

LEXICAL TONE, INTONATION, AND THEIR INTERACTION:  
A SCOPAL THEORY OF TUNE ASSOCIATION

A Dissertation

Presented to the Faculty of the Graduate School  
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of  
Doctor of Philosophy

by

Masayuki Katagiri Gibson

January 2013

© 2013 Masayuki Katagiri Gibson

LEXICAL TONE, INTONATION, AND THEIR INTERACTION:  
A SCOPAL THEORY OF TUNE ASSOCIATION

Masayuki Katagiri Gibson, Ph. D.

Cornell University 2013

There is still much to be learned regarding the nature of the interaction between lexical tone and utterance-level intonation. Previous studies in individual languages tend to be too narrow, focusing on ways to model the final  $F_0$  output without regard to cross-linguistic implications; studies mainly concerned with phonological patterns across languages tend to over-generalize, missing or glossing over many language-specific and category-specific phenomena. This dissertation attempts to address the gap left by these previous studies.

The first part of the dissertation presents results from a series of production and perception experiments conducted for a handful of tone languages, including Standard Mandarin, Henanhua, Cantonese, North Kyeongsang Korean, and Kansai Japanese (a family of dialects including Shiga, Kyoto, Osaka, and Kobe Japanese). The production experiments were designed to elicit multiple renditions of various lexical tones in declarative and echo question contexts, and the perceptual experiments were designed to test the degree of recoverability for each communicative function (lexical tone and utterance-type intonation) in the various conditions.

The second part of the dissertation considers the implications of the experimental results for building a comprehensive model of speech melody. First, by examining the behavior of intonation across tonal categories within each language, it is shown that there is evidence for unpredictable tone-dependent intonation implementation, suggesting that our model must allow for some interaction between the two at some level before phonetic implementation (a principle

taken for granted by some models and ruled out by others). In addition, the results are assessed cross-linguistically, characterizing the ways in which the model must be parameterized. Finally, an enhanced autosegmental-metrical model that meets both of the above demands is proposed. The model includes an autosegmental geometry that encodes lexical tones in languages like Mandarin and Cantonese as tones associated with syllables and so-called “accentual melodies” in languages like Kansai Japanese and NKK as tones associated with words. The scope of a given melodic unit’s effect on an utterance is determined by the level(s) of the prosodic constituent(s) with which it is associated.

## BIOGRAPHICAL SKETCH

Masayuki Gibson was born in Plainfield, NJ in 1980. He is the oldest child of Carolyn and Naotoshi Gibson and the older brother of Naotomo Gibson. His years at Scotch Plains-Fanwood High School were broken up by a one-year study-abroad at Tsuruga High School in Tsuruga, Japan. He graduated *magna cum laude* from Rutgers, the State University of New Jersey in 2003 with Bachelor of Arts degrees in linguistics and music as well as a minor in physics. After graduating from Rutgers, he spent one year in Ishiyama, Japan teaching English at Ishiyama High School and then one year as a substitute teacher at the Pingry School in Short Hills, NJ before commencing his graduate study in Linguistics at Cornell University in the fall of 2005. Currently, Masayuki works as a research associate at NovaSpeech LLC.

## ACKNOWLEDGEMENTS

I am profoundly grateful to my advisor, Michael Wagner, for being my guide through the concentric circles of Dissertation-land. He taught me the importance of keeping things in perspective, and he had a knack for making me feel like my questions and ideas were valuable and worth exploring. This gratitude also extends to the other members of my committee, Abby Cohn, John Whitman, and Draga Zec. Abby always kept me honest when it came to interpreting my data, framing my assumptions, and drawing conclusions. John brought a wealth of historical knowledge to the table, and his insights into how the melodic systems of various languages were traditionally analyzed by “native” linguists elucidated for me the sources of various theoretical quirks that persist in the literature cross-linguistically. Draga kept me disciplined about theoretical follow-through; if I had an idea, she always knew what questions to ask and what thought experiments to invoke in order to put that idea to the test.

Thanks to all of the other graduate students who came through the department during my time there; from day one it felt like, if I ever needed advice, I could walk into the basement of Morrill and get it. Thanks in particular to my classmates, Ásgrímur Angantýsson, Seongyeon Ko, Jiwon Yun, and Zhiguo (Henry) Xie, for helping me grow as a linguist while we were in the trenches together. Thanks also to Angie Tinti, the departmental Administrative Assistant and Graduate Field Assistant, for all of her help over the years in making the administrative aspects of my time with the department as easy and painless as possible.

In designing the experiments presented in Chapters 2 and 3 of the dissertation and in the interpretation of the results I leaned heavily on my language consultants—Hongyuan Dong (Mandarin), Aletheia Cui (Henanhua), Natalie Lui (Cantonese), Hye-Sook Lee (North Kyeongsang Korean), Naho Orita (Osaka Japanese), and Kiyoshi Takeda (Shiga Japanese)—for their expertise and native intuitions. I thank them for their boundless patience and intellectual generosity.

The North Kyeongsang Korean perceptual experiment, which was administered online, was made possible with help from Hye-Sook Lee, who assisted in the design of the experiment and acted as a guinea pig in the testing phase of the experiment; Eric Evans, who wrote the code for the website; Jiwon Yun, who also helped with some of the design and Korean translations; Seongyeon Ko, who helped me secure a local coordinator at Kyeongbuk National University; and Yeongkon Jeon, who was the local coordinator who acquired volunteers and facilitated the actual running of the perceptual test. I am equally indebted to my contacts in Japan who made my fieldwork there possible. I owe Haruo Kubozono thanks not only for helping me secure locations for recording sessions on the Kobe leg of my trip, but also for providing me with stimulating discussion and invaluable resources on tone and intonation in Japanese dialects. Also in Kobe, Naho Orita did so much of the legwork in securing volunteers for my study and scheduling their recording sessions that all I had to do was show up and hit “record”. Thanks to Francis Michinao Matsui and the Kobe Shoin Women’s College Phonetics Lab for allowing me to set up camp in their sound-proof booth for four days to make recordings. On the Shiga leg of my trip, the administration and faculty at Ishiyama High School generously allowed me to set up a recording space in the school. I was delighted to be reunited with my old friends and colleagues at Ishiyama High School, who took time out of their busy schedules to make me feel welcome, as if I had never left. Thanks specifically to Tsugio Kitazawa and his family for graciously letting me stay in their home during my time in Shiga. Finally, the entire visit to Shiga simply would not have happened without the tireless efforts of my dear friend Kiyoshi Takeda, who secured permission from the school, made travel and lodging arrangements, and harassed all of the other faculty members into volunteering for my experiment!

This dissertation has benefitted either directly or indirectly from ideas and questions that have come up during my interactions with many people over the years. Some of them are named elsewhere in these acknowledgements, but the others include Mary Beckman, Johanna Brugman, Marc Brunelle, Becky Butler, Yiya Chen, Adam Cooper, Cliff Crawford, Effi Georgala, Carlos Gussenhoven, Sue Hertz, Jonathan Howell, Hyun Kyung Hwang, Larry Hyman, Yosuke Igarashi,

Kiwako Ito, Natapon Kidrai, Bob Ladd, Julie McGory, James Mesbur, Amanda Miller, Yumiko Nishi, Janet Pierrehumbert, Pittayawat “Joe” Pittayaporn, Nikola Predolac, Peggy “Hank” Renwick, Hubert Truckenbrodt, Kazuha Watanabe, Yi Xu, and Jiahong Yuan. I am surely leaving out some names but that doesn’t make the contributions of the individuals denoted by those names any less valuable.

Thanks to Steven Ikier for proofreading the dissertation (at least the portions I was able to get to him before my filing deadline) as well as providing valuable comments on content and exposition. Any errors in the final draft are mine alone.

The following institutions provided funding for research and coursework that allowed me to complete this dissertation: the Cornell Graduate School, the Cornell Linguistics Department, the Cornell Phonetics Laboratory, the East Asia Program at Cornell, and the Foreign Language and Area Studies Fellowship Program at Cornell.

Thanks to Abby Smith for being there for me, for bringing a non-linguist’s perspective to the table, and for the free milk.

Finally, I want to thank my family for their love and support. I’m grateful to my parents for bestowing me with the gift of a bilingual household, which shaped my brain, my worldview, and my path in life (at times quite literally; it got me a job teaching English in Japan, which introduced me to the Shiga Japanese speakers who, years later, became the subjects of the experiments presented in this dissertation). My grandfather (Grandpa Bill) also supported me unconditionally over the years, not always with a full understanding of my plan—what does one do, after all, with undergraduate concentrations in linguistics, music, and physics?—but always with love and pride. I’m also grateful for my family’s practical help during the writing of this dissertation, the nature of which epitomizes their respective ways of loving: My father always made sure I was getting enough to eat, my mother accosted on my behalf all of the Mandarin- and Cantonese-speaking people she could find in the greater North Jersey area, and my brother and his wife provided me with a peaceful getaway at their San Diego condominium. All of these respective modes of support buoyed me through the thick of things and got me to the finish line.



## TABLE OF CONTENTS

Biographical Sketch .....	iii
Acknowledgements .....	iv
Table of Contents .....	vii
List of Figures .....	xiii
List of Tables .....	xxii
Chapter 1 : Introduction and Background.....	1
1.1 Definitions of <i>Lexical Tone</i> and <i>Intonation</i> .....	3
1.1.1 <i>Lexical tone</i> .....	3
1.1.2 <i>Intonation</i> .....	3
1.2 Overlay vs. Sequential Models.....	4
1.2.1 Overlay models.....	4
1.2.2 Sequential models.....	9
1.3 Issues Raised by a “ToBI Typology” .....	13
1.4 Toward a Melodic Typology.....	15
1.4.1 Declarative vs. echo question intonation.....	16
1.4.2 A new model.....	20
1.5 Overview of the Dissertation.....	22
1.5.1 Overview of Chapter 2 .....	22
1.5.2 Overview of Chapter 3 .....	23
1.5.3 Overview of Chapter 4 .....	23

1.5.4 Overview of Chapter 5 .....	24
1.5.5 Overview of Chapters 6 and 7 .....	24
Chapter 2 : Production Experiments .....	25
2.1 Questions about Production .....	25
2.2 Production in Mandarin.....	26
2.2.1 Overview of Mandarin tones .....	26
2.2.2 Subjects.....	28
2.2.3 Materials and procedure .....	28
2.2.4 Results .....	29
2.2.5 Discussion.....	37
2.3 Production in Henanhua .....	38
2.3.1 Overview of Henanhua tones .....	39
2.3.2 Subject .....	40
2.3.3 Materials .....	40
2.3.4 Results .....	42
2.3.5 Discussion.....	48
2.4 Production in Cantonese.....	51
2.4.1 Overview of Cantonese tones .....	51
2.4.2 Subjects.....	52
2.4.3 Materials .....	52
2.4.4 Differences from the Mandarin design.....	54
2.4.5 Results .....	54

2.4.6 Discussion.....	62
2.5 Production in North Kyeongsang Korean.....	67
2.5.1 Overview of NKK tonal classes.....	67
2.5.2 Subject.....	68
2.5.3 Materials.....	68
2.5.4 Differences from the Mandarin and Cantonese designs.....	69
2.5.5 Results.....	70
2.5.6 Discussion.....	80
2.6 Production in Shiga Japanese and Other Kansai Dialects.....	83
2.6.1 Overview of the Kansai tonal system.....	83
2.6.2 Subjects.....	85
2.6.3 Materials.....	85
2.6.4 Comparison with other experiments.....	87
2.6.5 Results.....	88
2.6.6 Discussion.....	97
2.7 Summary and Conclusion.....	103
2.7.1 Cross-linguistic differences.....	103
2.7.2 Tonal-category-dependent differences.....	107
2.7.3 Answers to questions about production.....	107
2.7.4 Conclusion.....	108
Chapter 3 : Perceptual Experiments.....	109
3.1 Questions about Perception.....	109

3.2 Perception in Mandarin .....	110
3.2.1 Subjects.....	110
3.2.2 Materials and procedure .....	111
3.2.3 Results .....	112
3.2.4 Discussion.....	118
3.3 Perception in Henanhua .....	120
3.3.1 Subject .....	120
3.3.2 Stimuli .....	120
3.3.3 Results .....	121
3.3.4 Discussion.....	124
3.4 Perception in Cantonese .....	126
3.4.1 Subjects.....	126
3.4.2 Stimuli .....	126
3.4.3 Differences from Mandarin design.....	128
3.4.4 Results .....	128
3.4.5 Discussion.....	137
3.5 Perception in North Kyeongsang Korean.....	140
3.5.1 Subjects.....	140
3.5.2 Stimuli .....	140
3.5.3 Differences from Mandarin and Cantonese designs.....	142
3.5.4 Results .....	143
3.5.5 Discussion.....	150

3.6 Perception in Shiga Japanese .....	153
3.6.1 Subjects.....	153
3.6.2 Stimuli .....	153
3.6.3 Comparison with other experiments.....	154
3.6.4 Results .....	155
3.6.5 Discussion.....	157
3.7 Summary and Conclusion .....	160
3.7.1 Answers to questions about perception .....	164
Chapter 4 : Tone-Dependent Intonation and Its Theoretical Consequences.....	167
4.1 Introduction.....	167
4.2 Accounting for Tone-Dependent Intonation in Syllable-Tone Languages .....	169
4.2.1 Mandarin.....	169
4.2.2 Henanhua .....	174
4.2.3 Cantonese.....	176
4.2.4 Interim summary and conclusion .....	183
4.3 Accounting for Tone-Dependent Intonation in Word-Tone Languages .....	183
4.3.1 NKK.....	184
4.3.2 Kansai Japanese.....	188
4.4 Summary and Conclusion .....	189
Chapter 5 : A Phonological Approach to Melodic Interactions.....	191
5.1 Introduction.....	191
5.2 Tone-Dependent Algorithms vs. Intonation-Dependent Allotones.....	192

5.3 The Autosegmental-Metrical Framework .....	198
5.3.1 AM analysis for NKK.....	200
5.3.2 AM analysis for Kansai Japanese.....	209
5.4 Conclusion.....	217
Chapter 6 : Toward a Unified Scopal Model of Speech Melody.....	219
6.1 Introduction.....	219
6.2 Autosegmental <i>Tune</i> Geometry.....	221
6.3 Where Tone and Intonation Meet.....	234
6.4 Phonetic Awareness of the Tree Structure .....	245
6.5 Summary and Conclusion .....	246
Chapter 7 : Conclusion.....	251
Appendix.....	254
References.....	256

## LIST OF FIGURES

Figure 1.1: Pitch tracks of disyllabic words in three languages (Mandarin, North Kyeongsang Korean, and Shiga Japanese), each with a “falling” lexical tone on the second syllable, in declarative (dark gray) and interrogative (light gray) contexts. ....	2
Figure 1.2: Schematic diagram of an overlay model. ....	5
Figure 1.3: Schematic diagram of a sequential model. ....	9
Figure 2.1: Mean durations for Mandarin. ....	30
Figure 2.2: Mean target-utterance duration ratios for declarative and echo question intonation, by tone (Mandarin A on the left and Mandarin B on the right). ....	31
Figure 2.3: Mean $F_0$ of the frame by sentence position. Frame 1 above and Frame 2 below, Mandarin A on the left and Mandarin B on the right. ....	32
Figure 2.4: Mean $F_0$ contours for Mandarin. ....	34
Figure 2.5: Multiple-contour plots for Mandarin. ....	35
Figure 2.6: Standard deviations of declarative, echo question, and overall $F_0$ , by tone, in Mandarin. ....	36
Figure 2.7: Mean durations for Henanhua. Results for the isolation condition are shown above and those for the frame condition are shown below. ....	43
Figure 2.8: Mean contours for all four tones in Henanhua in isolation on the syllable <i>wan</i> (left) and on the syllable <i>yan</i> (right). ....	44
Figure 2.9: Mean $F_0$ contours for tones in isolation in Henanhua. ....	45
Figure 2.10: Mean $F_0$ contours for tones in an all-T4 frame sentence in Henanhua. ....	45
Figure 2.11: Mean $F_0$ contours for tones in an all-T3 frame sentence in Henanhua. ....	46
Figure 2.12: Multiple-contour plots for Henanhua. ....	47
Figure 2.13: Standard deviations of declarative, echo question, and overall $F_0$ , by tone, in Henanhua (all-T4 frame on the left, all-T3 frame on the right). ....	47

Figure 2.14: Overlaid mean contours for T2 and T4 in Henanhua (below) and plots for the four tones in isolation (above).....	48
Figure 2.15: Henanhua declarative and echo-question contours and equivalent contours in Mandarin.....	49
Figure 2.16: Multiple-contour plots for Henanhua T3 and T4 (left) and Mandarin T1 and T4 (right).....	51
Figure 2.17: Mean durations in Cantonese.....	55
Figure 2.18: Mean target-utterance duration ratios for declarative and echo question intonation, by tone (Cantonese A on the left and Cantonese B on the right).....	56
Figure 2.19: Mean $F_0$ of the frame by sentence position. Frame 1 above and Frame 2 below, Cantonese A on the left and Cantonese B on the right.....	56
Figure 2.20: Mean $F_0$ contours for Cantonese.....	58
Figure 2.21: Representative contours for “level” tones (T1, T3, and T6) after Frame 1 and Frame 2, respectively, for Cantonese Speaker A.....	60
Figure 2.22: Multiple-contour plots for Cantonese.....	61
Figure 2.23: Standard deviations of declarative, echo question, and overall $F_0$ , by tone, in Cantonese.....	62
Figure 2.24: Mean $F_0$ plots and multiple-contour plots for T1 in Mandarin (left) and Cantonese (right).....	66
Figure 2.25: Mean durations for the isolation condition in NKK.....	70
Figure 2.26: Mean durations for the frame condition in NKK.....	70
Figure 2.27: Mean target-utterance duration ratios by intonation by tonal category (NKK).....	71
Figure 2.28: Syll-1 vs. syll-2 mean duration ratios for disyllabic initial-accented and final-accented target words by intonation in NKK.....	72
Figure 2.29: Mean $F_0$ contours (isolation) for NKK.....	73
Figure 2.30: Segmental anchoring of initial-accented (thick solid), double-accented (dotted), and final-accented (thin solid) disyllables in an utterance-medial context.....	74



Figure 2.31: Representative F <sub>0</sub> contours for declarative renditions of initial- ( <i>nam-i</i> ‘south-NOM’), double- ( <i>nam-i</i> ‘third party-NOM’), and final-accented ( <i>muni</i> ‘pattern’) disyllables in NKK. .....	74
Figure 2.32: Representative F <sub>0</sub> contours for echo question renditions of initial- ( <i>nam-i</i> ‘south-NOM’), double- ( <i>nam-i</i> ‘third party-NOM’), and final-accented ( <i>muni</i> ‘pattern’) disyllables in NKK.....	75
Figure 2.33: Mean F <sub>0</sub> contours (frame) for NKK. ....	76
Figure 2.34: Representative F <sub>0</sub> contours for initial- ( <i>nam</i> ‘south’) and double-accented ( <i>nam</i> ‘third party’) monosyllables in NKK.....	77
Figure 2.35: Multiple repetitions of initial- (dark gray) and double-accented (light gray) monosyllables in the declarative context, separated out by context and minimal pair. ....	78
Figure 2.36: Multiple repetitions of initial- (dark gray) and double-accented (light gray) monosyllables in the echo question context.....	78
Figure 2.37: Multiple-contour plots for NKK. Declarative contours are in light gray and echo question contours are in dark gray. ....	79
Figure 2.38: Standard deviations of declarative echo question, and overall F <sub>0</sub> , by tone, in NKK. .....	80
Figure 2.39: Mean F <sub>0</sub> contours for T1 in Cantonese (left) and initial-accented monosyllable in NKK (right).....	81
Figure 2.40: Mean F <sub>0</sub> contours for T4 in Mandarin (left) and initial-accented disyllable in NKK (right). ....	81
Figure 2.41: Multiple-contour plots for T4 in Mandarin (left) and initial-accented disyllables in NKK (right).....	82
Figure 2.42: Mean duration of target word by intonation type for SJ speakers.....	88
Figure 2.43: Syll-1 vs. syll-2 mean durations for SJ Subjects C, I, and J by intonation. ....	89
Figure 2.44: Duration difference between intonational types by accentedness by subject .....	90

Figure 2.45: F <sub>0</sub> at two sentence positions in the frame for Shiga C (declarative on the left, echo question on the right). .....	91
Figure 2.46: Three-way tonal contrast on monosyllables in SJ. ....	92
Figure 2.47: F <sub>0</sub> contours for representative disyllabic tokens in SJ. Dotted vertical lines indicate the onset-nucleus boundary of the second syllable. ....	93
Figure 2.48: F <sub>0</sub> contours for representative disyllabic tokens in Kyoto (left) and Osaka (right) Japanese. Dotted lines indicate the onset-nucleus boundary of the second syllable. ....	94
Figure 2.49: Representative F <sub>0</sub> contours for the four-syllable tokens <i>yamaimo</i> ('mountain yam'; H-unaccented) and <i>nagaimo</i> ('long yam'; L-unaccented) in a declarative context for Shiga (top), Kyoto (bottom left), and Osaka (bottom right) Japanese. Contours on the same plot are zero-aligned at the onset of the last syllable. ....	96
Figure 2.50: Representative F <sub>0</sub> contours for <i>yamaimo</i> ('mountain yam'; H-unaccented) and <i>nagaimo</i> ('long yam'; L-unaccented) in an echo question context for Shiga (top), Kyoto (bottom left), and Osaka (bottom right) Japanese. Contours on the same plot are zero-aligned at the onset of the last syllable. ....	97
Figure 2.51: Comparison of a level tone in Cantonese and an unaccented tone in Kansai. Dotted vertical lines indicate the left boundary of the rightmost syllable in each case. ....	98
Figure 2.52: Representative F <sub>0</sub> contours for analogous accented tonal categories in NKK and SJ. Dotted vertical lines indicate the onset-nucleus boundary of the rightmost syllable. ....	99
Figure 2.53: Preliminary surface-phonological representations for final-accented disyllables in NKK (left) and Kyoto (right). ....	100
Figure 2.54: Preliminary surface-phonological representations for accented monosyllables in NKK and Kyoto Japanese. ....	100
Figure 2.55: Preliminary surface-phonological representations for the L-final-accent in three Kansai dialects. ....	102
Figure 2.56: Representative declarative and echo question contours for <i>yamaimo</i> (H-unaccented) and <i>nagaimo</i> (L-unaccented) for two SJ speakers. ....	102

Figure 2.57: F <sub>0</sub> contours for the “falling tone” category in Cantonese, NKK, Mandarin, and SJ. Dotted vertical lines indicate the beginning of the onset of the rightmost syllable. ....	104
Figure 3.1: Sample answer sheet for the Mandarin perceptual test. ....	111
Figure 3.2: Tonal confusion matrices for Mandarin. ....	114
Figure 3.3: Multiple-contour plots for Mandarin T2 (left) and T3 (right), for speakers A (top) and B (bottom). Declarative contours are shown in dark gray and echo question contours in light gray. ....	115
Figure 3.4: The rate of response—or “bin test”—for each intonation type in Mandarin. ....	116
Figure 3.5: The rate of accuracy in judging intonational category, by tonal category of the stimulus, in Mandarin. ....	117
Figure 3.6: The rate of accuracy in judging intonational category, by tonal category of the stimulus, in Mandarin—the declarative condition above and the echo question condition below. ....	118
Figure 3.7: Tonal confusion matrices for Henanhua, by phrasal context and intonation type. A 17% difference represents a one-stimulus difference. ....	122
Figure 3.8: Bin test for intonation responses in Henanhua. ....	123
Figure 3.9: The rate of accuracy in judging intonation, by tonal category of the stimulus, in Henanhua—the declarative condition above and the echo question condition below....	123
Figure 3.10: Multiple-contour plots for Henanhua (left) and Mandarin (right), repeated from Chapter 2. ....	125
Figure 3.11: A sample answer sheet for the Cantonese perceptual test. ....	127
Figure 3.12: Rate of perceptual accuracy for tone, by tone and intonation type of the stimulus, in Cantonese. ....	130
Figure 3.13: The bin test for tone responses in a declarative context (above) and in an echo question context (below) in Cantonese. ....	132
Figure 3.14: Tonal confusion matrices for Cantonese. ....	133
Figure 3.15: Bin test for intonation responses in Cantonese. ....	135

