-starred syllables. For the pitch accents such as T*+T, a rightward interval spans from a starred syllable to its last following non-starred syllables.

$$\begin{aligned} \mathit{leftwardInterval}(X) \stackrel{\text{def}}{=} & \; \mathit{interval}(X) \wedge \exists x, y[X(x) \wedge X(y) \wedge \mathit{leftEdge}_{\sigma}(x) \wedge \mathit{rightEdge}_{\sigma^*}(y)] \\ & \wedge \neg\exists Y[\mathit{interval}(Y) \wedge x \prec_T Y \prec_T y]], \\ \mathit{rightwardInterval}(X) \stackrel{\text{def}}{=} & \; \mathit{interval}(X) \wedge \exists x, y[X(x) \wedge X(y) \wedge \mathit{leftEdge}_{\sigma^*}(x) \wedge \mathit{rightEdge}_{\sigma}(y)] \\ & \wedge \neg\exists Y[\mathit{interval}(Y) \wedge x \prec_T Y \prec_T y]], \end{aligned}$$

where $\mathit{leftwardInterval}(X)$ is an interval that starts from a non-starred syllable at the left edge, ends at a starred syllable at the right edge, and there shouldn't exist any other interval Y between x and y. $\mathit{rightwardInterval}(X)$ is an interval that starts from a starred syllable at the left edge, ends at a non-starred syllable at the right edge, and there shouldn't exist any other interval Y between x and y. These two predicates are illustrated in Figure 6.15.

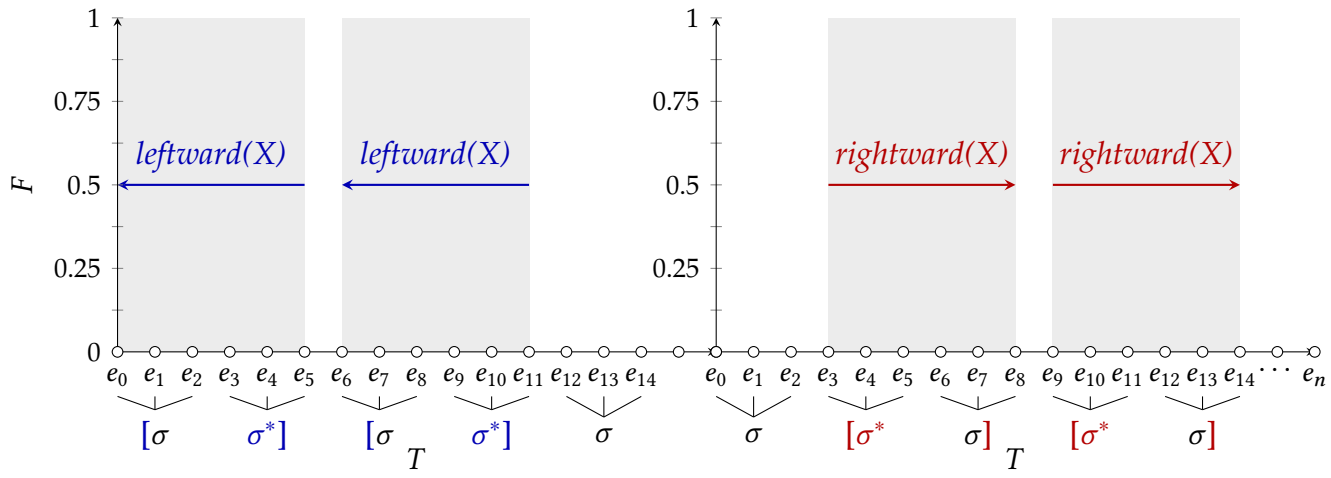


Figure 6.15: A schematic representation of *leftwardInterval(X)* and *rightwardInterval(X)*.

Discrete tonal targets Now that we have defined a search space for the pitch accents, *peaks* for the H targets and *troughs* for the L targets can be defined from an individual timepoint with the maximum f0 value and with the minimum f0 value within an interval, respectively. For this, we first need predicates that find the maximum or minimum f0 value on an interval, using the ordering relations $\prec_F$ and $\succ_F$ that refer to f0 values. The predicates are defined as follows:

$$max(x, X) \stackrel{\text{def}}{=} interval(X) \wedge X(x) \wedge \forall y[X(y) \rightarrow \neg\exists y\ [x \prec_F y \wedge x \neq y]]$$

$$min(x, X) \stackrel{\text{def}}{=} interval(X) \wedge X(x) \wedge \forall y[X(y) \rightarrow \neg\exists y\ [x \succ_F y \wedge x \neq y]],$$

where $max(x, X)$ is true iff the set X is an interval, x is an individual timepoint in that set X, and for all ys in the set X, there isn't any timepoint y whose f0 value is greater than or does not equal to the x's f0 value. Simply speaking, $max(x, X)$ refers to a maximum f0 point x in an interval X. Conversely, $min(x, X)$ is true if the set X is an interval, x is an individual timepoint in the set X, and for all ys in the set X, there isn't any timepoint y

whose f0 value is less than or does not equal to the x's f0 value. That is, $min(x, X)$ refers to a minimum f0 point x in an interval X.

Using these $max(x, X)$ and $min(x, X)$ predicates, *local peaks* and *troughs* are defined within a specific interval, leftward or rightward:

$$localPeak(x, X) \stackrel{\text{def}}{=} leftwardInterval(X) \vee rightwardInterval(X) \wedge max(x, X)$$

$$localTrough(x, X) \stackrel{\text{def}}{=} leftwardInterval(X) \vee rightwardInterval(X) \wedge min(x, X),$$

where $localPeak(x, X)$ is true if X is either a leftward or a rightward interval, and x is an individual timepoint whose f0 value is maximum within the interval X, while $localTrough(x, X)$ is true if X is either a leftward or a rightward interval, x is an individual timepoint whose f0 value is minumum within the interval X. As illustrated in Figure 6.16, for example, these two predicates can find the discrete tonal targets, Hs and Ls, within the leftward intervals.

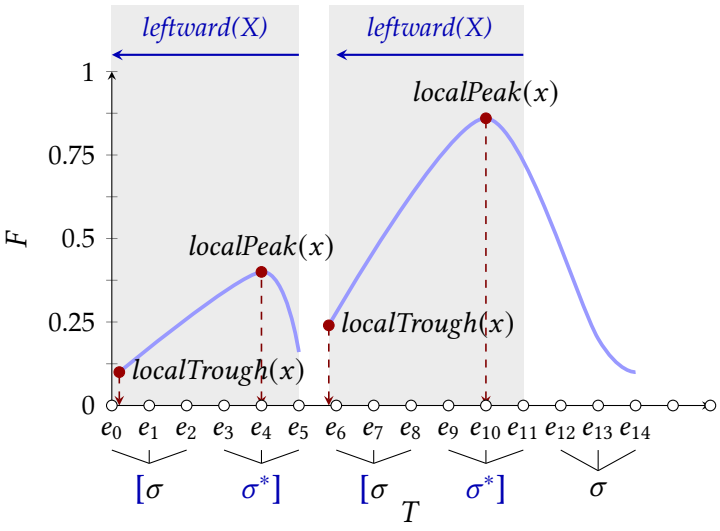


Figure 6.16: A schematic representation of $localPeak(x)$ and $localTrough(x)$.

Contour shapes and dynamics Lastly, the shapes and dynamics of an f0 contour can be defined by specifying the properties of the *sets* of timepoints in an interval. The rising and falling contour shapes can be represented with an increase or decrease of the f0 values, with respect to an increase in time within an interval, as defined with the following predicates:

$$\begin{aligned} \mathit{rise}(X) &\stackrel{\text{def}}{=} \mathit{leftwardInterval}(X) \vee \mathit{rightwardInterval}(X) \\ &\wedge \forall x, y[X(x) \wedge X(y) \wedge [x \prec_T y \rightarrow x \prec_F y]] \\ \mathit{fall}(X) &\stackrel{\text{def}}{=} \mathit{leftwardInterval}(X) \vee \mathit{rightwardInterval}(X) \\ &\wedge \forall x, y[X(x) \wedge X(y) \wedge [x \prec_T y \rightarrow x \succ_F y]], \end{aligned}$$

where $\mathit{rise}(X)$ is an interval X that is either a leftward or a rightward interval, and for all x, y in the set X, iff the timepoint x precedes the timepoint y, the f0 value of x is always greater than that of y. In contrast, $\mathit{fall}(X)$ is an interval X that is either a leftward or a rightward interval, and for all x, y in the set X, iff the timepoint x precedes the timepoint y, the f0 value of x is always less than that of y. These predicates are illustrated in Figure 6.17.

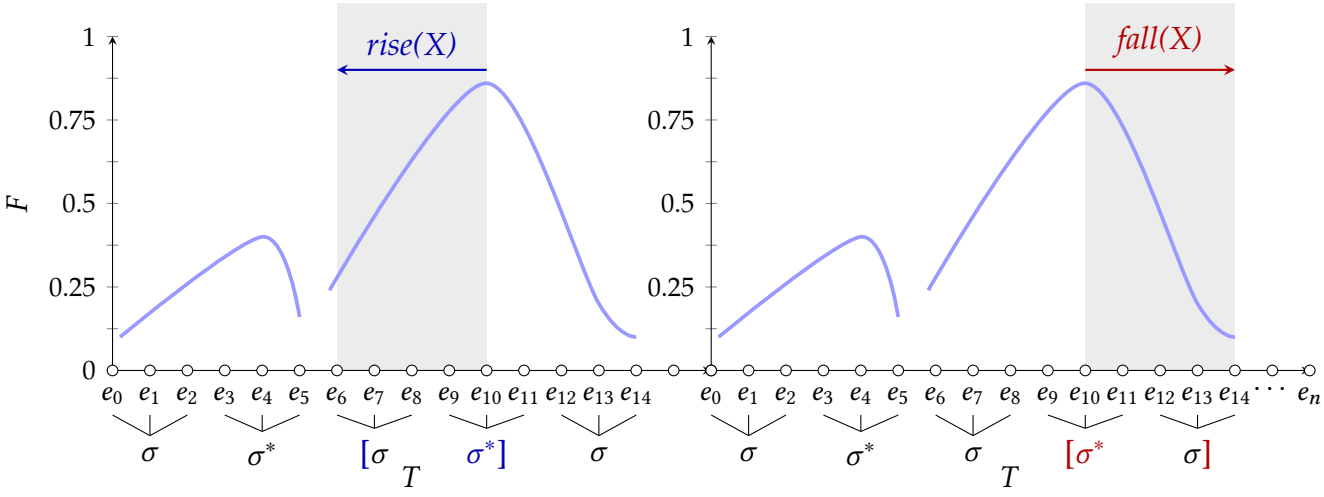


Figure 6.17: A schematic representation of *rise(X)* and *fall(X)*.

Importantly, more detailed information about an f0 contour shape, whether it is more steep or gradual, can be defined. Recall that in the input signature $\mathcal{S}_i$, we have the properties of P_{steep} and $P_{gradual}$. By using these shape primitives, we can represent steepness or gradualness with the following predicates:

$$steep(X) \stackrel{\text{def}}{=} interval(X) \wedge \forall x[X(x) \rightarrow P_{steep}(x)]$$

$$gradual(X) \stackrel{\text{def}}{=} interval(X) \wedge \forall x[X(x) \rightarrow P_{gradual}(x)],$$

where $steep(X)$ is an interval X such that for all timepoint xs in the set X, x has the property of steep, while $gradual(X)$ is an interval X such that for all timepoint xs in the set X, x has the property of gradual. These predicates are illustrated in Figure 6.18 and Figure 6.19, respectively.