Figure 3.16: The rate of accuracy in judging intonational category, by tonal category of the stimulus, in Cantonese—the declarative condition above and the echo question condition below.....	136
Figure 3.17: Rates of correct intonational category response by tone response, in the declarative condition (above) and the echo question condition (below), in Cantonese. ....	137
Figure 3.18: A sample screenshot from the NKK perceptual test. ....	142
Figure 3.19: Bin test for word responses in the NKK perceptual test. ....	146
Figure 3.20: Rate of perceptual accuracy for tone (excluding the monosyllabic condition), by tone of the stimulus, in NKK. ....	147
Figure 3.21: Tonal confusion matrices for the disyllabic condition in NKK. ....	148
Figure 3.22: Bin test for intonation responses in NKK. ....	148
Figure 3.23: Rate of intonational accuracy by intonation, syllable count, and tonal category for NKK.....	149
Figure 3.24: Schematization of slope and alignment of double-accented and final-accented contours on disyllables in an echo question context in NKK. ....	151
Figure 3.25: A sample of the answer sheet in the SJ perceptual test.....	154
Figure 3.26: Tonal confusion matrices for the disyllabic condition in SJ. ....	156
Figure 3.27: Bin test for intonation responses in SJ. ....	156
Figure 3.28: Pitch tracks for L-unaccented <i>asa</i> ‘hemp’ (thin black line) and L-final-accented <i>asa</i> ‘morning’ (thick gray line) in isolation in an echo question context, in SJ.....	158
Figure 3.29: Cross-linguistic comparison of perceptual accuracy for tone and intonation. ....	161
Figure 4.1: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for each of the four lexical tones in Frame 1, for Mandarin Speaker A, with predicted echo question contours (dotted light gray) for T2, T3, and T4 based on the hypothetical mechanism of phrase curves that diverge starting at the beginning of the last syllable (modeled after T1). ....	170

Figure 4.2: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for each of the four lexical tones in Frame 1, for Mandarin Speaker B, with predicted echo question contours (dotted light gray) for T1 and T2 based on the hypothetical mechanism of a multiplicative function, modeled after T3 and T4. .... 172

Figure 4.3: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for each of the four lexical tones in an all-T3 frame in Henanhua, with predicted echo question contours (dotted light gray) for T1 and T3 based on the hypothetical mechanism of a multiplicative function with a movable baseline, modeled after T2 and T4. .... 174

Figure 4.4: Mean declarative (dark gray) and echo question (light gray) pitch contours for each of the four lexical tones in an all-T3 frame in Henanhua. .... 175

Figure 4.5: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for Cantonese Speaker B's T1, T3, and T6, with a predicted echo question contour (dotted light gray) for T1 based on the hypothetical mechanism of a static high target approached after an initial steady-state pitch, modeled after T3 and T6. .... 177

Figure 4.6: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for Cantonese Speaker B's T2, T4, and T5, with a predicted echo question contour (dotted light gray) for T4 based on the hypothetical mechanism of a relative high target dependent on relative declarative pitch height, modeled after T2 and T5..... 178

Figure 4.7: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for all six tones for Cantonese Speaker B..... 180

Figure 4.8: T1 and T3 echo question contours compared to T2 declarative contours for Cantonese speakers A (left) and B (right)..... 180

Figure 4.9: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for all tonal categories on monosyllabic (top two) and disyllabic (bottom three) words in NKK..... 184

Figure 4.10: Schematization of a purely phonetic interaction between high and low targets in NKK..... 186

Figure 4.11: Individual declarative (dark gray) and echo question (solid light gray) pitch contours for an initial-accented monosyllable at a slow speech rate (above) and an initial-accented disyllable (below) in NKK.....	187
Figure 4.12: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for accented monosyllables (top) and disyllables (middle and bottom) in Kyoto Japanese (left) and Shiga Japanese (right). .....	189
Figure 5.1: Idealized contours associated with Class 1 words and Class 2 words in declarative and interrogative contexts in Arzbach and Cologne (from Köhnlein 2011).....	196
Figure 5.2: Kori’s schematization of the four phrase melodies in Osaka Japanese.....	210
Figure 6.1: Template for a lexical tune based on Bao’s (1990) autosegmental representation of lexical tone. T = <i>tune</i> , R = <i>register</i> , c = <i>contour</i> , and t = <i>contour endpoints</i> .....	222
Figure 6.2: Bao (1990)-style representations for the six Cantonese lexical tunes. The corresponding tune labels and Chao tone-letter-numerals are shown underneath.....	223
Figure 6.3: Representations for the four lexical tunes of Mandarin. ....	223
Figure 6.4: Representations for the four lexical tunes of Henanhua.....	224
Figure 6.5: A phonological representation of the segmental anchoring of tones utilizing secondary association lines (adapted from Ladd 2008 Figure 5.2). .....	225
Figure 6.6: Underlying forms for lexical tunes on disyllables in SJ.....	226
Figure 6.7: Surface forms for lexical tunes on disyllables in SJ.....	227
Figure 6.8: Surface representation for a high-beginning, second-accented trisyllable in SJ.....	228
Figure 6.9: Underlying representations for lexical tunes on disyllables in Tokyo Japanese.....	229
Figure 6.10: Surface representations for lexical tunes on disyllables in Tokyo Japanese. ....	229
Figure 6.11: Underlying representations for lexical tunes on double-, initial-, and final-accented disyllables in NKK.....	231
Figure 6.12: Surface representations for lexical tunes on disyllables in NKK. ....	231
Figure 6.13: Surface representations for lexical tunes on double- and initial-accented monosyllables in NKK.....	233

Figure 6.14: Underlying representation for the melody of the Cantonese utterance <i>Lei6lei6wa6laan4?</i> ('Leilei says 'orchid?').	236
Figure 6.15: Rearranged underlying representation for the melody of the Cantonese utterance <i>Lei6lei6wa6laan4?</i> ('Leilei says 'orchid?').	237
Figure 6.16: Surface representation for the melody of the Cantonese utterance <i>Lei6lei6wa6laan4?</i> ('Leilei says 'orchid?').	237
Figure 6.17: Underlying representation for the melody of the Mandarin utterance <i>Na4liang4nian4wan4?</i> ('Naliang reads 'ten thousand?').	239
Figure 6.18: Surface representation for the melody of the Mandarin utterance <i>Na4liang4nian4wan4?</i> ('Naliang reads 'ten thousand?').	239
Figure 6.19: Underlying representation for the melody of the NKK utterance <i>Eunhi-neun nam-i?</i> ('Eunhi-TOP horse-NOM?').	240
Figure 6.20: Surface representation for the melody of the NKK utterance	240
Figure 6.21: Proposed surface representations for three lexical tunes in an echo question context in NKK, with accompanying schematic representations for phonetic implementation.	241
Figure 6.22: Proposed surface representations for three lexical tunes in an echo question context in NKK (with a double-linked h representation for the double-accented tune), with accompanying schematic representations for phonetic implementation.	242
Figure 6.23: Alternative surface representations for three lexical tunes in an echo question context in NKK (with serial implementation of melodic units on the last syllable), with accompanying schematic representations for phonetic implementation.	243
Figure 6.24: Underlying representation for the melody of the SJ utterance <i>Aya-ga ame?</i> ('Aya-NOM rain?').	244
Figure 6.25: Surface representation for the melody of the SJ utterance <i>Aya-ga ame?</i> ('Aya-NOM rain?').	244
Figure 6.26: Schematic representation of prohibited crossing of association lines.	248

## LIST OF TABLES

Table 2.1: Descriptive labels for the four lexical tones in Mandarin.....	27
Table 2.2: Target-utterance duration ratio by intonation for Mandarin Speakers A and B. ....	30
Table 2.3: Descriptive labels for the four lexical tones in Henanhua. ....	39
Table 2.4: Target-utterance duration ratio by frame type by intonation for Henanhua. ....	43
Table 2.5: Descriptive labels for the six lexical tones in Cantonese. ....	52
Table 2.6: Target-utterance duration ratio by intonation for Cantonese Speakers A and B. ....	55
Table 2.7: Tone classes and equivalent Lee (2008)-style labels on disyllables in NKK.....	68
Table 2.8: Categorized word list for NKK production experiment .....	69
Table 2.9: Target-utterance duration ratio by intonation for NKK.....	71
Table 2.10: Haraguchi-style labels and example sequences for the four tone classes in Osaka Japanese. ....	84
Table 2.11: Mean durational difference between intonations by accentedness for SJ. ....	90
Table 2.12: realization of post-accentual fall in a phrase-final context in the Kansai dialects...	101
Table 3.1: Overall rates of perceptual accuracy in judging tonal and intonational categories in Mandarin.....	112
Table 3.2: Rate of perceptual accuracy in judging tonal category, by intonation type of the stimulus, in Mandarin. ....	113
Table 3.3: Rate of perceptual accuracy in judging intonational category, by intonational category of the stimulus, in Mandarin .....	115
Table 3.4: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations in Mandarin.....	120
Table 3.5: Overall rates of perceptual accuracy in judging tonal and intonational categories in Henanhua .....	121
Table 3.6: Rate of perceptual accuracy in judging intonational category, by intonational category of the stimulus, in Henanhua .....	122



Table 3.7: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations in Henanhua.....	125
Table 3.8: Overall rates of perceptual accuracy in judging tonal and intonational categories ...	128
Table 3.9: The rate of accuracy in judging tonal category, by intonational category of the stimulus, in Cantonese. ....	129
Table 3.10: Pairwise comparisons of tonal accuracy among tones in the echo question condition for Cantonese. ....	131
Table 3.11: Rate of perceptual accuracy in judging intonational category, by intonational category, in Cantonese.....	135
Table 3.12: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations in Cantonese .....	139
Table 3.13: Word list for NKK perceptual test.....	141
Table 3.14: Overall rates of perceptual accuracy in judging tonal and intonational categories in NKK.....	144
Table 3.15: Rate of perceptual accuracy in judging tonal category, by syllable count of the stimulus, in NKK .....	144
Table 3.16: Rate of perceptual accuracy in judging tonal category, by syllable count of the stimulus, for NKK speaker A (listening to her own speech).....	144
Table 3.17: Revised rates of perceptual accuracy in judging tonal and intonational categories in NKK (excluding monosyllables) .....	145
Table 3.18: The rate of perceptual accuracy in judging tonal category, by intonational category of the stimulus (excluding the monosyllabic condition), in NKK. ....	146
Table 3.19: Rate of perceptual accuracy in judging tonal category, by intonational category of the stimulus, in NKK .....	149
Table 3.20: Overall rates of perceptual accuracy in judging tonal and intonational categories in SJ.....	155

Table 3.21: The rate of perceptual accuracy in judging tonal category, by intonational category of the stimulus, for SJ.....	155
Table 3.22: Rate of perceptual accuracy in judging intonational category, by intonational category of the stimulus, in SJ.....	157
Table 3.23: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations in SJ .....	159
Table 3.24: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations across languages .....	163
Table 5.1: First attempt at derivations for NKK melodies.....	202
Table 5.2: Revised derivations for double-accented melodies in NKK.....	207
Table 5.3: First attempt at derivations for Kansai melodies .....	214

## CHAPTER 1: INTRODUCTION AND BACKGROUND

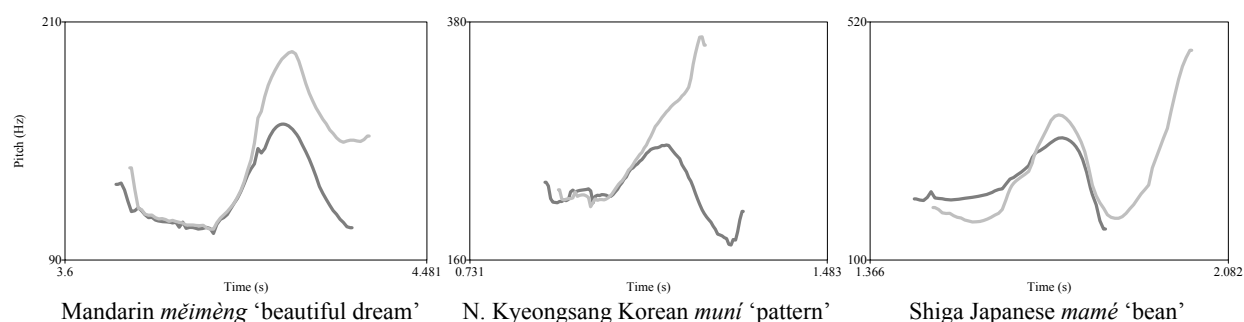
Variations in fundamental frequency—or pitch<sup>1</sup>—over time are an integral part of the speech signal. There are many different factors that can contribute to these pitch variations, including linguistic ones (lexical tone; phrasal intonation), paralinguistic ones (attitude; emotional state), and physiological ones (condition of the larynx; lung capacity). The linguistic factors are under speaker control; the physiological factors are not; the paralinguistic factors may be under speaker control or not (see Gussenhoven 2004 Ch. 5 for a discussion of paralinguistics and speaker control). We may refer to the linguistic functions and those paralinguistic functions that are under speaker control collectively as *communicative functions*. We may in turn refer to the pitch patterns governed by those communicative functions as *speech melody*<sup>2</sup>. That languages are able to encode several communicative functions in the speech melody is quite remarkable, as is the variety of ways in which different languages handle the encoding. As a simple illustration of this typological diversity, pitch tracks of utterances from three different tone languages are shown in Figure 1.1. In each case, a word that ends in a “falling” lexical tone has been produced once with declarative intonation and once with interrogative intonation, and the pitch contours for each intonational rendition are superimposed on one another. With just a brief glance at these pitch tracks it is obvious that the interaction of lexical tone and utterance-level intonation is different in each of these three languages. In Mandarin, the falling contour on the second syllable is preserved in the echo question context, but it is shifted into a higher register relative to its declarative counterpart. In North Kyeongsang Korean, the pitch rise that characterizes the

---

<sup>1</sup> Here and throughout the dissertation the terms *pitch*, *fundamental frequency*, and *F0* are used interchangeably.  $F_0$  is a commonly used abbreviation for fundamental frequency, but the lack of distinction between *fundamental frequency* and *pitch* is rather imprecise; the former refers to an acoustic property of the speech signal (measured in Hz) while the latter is its psychophysical correlate (Gussenhoven 2004; Ladd 2008). Given the scope of discussion in this dissertation, this indiscriminate use of the two terms should not detract from the comprehensibility of what is being conveyed in the discussion.

<sup>2</sup> Xu (2005) used this term, although it is not clear whether he was considering any paralinguistic functions to factor into speech melody.

interrogative intonation appears to “overwrite” the pitch fall that is apparent in the declarative context. Finally, in Shiga Japanese, the interrogative rendition of the word appears to be characterized by a sharply rising “tail” in its contour that is absent in the declarative rendition of the word.



**Figure 1.1: Pitch tracks of disyllabic words in three languages (Mandarin, North Kyeongsang Korean, and Shiga Japanese), each with a “falling” lexical tone on the second syllable, in declarative (dark gray) and interrogative (light gray) contexts.**

Despite the ease with which speakers of all different languages produce and interpret speech melody, the lack of a consensus on the best way to model it is a testament to how complex it is. Leaving the paralinguistic communicative functions aside<sup>3</sup>, there are at least two broad categories of models that attempt to capture the interaction of the various linguistic communicative functions in speech melody. Ladd (1996; 2008) referred to them as *overlay* models and *sequential* models, respectively<sup>4</sup>. Overlay models have often been deemed well-equipped to handle the melodic systems of languages like Mandarin, while sequential models have been used for those of languages like Japanese. In the following sections these two model types will be schematized and some examples will be briefly reviewed. First, though, a couple of other terminological clarifications are in order.

<sup>3</sup> See Ohala (1983), Ohala (1984), and Chen, Gussenhoven et al. (2004) for discussions on paralinguistic uses of pitch across languages.

<sup>4</sup> Ladd (2008) used these labels in a discussion on modeling intonation—not speech melody as a whole—but the models he cites all include components that handle lexical tonal specifications as well as intonational (i.e. postlexical melodic) ones.

## **1.1 Definitions of *Lexical Tone* and *Intonation***

### **1.1.1 *Lexical tone***

The term *lexical tone* (and, when it is unambiguous, just *tone*) covers any tonal specification that is part of the lexical specification of a word or a morpheme. This use of the term is in keeping with Hyman's (2001) definition of a tone language (also adopted by Yip 2002)), which crucially includes languages like Japanese, which are sometimes referred to as *pitch accent* languages. At points in the discussion of such languages, the term *lexical pitch accent* (or just *pitch accent*) will be used for the sake of convenience when making reference to existing analyses of the melodic systems of those languages, but eventually in Chapter 6 the classical notion of a lexical pitch accent will be discarded and the difference between Mandarin lexical tone and Japanese lexical tone will be captured as one of scopal domain of the lexical tone, which in turn will be modeled as a difference in level of association (the syllable for Mandarin and the prosodic word or accentual phrase for Japanese). This notion of associating lexical tones with different levels is not strictly new, although researchers have tended to stick to a single level of association as it suited the demands of their individual analyses; Leben (1973), whose analysis of tone as a suprasegmental feature was pivotal in ushering in the autosegmental revolution in the world of tone, advocated analyzing lexical tone as a feature on morphemes, and he also cited examples of other analyses that specified tone as a feature on the segment (Schrachter and Fromkin 1968; Woo 1969; Maddieson 1971), the syllable (Pike 1948; McCawley 1964; Wang 1967), the morpheme (Welmers 1962; McCawley 1970), and the phonological word (Rowlands 1959; Edmondson and Bendor-Samuel 1966). This notion of treating level-of-association as a typological parameter within a unified phonological component of speech melody will be fleshed out in Chapter 6.

### **1.1.2 *Intonation***

The term *intonation*, as it is used in this dissertation, refers to the collective components of speech melody other than the lexical-tonal component—that is, intonation is a strict sub-part of

speech melody in tone languages. This use of the term is in line with that of Ladd (1996; 2008) and Gussenhoven (2004) but contrasts with how it is defined by Gårding (1984), Laniran (1992), and others who use it in a sense that is synonymous with speech melody.

## **1.2 Overlay vs. Sequential Models**

In this section, the differences between overlay and sequential models will be laid out, with notable examples of each presented and briefly described. Examples of overlay models that will be covered include the Lund School model (Bruce and Gårding 1978; Gårding 1979; Gårding and Bruce 1981; Gårding 1983), the command-response model (Fujisaki and Hirose 1982; Fujisaki 1983; Gu, Hirose et al. 2006), the Stem-ML-based model for Mandarin (Shih and Kochanski 2000; Yuan, Shih et al. 2002; Kochanski, Shih et al. 2003), and the PENTA model (Xu 2005)<sup>5</sup>. Sequential models that will be covered include the tone levels model (Clements 1979), the register hierarchy model (Clements 1990), the Pierrehumbert and Beckman model for Japanese (Pierrehumbert and Beckman 1988). In addition to providing a brief overview of the different models, the different predictions that each type of model makes will be explicitly delineated.

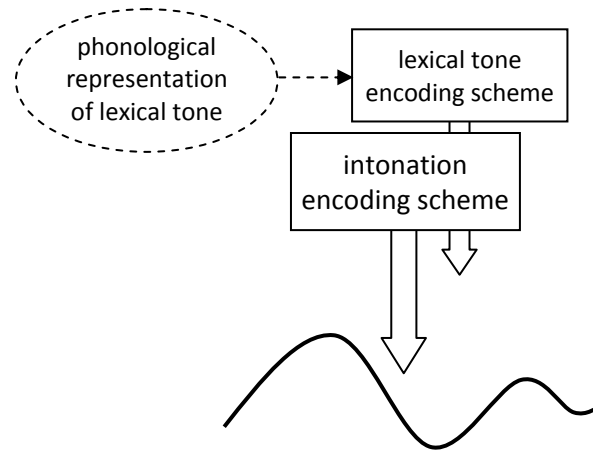
### **1.2.1 Overlay models**

An overlay model is schematized in Figure 1.2 where, for simplicity's sake, intonation is represented as a single communicative function, although in various models the category of intonation may be broken down further into subcategories like utterance-type intonation, focus, etc. In overlay models, most of the “work” is done by the phonetics. There may be some simple phonological processes—such as certain types of tone sandhi or tone deletion—that are accommodated in overlay models, but in practice most descriptions of these models make no

---

<sup>5</sup> Some of the names of the models mentioned here—both the overlay and sequential models—are not official names used by the original authors but either names that have been used in subsequent literature or names coined here for ease of reference.

mention of a phonological component. Lexical tones and intonational functions are encoded independently of one another and implemented in parallel to yield the final acoustic output



**Figure 1.2: Schematic diagram of an overlay model.**

(represented as a stylized contour in the diagram). The parallel implementation scheme that characterizes this type of model is graphically captured by Bolinger’s (1964) “ripples on waves on swells on tides” metaphor (a metaphor also espoused by Chao 1968).

### 1.2.1.1 The Lund School model

One example of an overlay model is the Lund School *grid* model (Bruce and Gårding 1978; Gårding 1979; Gårding and Bruce 1981; Gårding 1983). In this model, local tonal events, including lexical ones, are modeled in terms of *turning points* in the  $F_0$  contour, and global intonational effects are instantiated as a *grid* that forms the backdrop for these local turning points. The direction and degree of separation of the latitudinal lines that form the grid can change at various *pivots*. The grid model was originally introduced to describe the speech melody systems of various Swedish dialects, but Gårding, Zhang et al. (1983) later modeled Mandarin tone and intonation with the grid model.

### 1.2.1.2 The Command-Response model for Japanese

Another overlay model is the *command-response* model for Japanese (Fujisaki and Nagashima 1969; Fujisaki and Hirose 1984). In this model, *accent commands* and *phrase commands* are specified independently at various points in time, each with their own amplitude settings, and they are interpreted in parallel by a glottal oscillation mechanism that produces a continuous  $F_0$  contour as output. The accent commands corresponded to the more localized pitch movements associated with so-called *pitch accents* that are lexically specified in Japanese, while the phrase commands were tied to larger-scale intonational processes like declination. Ladd (1996) placed the command-response model squarely in the overlay camp, largely based on how the model handled declination independently of lexical pitch accents. Branching out from Japanese, Gu, Hirose et al. (2006) showed how the model could be made to handle the interaction of lexical tone and utterance-type intonation in Cantonese (using *tone commands* instead of accent commands). They were able to produce appropriate pitch contours for the lexical tones in an echo question context, but only by specifying a unique amplitude value for the final tone command as well as for the final phrase command in each case. In other words, although the phrase commands and tone commands were implemented in parallel, a unique combination of phrase-command and tone-command amplitudes was “hand-crafted” for each tone-intonation combination, in essence abandoning any implicit assumption of parallel encoding of the two melodic functions.

### 1.2.1.3 Stem-ML-based model for Mandarin

Kochanski and Shih (Kochanski and Shih 2000; Kochanski and Shih 2003) introduced a prosody modeling language called *Stem-ML* (*Soft Template Markup Language*) that was designed to function as the “prosody-generation” component to a text-to-speech system. Stem-ML in and of itself is not a model of speech melody, but it provides a set of mark-up tags that can be used to create the implementation component of such a model. A model for Mandarin speech melody that used Stem-ML tags was proposed by Kochanski, Shih and colleagues (Shih and Kochanski



2000; Yuan, Shih et al. 2002; Kochanski, Shih et al. 2003). This model implicitly assumes parallel encoding and maintains parallel implementation, just like the other overlay models discussed above. Templates for lexical tones are expressed as a set of parameters specified for every syllable, and two key components in the intonation implementation are a strength parameter that is set on every syllable and a phrase curve, which defines a global baseline for  $F_0$ .

#### **1.2.1.4 The PENTA model**

Xu (2005), building on Xu and Wang (2001), introduced the *PENTA (Parallel Encoding and Target Approximation)* model. Like the Stem-ML model for Mandarin, the PENTA model defines a finite set of parameters (called *melodic primitives*) that are specified for every syllable. Various communicative functions, including lexical tone and utterance-type intonation, contribute independently to the final value settings of these parameters. Once again, the PENTA model assumes a purely parametric approach to speech melody modeling, and it makes no reference to a phonological representation of intonation; therefore there is no phonological interaction between lexical tones and intonation in the model. The only indirect connection intonation has to any kind of phonological representation is that the target implementation, as well as parameter assignment, is syllable-synchronized. However, Xu (2005) defined the syllable as an articulatory unit, not a phonological one.