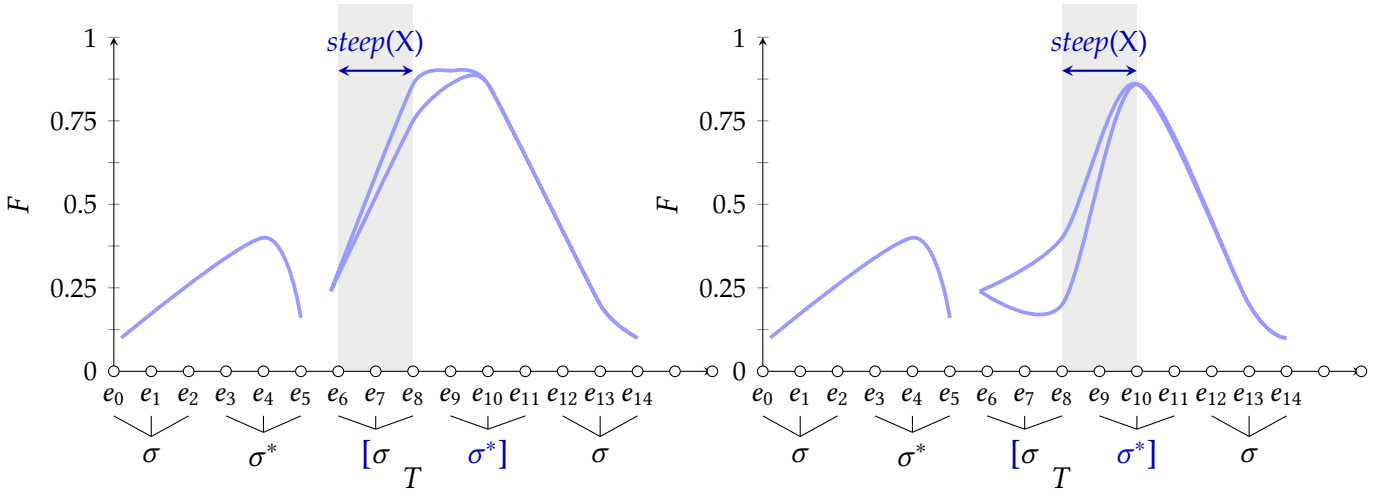


Figure 6.18: A schematic representation of *steep*(X).

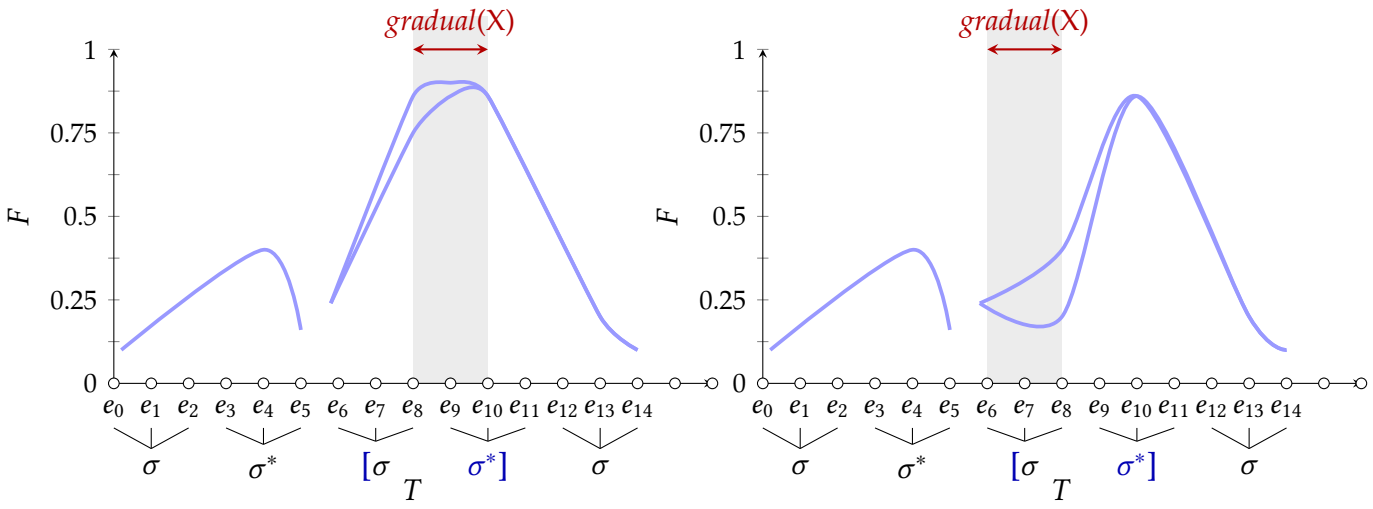


Figure 6.19: A schematic representation of *gradual*(X).

Finally, more dynamic aspects of f0 shape information, such as scooped or domed, can be represented using the set predicates, *steep*(X) and *gradual*(X). An ordering relation between these set predicates is used to represent the scooped or domed shapes. So, for the scooped shape, *gradual*(X) comes before *steep*(X), while for the domed shape, in a reverse order. For this, a new precedence relation between two sets is introduced: $\prec_S \stackrel{\text{def}}{=} \forall x, y[X(x) \wedge Y(y) \wedge (x \prec_T y)]$, where for all x and y, the timepoint x in the set X should precede the timepoint y in the set Y.

$$scooped(X) \stackrel{\text{def}}{=} leftwardInterval(X) \vee rightwardInterval(X) \wedge \exists Y, Z[\forall y, z[(Y(y) \rightarrow X(y)) \wedge (Z(z) \rightarrow X(z)) \wedge gradual(Y) \wedge steep(Z) \wedge Y \prec_S Z)]$$

$$domed(X) \stackrel{\text{def}}{=} leftwardInterval(X) \vee rightwardInterval(X) \wedge \exists Y, Z[\forall y, z[(Y(y) \rightarrow X(y)) \wedge (Z(z) \rightarrow X(z)) \wedge steep(Y) \wedge gradual(Z) \wedge Y \prec_S Z],$$

where *scooped*(X) refers to either a leftward or rightward interval X that consists of subintervals Y and Z such that Y with the property of gradual should come before Z with the

property of steep, whereas *gradual*(*X*) refers to a leftward or rightward interval *X* that is composed of subintervals *Y* and *Z* such that *Y* with the steep property always precedes *Z* with the gradual property. These two predicates characterize an interval in which a slice of an interval is gradual and then steep, or vice versa, as illustrated in Figure 6.20.

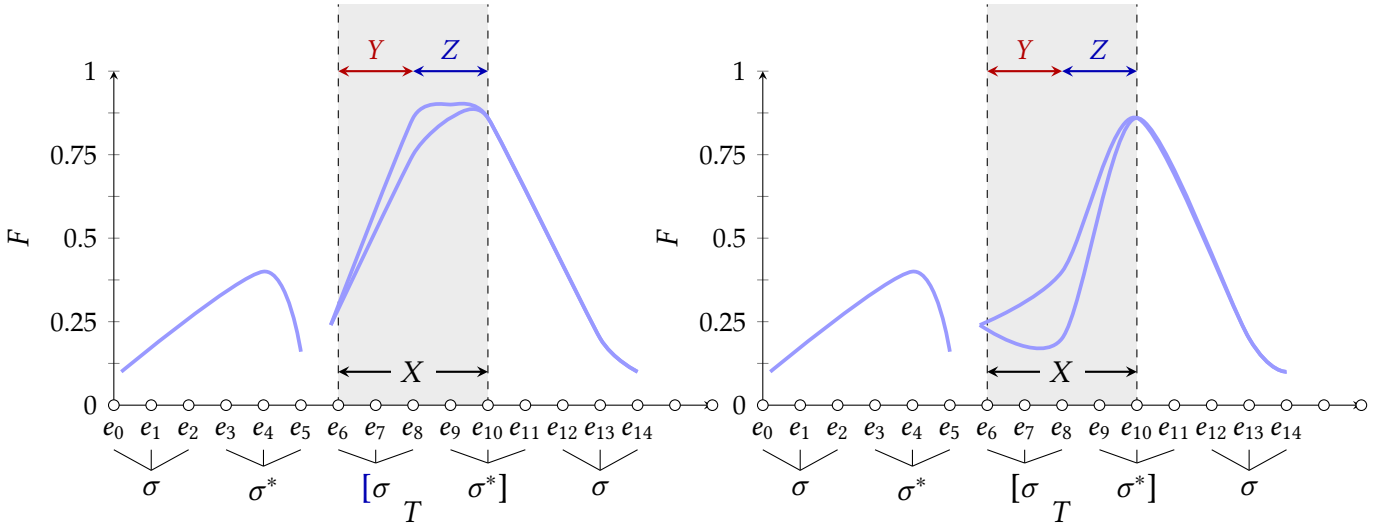


Figure 6.20: A schematic representation of *domed*(X) and *scooped*(X).

6.2.2 Definition of the pitch accents in American English

Pitch accents in American English can be mathematically defined with a two-step binary decision-making process. Importantly, this chapter focuses on determining the pitch accents into H* or L+H*. In the first step, the starred tones are defined using Boolean logic, FO and MSO logic, as proposed in 6.2.2.1. If the definition of the first step is satisfied, it is extended to fuzzy logic, where the unstarred tones and the gradient aspect of an f0 contour can be defined, as proposed in 6.2.2.2.

6.2.2.1 Step 1: Defining the pitch accents using Boolean logic

By using the basic predicates defined in 6.2.1.2, a pitch accent predicate, $T + H^*(x)$, defines all possible varying f0 contours of both H* and L+H*. Basically, this predicate is represented with a bitonal form, which consists of the specified starred tone (H*) and the unspecified leading tone (T). The predicate first defines a required H tone target (i.e., local peak) on a starred syllable. Then, it defines an optional preceding interval before the starred tone that has the properties of rise and scooped. Also, it adds an optional preceding local trough before the starred tone within that interval. Therefore, the pitch accent predicate for T+H* is defined as follows:

$$\begin{aligned} T + H^*(x) = {} & P_{\sigma^*}(x) \wedge \mathit{localPeak}(x, X) \\ & \wedge \exists X[\mathit{leftwardInterval}(X) \wedge \mathit{rise}(X) \wedge \mathit{scooped}(X) \\ & \wedge \exists y[P_{\sigma}(y) \wedge \mathit{localTrough}(y, X) \wedge y \prec_T x]], \end{aligned}$$

where $T + H^*(x)$ is true if the timepoint x is a starred syllable and a local peak in an interval X, there exist any X that is a leftward interval and has the properties of rise and scooped, and there exist an individual timepoint y that is a non-starred syllable and a preceding local trough before x in that inverval X. Crucially, this predicate determines whether a peak occurs at a metrically strong position and whether it is preceded by a trough that occurs at a non-metrically strong position.

A complete version of the pitch accent definition including the definition of other pitch accent predicates, $H^* + T(x)$, $L^* + T(x)$, and $T + L^*(x)$ is provided in Appendix B.

6.2.2.2 Step 2: Extending Boolean logic to fuzzy logic

After defining the pitch accents using Boolean logic in Step 1, the function $PitchAccent_{T+H^*}(x)$ converts $T + H^*(x)$ into a fuzzy predicate, $\mu_{T+H^*}(x)$, as follows:

$$PitchAccent_{T+H^*}(x) = \begin{cases} \mu_{T+H^*}(x), & \text{if } T + H^*(x) = 1 \\ 0, & \text{if } T + H^*(x) = 0, \end{cases}$$

where $PitchAccent_{T+H^*}(x)$ returns $\mu_{T+H^*}(x)$ if $T+H^*(x)$ is true, otherwise false. This $PitchAccent_{T+H^*}$ function extends the pitch accent definition defined within Boolean logic to that within fuzzy logic.

If $T + H^*(x)$ is true, $\mu_{T+H^*}(x)$ is defined as the following:

$$\mu_{T+H^*}(x) = \sup_{x:f(x_1,\cdots,x_6)=x} \Big(\min\big(\mu_{\sigma^*}(x_1), \mu_{localPeak}(x_2), \mu_{leftwardInterval}(x_3), \mu_{rise}(x_4), \mu_{scooped}(x_5), \mu_{localTrough}(x_6)\big)\Big),$$

where the membership values of $\mu_{\sigma^*}(x_1), \mu_{\text{localPeak}}(x_2), \mu_{\text{leftwardInterval}}(x_3), \mu_{\text{rise}}(x_4), \mu_{\text{scooped}}(x_5), \mu_{\text{localTrough}}(x_6)$ are combined into a minimum value, which is the supreme over all possible combinations of $x_1, \cdots, x_6$, where $f(x_1, \cdots, x_6) = x$. Among the predicates, $\mu_{\sigma^*}(x_1)$, $\mu_{\text{localPeak}}(x_2)$, $\mu_{\text{leftwardInterval}}(x_3)$, $\mu_{\text{scooped}}(x_5)$ are computed with Boolean values, 0 or 1, but for $T + H^*$, these predicates are all 1. But the fuzzy logic comes into play for $\mu_{rise}(x_4)$ and $\mu_{localTrough}(x_6)$, such that they are computed with gradient values between 0 and 1. Therefore, the output value of $\mu_{T+H^*}(x)$ can be boiled down into $\min\Big(\mu_{rise}(x_4), \mu_{localTrough}(x_6)\Big)$.

The calculation of $\mu_{rise}(x_4)$ and $\mu_{localTrough}(x_6)$ will be done through the fuzzy logical system, using the actual f0 values. The final output will be a combined value of the membership degree, which is determined between 0 and 1. Based on this output value, we can determine whether this output can be classified as H* or L+H*.

The following section shows how these two predicates are computed with actual f0 values within the fuzzy logical system.

6.3 Implementing a fuzzy logical system with American English pitch accent data

To extend the pitch accent predicates in American English defined in Boolean logic to those in fuzzy logic, the annotated real-world f0 data of American English was used to implement the fuzzy logical system. The actual f0 values for the two predicates, $\mu_{localTrough}(x)$ and $\mu_{rise}(x)$, were used as the input values for the fuzzy logical system. They were first fuzzified based on the fuzzy sets, $\mu_{localTrough}$ and μ_{rise}, and combined into a numeric value based on the linguistic rules formulated based on the f0 data. Each numeric output value was categorized as H* vs. L+H* based on the output fuzzy sets, which were compared with the perceptual judgement data. Lastly, the classification accuracy of the fuzzy logical system was measured and compared with several machine learning models to validate its performance.

6.3.1 Methods

6.3.1.1 Data

The actual f0 values for H* and L+H* pitch accents in American English were obtained from Steffman et al. (2024)'s production data[12]. The production data consists of 872 utterances from 56 American English speakers. It also included the perceptual judgement for each token, whether the accent is either H* or L+H*, annotated by the authors in Steffman et al. (2024).

The recorded sentences were "*She remained with Madelyn*" and "*They honored Melanie*", as shown in Table 6.3. The region of interest was the pitch accents of the last three syllables in the frame sentences, "*Madelyn*" and "*Melanie*", which were accented with either H* or L+H*. Due to the discontinuity of f0 contours, 130 tokens were discarded. A total of 742 tokens were used in the analysis of the fuzzy logical system of American English pitch accents.

	Test Sentence	**Target (*H** vs. *L* + *H**)**
(1)	*She remained with*	*Madelyn*.
(2)	*They honored*	*Melanie*.

Table 6.3: The target words within the test sentences.

6.3.1.2 Measurement

To compute the two predicates, $\mu_{localTrough}(x_6)$ and $\mu_{rise}(x_4)$, the actual f0 values for these predicates were measured and calculated within the target word for each token, which

[12]This data can be found in Steffman, Cole & Shattuck-Hufnagel's OSF repository with the following link: `https://osf.io/ehx7w/files/osfstorage`.

includes both H* and L+H*, as can be seen in Figure 6.21 and Figure 6.21, respectively. As for the $\mu_{localTrough}$ predicate, the minimum f0 value within the target was measured from each token, using the minimum function ("Get minimum pitch") in Praat (Boersma 2009). Moreover, the f0 range for each speaker was measured from the entire test sentences to define the range of the fuzzy set $\mu_{localTrough}$ (i.e., the universe of discourse, $\mathcal{U}_6$).

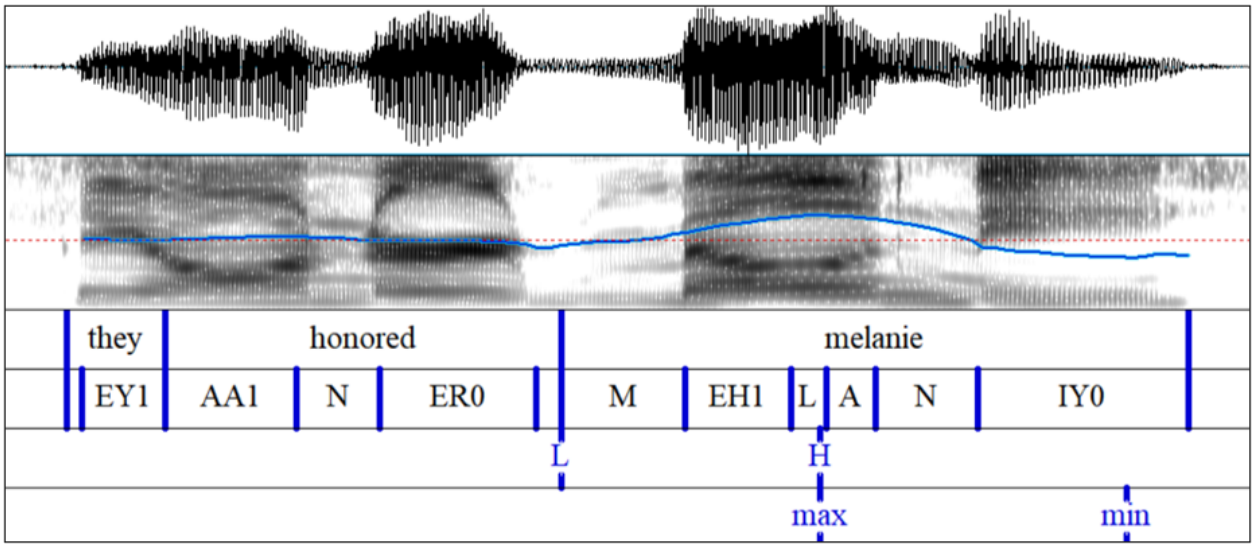


Figure 6.21: An example f0 contour for H*. *L* and *H* here indicate the minimum and maximum f0 values within the target word, respectively. *min* and *max* indicate the minimum and maximum f0 values within the entire sentence, respectively.

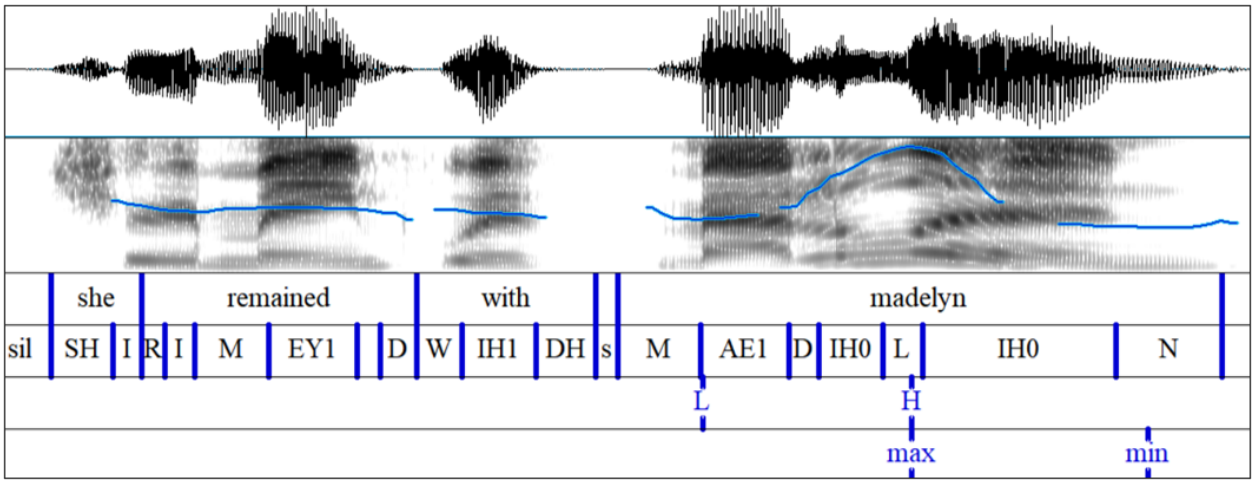


Figure 6.22: An example f0 contour for L+H*. *L* and *H* here indicate the minimum and maximum f0 values within the target word, respectively. *min* and *max* indicate the minimum and maximum f0 values within the entire sentence, respectively.

As for the $\mu_{rise}(x_4)$ predicate, both the minimum and maximum f0 values and their timepoints were measured within the targer word for each token using the minimum and maximum ("Get maximum pitch") functions in Praat. By using these f0 values and timepoints, the slope for each token was calculated with the following formula:

$$slope = \frac{f0max - f0min}{timepoint_{f0max} - f0min\text{-}timepoint_{f0min}}$$

Then, the slopes were normalized for each speaker to compare across speakers with the following formula (Rose 1987):

$$slope_{norm} = \frac{slope - \min(slope)}{\max(slope) - \min(slope)}$$

6.3.1.3 Rule formulation

Before implementing the measured f0 values for $\mu_{localTrough}$ and μ_{rise} into the fuzzy logical system of American English pitch accents, *if...then* rules were formulated to evaluate the input f0 values at the rule inference stage of the fuzzy logical system. These rules were derived from the distribution of the f0 values for $\mu_{localTrough}$ and μ_{rise}, with the perceptual judgment of either H* or L+H*, as shown Figure 6.23.

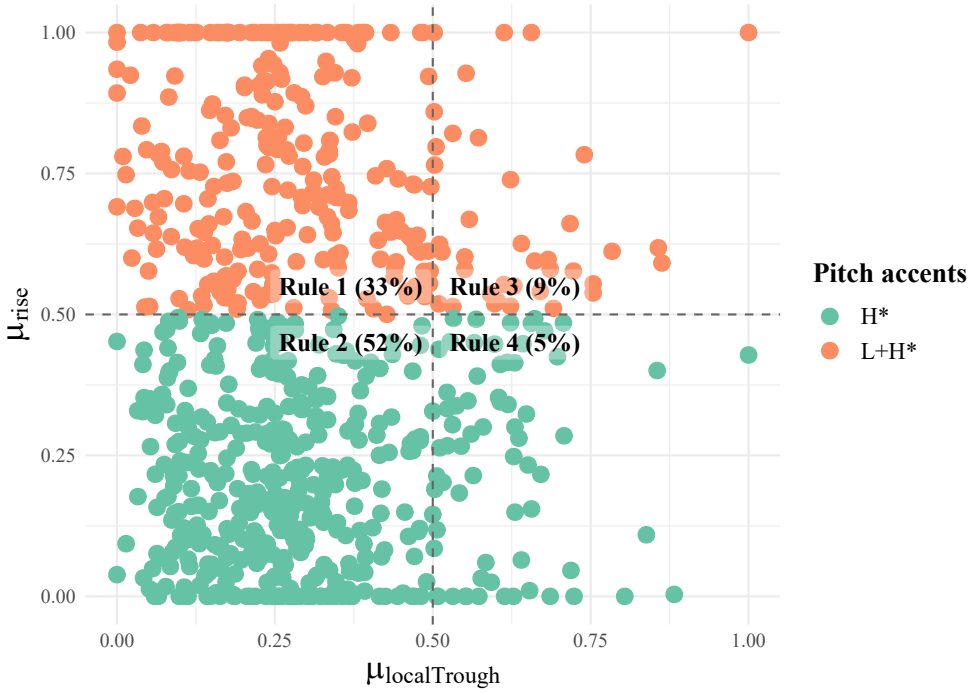


Figure 6.23: The H* and L+H* pitch accent distribution of $\mu_{localTrough}$ by μ_{rise}. The x-axis represents $\mu_{localTrough}$, while the y-axis represents μ_{rise}. The green dots indicate H* pitch accents, while the orange dots indicate L+H* pitch accents.