#### **1.2.1.5 Predictions of overlay models**

Before moving on to the other category of models, it should be reiterated that, because all of the overlay models described here explicitly assume or implicitly imply a direct parametric implementation of intonation, they do not allow for any interaction at the phonological level between lexical tones and intonational functions. In addition, since they all assume parallel implementation of the various melodic functions, there is no chance for interaction among the functions at the phonetic level, either. Because of this configuration, overlay models make the predictions shown in (1.1):

(1.1) Predictions of phonetics-only overlay models

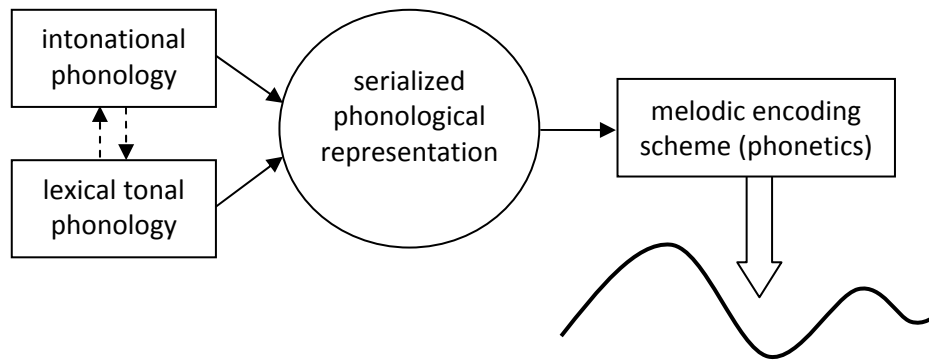
- I. The mutual alignment of pitch effects that occur in parallel should be unconstrained, except that the onsets or offsets of pitch effects may align with syllables for articulatory reasons.
- II. We should not find lexical-tone-specific effects in the implementation of intonational functions such as focus or utterance-type intonation.

These predictions will be important to keep in mind as the results from production experiments in several tone languages are discussed in Chapters 2 and 4 of this dissertation. It should be noted that many of the mechanisms underlying the models discussed above are compatible with models that allow for dependencies among melodic functions (either at the phonological level or between the phonological and phonetic level) and, as such, could be adapted for use in the phonetic components of such models. This point was discussed in the abstract by Kochanski and Shih (2003), who suggested that Stem-ML is versatile enough to be used as the prosody generation component in both overlay and sequential models. Gu, Hirose et al. (2006) adapted the command-response model developed by Fujisaki and Hirose (1984) to handle the interaction of lexical tones and echo question intonation in Cantonese. While the fundamental mechanism of the parallel implementation of tone commands and phrase commands was maintained for the Cantonese model, the amplitude values specified for the commands corresponding to the echo question intonation were unique depending on the identity of the lexical tone at the right edge of the utterance. In essence, then, the settings for the commands were “hand-crafted” according to the lexical tonal category, and any implicit notion of the independent encoding of tone and intonation was abandoned. The predictions in (1.1) would of course not apply to a model with the level of power built into that described by Gu, Hirose et al. (2006). Speech melody models of the type discussed in the next section—sequential models—allow for the interaction of

different melodic functions at the level of phonological representation, although the interactions may be constrained structurally.

### 1.2.2 Sequential models

A sequential model is schematized in Figure 1.3 where, once again, intonation is represented as a single function. In sequential models, the intonational functions, in addition to the lexical tonal



**Figure 1.3: Schematic diagram of a sequential model.**

ones, are specified in the phonology. Both are subject to phonological processes, and lexical tone and intonational mechanisms can interact at that stage. The output of the phonology is a serialized phonological representation (in which melodic “events” or “landmarks” have been encoded sequentially) that is then interpreted by a unified melodic encoding scheme to yield the final acoustic output. Note that the terms “serialized” and “sequential” are not meant to imply that the speech melody is solely represented as a linear sequence of targets. In these models, the phonological representation does include such sequences (and the targets can be a mixture of lexical tonal units and intonational units), but sequential models as defined here also allow for multiple melodic functions to affect the same stretch of the utterance simultaneously. Crucially, though, such interactions are mediated in the phonology via structural relations and the beginnings and endings of those stretches correspond to significant structural landmarks. A brief

look at the sequential model proposed by Clements (1979) drives this last point home. This and other sequential models are discussed below.

### **1.2.2.1 The Tone Levels model (Clements 1979)**

Examining data from Akan and Kikuyu (tone languages spoken in western and eastern Africa, respectively), Clements (1979) sketched out a model that involves a mechanism very much like the Lund School grid—a backdrop for lexical tones that he calls a *tone level frame*:

A tone level frame does not constitute a set of absolute acoustic parameters. Rather, it is subject to modification as a result of the intonational processes that apply to it. The identity of the frame itself, however, is not affected by these modifications. It might be compared to a grid drawn upon a flexible sheet, which retains its identity in spite of the distortions which result if the sheet is warped or twisted. (p. 549)

What sets Clements's (1979) model apart from the Lund School model is that, in addition to this backdrop for the lexical tones, it included a mechanism in the phonology that could trigger changes in the register of that backdrop. For example, he analyzed downstep as the insertion of a *downstep entity* (represented by the symbol !), and he posited SPE<sup>6</sup>-style phonological rules in which the triggering environments for the insertion of the downstep entity were various compositions of lexical tonal autosegments. Clements (1979) was restricting his discussion to the analysis of downstep and upstep in Akan and Kikuyu, but he noted that “this descriptive framework can be extended in a relatively straightforward fashion to take account of the additional factors that determine pitch” (p. 551-552).

### **1.2.2.2 The Register Hierarchy model (Clements 1990)**

A more nuanced register-shifting approach to downstep was formalized by Clements (1990) and embraced by Laniran (1992). This approach involved grouping sequences of lexical tones into *tonal feet* (based on language-specific criteria) that delimited the scope of each downstep or

---

<sup>6</sup> *The Sound Pattern of English*—Chomsky and Halle (1968).

upstep register. Downstep and upstep were then expressed in terms of a hierarchical register tree that allowed the relative register level of any given tonal foot to be derived arithmetically from the combination of *l* and *h* nodes dominating that foot.

### **1.2.2.3 Pierrehumbert and Beckman (1988)**

The quintessential example of a sequential model of speech melody that also included a detailed proposal for phonetic implementation is the underspecified autosegmental model proposed by Pierrehumbert and Beckman (1988) for Japanese speech melody. An extension of Pierrehumbert's (1980) model for English speech melody, their model specifies lexical tonal events and intonational events as sequences of high and low autosegments (abbreviated H and L, respectively) that associate with prosodic constituents in a prosodic hierarchy. The lexical tonal units are underlyingly associated with morae, and phrasal intonation units are underlyingly associated with higher-level constituents such as the "accentual phrase" and the "utterance". Before this string of melodic autosegments gets phonetically implemented, certain phonological processes apply, resulting in a surface-phonological representation that differs from the underlying representation. (Notably, some intonational units are secondarily associated with morae at edges that do not already have lexical units associated with them.) Phonetic implementation rules then interpret the surface string from left to right according to a principled set of phonetic rules, with the scaling and alignment of pitch targets depending on the type of melodic unit in question and the type of melodic unit immediately preceding it (in the case of things like melodic-sequence-dependent downstep, for example) as well as association configurations (i.e. status in the metrical prosodic hierarchy), and with the  $F_0$  of points between pitch targets arising from linear interpolation.

One conceit of the Pierrehumbert and Beckman (1988) model, and in fact all sequential models, is that there is a phonological component to speech melody that mediates between the communicative functions themselves and the phonetic implementation. Consequently, the phonetic implementation is blind to the sources of any melodic units in the surface representation.

In Pierrehumbert and Beckman (1988) terms, the phonetic implementation treats a H\* and a H- as different not because it knows that the source of one is lexical and that of the other is postlexical/intonational, but because the two Hs are structurally different—the H\* is coupled with a L melodic unit and the H\*+L sequence is associated with a mora, while the H- is associated with an accentual phrase node higher up in the prosodic tree. Laniran (1992) noted: “There is no reason to believe that the rules for tone implementation have access to where the tones come from, since implementation follows the same general principles independent of language type (stress, accent or lexical tones)” (p. 23). Ladd (2008) pointed out that this principle is in keeping with how linguists approach other areas of phonetics, a point that is sometimes overlooked:

In segmental phonetics, instrumental research is devoted to studying the physical cues to properties like voicing or vowel quality or nasality. Phoneticians do not try to study the physical cues to properties like plurality or verb aspect or negation—it seems obvious that it would be pointless to do so. The segmental categories investigated by instrumental phonetics are *phonological*, not lexical or grammatical. Yet one of the characteristic features of traditional instrumental research on intonation is that in many cases it attempts to identify direct physical correlates of meanings or linguistic functions, such as happiness or contrast or finality. (p. 17)

#### **1.2.2.4 Predictions of sequential models**

In contrast to strictly phonetic overlay models, then, the sequential models discussed above make the predictions shown in (1.2):

(1.2) Predictions of sequential models:

- I. Some phonetic pitch effects that occur in parallel are constrained in their relative alignment with one another and with respect to phonological constituents other than the syllable.

- II. The nature of the effects can be dependent on the phonological categories of the sources of those effects (i.e. we expect to find lexical-tone-specific effects in the implementation of intonational functions such as focus or utterance-type intonation).

Once again, results from the experiments presented in Chapter 2 and subsequent discussion in Chapter 4 will come to bear on these predictions (along with those presented in (1.1)). It will turn out that both predictions made by sequential models are borne out in the results.

### **1.3 Issues Raised by a “ToBI Typology”**

While there is evidence that sequential models are on the right track, there is not currently a template for a sequential model that makes it possible to compare easily across languages the nature of the interaction of various melodic functions. The framework in which the phonological component of the Pierrehumbert and Beckman (1988) model was couched has gotten us part of the way there in that it allows researchers to posit inventories of melodic units and to analyze certain interactions they observe as phonological processes (like the deletion or the delinking of a certain kind of melodic unit in a certain phonological environment, e.g.). This framework, based largely on the framework set forth in Pierrehumbert (1980) as well as concepts presented in Liberman (1975) and Bruce (1977) came to be known as the *Autosegmental Metrical* (henceforth AM) framework (coined by Ladd 1996).

When the ToBI (Tones and Break Indices) system arrived on the scene as a melodic transcription system for American English (Silverman, Beckman et al. 1992), it—and the AM framework in which it was couched—had a heavy influence on the analyses of many other languages, including some (other than Japanese) with lexical tone like Serbo-Croatian (Godjevac 2005), Cantonese (Wong, Chan et al. 2005), Mandarin (Peng, Chan et al. 2005), and Chicasaw (Gordon 2005). Some of these reports were compiled in a volume edited by Sun-Ah Jun entitled *Prosodic Typology: The Phonology of Intonation and Phrasing* (Jun 2005), along with some reports on non-tone languages. Despite the volume’s title, approaching this collection of

descriptive analyses as a meaningful typology from a phonological perspective is problematic—with respect to the tone languages in particular—for several reasons. These issues are listed in (1.3):

(1.3) Issues raised by a ToBI typology

- I. *Varying representations of lexical tone* – Lexical tonal units were not represented in ways that fostered meaningful comparisons across languages. Lexical tonal units in languages like Japanese, i.e. languages traditionally referred to as *pitch accent languages*, were expressed in terms of complexes of L and H units, but lexical tonal specifications for Mandarin and Cantonese were expressed in Chao (1930)-style “tone letter” values (in which the pitch-level contours of each tone are assigned a sequence of numerical values on a relative five-point scale).
- II. *Inert phonology* – In providing language-specific transcription schemata for these various languages, the authors were in essence positing surface-phonological representations for various types of melodies. However, many of the reports made no claims about underlying representations or the possible phonological processes that yielded the representations being posited (there were some exceptions—e.g. Gordon 2005). For example, in the analysis provided by Peng, Chan et al. (2005) for Mandarin, intonational units like boundary tones were placed on a tier separate from the tier specifying lexical tones, and the authors also posited tags like *%reset*, which indicated the beginning of a new pitch downtrend, and *%raise*, the beginning of a raised pitch range. On yet another tier, they posited four “stress” levels on syllables (independent of their inherent lexical tones). Clearly the authors were analyzing what they saw in the pitch contours as sequences of pitch events occurring in parallel with one another, but they were non-committal about whether those events were tied to anything structural other than the syllable (which, as we have seen from Xu (2005), may be defined as an articulatory unit) and how the events interacted with one another.



III. *Lack of clarity regarding phonetic interpretation* – The various analyses also made no claims about the phonetic interpretation of the surface phonological configurations they posited other than providing the  $F_0$  contours themselves. So, for example, Peng, Chan et al. (2005) posited for echo questions a representation in which the final syllable in an utterance was given a lexical tonal specification on one tier and a H boundary tone on another tier, but it is unclear how they envisioned the phonetics interpreting this configuration. Likewise, Wong, Chan et al. (2005) modeled echo question intonation in Cantonese with a H boundary tone that was placed after the rightmost lexical tone on the same tier, but the closest they came to describing the implementation of this sequence of melodic units was to describe the effect of the H boundary tone as a “rise from the final lexical tone” (p. 287). The lack of discussion on phonetic interpretation in a volume purportedly focusing on phonology is certainly not surprising, but it dampens the explanatory power of a supposedly “language-independent” transcription system if we don’t know which acoustic differences reflect language-specific structural differences and which ones are being chalked up to language-specific phonetics.

One of the aims of this dissertation is to try to provide a more explicit basis for the typological comparison of the melodic systems of various languages. Laid out in the following section are an overview of the general experimental paradigm used in the current study and a preview of the model born from the results of the study.

#### **1.4 Toward a Melodic Typology**

This dissertation presents a series of production and perception experiments that are designed to highlight how two different melodic functions interact in various languages. All of the languages under investigation make use of lexical tone; the other melodic function that is manipulated in the experiments is utterance-type intonation. Based on the results of these experiments, as well as on some other observations in the literature, a new type of sequential

model is proposed in Chapter 6—a model that fosters more meaningful cross-linguistic comparisons of melodic systems.

### **1.4.1 Declarative vs. echo question intonation**

In the experiments presented in Chapters 2 and 3 of this dissertation, the objective was to explore the nature of the interaction between the melodic manifestations of two different communicative functions in each language. The term *melodic functions* has been used thus far in this chapter as a cover term for these different melodic manifestations, that is, it is intended to be an umbrella term for lexical tone and various intonational functions. In the context of the experimental results, the melodic functions will sometimes be categorized in terms of their corresponding communicative functions, but this should not be taken as a theoretical stance. Certainly, the fact that echo questions were elicited from speakers of these various languages should not be interpreted as a contention that echo questions are encoded by the same type of melodic function in all languages. The term *echo question intonation* should be taken as shorthand for “the melodic function that encodes echo questions in the language at hand”.

#### **1.4.1.1 Working definition of *echo question***

The kind of question being referred to here as an *echo question* is a subclass of what Fiengo (2007) dubs *confirmation yes-no questions*<sup>7</sup>. Such questions have the same syntax and lexical composition as an assertion, but they are used to seek confirmation regarding a belief related to that assertion or to express a lack of confidence in such a belief. Fiengo (2007) identifies several different sub-types of confirmation yes-no questions, based on the various belief scenarios in which the questions may be used. In one scenario the speaker may have a belief but is not confident enough in her belief to express it as an assertion, so she expresses it as a confirmation yes-no question to elicit confirmation from the listener. For example, someone may utter “It’s

---

<sup>7</sup> It should be noted that Fiengo (2007) actually equates the term *echo question* with a different type of question—namely one in which the speaker is asking someone to repeat something by means of a construction that is not lexically equivalent to the original assertion but rather is incomplete or contains a *wh*-element—for example, “You just bought what?” in response to “I just bought a horse.” (p. 77)

raining?” in response to seeing someone walk in off the street soaking wet. In another scenario, a student in a chemistry class might say, “Brass is an element.” The teacher might then respond, “Brass is an element?” to communicate that she lacks a belief in the proposition that brass is an element that is strong enough to warrant an assertion. This second type of confirmation yes-no question is what will be referred to as an echo question in this dissertation. Fiengo (2007) notes that in scenarios like the one described above, elements of sarcasm or surprise may enter into the use of the echo question. For example, the teacher may not be merely expressing a lack of confidence in being able to assert that brass is an element but rather expressing her strong belief that brass is *not* an element and in essence asserting that the student is wrong. From a semantic point of view, it is clear that “how confident a speaker is in a belief” is a continuous variable, but defined as it is by Fiengo (2007), an echo question is something a speaker decides to *use* or *not use*, depending on her own threshold of confidence. The contrast of declarative statement (i.e. assertion) vs. echo question is therefore a categorical one, and we can safely assume that, if the contrast is encoded in the phonology, it will be encoded categorically as well (see Cohn 2006 for a discussion on categorical vs. gradient phonology). It is possible, of course, that *within* the range of levels of confidence in the belief that warrant the use of an echo question (from, say, “I have little confidence in that belief” to “I know that belief to be utterly wrong!”) the level of confidence in the belief (or, perhaps more appropriately, in the disbelief) may be encoded phonologically in the form of level of emphasis, which may be a gradient factor. Additionally, if there is a particular element of the original assertion to which the speaker objects (i.e. if the speaker believes that the truth value of the proposition in the original assertion could be changed from false to true by replacing one of the elements with a different one), it is possible that the speaker may employ a kind of contrastive focus prominence on the part of the utterance corresponding to that element. For example, in the chemistry class scenario, the teacher may place focus on the word *element*—as in “Brass is an ELEMENT?”—in order to highlight the fact that replacing the word *element* with another word such as *alloy* would render the proposition true. It is important to note here that, while the semantic interpretation of focus prominence may

differ slightly depending on whether the focus is employed in an assertion or in an echo question (see Rooth 1992 for a detailed account of focus interpretation), whether focus prominence is employed or not (and if so, where in the utterance it falls) and whether an utterance is an assertion or an echo question are orthogonal parameters. In other words, regardless of focus conditions, the utterance type of “Brass is an element?” is unambiguously an echo question as opposed to a declarative statement.

Note that, according to the above definitions, what are called “yes-no questions” or “polar questions” or even just “questions” in a lot of the literature on intonation correspond to the broader category of confirmation yes-no questions. Whether they can be further characterized as a strict subclass of this category actually depends on the context in which the questions are uttered. For example, in one production experiment presented by Yuan, Shih et al. (2002), different types of utterance-type intonation were elicited from a speaker of Mandarin by having him read randomized sentences that were presented on a computer screen with no context, some of which ended in a period and some of which ended in a question mark. For the sentences that ended in a question mark, the speaker would have had to conjure his own context to justify uttering the sentence as a confirmation yes-no question, and there is no way to know what context he had in mind for any given sentence. It is possible that he uttered some of the questions as echo questions and some of them as some other kind of confirmation yes-no question.

#### **1.4.1.2 Why echo questions?**

So why choose to focus on echo question intonation and the contrast between echo questions and declarative statements? There are three main reasons, which are presented in (1.4):

##### (1.4) Methodological justifications for using echo questions

- I. *Intonation and Only Intonation* – Echo questions tend to be manifested solely intonationally and they are easy to elicit naturally. All of the languages under

investigation make use of utterance- and phrase-final particles for the encoding of certain semantic and pragmatic distinctions, including for certain types of polar questions and in some cases even confirmation yes-no questions. For Cantonese in particular, the speakers reported that the only natural context for using an intonational confirmation question without a final particle was the echo question context. It is possible in some languages for the same intonational pattern to be observed on confirmation questions whether they end in a particle or not—Lee (2008) contends that there is no effect of the presence of a particle on the intonational realization of a question in North Kyeongsang Korean—but in other languages this is not the case—Pittayaporn (2005) argues that, if an interrogative final particle in Thai is specified for lexical tone, the lexical specification trumps any intonational specification at the right edge. Moreover, since a subset of the recordings made during the production experiments were intended for use in perceptual experiments, it was important that there not be any other cues—syntactic, morphological, semantic, or otherwise—to the categories involved other than melodic cues.

II. *A Categorical Contrast* – Secondly, as noted in the previous section, the declarative-vs.-echo-question contrast is a categorical one, which provides a two-way categorical melodic contrast that can be cross-cut with the categorical lexical contrasts embodied by lexical tone. This becomes especially important in perceptual experiments, in which we want to be able to ask listeners to assign what they are hearing to distinct categories. Of course, some semantic variables such as contrastivity and degree of “disbelief” (discussed in the previous section) were not controlled for, so there may have been some variation with regard to whether or how these functions were encoded in the intonation by the various speakers, but at the very least we can be reasonably confident that the declarative-echo-question contrast was encoded categorically.

III. *Echo Questions Are “High” in the Relevant Languages* – Previous literature gives one the impression that the “broad strokes” of the phonetic implementation of echo question intonation are similar in the various languages discussed here (and related languages).

Echo questions have been characterized as “rising” or having a “higher pitch” relative to their declarative counterparts in these languages, and in analyses that assume a phonological component to speech melody, echo question intonation is often analyzed as being encoded with some kind of “high” melodic unit for these languages. As a concrete example, AM-based analyses of echo questions in Mandarin (Peng, Chan et al. 2005), Cantonese (Wong, Chan et al. 2005), NKK (Lee 2008), and Kansai Japanese (Kori 1987; Pierrehumbert and Beckman 1988) have all involved a H boundary tone at the right edge of the utterance.

For the reasons outlined above, echo questions are an ideal utterance type to contrast with declarative statements in the context of the experiments presented in this dissertation.

#### **1.4.2 A new model**

The *scopal* model introduced in Chapter 6 of this dissertation is an attempt to exploit the clear advantages of an AM-based sequential model over overlay models and to augment it with additional mechanisms to endow it with the power to express more of the differences observed among melodic systems across languages in a structural way. In particular, it attempts to address the concerns raised in Section 1.3 regarding the “ToBI typology” as outlined in (1.5):

##### (1.5) Attributes of the scopal model

- I. *Lexical tune representation* – Building on the types of representations proposed by Clements (1985), Yip (1989), Bao (1990), Duanmu (1990), and others, it offers a common geometry for representing lexical tones as *lexical tunes* in all different languages—including what will be coined *word-tune languages* like Japanese and *syllable-tune* languages like Mandarin—with elements from a common set of primitive tonal units.
- II. *Associations with many levels*

- a. It conflates the concept of intonational units associating with higher-level (i.e. non-terminal) nodes in the prosodic tree (which is allowed in the Pierrehumbert and Beckman (1988) model) and the concept of lexical tunes doing the same (in a sense similar to that proposed by Leben (1973) and references cited therein) in the underlying representation.
  - b. In the surface representation, it allows for the secondary association of subconstituents and terminal nodes of lexical tunes with nodes in the prosodic tree, thereby *anchoring* those subconstituents and terminals to certain parts of the prosodic structure. This anchoring mechanism is an extension of the concept of segmental anchoring discussed by Arvaniti, Ladd et al. (1998), Ladd, Mennen et al (2000), and Dilley, Ladd et al. (2005).
  - c. Also in the surface representation, it allows any melodic units associated high in the tree to secondarily associate (or re-associate) with nodes dominated by the node with which it is associated underlyingly. Secondary association is allowed in the Pierrehumbert and Beckman (1988) model, but the idea of a secondary association becoming the sole association for a melodic unit on the surface is a deviation from that model. Another way to view these level-changing re-associations is as an extension (perhaps a “3D” version) of the commonly exploited reassociation mechanism that is an integral part of many traditional autosegmental analyses (see, for example, Haraguchi 1977 Ch. 1 for an analysis of "accent slide" in words with accented syllables containing voiceless vowels as a case of tone reassociation).
- III. *Unified phonetic interpretation of autosegmental structure* – The primary mechanism of phonetic interpretation of the tree is one of *association-dependent tonal scope*. That is, the scope of a given tonal unit’s effect on an utterance is determined by its level(s) of association within the prosodic tree. The notion of tonal registers applying to prosodic constituents has precedents in Clements (1990) and Laniran (1992), as discussed in Section 1.2.2.2 above. This mechanism of phonetic interpretation ultimately allows us to

capture some of the differences among the melodic systems investigated as differences in register scope; in particular it allows us to reconcile the more parallel interactions between lexical tone and utterance-level intonation observed in languages like Mandarin (an insight that is at the forefront of many analyses promoting overlay models) with the more serial interactions between the two observed in languages like Japanese. In ToBI terms, it allows us to reconcile the multiple-tier approach employed by Peng, Chan et al. (2005) with the single-tier approach used by Wong, Chan et al. (2005). In addition to capturing a certain class of language-specific phenomena in the phonology, the structural parameters of the model, if adopted, make clear what language-specific phenomena must then be handled in the phonetics.