First, the upper-left quadrant in Figure 6.23 shows lower local troughs and steeper slopes, which were mainly categorized as L+H*. This means that the existence of a low tone before a peak should contribute to being identified as L+H*. This quadrant, which accounts for 33% of the data, leads to the formulation of Rule 1: *If* $\mu_{localTrough}$ *is Low &* μ_{rise}

is Steep, then the output is L+H.*

Second, the lower-left quadrant shows lower local troughs and more gradual slopes, which were mainly classified as H*. This could be the case where the preceding f0 context is very low, and there is a slight rise for H*. This suggests that if the slope is not sufficiently steep, it is difficult to be categorized as L+H*. This quadrant, which accounts for 52% of the data, leads to the formulation of Rule 2: *If $\mu_{localTrough}$ is Low & μ_{rise} is Gradual, then the output is H*.*

Third, the upper-right quadrant shows higher local troughs and steeper slopes, which were mainly categorized as L+H*. Despite the higher f0 value of the preceding troughs, this quadrant satisfies the steeper slopes, which is the primary property of L+H*, leading to L+H* responses. This quadrant accounts for 9% of the data, resulting in Rule 3: *If $\mu_{localTrough}$ is High & μ_{rise} is Steep, then the output is L+H*.*

Lastly, the lower-right quadrant shows higher local troughs and more gradual slopes, which were mainly categorized as H*. This is similar to Rule 2, so that no matter how low the local trough is, if the slope is not steep enough, then it is less likely to be categorized as L+H*. This quadrant accounts for 5% of the data, outputting Rule 4: *If $\mu_{localTrough}$ is High & μ_{rise} is Gradual, then the output is H*.*

Hence, the following rules will be used for distinguishing H* versus L+H* in the fuzzy logical system in the following section:

1. If $\mu_{localTrough}$ is Low & μ_{rise} is Steep, then the output is L+H*.
2. If $\mu_{localTrough}$ is Low & μ_{rise} is Gradual, then the output is H*.
3. If $\mu_{localTrough}$ is High & μ_{rise} is Steep, then the output is L+H*.

4. If $\mu_{localTrough}$ is High & μ_{rise} is Gradual, then the output is H*.

6.3.2 Analysis: Fuzzy logical system

The minimum f0 values for $\mu_{localTrough}$ and the slope values forμ_{rise} measured from the production data were used in the fuzzy logic system, as shown in Figure 6.24. This is implemented in the Fuzzy Logic Toolbox (The MathWorks Inc. 2024) in MATLAB (The MathWorks Inc. 2022). Note that each token went through this fuzzy logical system, resulting in a total of 742 systems being run. While this system is designed based on the Fuzzy Logic Toolbox in MATLAB, due to the large amount of data, it was implemented using the Python Scikit-Fuzzy (Warner et al. 2024).[13]

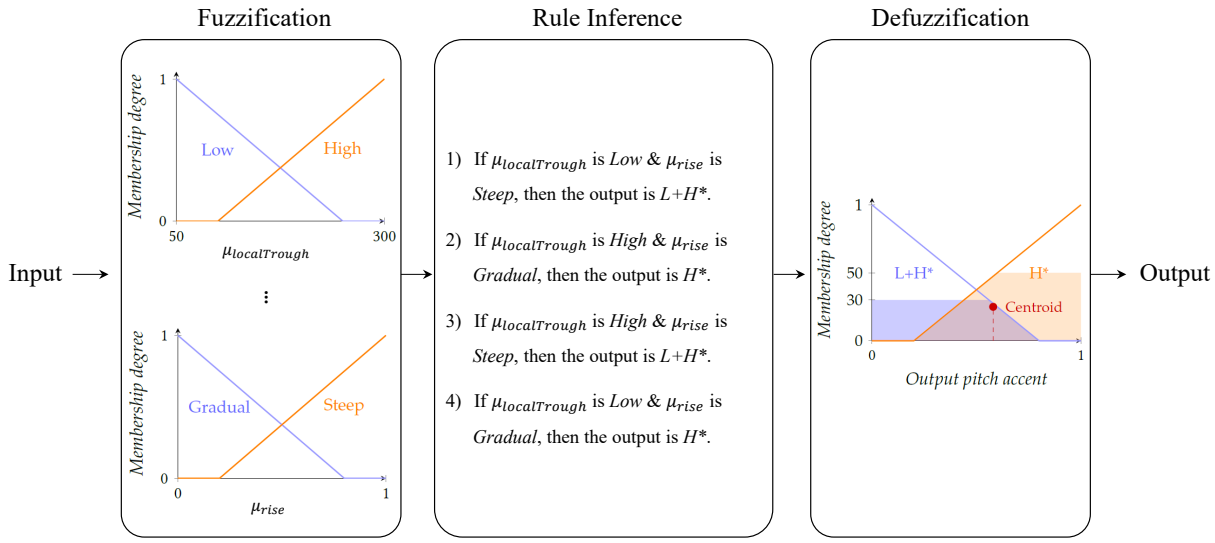


Figure 6.24: The fuzzy logic system for the pitch accent categorization of American English.

To distinguish American English pitch accents H^* and $L+H^*$, two linguistic variables, $\mu_{localTrough}$ and μ_{rise}, were defined in the fuzzy logical system. $\mu_{localTrough}$ consists of two

[13]The Python code is available in the GitHub repository of this dissertation: `https://github.com/hyunjungjoo/fuzzy-logic`.

fuzzy sets, {low, rise}, as shown in Figure 6.25, while μ_{rise} consists of two fuzzy sets, {gradual, steep}, as in Figure 6.26.

As for $\mu_{localTrough}$, the universe of discourse $\mathcal{U}_6$ was each participant's pitch range, while as for μ_{rise}, the universe of discourse $\mathcal{U}_4$ was the normalized slope values ranging from 0 to 1 for each participant.

In this dissertation, triangular shapes were used to define the fuzzy sets for the sake of simplicity, since they can only define the membership degree with the smallest parameters (i.e., only two lines) (Zadeh 2023). But other types of fuzzy sets, such as trapezoidal and Gaussian fuzzy sets, were also tested, resulting in comparable results to the triangular fuzzy sets. Examples of trapezoidal and Gaussian fuzzy sets and their accuracy rate for the pitch accent classification are provided in Appendix C.

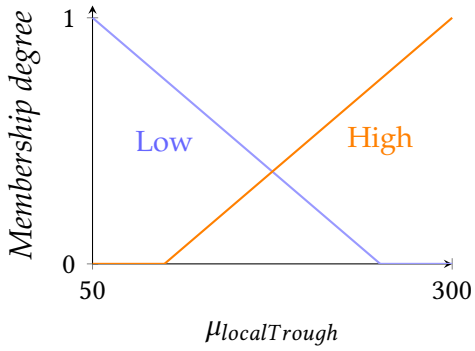


Figure 6.25: Fuzzy input: local trough

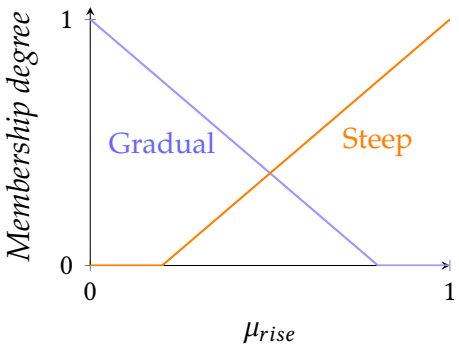


Figure 6.26: Fuzzy output: slope

At the *fuzzification* stage, the lowest f0 values and slope values in the inputs were interpreted in terms of the fuzzy sets, $\mu_{localTrough}$ and μ_{rise}, respectively. For $\mu_{localTrough}$, the degree of membership was defined based on an f0 value of the local trough within each participant's pitch range, while for μ_{rise}, it was defined based on a normalized slope value.

At the *rule inference* stage, the fuzzified values were calculated based on the linguistic

rules formulated in 6.3.1.3.

At the final stage, *defuzzification*, a numeric output value is calculated based on the fuzzy set of the output pitch accent, as shown in Figure 6.27. This fuzzy set consists of {L+H*, H*} in the discourse of universe $\mathcal{U}_o$ ranging from 0 to 1. The regions below the fuzzified values for each fuzzy subset are combined and then used to calculate a centroid. If the output values are below 0.5, they are considered to be classified as L+H*, while above 0.5 as H*.

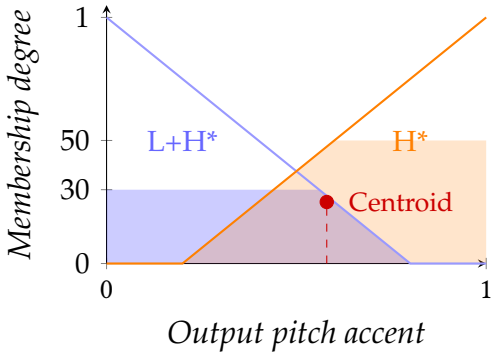


Figure 6.27: Output pitch accent: L+H* versus H*.

Figure 6.28 shows how we can get the output values based on the actual f0 minimum and slope values in the Fuzzy Logic Toolbox in MATLAB. In the left panel, if we put the input values, 184 for $\mu_{localTrough}$ and 0.202 for μ_{rise}, the output pitch accent value is 0.677, which can be categorized as H*. In contrast, for the input values, 189 for $\mu_{localTrough}$ and 1 for μ_{rise}, the output value is 0.399, being categorized as L+H*.

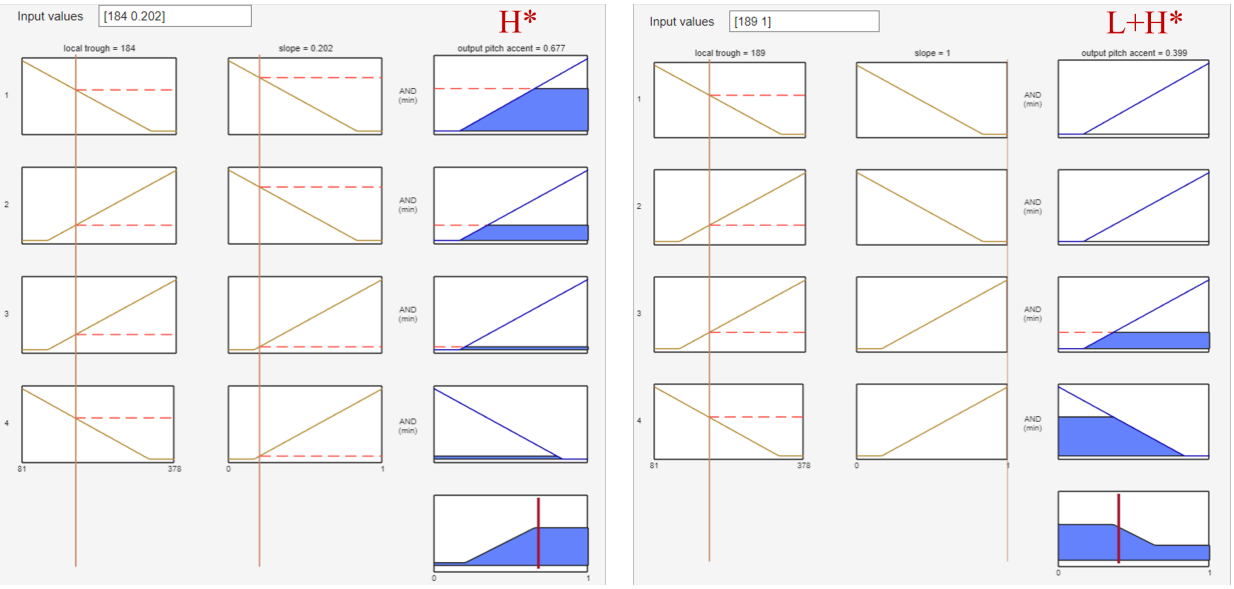


Figure 6.28: A screenshot of the *defuzzification* stage in the Fuzzy Logic Toolbox in MATLAB. The output pitch accent values obtained by combining $\mu_{localTrough}$ and μ_{rise}.