In addition to facilitating cross-linguistic comparison as outlined above, the structural limitations imposed by the current version of the model make certain predictions about the types of tonal and intonational inventories that we should observe in the world's languages.

## **1.5 Overview of the Dissertation**

In the remainder of this dissertation, data from several tone languages will be presented and used to underscore the need for a phonological component in a comprehensive model of speech melody, a need that is fulfilled by sequential models but not overlay models. Some of the results will highlight certain inadequacies of the AM analyses in the literature, both in terms of their ability to capture some of the facts within melodic systems and in terms of capturing typological generalizations in a structural way. These inadequacies will motivate the development of a new model that can serve as a typological template for modeling the phonological component of a melodic system in a number of different types of tone languages.

### **1.5.1 Overview of Chapter 2**

Chapter 2 of this dissertation presents findings from a series of production experiments designed to reveal systematically the types of interactions between lexical tone and utterance intonation



that occur at the right edge of utterances in several different tone languages. The languages surveyed in this chapter include Standard Mandarin (Putonghua) (Section 2.2), Henanhua (a non-standard dialect spoken in the Henan Province of China) (Section 2.3), Cantonese (Section 2.4), North Kyeongsang Korean (Section 2.5), and a family of Kansai Japanese dialects including Shiga Japanese (Section 2.6). Eliciting analogous combinations of conditions (every lexical tonal category and two utterance intonation categories, i.e. declarative and echo question) in each language makes it possible to make cross-linguistic comparisons that elucidate the various types of melodic interactions a general model of speech melody must be equipped to handle.

### **1.5.2 Overview of Chapter 3**

Chapter 3 presents findings from a series of perceptual experiments designed to draw out cross-linguistic differences in the capacity of melodic systems for encoding tonal and intonational contrasts in syntactically and semantically uninformative contexts. The same languages represented in Chapter 2 are investigated from this perceptual standpoint in Chapter 3: Standard Mandarin (Putonghua) (Section 3.2), Henanhua (Section 3.3), Cantonese (Section 3.4), North Kyeongsang Korean (Section 3.5), and Shiga Japanese (Section 3.6).

### **1.5.3 Overview of Chapter 4**

Whereas the discussions in Chapters 2 and 3 highlight language-, dialect-, and speaker-specific differences that are apparent in the experimental results, Chapter 4 focuses on the dialect-internal phenomenon of lexical-tone-dependent intonation, i.e. the realization of utterance-type intonation being dependent on the lexical tonal category specified at the right edge of an utterance. These types of results are pulled out to drive home the fact that the strictly parallel encoding of communicative functions that is assumed in overlay models is not tenable. Several examples of lexical-tone-dependent intonation are extracted from the experimental results in previous chapters, first for Mandarin-style tone languages, referred to here as *syllable-tone* languages

(Section 4.2), and then for so-called “pitch accent” languages, referred to here as *word-tone languages* (Section 4.3).

#### **1.5.4 Overview of Chapter 5**

In Chapter 5, the interactions of lexical tone and utterance-level intonation are explored from a phonological perspective. In Section 5.2, the concept of intonation-dependent allotony is introduced and its effectiveness in handling at least a subset of the melodic irregularities highlighted in Chapter 4 is discussed. In Section 5.3, the traditional autosegmental-metrical framework is considered as a candidate for modeling speech melody from a phonological perspective; in Sections 5.3.1 and 5.3.2, respectively, traditional AM treatments of North Kyeongsang Korean and Kansai Japanese are critically examined in light of the facts presented for those languages in earlier chapters.

#### **1.5.5 Overview of Chapters 6 and 7**

In Chapter 6, a new, scopal model of speech melody will be presented. This phonological model exploits the clear advantages of an AM-based sequential model over overlay models, but it has been augmented with additional mechanisms in order that it might express more of the differences observed among melodic systems across languages in a structural way. Chapter 7 concludes the dissertation.

## CHAPTER 2: PRODUCTION EXPERIMENTS

### 2.1 Questions about Production

In this chapter the interaction between lexical tone and utterance intonation will be explored. In particular, answers to the questions given in (2.1) will be sought:

(2.1) Questions regarding melodic perception in tone languages

**Q1 – *Language-General Tone-Intonation Interaction*:** Does intonation interact with lexical tone in the same way in all tone languages?

**Q2 – *Dimensions of Language-Specific Variation*:** If not, what are the dimensions along which melodic systems may vary in a language-specific way?

**Q3 – *Declarative-to-Echo-Question Mapping*:** Is the  $F_0$  contour of a tone in the echo question context predictable purely from the phonetic attributes of the tone in the declarative context (i.e. is there an algorithm that can take just phonetic pitch parameters of the declarative form as input and successfully predict the  $F_0$  contour of the echo question form, or vice versa)?

**Q4 – *Structural Correlation of Tone-Intonation Interaction*:** Is there any way to predict certain aspects of tone-intonation interaction based on given characteristics of a language (e.g., the structure or inventory of its tonal system)?

In an attempt to answer the above questions, a series of production experiments was carried out. The recording materials consisted of target words that were uttered either in isolation or at the end of a short frame sentence. The word lists were constructed to include representative words from each tonal category. Words were kept segmentally identical when possible (i.e. in Mandarin and Cantonese) and syllable/mora count was varied in those languages where word tone plays a role (NKK and SJ). If a frame sentence was used, care was taken to

ensure that intonation would be the only available tool for conveying whether the sentence was a statement or a question. For example, while all of the languages in question employ utterance-final particles in some contexts, often providing morphological cues to utterance type and reducing the amount of overlap between intonational events at the right edge of the utterance and lexical tonal events in the rightmost “content word”, such constructions were not included in any of the recorded materials. The basic task in each experiment was as follows: the speaker was given a word list and asked to utter each word in a variety of contexts, crucially including a declarative context and an echo question context. Detailed descriptions of the individual experiments and results are provided in the following sections, starting with Mandarin in Section 2.2.

## **2.2 Production in Mandarin**

In this section and throughout the rest of the dissertation, the label “Mandarin” will be used to refer to the “standard” dialect of Mandarin—alternatively known as *Putonghua*—which is spoken in Mainland China.

### **2.2.1 Overview of Mandarin tones**

Mandarin is traditionally analyzed as having four lexical tonal categories plus a fifth category that is given the label *neutral tone*. This fifth category is often described as “neutral” in that its surface realization is fully dependent on the realizations of neighboring tones (Chao 1968; Yip 1980; Shih 1987), although Chen and Xu (2006) argued, based on acoustic evidence from utterances containing strings of successive neutral-toned syllables, that this fifth tone has a fully specified underlying pitch target. This fifth tonal category will not be discussed further here, but its interaction with the other tones and with utterance intonation is not trivial and warrants attention in future studies (see Lee 2005 for a comprehensive study of question types in Mandarin that, among other things, systematically compares questions formed with the neutral-toned final question particle *-ma* and echo questions, which lack the particle).

The other four tones in Mandarin, which were included in the recorded materials of the production experiment, are conventionally given the non-descriptive labels “Tone 1”, “Tone 2”, “Tone 3”, and “Tone 4”, respectively. These labels will be used (along with their abbreviated counterparts *T1*, *T2*, *T3*, and *T4*) in this section and throughout the dissertation. In traditional grammars and language classes, the citation forms of these four tones are often described as being “high-level”, “high-rising”, “low-dipping”, and “high-falling”, respectively (Chao 1968), or simply as “high”, “rising”, “low”, and “falling” (Kratochvil 1968). It has been noted that Tone 3 surfaces as “dipping”, i.e. falling and then rising, in citation contexts (utterance-final position and isolation) only, and in utterance medial position it surfaces as “low-falling” except when it occurs before another Tone 3, in which case it is realized as “high-rising”, just like Tone 2 (Chao 1968). Table 2.1 gives a summary of each of the four Mandarin tones and their descriptive labels, along with their traditional Chao (1930)-style “tone letter” values (in which the pitch-level contours of each tone are assigned a sequence of numerical values on a relative five-point scale).

**Table 2.1: Descriptive labels for the four lexical tones in Mandarin.**

tone	citation description	tone letter value
Tone 1	high-level	55
Tone 2	high-rising	35
Tone 3	low-dipping	213
Tone 4	high-falling	51

One final note regarding Tone 3: Because it is often realized so low in the pitch range of the speaker, it often gets heavily glottalized, especially in utterance-final syllables (Chao 1968; Davidson 1991; Belotel-Grenié and Grenié 2004). In the current study, this glottalization (or “creakiness”) resulted in “gaps” in the measurable pitch contour.

### 2.2.2 Subjects

Two speakers were recorded, one female and one male, both in their twenties. The female was born and raised in Beijing. The male was born in Hohhot and moved to Tianjin at age 18. He lived in Tianjin for 4 years and then Beijing for 3 years.

### 2.2.3 Materials and procedure

All materials were recorded in a sound-attenuated booth with a Plantronics-500 DSP headset microphone connected to a laptop computer. The conditions included in the recording materials are given in (2.2):

(2.2) Mandarin production experiment conditions

4 tonal categories (Tone 1, Tone 2, Tone 3, and Tone 4)

2 intonational categories (declarative statement vs. echo question)

6 repetitions

2 frame sentences (one with all Tone 4 syllables and one with a T2-T2-T4 sequence)

The four target words, given in (2.3), constituted a minimal set (homophonous except for their tones):

(2.3) Mandarin target words

*wan1* ‘bay’,

*wan2* ‘pill’,

*wan3* ‘bowl’

*wan4* ‘ten thousand’

The two frame sentences given in (2.4) were used (*Naliang* and *Mingning* are names):

(2.4) Mandarin frame sentences

*Na4liang4 nian4 X* ‘Naliang reads (the word) X’

*Ming2ning2 nian4 X* ‘Mingning reads (the word) X’

The first one was constructed such that every syllable leading up to the target word would bear the same tone, in order to highlight any global  $F_0$  trends brought on by utterance intonation. The second sentence was included in order to determine if any global effects of intonation observed in the earlier part of the sentence were tone-dependent. Both the target words and the frame sentences were comprised solely of sonorant segments in order to maximize the probability of eliciting uninterrupted pitch curves.

Each recording session was divided into four phases, as described in (2.5):

(2.5) Mandarin production experiment phases

Phase 1: frame 1; 3 declaratives in a row followed by 3 echo questions in a row

Phase 2: frame 1; alternating declarative / echo question 3 times

Phase 3: frame 2; 3 declaratives in a row followed by 3 echo questions in a row

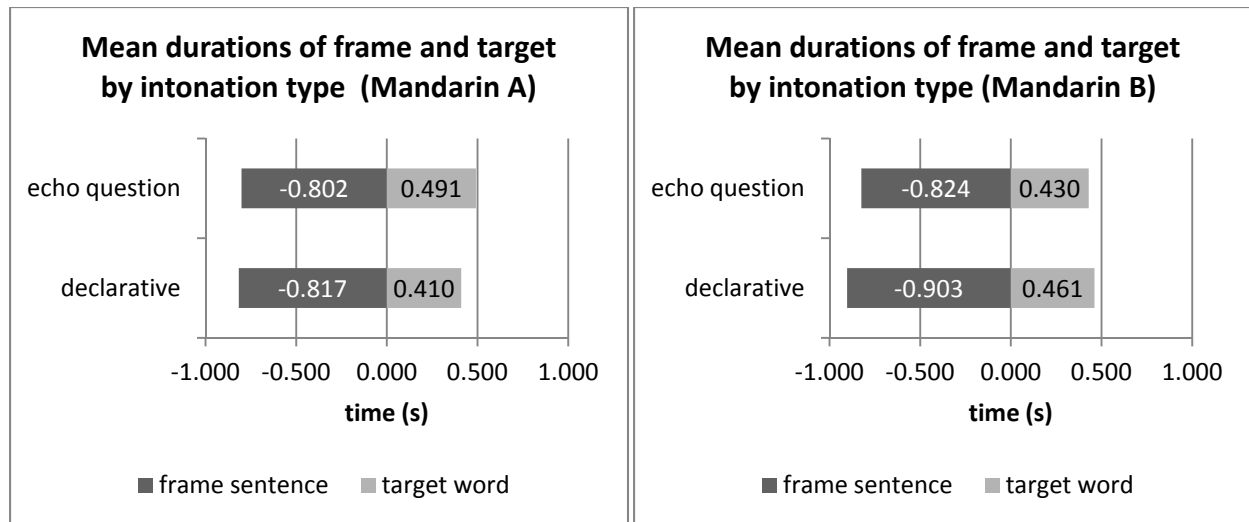
Phase 4: frame 2; alternating declarative / echo question 3 times

Each session took approximately 30 minutes.

## 2.2.4 Results

The two dimensions along which measurements were taken were duration and  $F_0$ . The duration results are presented first, in Figure 2.1. Here and in figures and tables throughout the dissertation, *Mandarin A* refers to Mandarin Speaker A (i.e. the first of the two Mandarin speakers) and *Mandarin B* refers to Mandarin Speaker B (i.e. the second Mandarin speaker). 48 tokens were measured in each intonational category for each speaker. The stacked bars in the graphs are aligned such that the boundary between the frame sentence and the target word is located at time 0. It is apparent that the duration effects were slightly different for the two

speakers. For Speaker A, while the mean duration of the frame sentence was about the same in the two intonational contexts ( $p = .938$  according to an ANOVA run in SPSS that treated each token as an independent measure), the target word was longer in the echo question context ( $p < .001$ ). For Speaker B, on the other hand, the mean duration of the frame sentence was



**Figure 2.1: Mean durations for Mandarin**

significantly shorter in the echo question context ( $p = .005$ ) but the mean target word duration was virtually the same ( $p = .863$ ). It seems reasonable to speculate that this shortening effect in the echo question conditions for Speaker B is traceable to a reflex of some sort of givenness (cf. Bard, Anderson et al. 2000), since the declarative versions of each sentence were uttered first and the echo questions came second in each trial. To factor out this effect so as to see any additional durational effect on the target word, the mean ratio of target word duration to total utterance duration was calculated for each intonational category for each speaker. These ratios are displayed as percentages in Table 2.2 (p-values returned by the ANOVA are given in the rightmost column).

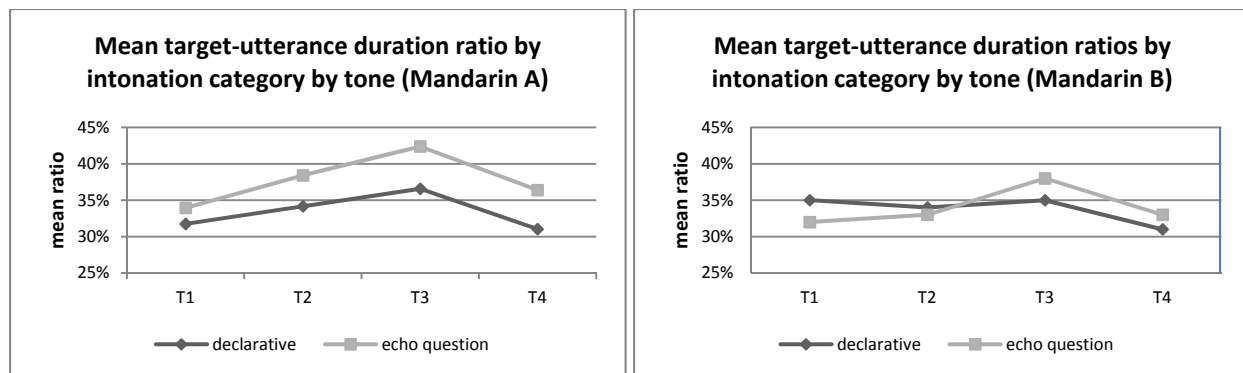
**Table 2.2: Target-utterance duration ratio by intonation for Mandarin Speakers A and B.**

Speaker	declarative %	echo question %	sig.
Mandarin A	33.38	37.79	.000
Mandarin B	33.78	34.23	.007



Apparently, the proportion of the utterance taken up by the target word was on average significantly greater in the echo question conditions for both speakers.

In order to see if this durational effect was observed across all tonal categories, the proportion differences are broken down by tonal category and re-displayed as line graphs in Figure 2.2. It is clear that the differences are not the same across all categories for either speaker.

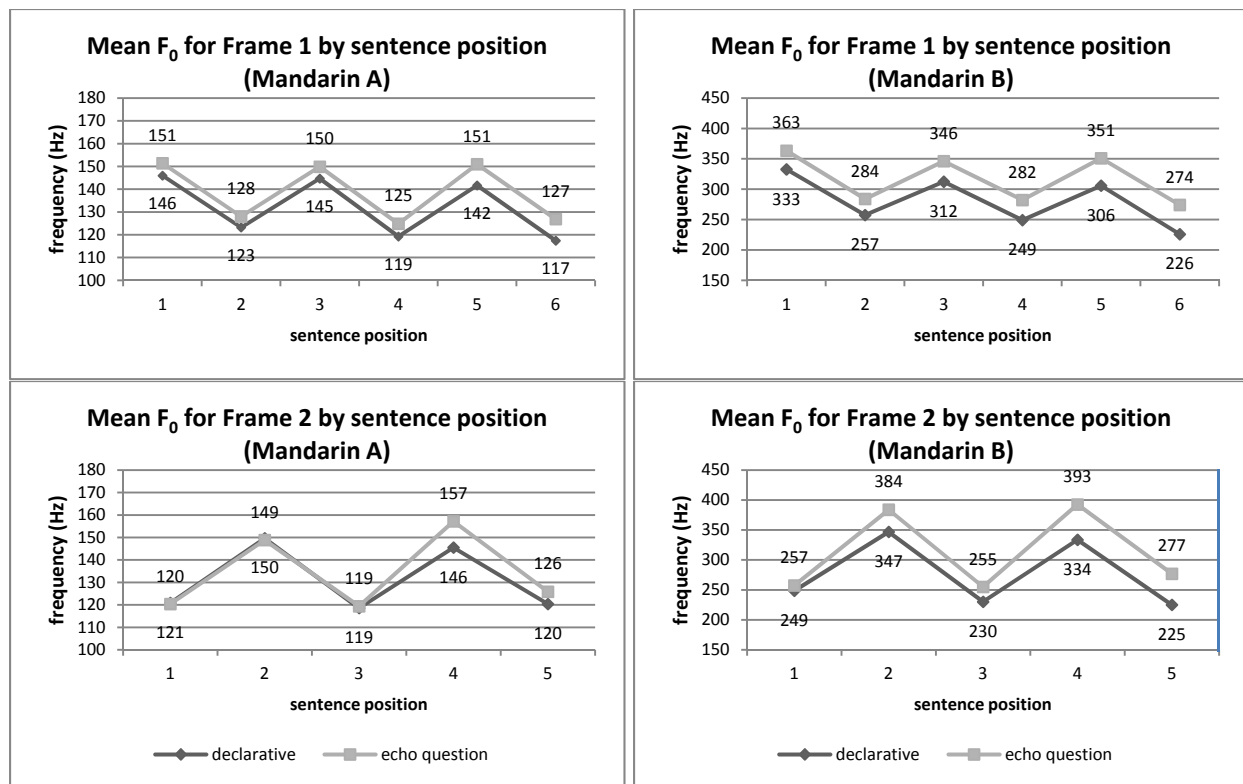


**Figure 2.2: Mean target-utterance duration ratios for declarative and echo question intonation, by tone (Mandarin A on the left and Mandarin B on the right).**

For Speaker A, the target word was proportionally longer in the echo question context in all four tonal categories ( $p = .018$  for T1;  $p < .001$  for all other tones). For Speaker B, T1 targets were actually shorter in the echo question context ( $p < .001$ ), T2 targets were about the same ( $p = .500$ ), and only T3 and T4 targets were longer ( $p < .001$  and  $p = .007$ , respectively).

Let us turn to the  $F_0$  results. Previous descriptions of Mandarin interrogative intonation have noted the tendency for the overall pitch range of an intonational (i.e. non-morphologically marked) question to be higher relative to the corresponding statement (Yuan, Shih et al. 2002; Liu and Xu 2005; Peng, Chan et al. 2005; Xu 2005). While the current study will focus mainly on the effect intonation has on the final syllable, let us briefly examine the  $F_0$  results for the frame sentences leading up to the target words.  $F_0$  was measured at successive local minima and maxima leading up to the target word—six measurement points for Frame 1 and five for Frame 2. The mean  $F_0$ s for these measurement points in each frame are given in Figure 2.3 for both speakers. In each plot the data series for the declarative condition is displayed in dark gray with

diamond-shaped data points while the data series for the echo question condition is displayed in light gray with square-shaped data points.



**Figure 2.3: Mean F<sub>0</sub> of the frame by sentence position. Frame 1 above and Frame 2 below, Mandarin A on the left and Mandarin B on the right.**

Note that the last two positions in each plot correspond to the same word with the same tone—*nian4*. It is apparent that there is indeed a global effect of intonation on pitch register. According to the post-hoc analyses for ANOVAs run in SPSS that treated each F<sub>0</sub> measurement for a given sentence position as an independent measure, the F<sub>0</sub> difference between corresponding points was significant for all positions in Frame 1 for both speakers ( $p < .01$  for Speaker A and  $p < .001$  for Speaker B after a Bonferroni adjustment). As for Frame 2, only the last two positions showed a significant difference across intonations for Speaker A (first:  $p = .832$ ; second:  $p = .714$ ; third:  $p = .714$ ; fourth:  $p < .001$ ; fifth:  $p = .012$ ), while for Speaker B the first position did not display a significant difference ( $p = .122$ ) and the last four displayed a highly significant difference ( $p < .001$ ). Although a broader study is called for, it would seem

that the global effect of intonation manifests itself differently on different tones (cf. Connell and Ladd 1990) and (Laniran 1992) for discussions on tone-specific declination effects in Yoruba, and (Yuan 2004) for the observation that T3 in Mandarin may resist the global raising effects associated with questions).

Looking at the plots for Frame 1, which was composed of three T4 syllables in a row, we can say descriptively that the points diverge more towards the end of the frame as compared to the beginning. The ANOVA results confirm this generalization to a certain degree. For Speaker A, the  $F_0$  difference at Position 6 was significantly greater than that at Positions 2 and 3 ( $p = .024$  and  $p = .038$ , respectively), just on the cusp of being significantly different from that at Positions 1 and 4 ( $p = .051$  and  $p = .059$ , respectively), and not significantly different from that at Position 5 ( $p = .935$ ). For Speaker B, the difference at Position 6 was significantly different from that at the first four positions (first:  $p = .017$ ; second:  $p = .004$ ; third:  $p = .049$ ; fourth:  $p = .040$ ) but not from that at the fifth position ( $p = .670$ ). These results are in line with those given by Yuan, Shih et al. (2002) and Xu (2005), although Xu (2005) controlled for focus and concluded that the divergence between declarative and interrogative utterance curves starts at a narrowly focused syllable.

Next, let us break up the results by tonal category of the target word and include the  $F_0$  results for the target words themselves in the plots. Traditional treatments of Mandarin intonation characterize the effect of intonation on tone as affecting the range or register of tonal contours, echoing Chao's (1968) well-known "small ripples riding on large waves" metaphor. Yuan (2004) noted that, while the generalization of a raised register and increased range is largely supported in his results, there are also some details that differ from tone to tone. His findings will be compared to the results of the current study in Section 2.4.6.