In the following section, the output pitch accents (< 0.5 as L+H* versus > 0.5 as H*) will be compared with the perceptual judgments (H* versus L+H*) from Steffman et al. (2024) to determine how accurately the fuzzy logical system predicts the pitch accent categorization.

6.3.3 Results

The results showed an overall mean accuracy of 81.9% in predicting the output pitch accents as H* versus L+H* using the fuzzy logical system of American English pitch accents, as shown in Figure 6.29. That is, the two linguistic variables, $\mu_{localTrough}$ and μ_{rise}, were able to account for the pitch accent categories, H* versus L+H*, correctly. Figure 6.29 also shows the classification accuracy of all participants was above chance level. These results indicate that the proposed fuzzy logical model performs well in categorizing the pitch

accents in American English.

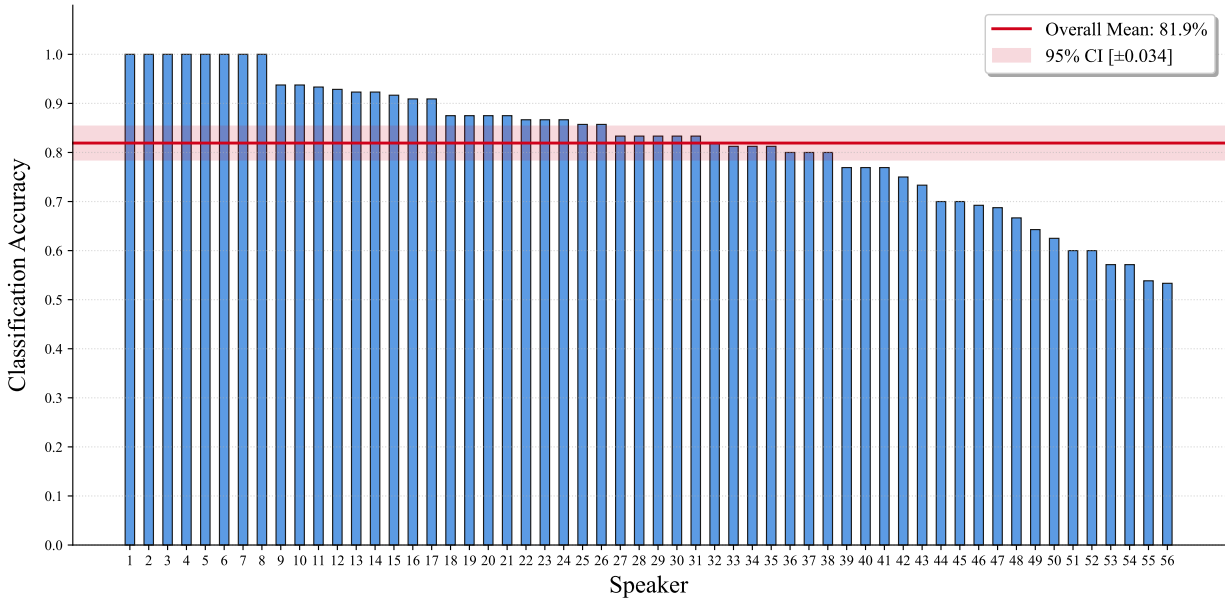


Figure 6.29: Classification accuracy across speakers. The red line indicates the overall mean, 81.9%, and the shaded region around it indicates the 95% confidence interval.

6.3.4 Model comparison

To evaluate the performance of the fuzzy logical system in American English pitch accents, the classification accuracy was compared with several machine learning models. For the machine learning models, logistic regression, decision tree, support vector machine, and random forest were chosen, which are used as standard methods for comparing model performance in the field of machine learning (e.g., Caruana & Niculescu-Mizil 2006, Fernández-Delgado et al. 2014, Strobl et al. 2009).

In order to obtain classification accuracy from these models, each model went through the training and testing phase. Note that the entire dataset was randomly splitted into training and test sets at an 8:2 ratio. The predicted values (i.e., H* versus L+H*) were compared with the annotated pitch accents from the test set. For reliable results, this process

was repeated 30 times using different random seeds, and the final average performance was reported.

The hyperparameters of each model were set as the following: As for the logistic regression, the model was trained without any hyperparameter tuning. As for the decision tree, the model was implemented with a minimum of 10 observations required to split an internal node (minsplit = 10) and a minimum of 7 observations in any terminal node (minbucket = 7). The complexity parameter (cp) was set to 0.01 to prevent overfitting, and the maximum tree depth was set to 30 (maxdepth = 30). For the support vector machine, radial basis function kernel was used in training. The cost parameter (C), which determines the penalty for misclassification was set to 1. Lastly, random forest was built with 100 trees (ntree = 100). At each node, the number of randomly selected candidate features was set to the $\sqrt{p}$. The minimum size of terminal nodes was set to 1.

Now let us look at the results of the model comparison. As shown in Figure 6.30, the fuzzy logical system (81.9%) performed well, comparable to other machine learning models, logistic regression (82.77%), decision tree (80.99%), support vector machine (82.59%), and random forest (80.16%). It is important to note that the fuzzy logical system of American English pitch accent is based on the linguistic rules that explicitly explain the logic of the pitch accent classification, while the machine learning models do not provide such explicit linguistic rules. By showing results very similar to those of the data-driven machine learning models, this logical system proved that it effectively captured the pitch accent categorization in American English.

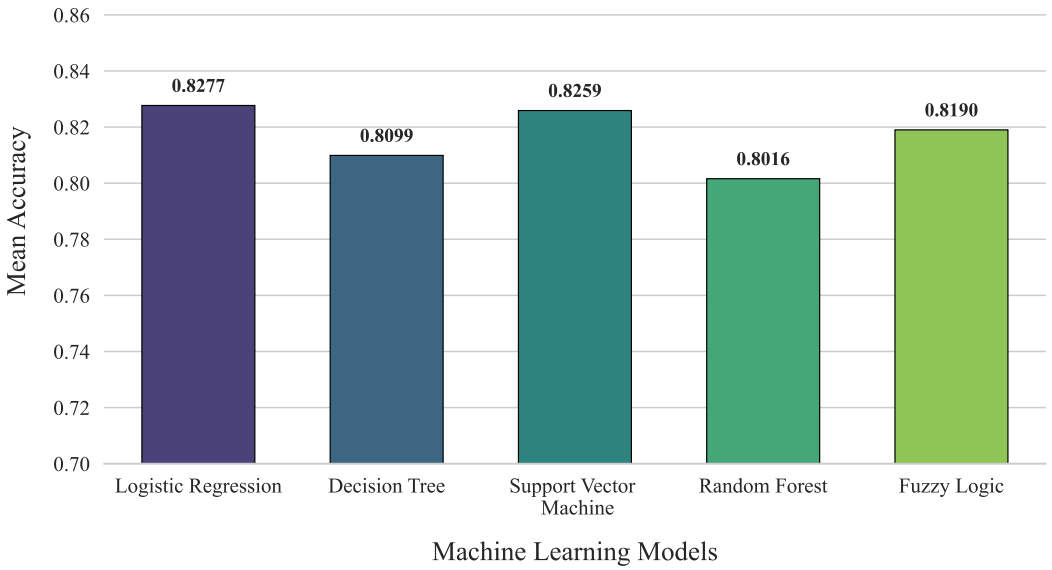


Figure 6.30: Comparison of the proposed fuzzy logical system with machine learning models.

6.4 Summary and Conclusion

In this chapter, pitch accents in American English were mathematically defined by connecting discrete and continuous f0 information of the pitch accents in American English, using two different types of logic, *Boolean* and *fuzzy logic*, within the model-theoretic framework. Based on First-Order and Monadic Second-Order logic, both discrete tonal targets and continuous f0 information were defined with perceptual primitives of intonation over the temporal domain, but only referencing f0 information. Then, the definitions were connected with the annotated f0 data by extending Boolean logic to fuzzy logic. The results showed an accuracy of 81.9% using those definitions, comparable to the performance of several machine learning models. By combining Boolean and fuzzy logic, we were able to *bridge* the discrete and continuous information of intonation.

Chapter 7

Discussion and Conclusion

7.1 Discussion

This dissertation aimed to investigate how discrete and continuous f0 information is encoded in the representation and computation of intonation using mathematical logic and speech perception data. Specifically, the following research questions have been examined: 1) What kind of f0 information is crucial for defining intonation in the phonological representation?; 2) How can we mathematically define discrete tonal targets, Hs and Ls, in intonation?; 3) How can we connect the discrete and continuous f0 information in intonation? The chapters in this dissertation answered these questions by providing experimental data and formalizing the intonational primitives.

For the first research question, a perceptual experiment was conducted to figure out the perceptual primitives in the lexical pitch accent in South Kyungsang Korean. In order to test two representational approaches, the AM and configurational models, the f0 peak timing (discrete tonal targets) and the f0 rise shape (continuous f0 information) were used

for distinguishing H versus LH pitch accent in South Kyungsang Korean. Although the results showed the significant role of the f0 rise shape for the pitch accent distinction, the results for the f0 peak timing were difficult to interpret. This requires further investigation of the subtle f0 variabilities, such as the initial f0 value or the amount of f0 dynamic that need to be included in the TBU. Also, further studies are needed to generalize these results to other lexical pitch accent languages.

Second, the intonational structures were mathematically defined with discrete tonal targets, Hs and Ls, using logical transductions. The results showed that intonation can be viewed as a QF logical interpretation of a metrical and prosodic structure. This QF characterization enabled us to impose strong restrictions on the representation and computation of intonation, just like other phonological processes that fall in the regular upper bound of phonology.

However, it should be noted that this dissertation only focused on the basic phonological patterns of intonation, by only dealing with the tone-TBU association, but not including the variabilities of the actual f0 realization. For example, the pitch accents may not always be realized at the prominent syllables, but sometimes the f0 peak can be delayed to the following syllables depending on the f0 context and segmental information. Also, due to the tonal crowding of phrasal and boundary tone at the end of the phrase, the actual f0 realization may possibly show tonal undershoot. However, the QF logic used in the logical transductions is too powerful to account for these variabilities. In order to deal with the continuous f0 information, the MSO logic was used in the following chapter, but it needs further investigation to determine whether the MSO logic can account for those f0 variabilities.

Also, we should account for the prominence realization under focus, such as broad focus versus narrow (contrastive) focus. When the prominence-induced focus is introduced to a prosodic structure, the entire tone-TBU mappings may override other tone-TBU mappings. One of the possibilities is to use another transduction to account for the prominence-related focus, but it needs further investigation since this focus-induced transduction may not be restricted within the QF logic, but may need to be characterized with less restrictive logic.