Figure 2.4 shows mean  $F_0$  contours for the declarative and echo question conditions for each tone, for each speaker, for the frame 1 condition only (where the frame sentence consisted of three T4—i.e. falling—syllables in a row leading up to the target word). The contours were constructed by taking time and  $F_0$  measurements at strategic points (mostly at local maxima and

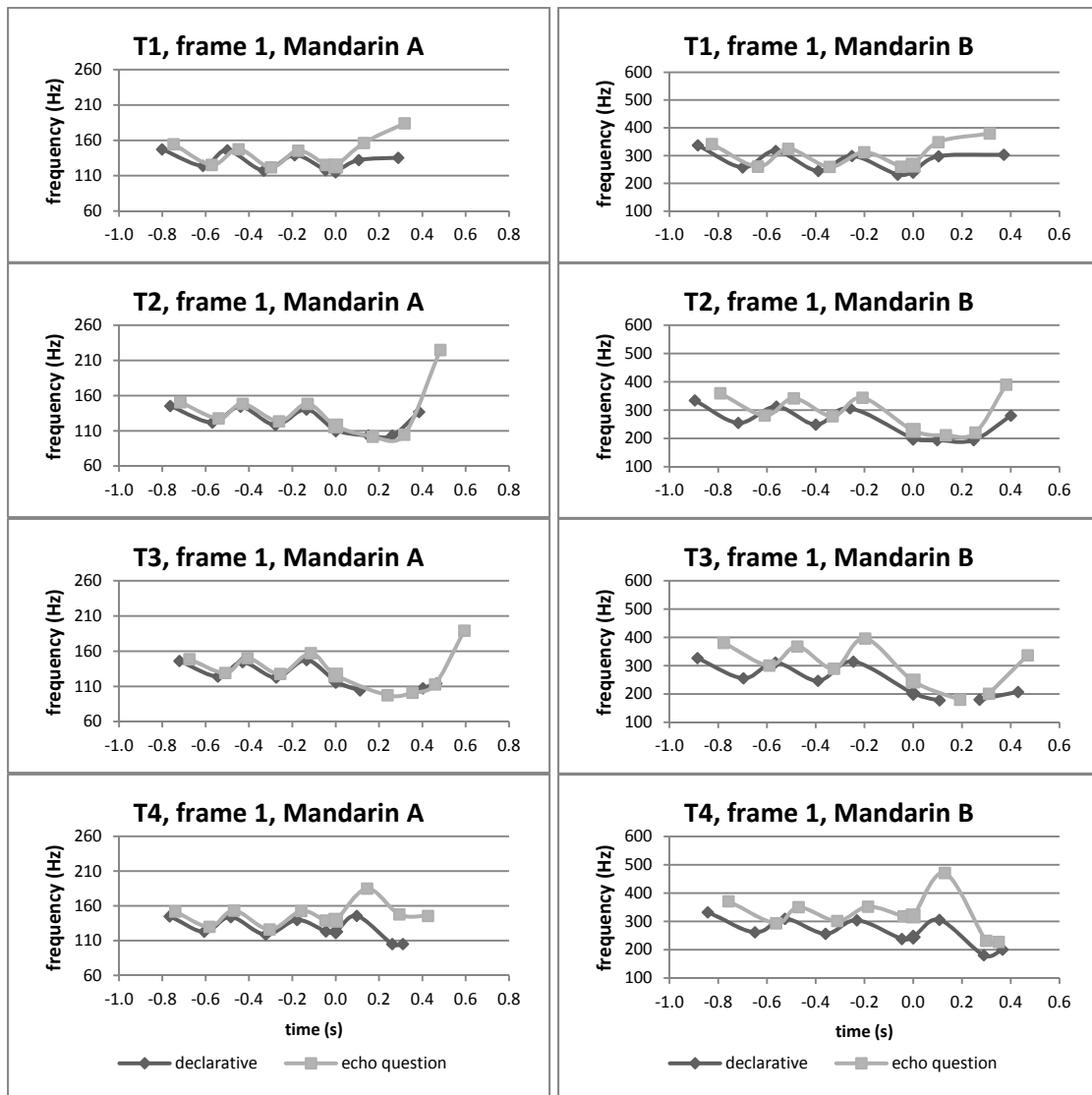


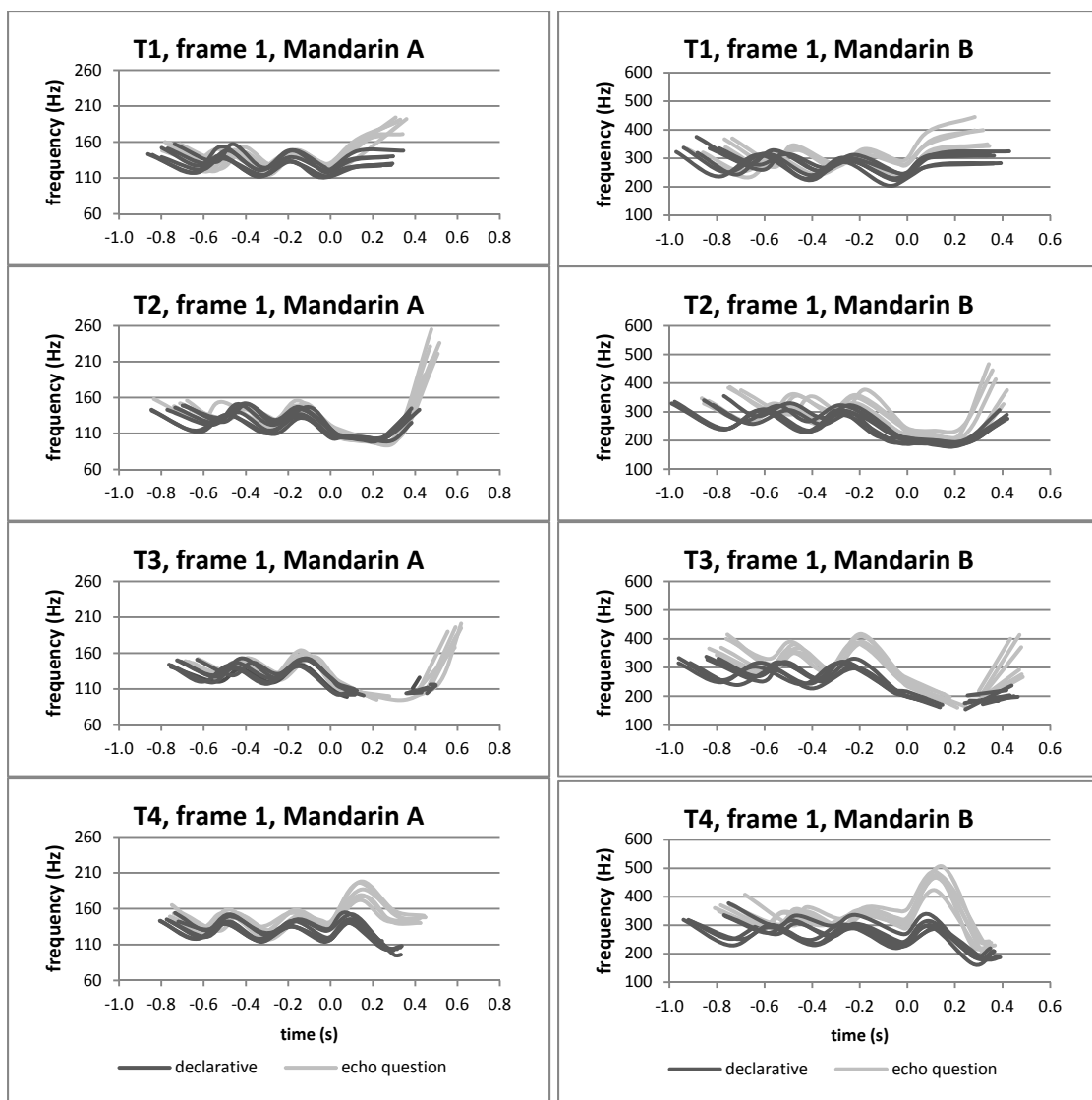
Figure 2.4: Mean  $F_0$  contours for Mandarin

minima) for each token, plotting the averages of those points, and interpolating them with a smooth curve<sup>8</sup>. The plots are time-aligned such that the boundary between the frame sentence and the target word, indicated with an enlarged data point, falls at time 0 in every case. Speaker A's results are displayed on the left and Speaker B's results are on the right. There are clear effects of intonation category on the  $F_0$  contour of the last syllable; broadly speaking, the high point of the contour is considerably higher in the echo question context, resulting in a wider

<sup>8</sup> Note that there are some gaps in some of the T3 contours. This is due to the fact that many of the low points in the T3 contour were glottalized, thus making it impossible to get accurate  $F_0$  measurements at those points.

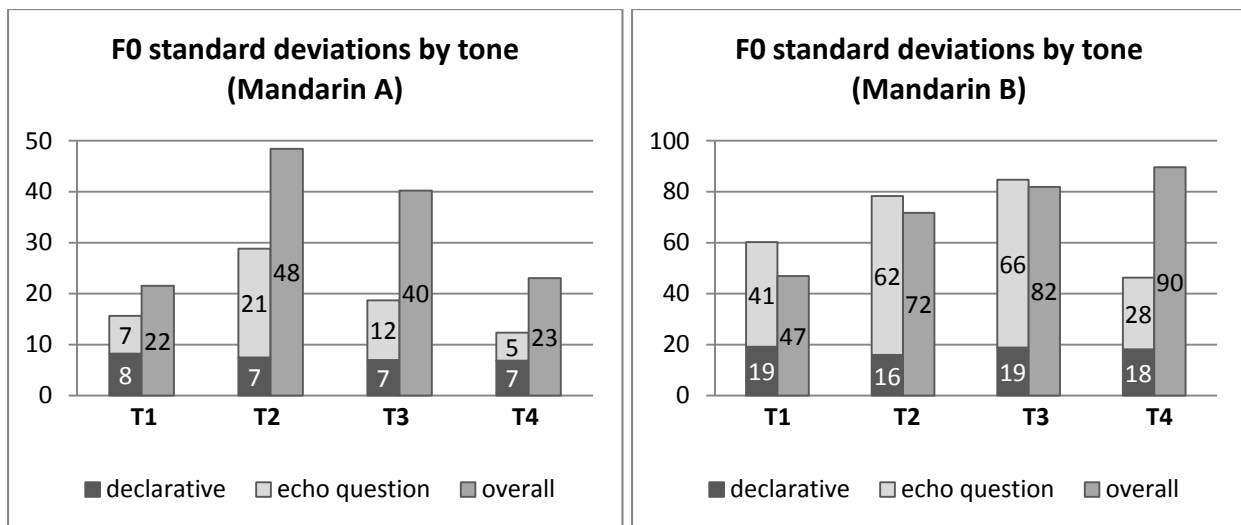
overall pitch range on the final syllable and a more “exaggerated” realization of the tone-specific contour shape.

It is useful to see the mean contours of the respective intonation conditions, as shown in Figure 2.4, in order to get a clear sense of general trends. In doing so, however, we lose some information regarding variability that we would see if we were to plot all of the individual tokens on the same graph. This type of “multiple-contour plot” is shown in Figure 2.5 (all of the individual tokens are still time aligned such that their frame-target boundaries fall at time 0).



**Figure 2.5: Multiple-contour plots for Mandarin**

Figure 2.5 gives a sense of the range of pitch contours produced by each speaker. Eyeballing the figure, it appears that Speaker A maintained more of a distinction across intonational categories in general than Speaker B, although the separation was perhaps the least distinct for Speaker A's T1. Speaker B, while displaying less separation overall, displayed the sharpest intonation distinction for T4. In an attempt to quantify this metric of intonation-dependent "pitch range distinctiveness", the  $F_0$  measurement point at which the mean  $F_0$ s of the declarative and echo question tokens were maximally different from one another was determined (e.g. in the case of T1 for both speakers it was the rightmost  $F_0$  measurement point) and the standard deviations of the declarative, echo question, and overall  $F_0$  measures, respectively, were calculated. The results of these calculations are shown in Figure 2.6. The thing to pay attention to in the graphs in Figure 2.6 is how the stacked declarative-echo question bars compare with their "overall"



**Figure 2.6: Standard deviations of declarative, echo question, and overall  $F_0$ , by tone, in Mandarin**

counterparts. For a given tone, if the former approaches being the same size or bigger than the latter, it is an indication that the  $F_0$  ranges of the intonational groups were minimally distinct for that tone. A shorter stacked bar with a taller singleton partner indicates more intonation-dependent clustering and a greater separation between intonational groups (since, in a bimodal distribution, the standard deviations of the respective modes are smaller than the standard

deviation of the pooled data, the mean of which is somewhere between the two modes and not representative of either of them). We see, then, that the qualitative observations made with respect to Figure 2.5 are supported by these quantitative results. Speaker A indeed shows greater separation overall (all of the stacked bars are shorter than the singleton bars) but the lowest degree of distinctiveness is seen in the T1 condition. Speaker B shows less separation for the first three tonal conditions but a degree of separation comparable to that of Speaker A in the T4 condition (comparable in that the sum of the intonation-specific standard deviations is about half of the overall standard deviation in each case). In the next section these results, as well as the rest of the production results, will be discussed in the context of the mechanisms that underlie the declarative-vs.-echo-question realization differences as well as their possible ramifications for perception.

### **2.2.5 Discussion**

In the broadest terms, the realization of echo question intonation relative to declarative intonation in Mandarin involves a higher  $F_0$  peak on the rightmost syllable as well as a more global positive  $F_0$  difference over the entire utterance (the magnitude of the latter varying by speaker and repetition). Yuan (2004) proposes the following mechanisms to account for question intonation in Mandarin: “An overall higher phrase curve, higher strengths<sup>9</sup> of sentence final tones, and a tone-dependent mechanism that flattens the falling slope of the final falling tone and steepens the rising slope of the final rising tone”. He also mentions that there is one additional tone-dependent mechanism whereby T3 can sometimes “pull the question curve down to the statement curve”. The first two mechanisms—the higher phrase curve and higher strengths of sentence-final tones—are generally supported by the results of the current study, although the phrase curve difference is more pronounced for Speaker B in general, and for T3 and T4 for both speakers. The third mechanism—a flattened final slope for T4 and a steepened slope for T2—is

---

<sup>9</sup> The term “strength” here refers to the strength parameter in the modeling language Stem-ML (see Yuan, Shih et al. 2002), which in the context of  $F_0$  translates to a more faithful rendition of a contour tone and a greater  $F_0$  excursion from the phrase curve.

supported, with one exception: T4 for Speaker B. Shown in the lower right graph in Figure 2.4, it is clear that the slope of the falling tone is steepened rather than flattened for Speaker B. Finally, the observation Yuan (2004) makes for T3 seems to extend to T2 for Speaker A in the current study; that is, both T3 and T2 appear to “pull the question curve down to the statement curve” in the case of Speaker A.

The issue of modeling these different mechanisms will be addressed in Chapter 4, but for now it suffices to say that the presence of all of the tone-dependent mechanisms highlighted in this section implies that the answer to Q3 (*Declarative-to-Echo-Question Mapping*) at the beginning of this chapter is “no”; any algorithm for mapping from the declarative contour to the echo question contour would need the tonal category as part of its input.

Returning to the  $F_0$  patterns on the target words themselves, we see some interesting results in the multiple-contour plots shown in Figure 2.5 and the standard deviation graphs in Figure 2.6. T4 shows a relatively pronounced intonation-dependent difference for both speakers. In the perceptual results in Chapter 3, we will see that the T4 condition yielded the highest rate of perceptual accuracy for intonation. In the section on Cantonese production (Section 2.4) it will be suggested that the degree of intonation-dependent distinctiveness may be a dimension along which the language-specific phonetics of speech melody may vary, an idea that is pertinent to Q2 (*Dimensions of Language-Specific Variation*) at the beginning of this chapter. Specifically, it will be shown that the mean contours for Cantonese T1 look quite similar to those of Mandarin T1, but the intonation-dependent pitch range difference on the final syllable is much more robust in Cantonese than in Mandarin.

### **2.3 Production in Henanhua**

Some preliminary results from a pilot experiment for a variety of Henanhua (a dialect spoken in the Henan Province, south of Beijing) are presented here. Although the results may not be representative of a “prototypical” Henan dialect, since the speaker was raised bilingual in



Henanhua and Mandarin, they are included here nevertheless because of their importance for the overall typological picture being presented in this chapter.

### 2.3.1 Overview of Henanhua tones

There are limited sources on dialects of the Henan Province. Zhan, Chen et al. (1993) provided a typological survey in which they listed impressionistic descriptions of the four tones in these dialects. Table 2.3 shows the tone-letter values given by Zhang, Chen, et al. (1993)<sup>10</sup> for the dialects spoken both in Zhengzhou and Zhumadian, the childhood homes of the speaker who was

**Table 2.3: Descriptive labels for the four lexical tones in Henanhua.**

tone	citation description	tone letter value
Tone 1	rising	24
Tone 2	high-falling	42
Tone 3	high-level	55
Tone 4	low-falling	31

recorded for the production experiment presented in this section. It should be noted that the data presented in Zhang, Chen et al. (1993) were compiled from fieldwork that was done going back as far as 1960, so their observations may not be representative of the current dialects spoken in those cities. The tone labels *Tone 1*, *Tone 2*, etc. were assigned according to the traditional tonal category labels in Mandarin to which each of the respective Henanhua tones corresponds; so, for example, a given lexical item that bears a Tone 1 in Mandarin will generally<sup>11</sup> bear what is here being called a Tone 1 in Henanhua.

<sup>10</sup> I am indebted to Aletheia Cui for pointing me to this reference, which is written in Chinese, and interpreting the data contained therein.

<sup>11</sup> These correspondences held for a majority of the words tested for the speaker who was recorded. Exceptions included *la* ‘spicy’ and *mai* ‘wheat’, both of which bear a Tone 4 in Mandarin; the speaker, who is bilingual in Henanhua and Mandarin, pronounced these words with a Tone 4 (i.e. a high falling tone) when prompted to say them in Mandarin but with a Tone 1 (i.e. rising tone) when prompted to say them in Henanhua. There were exceptions in other tonal categories as well. The fact that the lexical tonal correspondences do not hold across the board for this speaker is perhaps a sign that the speaker is indeed operating with a native melodic system in each language, with separate lexical entries in each language for every word.

### **2.3.2 Subject**

One 20-year-old female speaker of Henanhua was recorded. She was born in the Henan Province, where she lived until age 9 (Zhengzhou until age 5 and then Zhumadian until age 9). She then moved to Shenzhen in Guangdong Province for two years, then to Belmont, Massachusetts for six years, and finally to Ithaca, New York for three years. Both of her parents were native Henan speakers, but she was raised speaking both Henanhua and Mandarin at home (her grandmother speaks only Henanhua).

### **2.3.3 Materials**

All materials were recorded in a quiet room with a Sennheiser PC 156 headset microphone connected to a laptop computer. The design of the recording materials was very similar to that for Mandarin, with the exception of the number of repetitions (three instead of six) and the phrasal contexts (isolation vs. frame instead of two different frames). Two separate recording sessions were held because the results from the first one revealed a possible complex interaction between the tone of the penultimate syllable in the frame sentence with that of each target word. The conditions for both recording sessions are summarized in (2.6):

(2.6) Henanhua production experiment conditions

- 4 tonal categories (Tone 1, Tone 2, Tone 3, and Tone 4)
- 2 intonational categories (declarative statement vs. echo question)
- 3 repetitions
- 2 phrasal contexts (isolation vs. frame-final)

The four target words in the first recording session—shown in (2.7)—were the same as those used for Mandarin:

(2.7) Henanhua target words

*wan1* ‘bay’,

*wan2* ‘pill’,

*wan3* ‘bowl’

*wan4* ‘ten thousand’

The frame sentence given in (2.8) was used for the frame conditions in the first recording session (*Na* is a given name):

(2.8) Henanhua frame sentence

*Na4yao4 nian4 X* ‘Na must read (the word) X’

Note that this frame differs by one syllable from the frame used in the Mandarin experiment. Preliminary recordings for Henanhua had revealed that T4 and T2 (both normally falling tones) followed by T4 in certain restricted environments (most likely word-internal) changed from being falling tones to being rising tones<sup>12</sup>. To avoid this sandhi from applying in the frame sentence, the second part of the given name *Na4liang4* that had been used in the Mandarin experiment was replaced by the auxiliary *yao4* (‘want to’, ‘must’, ‘will’). A couple of test recordings showed that, while T4 seemed to level out in general phrase medially (a behavior seen among other tones as well in this dialect), the word-internal sandhi did indeed appear to be blocked by changing the second syllable to an auxiliary verb. Thus, the frame sentence above, comprising all T4s, was settled upon for the first recording session.

Once initial analyses of the results indicated that the T4 frame might be triggering some other (i.e. non-word-internal) process before the T4 target word, a different frame, comprising all T3s, was chosen for the second recording session. This new frame is given in (2.9):

---

<sup>12</sup> This productive sandhi in Henanhua is quite interesting in that it is reminiscent of the well-known, non-productive sandhi in standard Mandarin that is triggered when the tone on select lexical items such as *bu4* (verb negator) changes to a rising tone when followed by a T4-bearing word (e.g. *bu2hui4* ‘cannot’).

(2.9) Revised Henanhua frame sentence

*Li3wei3mai3 X* ‘Liwei buys X’

In order to keep some semblance of semantic plausibility, a new minimal set of words was chosen for the target words. This revised word list is given in (2.10):

(2.10) Revised Henanhua target words

*yan1* ‘tobacco’

*yan2* ‘salt’

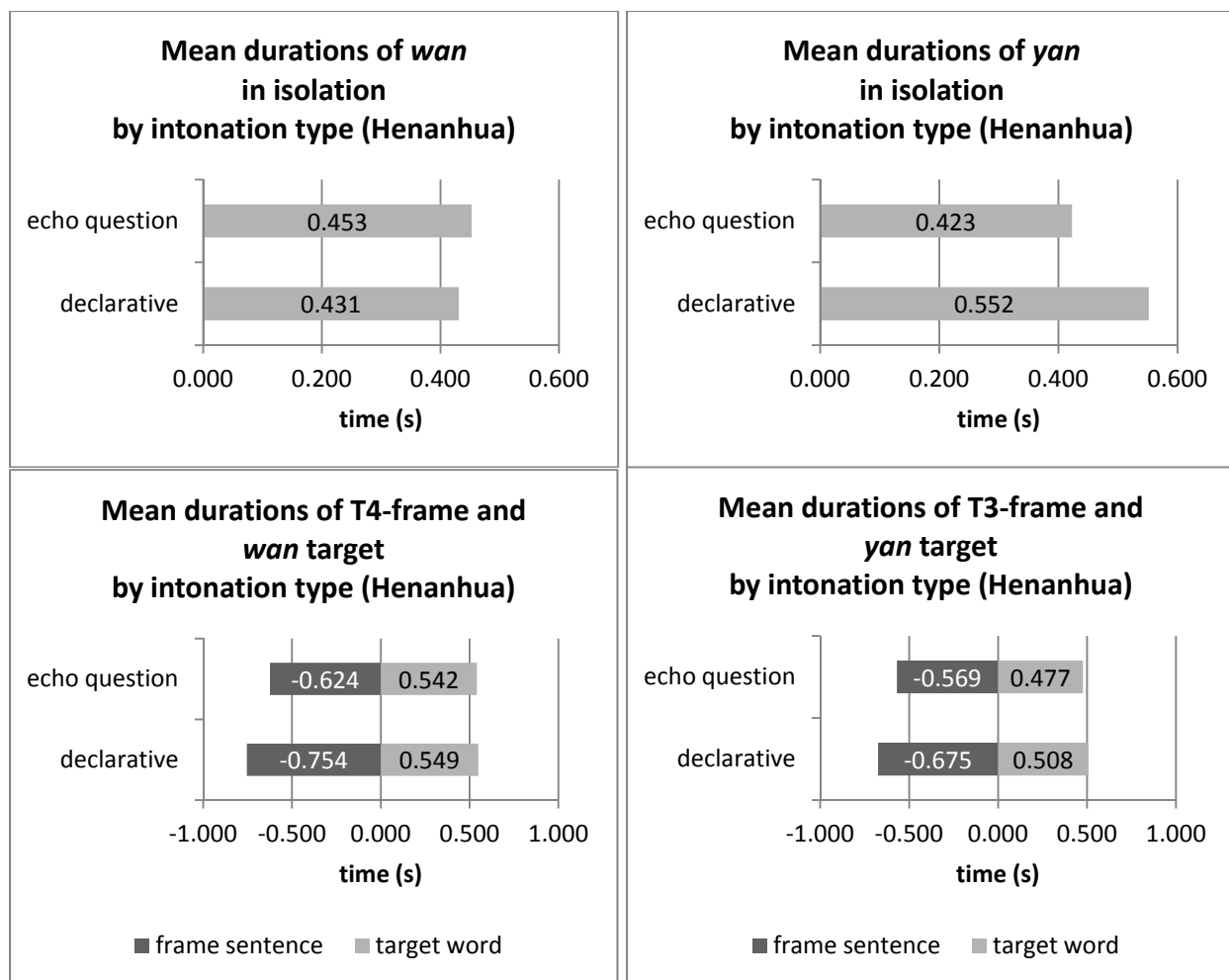
*yan3* ‘eye’

*yan4* ‘swallow’ (bird family)

Each recording session took approximately 15 minutes, and all of the declarative conditions were recorded first, followed by the echo question conditions. The second recording session took place about three weeks after the first one.

### 2.3.4 Results

The absolute duration results for the Henanhua speaker are shown in Figure 2.7, separated out by recording set (target word set—*wan* vs. *yan*—and frame type—all-T4 vs. all-T3) and context. It should be kept in mind that only 12 tokens were measured for each intonation in each context. It is apparent that the frame, when there was one, was shorter in the echo question context ( $p < .001$  for both frame types); this result was similar to that for Mandarin Speaker B, whose echo question frames were shorter compared to their declarative counterparts on average. As for the words in isolation, the *wan* set was quite similar across intonations ( $p = .059$ ) and the *yan* set was shorter in the echo question context ( $p < .001$ ). In order to give a sense of the proportional



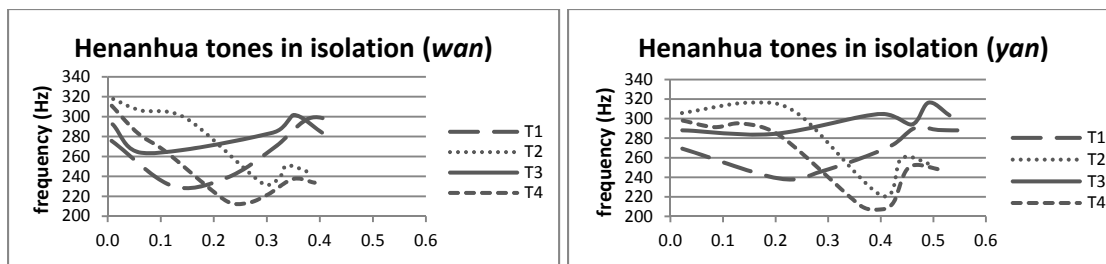
**Figure 2.7: Mean durations for Henanhua. Results for the isolation condition are shown above and those for the frame condition are shown below.**

differences across intonational categories, the frame condition results are re-displayed in Table 2.4 as ratios (as before, p-values are given in the rightmost column). Although the sample size was much smaller than for Mandarin, the results suggest that intonation had the same effect on the proportional duration of the target word.

**Table 2.4: Target-utterance duration ratio by frame type by intonation for Henanhua.**

frame	declarative %	echo question %	sig.
all-T4	42.92	45.62	.000
all-T3	42.11	46.45	.001

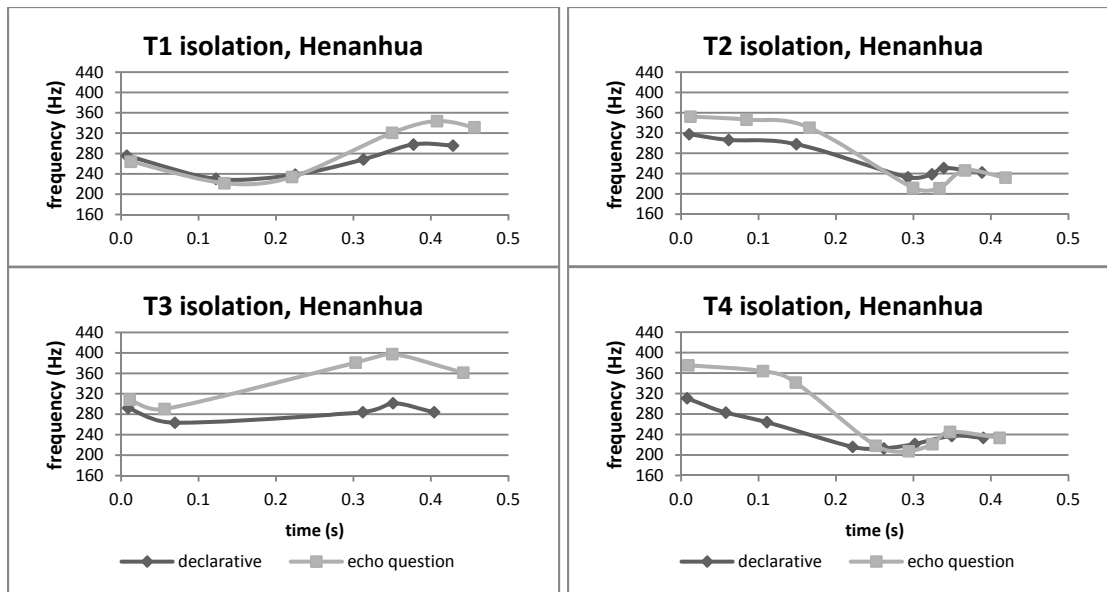
Let us turn to the  $F_0$  results. Figure 2.8 shows all four tones produced in isolation (in a declarative context), for each of the minimal sets. The contours were constructed by taking time



**Figure 2.8: Mean contours for all four tones in Henanhua in isolation on the syllable *wan* (left) and on the syllable *yan* (right).**

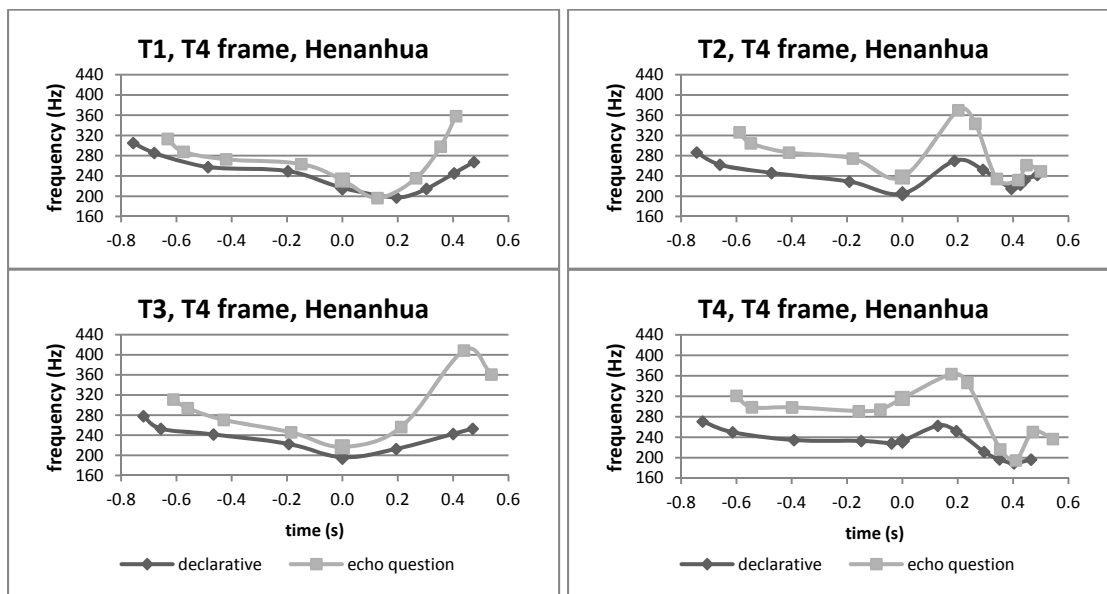
and  $F_0$  measurements at strategic points (mostly at local maxima and minima) for each token, plotting the averages of those points, and interpolating them with a smooth curve. In isolation, T1 and T3 are both rising, although T1 dips down before rising back up while T1 is nearly level with just a gentle rise. T2 and T4 are both falling, but T2 tends to fall later than T4 and possibly in a higher register. Unlike in Mandarin, then, where there is only one falling tone, there are *two* falling tones in this system, which makes it a typologically interesting system to investigate. Note that Zhang, Chen et al. (1993) captured the difference between the two falling tones as a register difference, but it is not clear that that is the main salient difference between the tones for the speaker in the current study. Also, they characterized T3 as a high level tone, but it was always rising utterance-finally in the current results.

Let us move on to the echo question results. Figure 2.9 shows mean  $F_0$  contours for the declarative and echo question conditions for each tone in the isolation condition for the *wan* set. In each plot the data series for the declarative condition is displayed in dark gray with diamond-shaped data points while the data series for the echo question condition is displayed in light gray with square-shaped data points. The echo question plots for the *yan* isolation set were largely



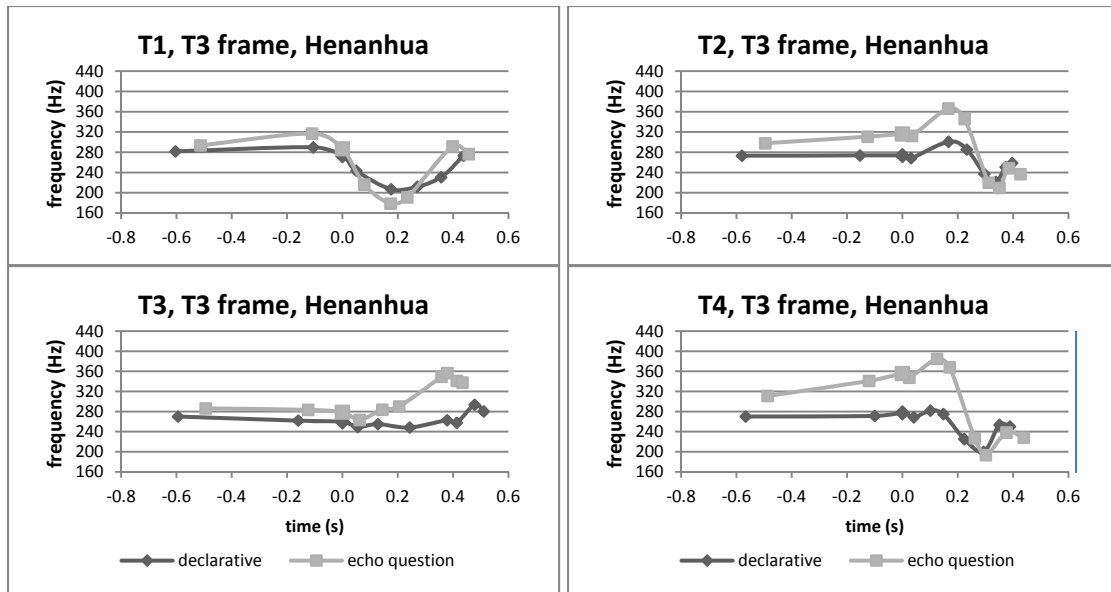
**Figure 2.9: Mean F<sub>0</sub> contours for tones in isolation in Henanhua.**

similar and will not be shown. The frame results for the *wan* set are shown in Figure 2.10. The two “falling” tones get realized as a rise and a fall on the last syllable, just as in Mandarin. This is not surprising, as the continuous nature of the F<sub>0</sub> contour demands a transition from target to



**Figure 2.10: Mean F<sub>0</sub> contours for tones in an all-T4 frame sentence in Henanhua.**

target and such a “carry-over effect” has been widely observed (Chao 1968; Shen 1990; Xu 1994). Unlike in Mandarin, however, the contour of the frame sentence is relatively flat in most cases, despite the fact that the frame sentence is composed of three T4s (falling tones) in a row. Also note that, for T1, T2, and T3, there is a fall on the third syllable leading into the target word, but this fall is noticeably absent when the target word bears T4. Figure 2.11 shows the results for the all-T3 frame. The use of the all-T3 frame appears to have allowed us to circumvent this



**Figure 2.11: Mean F<sub>0</sub> contours for tones in an all-T3 frame sentence in Henanhu.**

T4-specific effect. Some crucial comparisons among certain tone-intonation combinations will be discussed in the next section.

Finally, the multiple-contour plots for the two frame conditions are shown in Figure 2.12 and the corresponding standard deviation graphs are shown in Figure 2.13. Although these figures are not as informative as those shown in Section 2.2.4 for Mandarin, since these only include data points from three repetitions where the Mandarin figures included data from six, they are nonetheless given here for comparison. Note that, with the possible exception of T1 (especially in the all-T3 frame), all of the declarative-echo question distinctions are quite sharp.



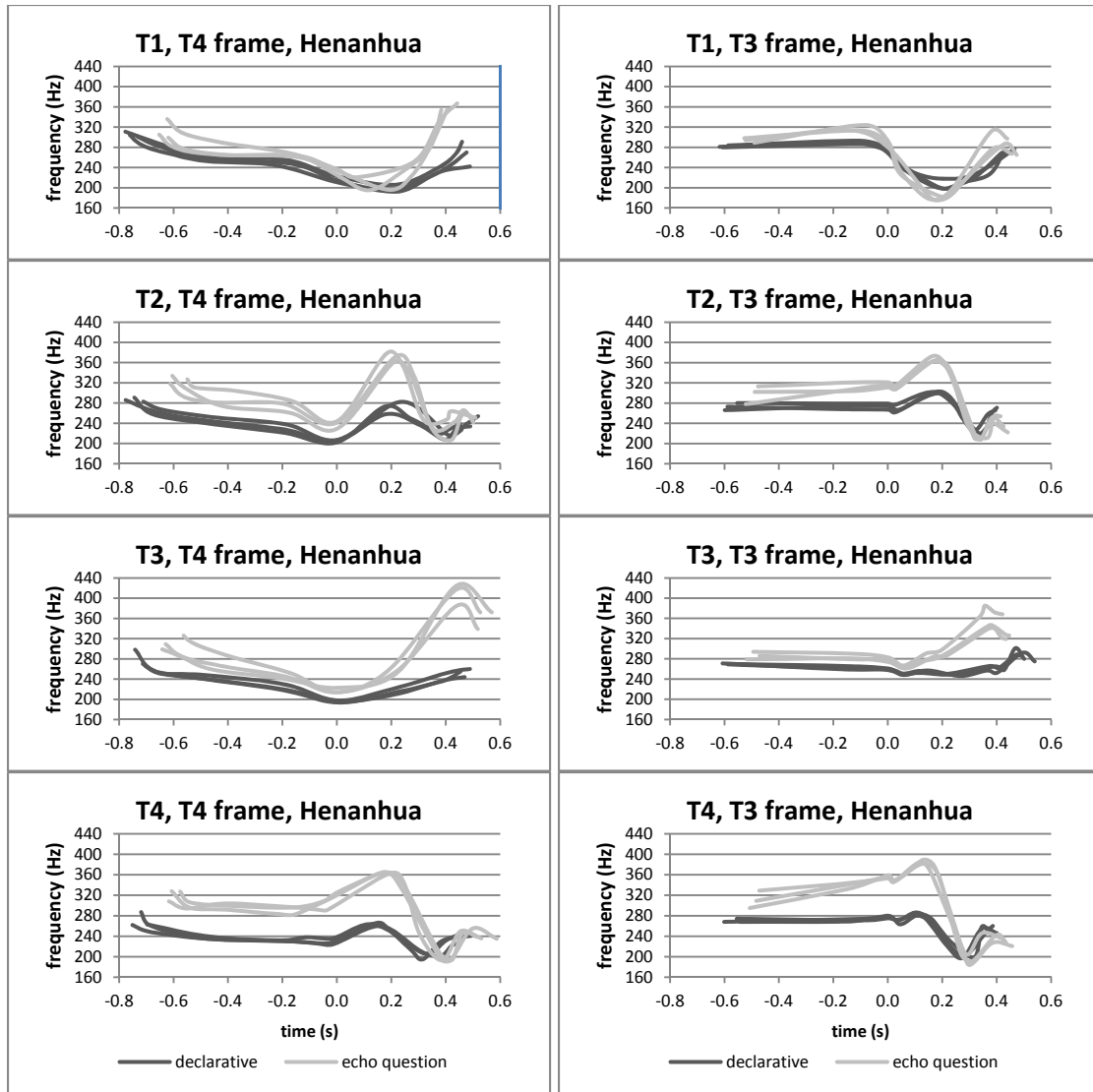


Figure 2.12: Multiple-contour plots for Henanhu.

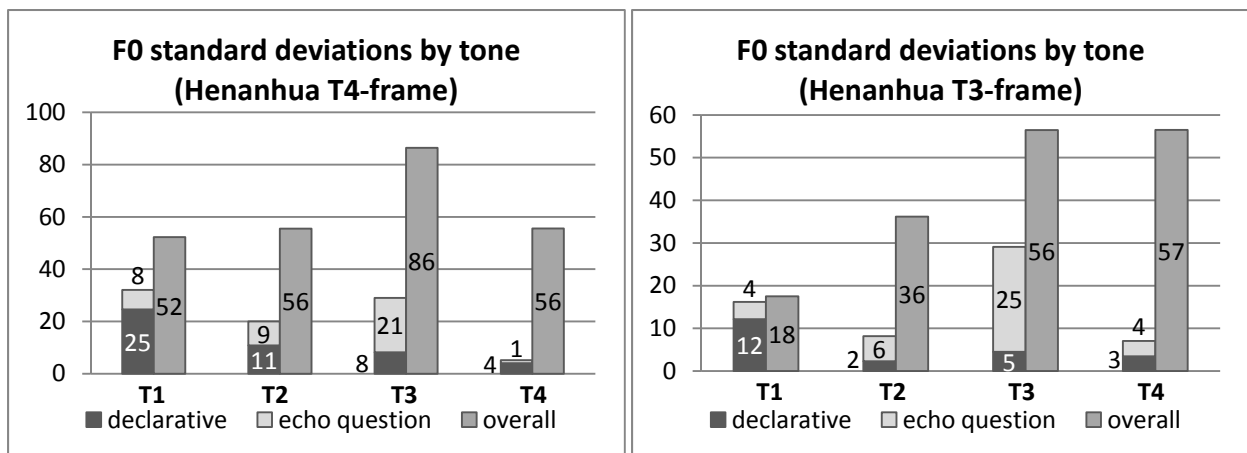


Figure 2.13: Standard deviations of declarative, echo question, and overall  $F_0$ , by tone, in Henanhu (all-T4 frame on the left, all-T3 frame on the right).

### 2.3.5 Discussion

Perhaps even more so than in isolation, T2 and T4 in a frame are distinguishable in large part by the relative alignments of their maxima and subsequent falls. To make these alignment differences within the frame contexts more obvious, they are plotted on the same graphs for ease of comparison (the plot points have been left out). This is shown in Figure 2.14, with the plots for all four tones in isolation also shown for ease of reference. We will see very similar tone-dependent alignment differences for NKK tones on disyllables in Section 2.5.5. This similarity across languages is striking given that NKK is traditionally analyzed from a pitch accent perspective and Mandarin is analyzed from a syllable-tone perspective. We will see in Section 2.5.5, however, that there are some crucial differences between the tonal systems of the two languages, including the nature of these alignment contrasts on a single syllable and the manner in which echo question intonation is realized on these “analogous” tones. Perceptual test results in Chapter 3 will further highlight some differences between the two systems.

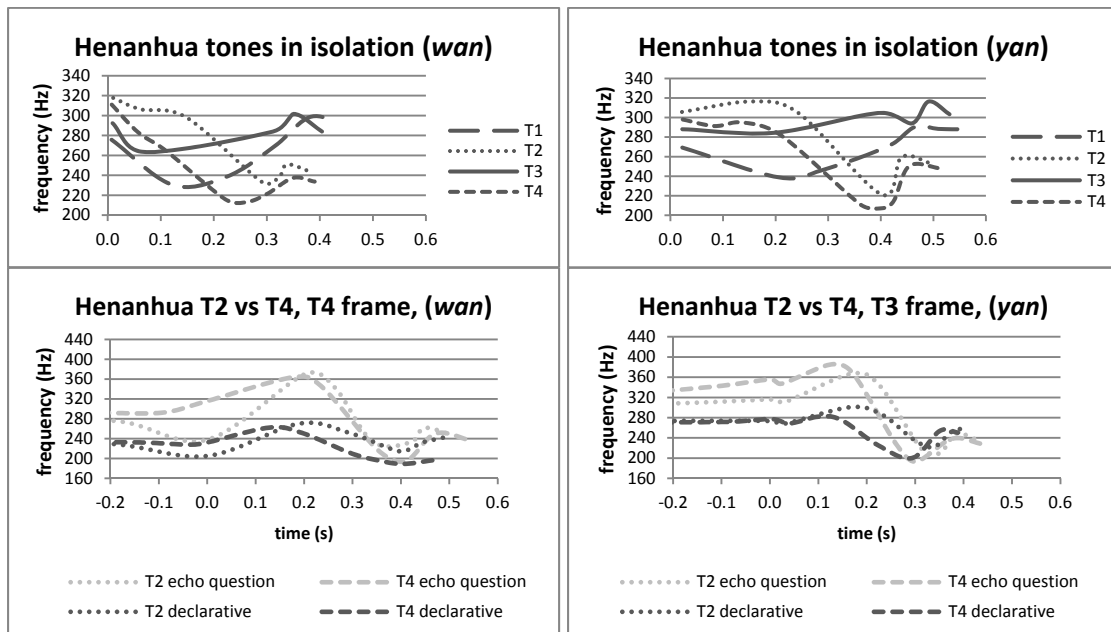
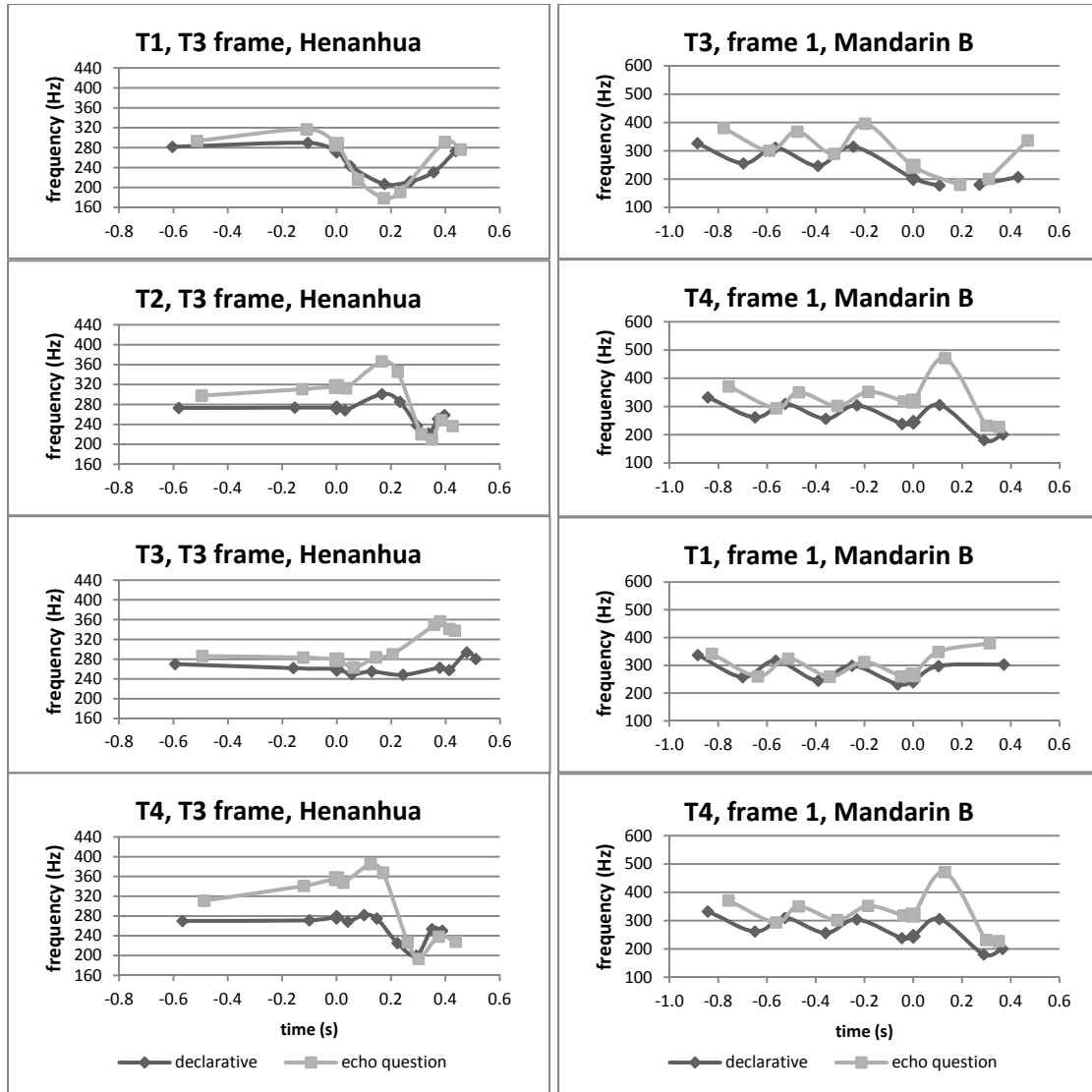


Figure 2.14: Overlaid mean contours for T2 and T4 in Henanhua (below) and plots for the four tones in isolation (above).