Lastly, this dissertation connected the discrete and continuous f0 information of the pitch accents, H* versus L+H*, in American English, using two different types of logic, Boolean and fuzzy logic. It should be noted that by defining the signatures with $\prec_F$, we were able to only refer to the f0 ordering, without accounting for the entire continuous space of f0, which cost a lot of computational capacity. However, we haven't looked at other possible pitch accent patterns, such as H* versus H*+L, so we need further investigation by characterizing other pitch accents with their dynamic properties.

As for the definitions using Boolean logic, the intervals to search for the leading or trailing tones were defined based on FO and MSO logic. However, the leftward or rightward interval can only deal with the cases of non-prominent and prominent syllables, [σ σ^*] or [σ^* σ], but there could be some cases where two prominent syllables are right next to each other. This means that the definition of the interval needs to be further defined to allow for the variability of TBU contexts.

As for the definitions using fuzzy logic, the fuzzy logical systems were implemented for the pitch accent patterns in American English using the actual f0 data. We need to see whether the linguistic rules in the fuzzy logic are generalizable to other tonal patterns,

such as whether this system can account for the downstep patterns by properly defining the interval.

Moreover, we need to see whether the linguistic rules formulated in the fuzzy logical system are generalizable to other languages. A quick run for using H vs. LH data of a lexical pitch accent language, South Kyungsang Korean, showed that the American English pitch accent rules cannot be applied to this H vs. LH distinction. Also, the triangular shapes were used to define the fuzzy sets, since they were computationally simpler and showed comparable results with other trapezoidal and Gaussian shapes. However, further studies are needed to account for what kind of fuzzy set representation is needed to explain different kinds of linguistic variables.

Lastly, this dissertation offers a novel computational perspective on the representation of intonation by showing the interpretability between continuous and discrete f0 information. This may shed light on extending to other phenomena that need to be accounted for within the phonology-phonetics interface. This combined formal perspective may provide a way to explicitly and precisely define the interplay between phonetics and phonology.

7.2 Conclusion

This dissertation examined how discrete and continuous f0 information can be encoded in the representation and computation of intonation using mathematical logic and speech perception data. Specifically, this dissertation provided the definition of intonation by connecting both discrete and continuous information using model theory and logic.

This dissertation contributed to providing an important experimental finding that continuous f0 shape information plays a significant role in distinguishing the lexical pitch accents in South Kyungsang Korean. This supported existing findings that emphasize the importance of f0 shape information for phonological contrast.

Also, this dissertation showed a valuable result that intonation can be mathematically viewed as a quantifier-free logical interpretation of a metrical and prosodic structure. The discrete tonal targets were found to be literal copies of prosodic elements, such as accented syllables/moras or phrasal boundaries. Importantly, they were always linked locally to their tone-bearing units, which were similar to other phonological patterns that fall in the regular upper bound of phonology. Also, the typological patterns of intonation were explicitly captured by interpreting the copied heads and edges of a constituent directly or indirectly to the discrete tonal targets. This provided a new perspective on accounting for intonational typology.

Lastly, this dissertation connected discrete and continuous f0 information of the pitch accents in American English, using two different types of logic, Boolean and fuzzy logic, within the model-theoretic framework. Based on First-Order and Monadic Second-Order logic, both discrete tonal targets and continuous f0 information were defined with perceptual primitives of intonation over the temporal domain, but only referencing f0 information. Then, the definitions were connected with the annotated real-world f0 data of American English by extending Boolean logic to fuzzy logic. The results showed an accuracy of 81.9% using those definitions, comparable to the performance of several machine learning models.

To conclude, this dissertation offers a novel computational perspective on the repre-

sentation of intonation by showing the interpretability between continuous and discrete f0 information.

References

Arnhold, Anja. 2014. Prosodic structure and focus realization in west greenlandic. *Prosodic typology* 2.

Arvaniti, Amalia. 2009. Intonational primitives. *Companion to Phonology, Wiley-Blackwell* .

Arvaniti, Amalia. 2022. The autosegmental-metrical model of intonational phonology .

Arvaniti, Amalia, Gina Garding et al. 2007. *Dialectal variation in the rising accents of american english*. Mouton de Gruyter Berlin.

Arvaniti, Amalia, D Robert Ladd & Ineke Mennen. 1998. Stability of tonal alignment: the case of greek prenuclear accents. *Journal of phonetics* 26(1). 3–25.

Atterer, Michaela & D Robert Ladd. 2004. On the phonetics and phonology of "segmental anchoring" of f0: evidence from german. *Journal of phonetics* 32(2). 177–197.

Barnes, Jonathan, Alejna Brugos, Nanette Veilleux & Stefanie Shattuck-Hufnagel. 2021. On (and off) ramps in intonational phonology: Rises, falls, and the tonal center of gravity. *Journal of Phonetics* 85. 101020.

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2010. The effect of global f0 contour shape on the perception of tonal timing contrasts in american english intonation. *Speech Prosody 2010* 100445. 1–4.

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2012. Tonal center of gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2). 337–383.

Barr, Dale J, Roger Levy, Christoph Scheepers & Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* 68(3). 255–278.

Beckman, Mary E. 1988. *Japanese tone structure*. MIT Press.

Beckman, Mary E, Manuel Díaz-Campos, Julia Tevis McGory & Terrell A Morgan. 2002. Intonation across spanish, in the tones and break indices framework .

Beckman, Mary E & Julia Hirschberg. 1994. The tobi annotation conventions. *Ohio State University* .

Beckman, Mary E & Janet B Pierrehumbert. 1986. Intonational structure in japanese and english. *Phonology* 3. 255–309.

Boersma, Paul. 2009. Praat: doing phonetics by computer (version 5.1. 05). *http://www.praat. org/* .

Bolinger, Dwight L. 1951. Intonation: levels versus configurations. *Word* 7(3). 199–210.

Browman, Catherine P & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49(3-4). 155–180.

Calhoun, Sasha. 2004. Phonetic dimensions of intonational categories-the case of l+ h* and h .

Caruana, Rich & Alexandru Niculescu-Mizil. 2006. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on machine learning*, 161–168.

Chandlee, Jane. 2014. *Strictly local phonological processes*. University of Delaware.

Chandlee, Jane & Jeffrey Heinz. 2018. Strict locality and phonological maps. *Linguistic Inquiry* 49(1). 23–60.

Chandlee, Jane & Adam Jardine. 2019a. Autosegmental input strictly local functions. *Transactions of the Association for Computational Linguistics* 7. 157–168.

Chandlee, Jane & Adam Jardine. 2019b. Quantifier-free least fixed point functions for phonology. In *Proceedings of the 16th meeting on the mathematics of language*, 50–62.

Chandlee, Jane & Steven Lindell. 2021. Logical perspectives on strictly local transformations. *Doing Computational Phonology. Oxford University Press, forth* .

Chang, Seung-Eun. 2007. *The phonetics and phonology of south kyungsang korean tones*. The University of Texas at Austin.

Chang, Seung-Eun. 2013. Effects of fundamental frequency and duration variation on the perception of south kyungsang korean tones. *Language and speech* 56(2). 211–228.

Cho, Taehong. 2016. Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass* 10(3). 120–141.

Cho, Taehong, Dong Jin Kim & Sahyang Kim. 2019. Prosodic strengthening in reference to the lexical pitch accent system in south kyungsang korean. *The Linguistic Review* 36(1). 85–115.

Cohen, Antonie & JT Hart. 1968. On the anatomy of intonation. *Lingua* 19(1-2). 177–192.

Courcelle, Bruno. 1994. Monadic second-order definable graph transductions: a survey. *Theoretical Computer Science* 126(1). 53–75.

Crystal, David. 1969. Prosodic systems and intonation in english .

Czarnecki, Vincent. 2025. The logic of linearization: Interpretations of trees via strings. In *Proceedings of the society for computation in linguistics 2025*, 123–132.

Dainora, Audra. 2002. An empirically based probabilistic model of intonation in american english. .

D'Imperio, Mariapaola. 2000. *The role of perception in defining tonal targets and their alignment*. The Ohio State University.

D'Imperio, Mariapaola & David House. 1997. Perception of questions and statements in neapolitan italian. In *Eurospeech*, vol. 97 5, 22–25.

Do, Young Ah & Michael Kenstowicz. 2010. A note on phonological phrasing in south kyungsang .

Do, Youngah, Chiyuki Ito & Michael Kenstowicz. 2014. Accent classes in south kyengsang korean: Lexical drift, novel words and loanwords. *Lingua* 148. 147–182.

Dorokhova, Lydia & Mariapaola D'Imperio. 2019. Rise dynamics determines tune perception in french: The case of questions and continuations. In *International congres of phonetic science icphs2019*, .

Elordieta, Gorka. 1998. Intonation in a pitch accent variety of basque. *Anuario del Seminario de Filología Vasca" Julio de Urquijo"* 32(2). 511–569.

Enderton, Herbert B. 2001. *A mathematical introduction to logic*. Elsevier.

Engelfriet, Joost & Hendrik Jan Hoogeboom. 2001. Mso definable string transductions and two-way finite-state transducers. *ACM Transactions on Computational Logic (TOCL)* 2(2). 216–254.

Face, Timothy L. 2007. The role of intonational cues in the perception of declaratives and absolute interrogatives in castilian spanish. *Journal of Experimental Phonetics* 16. 185–225.

Fernández-Delgado, Manuel, Eva Cernadas, Senén Barro & Dinani Amorim. 2014. Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research* 15(1). 3133–3181.

Filiot, Emmanuel & Pierre-Alain Reynier. 2016. Transducers, logic and algebra for functions of finite words. *ACM SIGLOG News* 3(3). 4–19.

Goldsmith, John. 1981. English as a tone language. *Phonology in the 1980's* 287–308.

Goldsmith, John Anton. 1976. *Autosegmental phonology*: Massachusetts Institute of Technology dissertation.

Grabe, Esther, Greg Kochanski & John Coleman. 2004. Quantitative modelling of intonational variation .

Graf, Thomas. 2010. Comparing incomparable frameworks: A model theoretic approach to phonology .

Grice, Martine, Michelina Savino et al. 1995. Intonation and communicative function in a regional variety of italian. *Phonus* 1. 19–32.

Halliday, Michael Alexander Kirkwood & William S Greaves. 2008. Intonation in the grammar of english .

Hart, Johan't, JT Hart, R Collier, A Cohen, Rene Collier & Antonie Cohen. 2003. *Perceptual study of intonation*. Cambridge.

Heinz, Jeffrey. forthcoming. Doing computational phonology.

Hirst, Daniel & Albert Di Cristo. 1998. A survey of intonation systems. *Intonation systems: A survey of twenty languages* 144. 152–166.

Iskarous, Khalil & Jennifer Cole. 2026. A quantal dynamical theory of f0 contours: Bridging the phonetics and phonology of intonation in: Developments in the modeling of speech prosody .

Jardine, Adam. 2016. *Locality and non-linear representations in tonal phonology*. University of Delaware.

Jardine, Adam. 2017. On the logical complexity of autosegmental representations. In *Proceedings of the 15th meeting on the mathematics of language*, 22–35.

Jardine, Adam, Nick Danis & Luca Iacoponi. 2021. A formal investigation of q-theory in comparison to autosegmental representations. *Linguistic Inquiry* 52(2). 333–358.

Joo, Hyunjung & Mariapaola D'Imperio. 2025. The perception of lexical pitch accent in south kyungsang korean: The relevance of accent shape. *Language and Speech* 00238309251368294.

Joo, Hyunjung & Adam Jardine. 2025. Intonation as a quantifier-free logical interpretation of metrical and prosodic structure. In *Proceedings of the society for computation in linguistics 2025*, 261–270.