The two different falling tones aside, the echo question effects on the Henanhua tones are quite similar to those for equivalent tones in Mandarin. These are displayed in Figure 2.15 (Mandarin T4 is displayed twice, for comparison with each of the falling Henanhua tones).



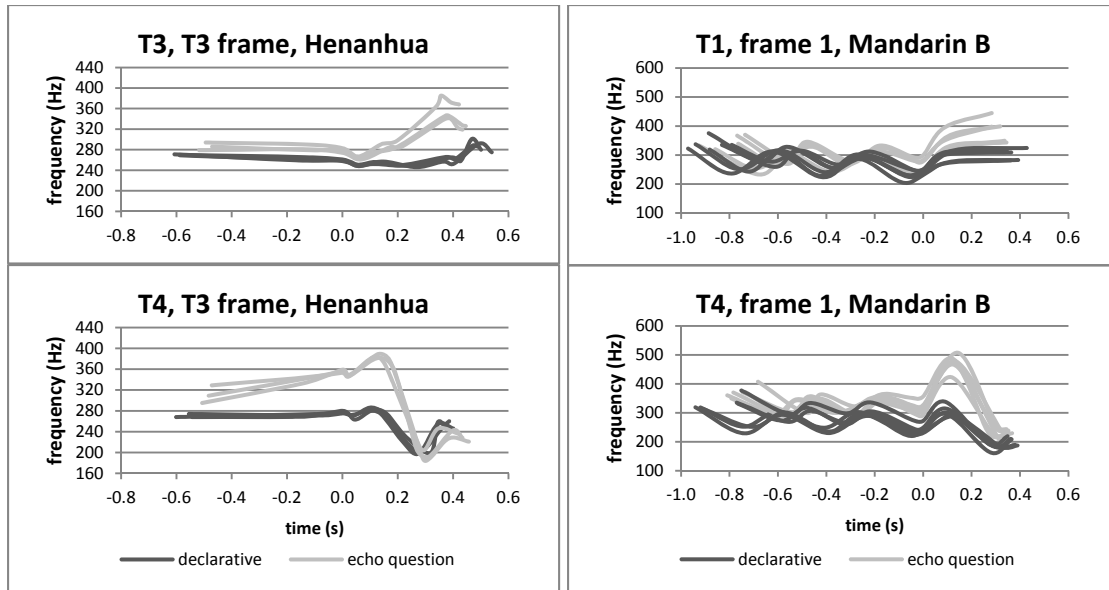
**Figure 2.15: Henanhua declarative and echo-question contours and equivalent contours in Mandarin.**

Certain crucial similarities are worth noting. First, the overall  $F_0$  range of the echo question is shifted upwards from that of the declarative utterance, but the “dipping” tones in the two dialects have the same effect on the echo question contour—at their low point they “pull it down” to the level of the declarative contour (Actually, in the case of Henanhua the echo question contour

dips *below* the declarative contour at this point). As for the falling tones in Henanhua, their intonational behavior is similar to that for Mandarin T4 (especially for Mandarin Speaker B)—the maximum cross-intonational  $F_0$  difference is seen at the peak of the contour, and then the subsequent fall is steepened in the echo question context such that it reaches an endpoint comparable to that of the declarative contour. Finally, Henanhua T3 is comparable to Mandarin T1—relatively level in the declarative context, it is realized with a steeper, linearly rising contour in the echo question context.

What about the differences? The fact that the echo question rendition of T1 dips below the declarative rendition was already mentioned above (though this sub-declarative dip is not observed after the all-T4 frame). There also seem to be more global co-articulatory effects in the echo question context in Henanhua, i.e. the  $F_0$  trend over the whole utterance leading up to the target word seems dictated by the tonal category of the target word, implying a kind of look-ahead effect. In particular, the slope of the frame contour is steepest for T4, resulting in a relative  $F_0$  at time 0 that is markedly higher than that for T2. This is surprising given the initial observation that T2 tends to be realized at a higher relative  $F_0$  than T4 in the declarative context, especially in isolation. Nevertheless, the effect appears robust, showing up in both frame contexts as well as in isolation: the declarative-vs.-echo question intonation distinction is such that the T4 peak difference is consistently greater than the T2 peak difference. This tone-specific intonational phenomenon is more evidence for a negative response to Q3 (*Declarative-to-Echo-Question Mapping*) at the beginning of the chapter.

One other comparison worth making is between the multiple-contour plots for the two dialects, shown side-by-side for two analogous tones in Figure 2.16. The intonation-dependent  $F_0$  difference for T3 in Henanhua was more robust than that for T1 in Mandarin, while the degree of separation for T4 in the two languages was more similar. Perhaps this is not a fair comparison, since only three repetitions of each intonation were elicited in Henanhua where six were elicited



**Figure 2.16: Multiple-contour plots for Henanhua T3 and T4 (left) and Mandarin T1 and T4 (right).**

in Mandarin (in two separate batches of three each). Nevertheless, the comparison is made here so that the perceptual results for the two dialects, presented in the next chapter, can be analyzed in light of it.

## 2.4 Production in Cantonese

The variety of Cantonese that will be examined in this section is that spoken in Hong Kong. Henceforth the label “Cantonese” will be used to refer to this dialect.

### 2.4.1 Overview of Cantonese tones

In traditional grammars, Cantonese is analyzed as having nine distinct lexical tones. However, three of the nine tones appear only on closed syllables and can be analyzed as predictable variants of the so-called “level” tones that appear on open syllables. Accordingly, most modern analyses only treat six of the Cantonese tones as distinctive (Matthews and Yip 1994). Table 2.5 gives descriptions and tone-letter values to each of these six tones (adapted from Matthews and

Yip 1994). Note that register plays a crucial role in this traditional analysis; three (and Note that register plays a crucial role in this traditional analysis; three (and sometimes four<sup>13</sup>) registers

**Table 2.5: Descriptive labels for the six lexical tones in Cantonese.**

tone	citation description	tone letter value
Tone 1	high-level	55
Tone 2	high-rising	35/25
Tone 3	mid-level	33
Tone 4	low-falling	21/11
Tone 5	low-rising	23/13
Tone 6	low-level	22

of level tones are recognized and two registers of rising tones are recognized. It is also worth noting that there is no high-falling tone recognized in this dialect. Historically there was a distinction between a high-level tone and a high-falling tone, but this distinction is not maintained by most speakers, and most younger speakers are reported to have merged them into a single high-level tone (Cheung 1986; Matthews and Yip 1994). One consequence of this fact is that the frame sentence used in the current experiment could not be composed of a sequence of falling tones as it was in the Mandarin experiment.

#### **2.4.2 Subjects**

Two female speakers born and raised in Hong Kong were recorded, one in her twenties and the other in her thirties.

#### **2.4.3 Materials**

All materials were recorded in a sound-attenuated booth with a Plantronics-500 DSP headset microphone connected to a laptop computer. The conditions listed in (2.11) were included:

<sup>13</sup> For some speakers, Tone 4 is realized as a level tone that is lower than Tone 6. For Speaker B in the current study, Tone 4 falls for a bit and then levels out.

(2.11) Cantonese production experiment conditions

6 target words (corresponding to the six lexical tones)

2 intonational categories (declarative statement vs. echo question)

2 frame sentences (one with all Tone 6 syllables and one with a T1-T1-T6 sequence)

6 repetitions

The 6 target words listed in (2.12) constituted a minimal set, being homophonous except for their tones:

(2.12) Cantonese target words

*laan1* ‘market’

*laan2* ‘make unfounded allegations’

*laan3* ‘skin rash’

*laan4* ‘orchid’

*laan5* ‘lazy’

*laan6* ‘broken’; ‘decay’

The two frame sentences given in (2.13) were used:

(2.13) Cantonese frame sentences

*Leiblei6 waa6 X* ‘Leilei says X’

*Yinglying1 waa6 X* ‘Yingying says X’

As in the Mandarin experiment, the first frame sentence was constructed such that every syllable leading up to the target word would bear the same tone, in order to highlight any global  $F_0$  trends brought on by utterance intonation. The second sentence was included in order to determine if any global effects of intonation on the earlier part of the utterance were tone-dependent. Both the target words and the frame sentences were comprised solely of sonorant segments in order to

maximize the probability of eliciting uninterrupted pitch curves. Each recording session was divided into four phases, as described in (2.14):

(2.14) Cantonese production experiment recording phases

Phase 1: frame 1; 3 declaratives in a row followed by 3 echo questions in a row

Phase 2: frame 1; alternating declarative / echo question 3 times

Phase 3: frame 2; 3 declaratives in a row followed by 3 echo questions in a row

Phase 4: frame 2; alternating declarative / echo question 3 times

Each session took approximately 50 minutes, including a brief (couple-minute) break taken between Phase 3 and Phase 4.

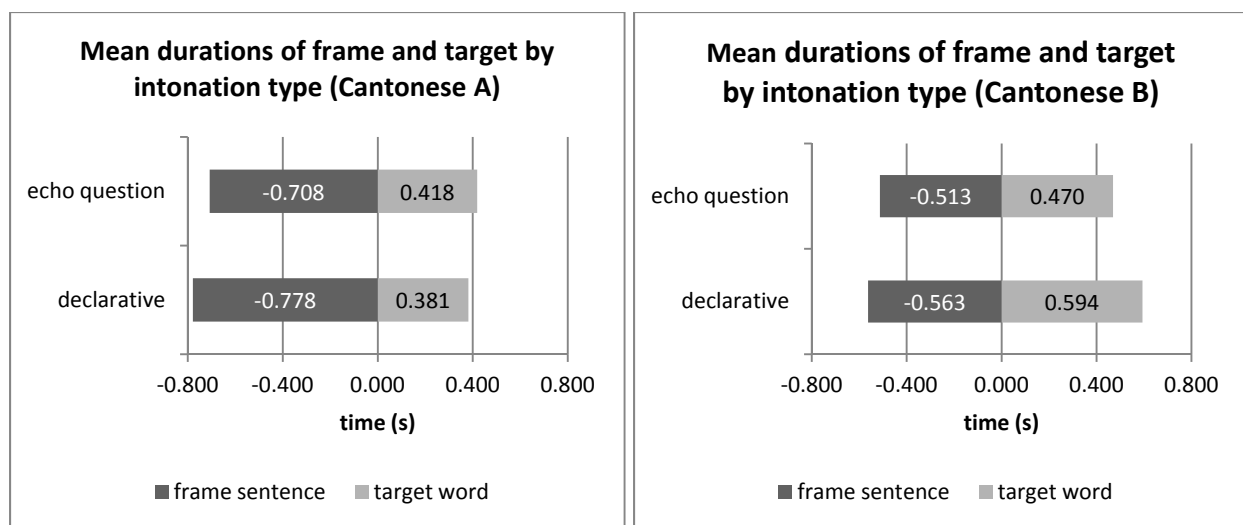
#### **2.4.4 Differences from the Mandarin design**

While the Mandarin and Cantonese production experiments were quite similar in design, there are a few differences worth noting. First, since Cantonese has six contrastive tones compared to only four in Mandarin, the total number of recorded tokens was greater for Cantonese, which made the whole recording session take longer (about 45 minutes as opposed to 30 minutes for Mandarin). Additionally, whereas the first frame sentence for Mandarin consisted of three syllables bearing a falling tone (Tone 4 in Mandarin), the first frame sentence for Cantonese consisted of three syllables bearing a low level tone (Tone 6 in Cantonese). As a result, the tonal context leading up to the target word was slightly different in the two experiments.

#### **2.4.5 Results**

The duration results for Cantonese are shown in Figure 2.17. 72 tokens were measured in each intonational category for each speaker. Cantonese Speaker A's results were similar to those for the speakers of other languages shown thus far in that the frame sentence portion of the utterance was shorter in the echo question condition ( $p < .001$ ) while the target word was longer in that context ( $p < .001$ ). Displayed in this way, it is less clear whether Cantonese Speaker B's results





**Figure 2.17: Mean durations in Cantonese.**

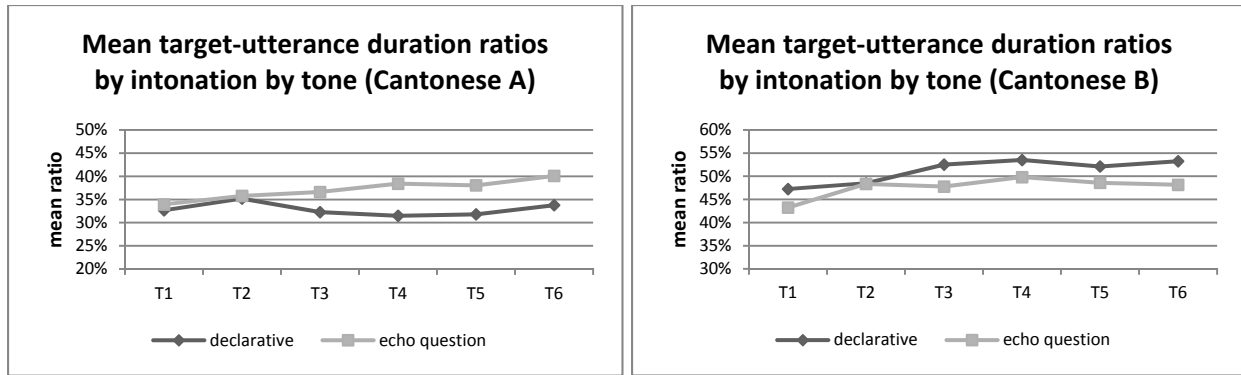
are in line with those of other speakers, but the proportional duration results, shown in Table 2.6, indicate that they are not. Of the speakers discussed thus far, it seems that Cantonese Speaker B is the only one whose target words were proportionally shorter in the echo question context on average.

**Table 2.6: Target-utterance duration ratio by intonation for Cantonese Speakers A and B.**

Speaker	declarative %	echo question %	sig.
Cantonese A	38.14	41.84	.000
Cantonese B	59.36	46.96	.000

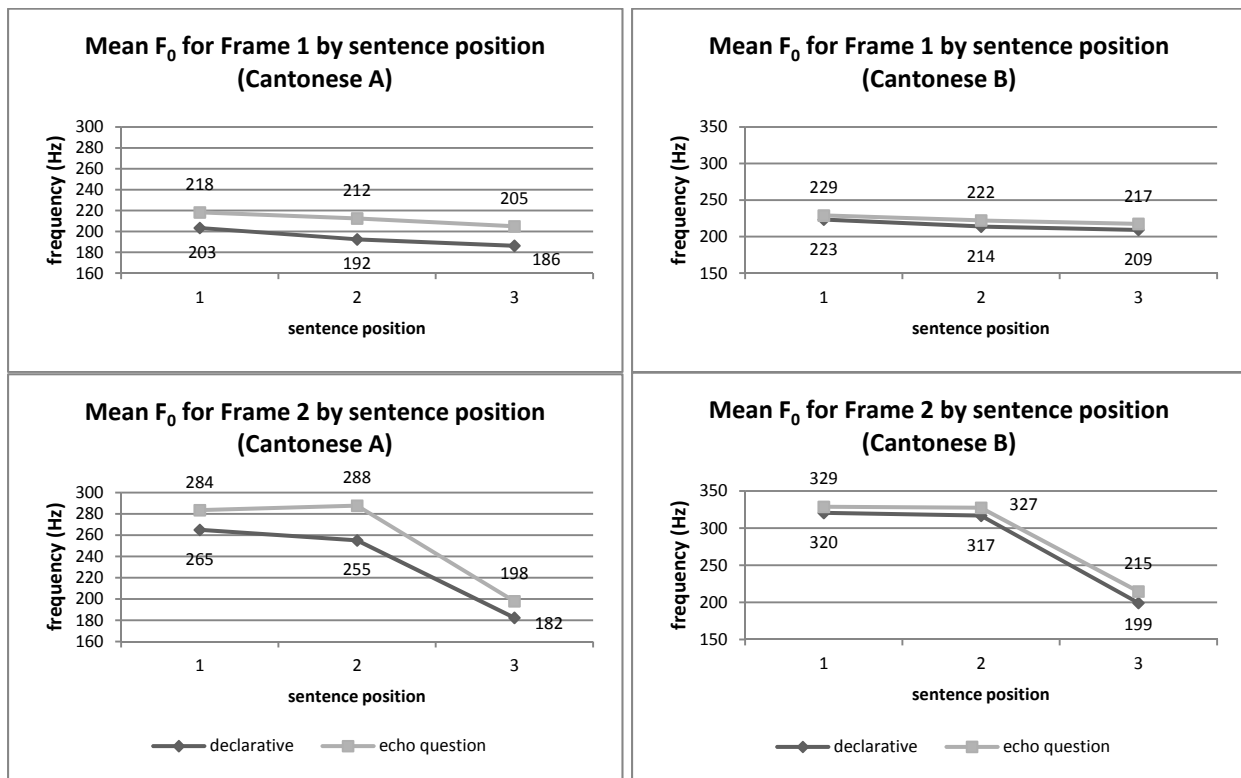
The duration results are broken down by tonal category in Figure 2.18. We see that the overall trend for each speaker is reflected, for the most part, in each tonal category. For Speaker A, T1 and T2 were the exceptions ( $p = .284$  for T1;  $p = .615$  for T2;  $p < .001$  for other tones). For Speaker B, T2 was the sole exception ( $p = .878$  for T2;  $p < .001$  for all other tones).

Let us turn to the  $F_0$  results. Previous descriptions of Cantonese echo question intonation have noted a tendency for the overall  $F_0$  of the entire utterance to be raised in echo questions relative to their declarative counterparts (Wu 1990; Ma, Ciocca et al. 2006), although Xu and Mok (2011) reported that this was only a tendency and that it did not necessarily occur in



**Figure 2.18: Mean target-utterance duration ratios for declarative and echo question intonation, by tone (Cantonese A on the left and Cantonese B on the right).**

utterances with certain lexical tonal compositions. The mean  $F_0$  plots for the frame sentence positions are shown in Figure 2.19; since the frame sentences consisted of level tones in each case,  $F_0$  measurements were taken at the beginning of the nucleus of each successive syllable.



**Figure 2.19: Mean  $F_0$  of the frame by sentence position. Frame 1 above and Frame 2 below, Cantonese A on the left and Cantonese B on the right.**

The raw numbers suggest that the global intonational effect was a bit more robust for Cantonese Speaker A than for Cantonese Speaker B. The ANOVA results, with a Bonferroni

correction, indicate that the differences are highly significant across the board for Speaker A ( $p < .001$  at all positions in both frames) but they also indicate that, for Speaker B, the differences are likewise significant for both frames ( $p < .001$  for all three positions in Frame 1;  $p = .029$ ,  $p = .005$ ,  $p < .001$  for the three respective positions in Frame 2). Whether Speaker B's differences are perceptually relevant is another question (one that unfortunately will not be addressed in the current study). Unlike in the Mandarin all-T4 frame results, there is no evidence here for either speaker that the Cantonese all-T6 frame shows any kind of successive divergence through the course of the frame; that is, the pitch gap between the declarative and echo question conditions does not widen later in the frame. Also, at least for the two tone-types tested—Tone 1 (high level) and Tone 6 (low level)—there is no evidence for tone-specific global intonation effects.

Let us move on to the target word results. Previous studies have reported that echo question intonation changes the final tone in an utterance into a rising tone (Wu 1990; Yip 2002; Ma, Ciocca et al. 2006). Wu (1990) asserted that, in an echo question context, if a tone is inherently rising the rise is “intensified” while if it is inherently falling the rise begins after the fall (he did not mention the inherently level tones). Yip (2002) asserted that a high tone combines with the final lexical tone to form a new “tone” that starts on the pitch of the original tone and ends “high” (so echo questions generally end with a rising tone unless the final lexical tone is T1, in which case they end with a high level tone). Gu, Hirose et al. (2006) performed a more detailed assessment of each tone in the echo question context and note the lack of correlation between the “inherent” pitch excursion of the declarative renditions of the tones and that of the echo question rendition of the tones. The results of the current study are compatible with most of the observations and claims above, with the exception of the claim by Yip (2002) that echo questions ending in T1 end with a level contour.