Jun, Sun-Ah. 2000. K-tobi (korean tobi) labelling conventions. *UCLA working papers in phonetics* 99. 149–173.

Jun, Sun-Ah. 2006a. Intonational phonology of seoul korean revisited. *Japanese-Korean Linguistics* 14. 15–26.

Jun, Sun-Ah. 2006b. *Prosodic typology: The phonology of intonation and phrasing*, vol. 1. Oxford University Press.

Jun, Sun-Ah. 2014. Prosodic typology: By prominence type, word prosody, and macrorhythm. *Prosodic typology II: The phonology of intonation and phrasing* 520539. 520–539.

Jun, Sun-Ah. 2025. Prosodic typology: Intonational tone types and functions. *Contemporary linguistics: Integrating languages, communities, and technologies* 7. 93–111.

Keyser, Samuel Jay & Kenneth N Stevens. 2006. Enhancement and overlap in the speech chain. *Language* 82(1). 33–63.

Kim, Jieun & Sun-Ah Jun. 2009. Prosodic structure and focus prosody of south kyungsang korean. *Language Research* .

Kimball, Amelia E & Jennifer Cole. 2016. Pitch contour shape matters in memory. In *Proceedings of the international conference on speech prosody*, vol. 8, 1171–75. International Speech Communication Association.

Kimia Lab. 2019. Machine intelligence - lecture 17 (fuzzy logic, fuzzy inference) [video]. `https://www.youtube.com/watch?v=TReelsVxWxg&t=1623s`.

Kingdon, Roger. 1958. The groundwork of english intonation. *(No Title)* .

Koser, Nate, Chris Oakden & Adam Jardine. 2018. Tone association and output locality in non-linear structures. In *Proceedings of the annual meetings on phonology*, .

Ladd, D Robert. 2008. *Intonational phonology*. Cambridge University Press.

Ladd, D Robert & Rachel Morton. 1997. The perception of intonational emphasis: continuous or categorical? *Journal of phonetics* 25(3). 313–342.

Ladd, D Robert & Astrid Schepman. 2003. "sagging transitions" between high pitch accents in english: Experimental evidence. *Journal of phonetics* 31(1). 81–112.

Leben, William Ronald. 1973. *Suprasegmental phonology.*: Massachusetts Institute of Technology dissertation.

Lee, Hyunjung & Jie Zhang. 2014. The nominal pitch accent system of south kyungsang korean. *Journal of East Asian Linguistics* 23(1). 71–111.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic inquiry* 8(2). 249–336.

Liberman, Mark Yoffe. 1975. *The intonational system of english.*: Massachusetts Institute of Technology dissertation.

Libkin, Leonid. 2004. *Elements of finite model theory*, vol. 41. Springer.

Massaro, DOMINIC W. 1989. A fuzzy logical model of speech perception. In *Proceedings of xxiv international congress of psychology. human information processing: Measures, mechanisms and models (d. vickers and p. smith, eds.),(amsterdam, north holland)*, 367–379.

McNaughton, Robert & Seymour A Papert. 1971. *Counter-free automata (mit research monograph no. 65)*. The MIT Press.

Nelson, Scott. 2022a. Are representations in articulatory and generative phonology so different? In *Annual meeting on phonology 2022 (amp)*, .

Nelson, Scott. 2022b. A model theoretic perspective on phonological feature systems. In *Proceedings of the society for computation in linguistics 2022*, 1–10.

Nelson, Scott. 2023. Model theoretic phonology and theory comparison: Segments, gestures, and coupling graphs. In *North american phonology conference 12 (naphc)*, .

Niebuhr, Oliver & Klaus J Kohler. 2004. Perception and cognitive processing of tonal alignment in german. In *Proceedings of the international symposium on tonal aspects of languages: Emphasis on tone languages*, 155–158.

Nolan, Francis. 2022. The rise and fall of the british school of intonation analysis .

OConnor, Joseph Desmond & Gordon Frederick Arnold. 2004. *Intonation of colloquial english*. РГБ.

Peirce, Jonathan W. 2009. Generating stimuli for neuroscience using psychopy. *Frontiers in neuroinformatics* 2. 343.

Pierrehumbert, Janet. 1981. Synthesizing intonation. *The Journal of the Acoustical Society of America* 70(4). 985–995.

Pierrehumbert, Janet & Julia Hirschberg. 2026. The meaning of intonational contours in the interpretation of discourse. In *Pronunciation*, 256–295. Routledge.

Pierrehumbert, Janet B & Shirley A Steele. 1989. Categories of tonal alignment in english. *Phonetica* 46(4). 181–196.

Pierrehumbert, Janet Breckenridge. 1980. *The phonology and phonetics of english intonation*: Massachusetts Institute of Technology dissertation.

de Pijper, Jan-Roelof. 1983. Perceptual evaluation of some proposed models of intonation. *Sound Structures: Studies for Antonie Cohen* 13. 205.

Pitrelli, John, Mary Beckman & Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling. In *Proc. icslp*, 123–126.

R Core Team. 2020. Ra language and environment for statistical computing, r foundation for statistical. *Computing* .

Rose, Phil. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech communication* 6(4). 343–352.

Sobhy, Sameh Mohamed & Wael Mohamed Khedr. 2015. Developing of fuzzy logic controller for air condition system. *International Journal of Computer Applications* 126(15). 1–8.

Steffman, Jeremy, Jennifer Cole & Stefanie Shattuck-Hufnagel. 2024. Intonational categories and continua in american english rising nuclear tunes. *Journal of Phonetics* 104. 101310.

Stevens, Kenneth Noble & Samuel Jay Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of phonetics* 38(1). 10–19.

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods* 14(4). 323.

Strother-Garcia, Kristina. 2019. *Using model theory in phonology: a novel characterization of syllable structure and syllabification*. University of Delaware.

Strother-Garcia, Kristina & Jeffrey Heinz. 2017. Logical foundations of syllable representations. In *Poster presented at the 5th annual meeting on phonology, new york university, new york city*, .

The MathWorks Inc. 2022. Matlab version: 9.13.0 (r2022b). `https://www.mathworks.com`.

The MathWorks Inc. 2024. Fuzzy logic toolbox 24.2 (r2024b). `https://www.mathworks.com`.

Veilleux, Nanette, Stefanie Shattuck-Hufnagel & Alejna Brugos. 2006. Transcribing prosodic structure of spoken utterances with tobi. *Creative Commons BY-NC-SA* .

Venditti, Jennifer J. 2005. The j_tobi model of japanese intonation. *Prosodic typology: The phonology of intonation and phrasing* 172–200.

Warner, Josh, Jason Sexauer, Wouter Van den Broeck, Bruno P Kinoshita, Jakub Balinski, Christian Clauss, Aishwarya Unnikrishnan, Marco Miretti, Guilherme Castelão, Felipe Arruda Pontes et al. 2024. Jdwarner/scikit-fuzzy: Scikit-fuzzy 0.5. 0. *Zenodo* .

Welby, Pauline. 2003. Effects of pitch accent position, type, and status on focus projection. *Language and Speech* 46(1). 53–81.

Wells, John Christopher. 2006. *English intonation pb and audio cd: An introduction*. Cambridge University Press.

Williams, Edwin S. 1976. Underlying tone in margi and igbo. *Linguistic Inquiry* 463–484.

Xu, Yi. 2004. Transmitting tone and intonation simultaneously-the parallel encoding and target approximation (penta) model. In *International symposium on tonal aspects of languages: With emphasis on tone languages*, 215–220.

Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech communication* 46(3-4). 220–251.

Xu, Yi & Q Emily Wang. 2001. Pitch targets and their realization: Evidence from mandarin chinese. *Speech communication* 33(4). 319–337.

Zadeh, Lotfi A. 1965. Fuzzy sets. *Information and control* 8(3). 338–353.

Zadeh, Lotfi A. 2008. Is there a need for fuzzy logic? *Information sciences* 178(13). 2751–2779.

Zadeh, Lotfi A. 2023. Fuzzy logic. In *Granular, fuzzy, and soft computing*, 19–49. Springer.

Zadeh, Lotfi Asker. 1975. The concept of a linguistic variable and its application to approximate reasoning—i. *Information sciences* 8(3). 199–249.

Zadeh, Lotfi Asker, George J Klir & Bo Yuan. 1996. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers*, vol. 6. World scientific.

A Melodic transductions of intonation in multiple phrases

This section shows the melodic transductions of intonation in multiple phrases in American English, Seoul Korean, and Tokyo Japanese. This is to show that the melodic transductions for each language defined in Chapter 5 work across multiple phrases within an Intonational Phrase.

A.1 American English

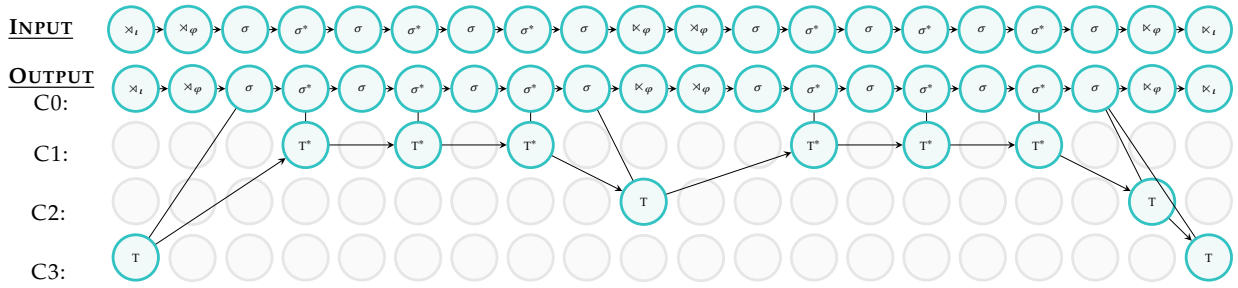


Figure 1: A graph illustrating melodic transduction of multiple phrases in American English intonation.

A.2 Seoul Korean

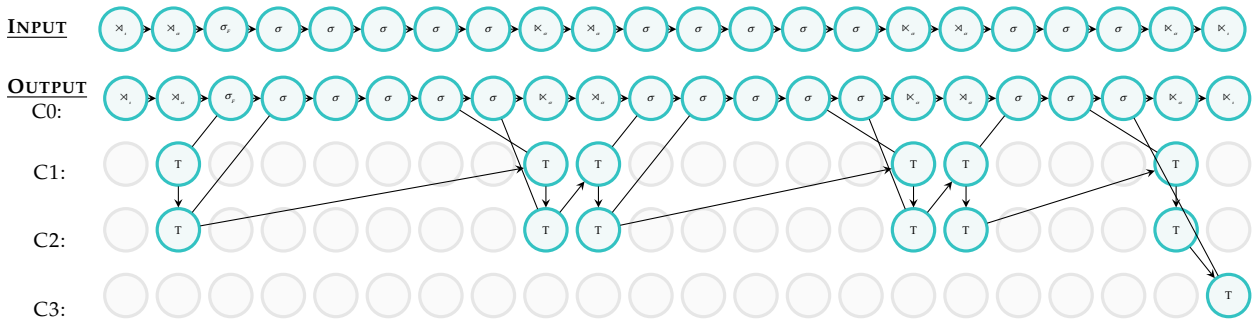


Figure 2: A graph illustrating melodic transduction of multiple phrases in Seoul Korean intonation.

A.3 Tokyo Japanese

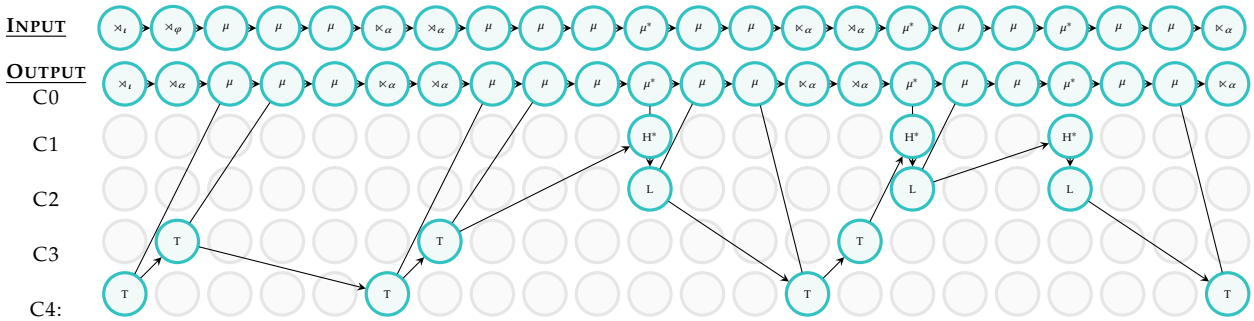


Figure 3: A graph illustrating melodic transduction of multiple phrases in Tokyo Japanese intonation.

B A complete version of the definition of the pitch accents in American English

Possible pitch accents in American English can be defined with the following definitions in the two-step process in Figure 6.12. In the first step, the pitch accent predicates using Boolean logic are defined based on the formulas in B.1. Then, the predicates are extended to fuzzy logic using the definitions provided in B.2.

B.1 Step 1: Defining the pitch accents using Boolean logic

$$
\begin{aligned}
T + H^*(x) &= P_{\sigma^*}(x) \wedge localPeak(x, X) \\
&\quad \wedge \exists X[leftwardInterval(X) \wedge rise(X) \wedge scooped(X) \\
&\quad \wedge \exists y[P_\sigma(y) \wedge localTrough(y, X) \wedge y \prec_T x]] \\
H^* + T(x) &= P_{\sigma^*}(x) \wedge localPeak(x) \\
&\quad \wedge \exists X[rightwardInterval(X) \wedge fall(X) \\
&\quad \wedge \exists y[P_\sigma(y) \wedge localTrough(y, X) \wedge x \prec_T y]] \\
T + L^*(x) &= P_{\sigma^*}(x) \wedge localTrough(x) \\
&\quad \wedge \exists X[rightwardInterval(X) \wedge fall(X) \\
&\quad \wedge \exists y[P_\sigma(y) \wedge localPeak(y, X) \wedge x \prec_T y]] \\
L^* + T(x) &= P_\sigma(x) \wedge localTrough(x) \\
&\quad \wedge \exists X[\wedge rightwardInterval(X) \wedge rise(X) \wedge domed(X) \\
&\quad \wedge \exists y[P_\sigma(y) \wedge localPeak(y, X) \wedge y \prec_T x]]
\end{aligned}
$$

B.2 Step 2: Extending Boolean logic to fuzzy logic

$$
PitchAccent_{T+H^*}(x) = \begin{cases} \mu_{T+H^*}(x), & \text{if } T + H^*(x) = 1 \\ 0, & \text{if } T + H^*(x) = 0 \end{cases}
$$

$$PitchAccent_{H^*+T}(x) = \begin{cases} \mu_{H^*+T}(x), & \text{if } H^* + T(x) = 1 \\ 0, & \text{if } H^* + T(x) = 0 \end{cases}$$

$$PitchAccent_{T+L^*}(x) = \begin{cases} \mu_{T+L^*}(x), & \text{if } T + L^*(x) = 1 \\ 0, & \text{if } T + L^*(x) = 0 \end{cases}$$

$$PitchAccent_{L^*+T}(x) = \begin{cases} \mu_{L^*+T}(x), & \text{if } L^* + T(x) = 1 \\ 0, & \text{if } L^* + T(x) = 0 \end{cases}$$

C Examples of trapezoidal and Gaussian fuzzy sets and their accuracy rate for the pitch accent classification

C.1 Examples of trapezoidal and Gaussian fuzzy sets

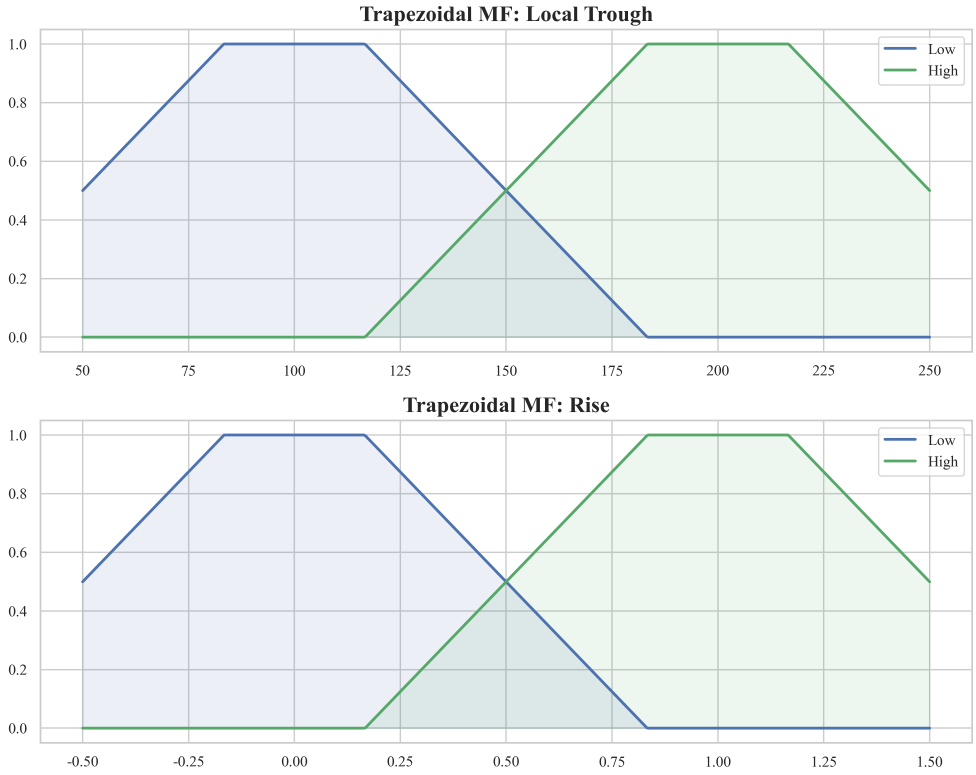


Figure 4: An example of trapezoidal fuzzy sets.

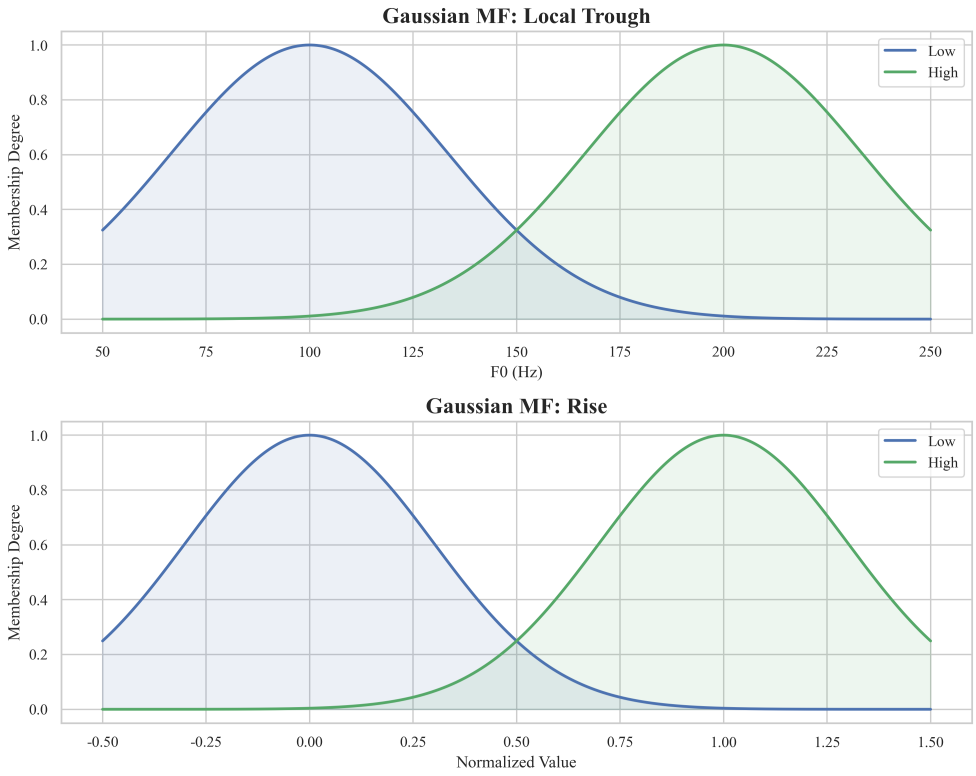


Figure 5: An example of Gaussian fuzzy sets.

C.2 Accuracy rates for the pitch accent classification using trapezoidal and Gaussian fuzzy sets

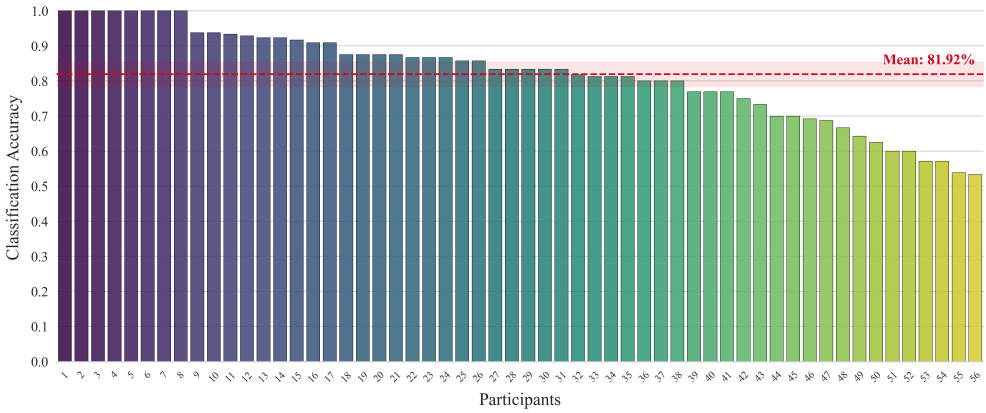


Figure 6: Accuracy rate of trapezoidal fuzzy sets.

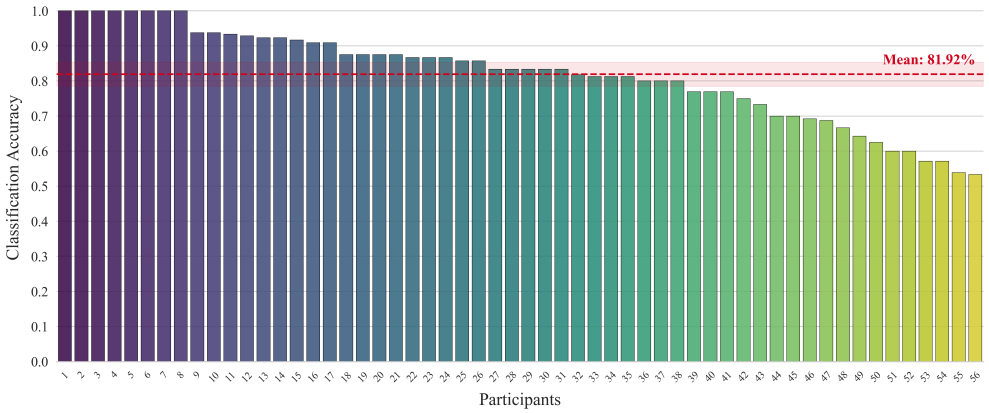


Figure 7: Accuracy rate of Gaussian fuzzy sets.