Figure 2.20 shows mean  $F_0$  contours for the declarative and echo question conditions for each tone, for each speaker, for the frame 1 condition only (where the frame sentence consisted of three T6—i.e. low level—syllables in a row leading up to the target word). As was done for

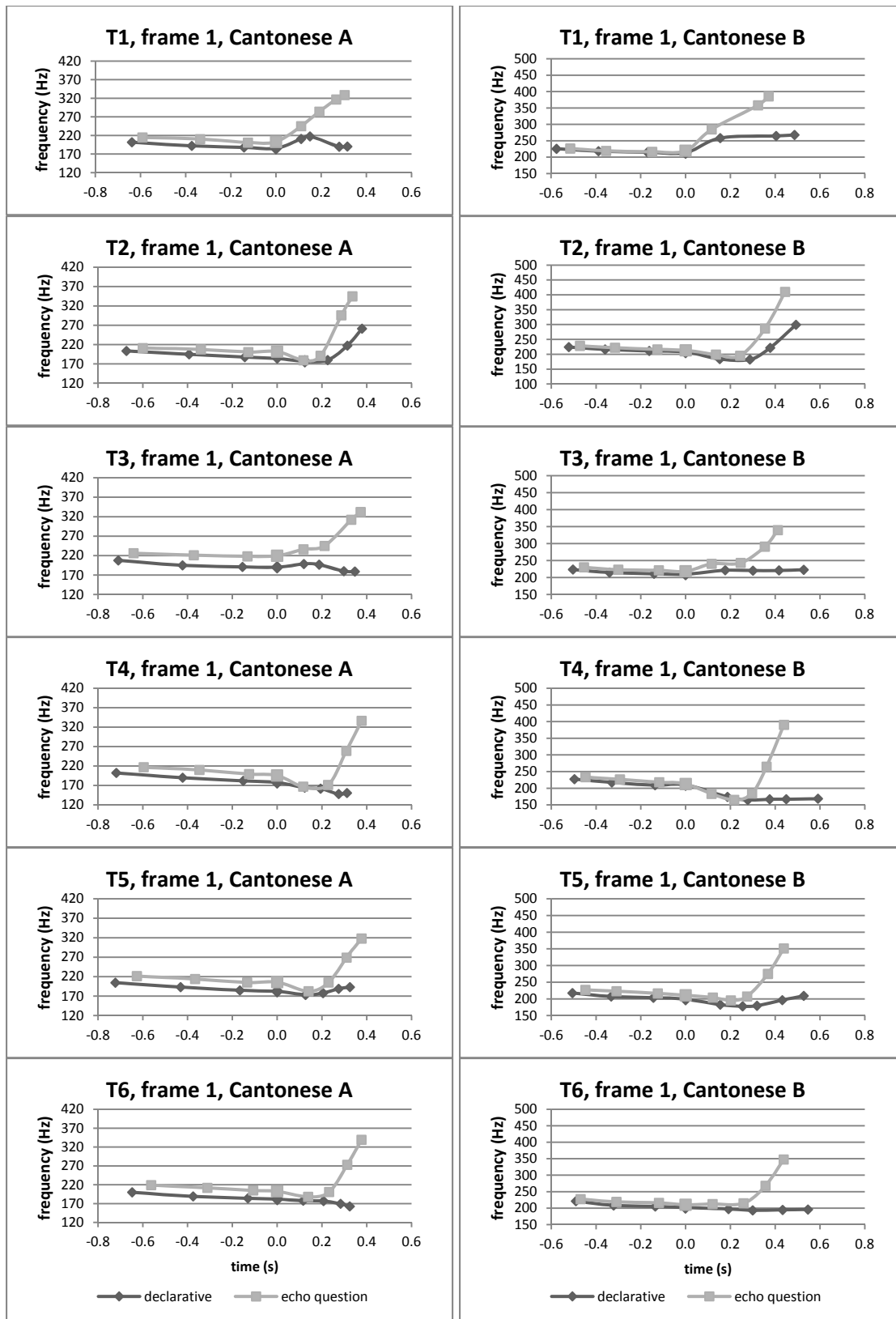
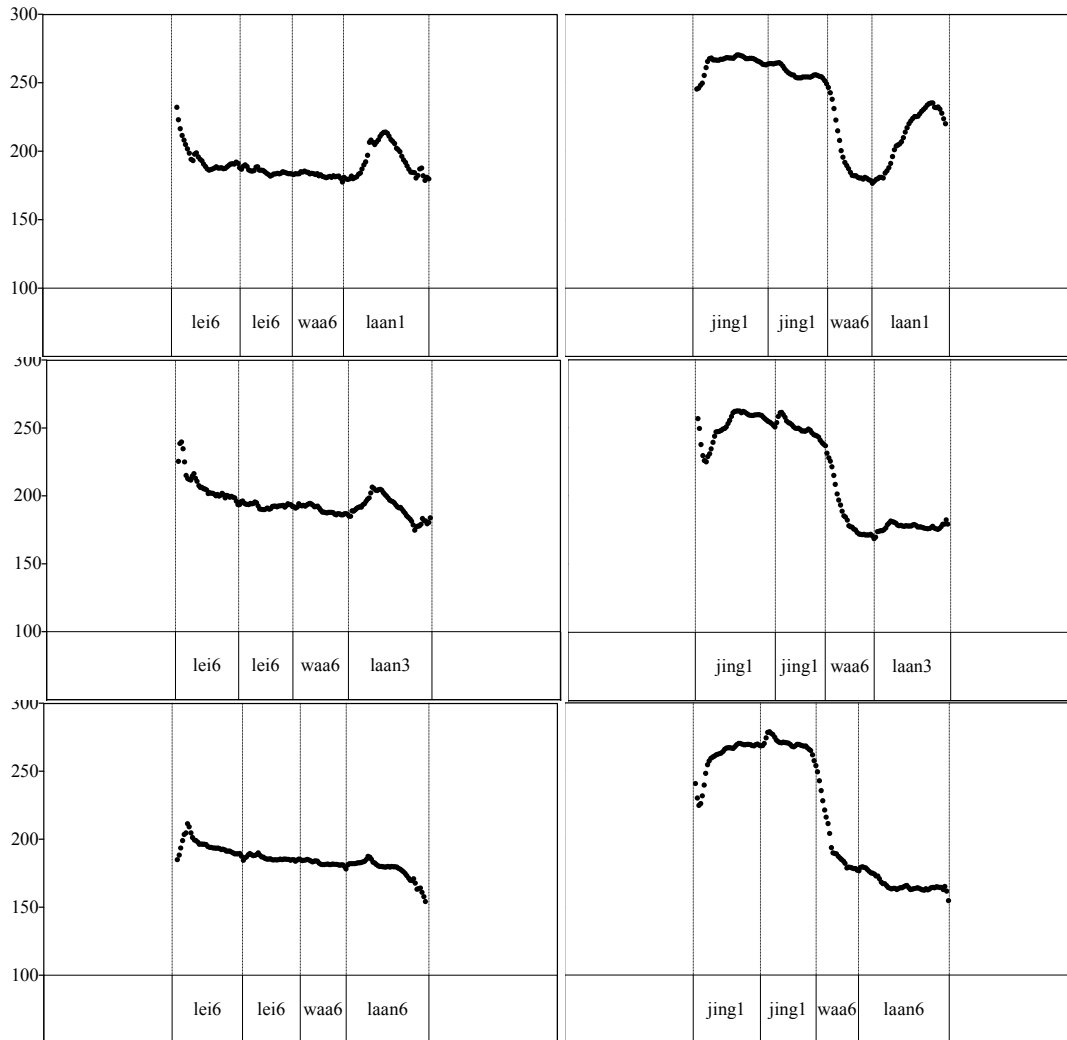


Figure 2.20: Mean  $F_0$  contours for Cantonese.

the previous languages, the contours were constructed by taking time and  $F_0$  measurements at strategic points for each token, plotting the averages of those points, and interpolating them with a smooth curve. The plots are time-aligned such that the boundary between the frame sentence and the target word, indicated with an enlarged data point, falls at time 0 in every case. Speaker A's results are displayed on the left and Speaker B's results are on the right. Broadly speaking, there are clear effects of intonation type on the surface contours of the utterances, and most of these effects are visible on utterance-final syllables. While the pitch contour on the final syllable is variably falling, level, or rising in the declarative context, depending on the lexical tone on that final syllable, it is steeply rising in the echo question context, regardless of the lexical tonal category.

The main focus of this section is the difference between the declarative and echo question contours, but it is worth taking a moment to note a particular cross-speaker difference in the realizations of the so-called “level” tones—T1, T3, and T6—in the declarative context. While Speaker B produced these tones with a very level pitch in the declarative context, Speaker A tended to produce them with a falling pitch in that context after Frame 1 only. Although the overall Frame 2 results will not be presented in this section (because they generally fall in line with the Frame 1 results), some representative contours of T1, T3, and T6 targets in each of the frame contexts, produced by Speaker A, are shown side by side in Figure 2.21 for comparison. Although both frames end with the same T6-bearing word (*waa6*), there is a striking contrast across the frame conditions in the realizations of these tones. T3 and T6 were quite level following Frame 2; T1 was actually rising through most of the syllable and then an abrupt (i.e. glottalized) offset caused the pitch trace to fall off a bit in the final nasal. The contours shown here for Frame 2 are quite representative of all of the repetitions of those respective tones that were recorded in the Frame 2 conditions. A brief discussion of this phenomenon will be undertaken in Section 2.4.6, but for now it is worth noting that this seemingly free variation may have to do with the fact that there is no level-vs.-falling distinction made in the Cantonese tonal system.



**Figure 2.21: Representative contours for “level” tones (T1, T3, and T6) after Frame 1 and Frame 2, respectively, for Cantonese Speaker A.**

Let us turn to the multiple-contour plots and  $F_0$  standard deviation graphs for both Cantonese speakers. These are shown in Figures 2.22 and 2.23, respectively. In general, we see very pronounced intonation-dependent separation on the last syllable, the exception being T2 for both speakers (the T2 multiple-contour plots for both Cantonese speakers actually look very similar to the T2 multiple-contour plot for Mandarin Speaker B). In the next section these and the other production results will be discussed in more depth, but for now it is worth noting that the T2 conditions received the lowest rate of intonational accuracy in the perceptual study presented in Chapter 3.

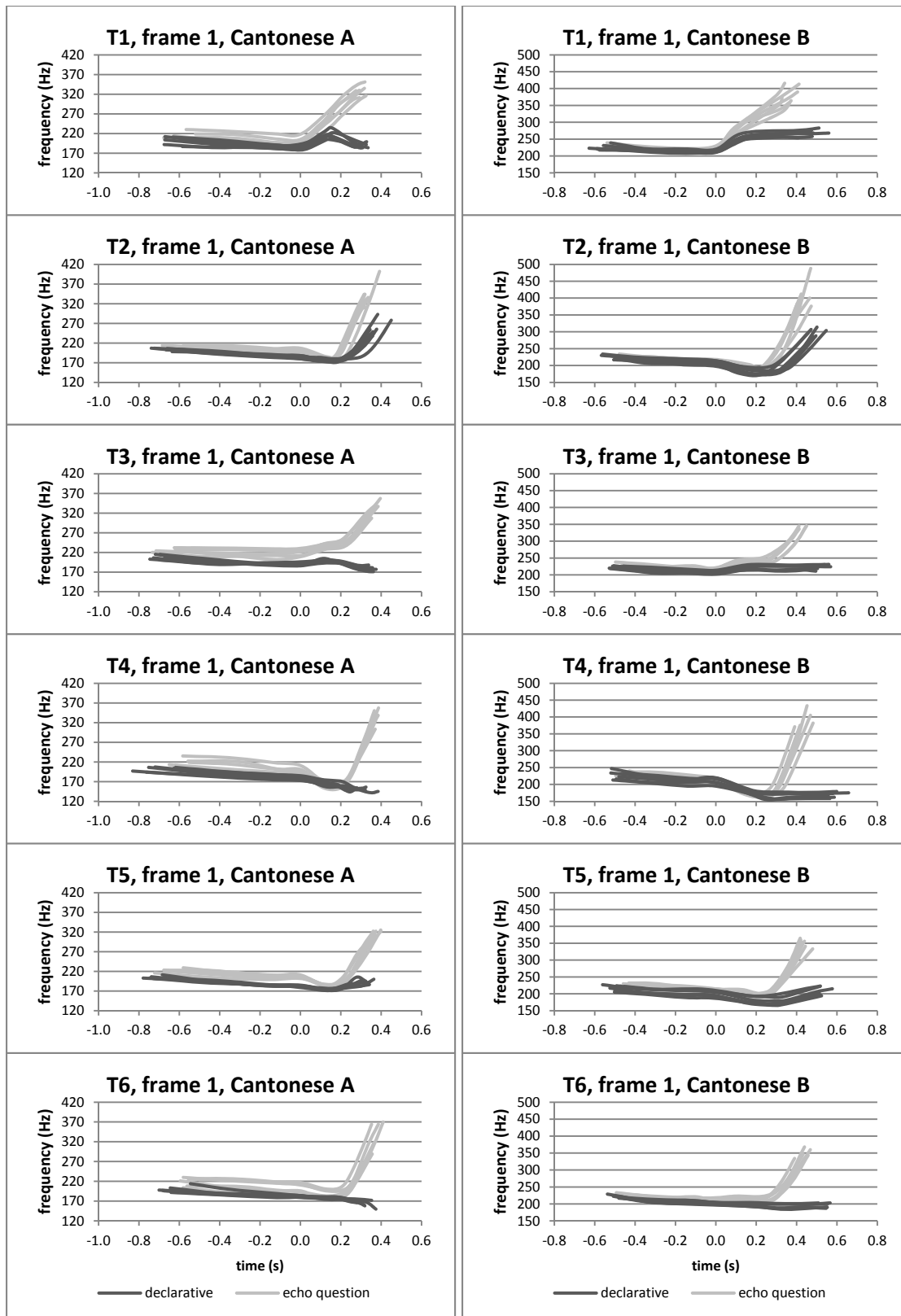
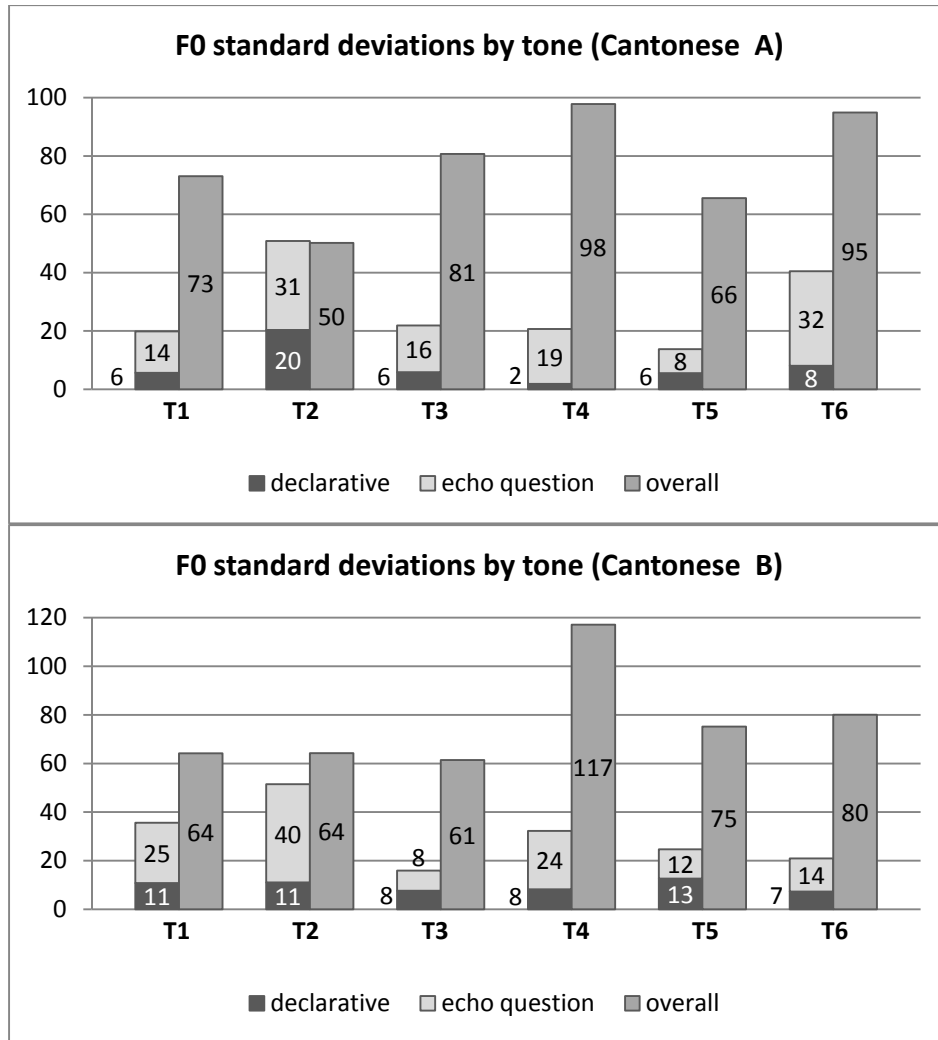


Figure 2.22: Multiple-contour plots for Cantonese.



**Figure 2.23: Standard deviations of declarative, echo question, and overall F<sub>0</sub>, by tone, in Cantonese.**

### 2.4.6 Discussion

Before commencing the main discussion of the results, let us briefly return to the issue of Speaker A’s varying realizations of the “level” tones. During a follow-up interview, these different realizations were pointed out to Speaker A, who readily perceived the acoustic differences. When asked to characterize the difference in meaning, scope, nuance, etc. conveyed by this melodic difference, if any, she was unable to identify any such difference. When the different tunes were pointed out to Speaker B (who always produced these tones with a level pitch in both frame contexts), she also acknowledged the melodic difference but insisted that



both were acceptable realizations of the same thing. One possible explanation for the observed acoustic differences is that Speaker A was utilizing an optional local falling tune associated with declarative intonation during Phases 1 and 2 but not during Phases 3 and 4 (recall that a brief break was taken between Phases 2 and 3). The fact that, historically, the Cantonese “high-level” and “high-falling” tones merged and are no longer distinctive (Cheung 1986; Matthews and Yip 1994) might also give license to the current generation of speakers to produce a range of realizations of the so-called “level” tones. Indeed, although both Cantonese Speakers A and B in the current study were able to perceive the acoustic difference between the falling pitch contours and the level ones, Wong (1982) noted a study reported on by Xie (1974) in which native Cantonese speakers who were learning Mandarin had trouble both in consistently producing appropriate contours for and perceiving the difference between Mandarin T1 and T4 (the “high level” and “falling” tones, respectively).

The main effect of the intonational category is observed on the last syllable. In general the  $F_0$  curve for the echo question condition starts at the beginning of the syllable in the vicinity of where the  $F_0$  curve for the declarative condition starts and then “peels away” from it partway through, always resulting in a final rise but giving a different overall shape to the  $F_0$  curve on the final syllable. Consequently, the shape of the echo question contour for all of the lexical tonal categories resembles either the declarative or echo question contour for T2 (in fact, Law 1990 claimed that most of the lexical tones get phonologically converted into a T2 in this environment), and to a lesser extent the more gently rising contour for T5. In the perceptual results for Cantonese in Chapter 3, these similarities will be shown to elicit a bias towards the perception of T2 in the echo question context.

It is worth taking a moment to compare these results to the descriptive observations made in the literature. Recall Wu’s (1990) observations that, in an echo question context, if a tone is inherently rising the rise is “intensified” while if it is inherently falling the rise begins after the fall. The results of the current study support those observations. T2 and T5 are both rising tones in citation form, and in the echo question context their  $F_0$  slope steepens and they reach a higher

final  $F_0$  than in the declarative context. Meanwhile T4, the low falling tone, does fall initially before rising in the echo question context. What about Yip's (2002) assertion that a high tone combines with the final lexical tone to form a new tone that starts on the pitch of the original tone and ends "high"? She claims that, as a result, echo questions end with a high level tone if the final lexical tone is the high tone (T1). Theoretical assumptions aside, this characterization is clearly not supported by the current results. It is true that the  $F_0$  contour on the final syllable generally starts in the range of the original tone, but there does not seem to be a static high target that is approached at the end of the syllable. T1 in the echo question condition is clearly *rising*, not level, and the different tones reach different final  $F_0$  values at the end of the syllable. One would be hard-pressed to attribute this variation to various degrees of undershoot, since T4 reaches a higher final  $F_0$  on average than T3 or T5, despite the fact that T3 and T5 start rising from a higher initial  $F_0$  than that of T4. Gu, Hirose et al. (2006) noticed this phenomenon as well, and they attribute it to a "compensation" effect, noting that it rules out the possibility of the lexical tone mechanism and the echo question intonation mechanism being additive<sup>14</sup>. These two points—that echo question interaction is neither a static target following the lexical tonal target nor an  $F_0$  function that interacts additively in parallel with the lexical tone—are crucial for ruling out certain models of speech melody, and they will be revisited in Chapter 4.

There are some superficial similarities between the tone-intonation interaction in Cantonese and that in Mandarin. Durational effects appear to be minimal in both languages. The duration results from the two Cantonese speakers vary along the same lines as those from the two Mandarin speakers, and a reasonable explanation for the variation is distinctive focus assignment by each of the respective speakers—narrow focus on the target word for Speaker A and broad (i.e. "neutral") focus for Speaker B. In both languages the difference between declarative and echo question intonation involves a difference of  $F_0$ . This  $F_0$  difference is most extreme on the final syllable, but there are subtler global effects on the frame sentence leading up

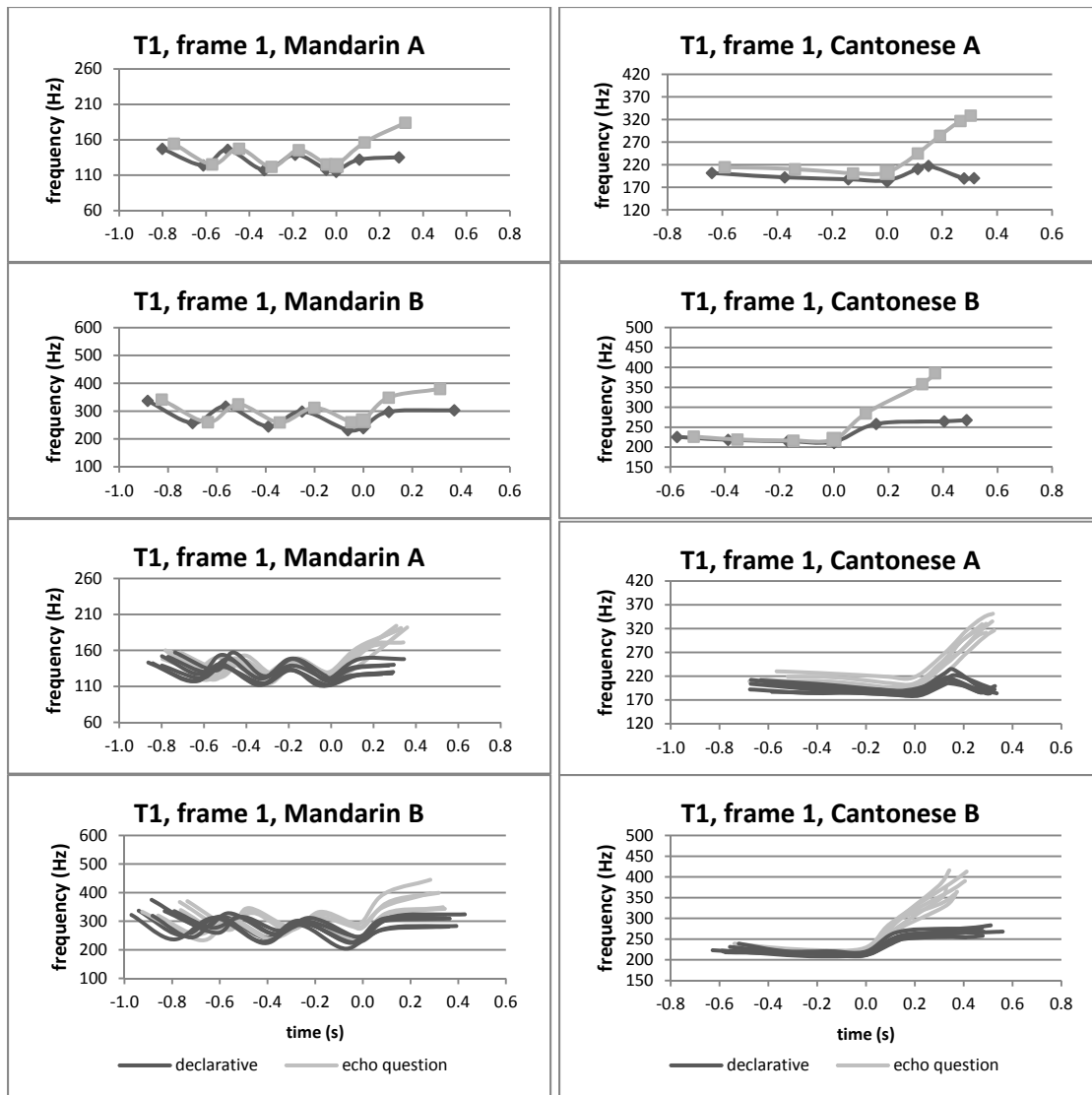
---

<sup>14</sup> They are working with a command-response model, so their discussion is in terms of the respective tone commands of the lexical tones and the echo question intonation.

to the final syllable as well. In Cantonese, just as in Mandarin, this global  $F_0$  difference was more robust for one of the two speakers and may be related to the scope of focus in the echo questions for that speaker.

The biggest difference between the Mandarin and Cantonese melodic systems is observed within the domain of the rightmost syllable in the utterance. From a purely quantitative standpoint, it can be said that the interaction of echo question intonation with any of the six lexical tones in Cantonese results in the  $F_0$  contour ending with a strongly positive slope, regardless of the sign of the tones' inherent slopes or of their slopes in the declarative context. In Mandarin, meanwhile, the inherent sign of the slope for T4, the falling tone, is preserved in the echo question context (i.e. it is still falling). The two languages also show distinct patterns when it comes to how robust the intonation-dependent pitch range differences are on the last syllable. This difference is strikingly illustrated by comparing, side by side, the mean  $F_0$  plots and the multiple-contour plots for T1 in each language. This is shown in Figure 2.24. While the respective mean  $F_0$  contours for the declarative and echo question renditions of T1 look quite similar across-the-board, the multiple-contour plots tell a different story. The range of  $F_0$  on the final syllable is more distinct for Cantonese than for Mandarin. A broader study with more speakers is needed to confirm that this is in fact a language-specific phenomenon and not a coincidental speaker-specific phenomenon, but it is an interesting result nonetheless; we will see in the next chapter that these differences are reflected in the different perceptual results attained for the respective languages.

As for Q1 (*Language-General Tone-Intonation Interaction*) at the beginning of this chapter, it is clear that intonation does *not* interact with lexical tone in the same way in all tone languages. As for Q2 (*Dimensions of Language-Specific Variation*), we have seen that the  $F_0$  functions may interact in a way that appears more “translational” (either additive or multiplicative), as in Mandarin, or one may “overwrite” the other, as in Cantonese. It also



**Figure 2.24: Mean  $F_0$  plots and multiple-contour plots for T1 in Mandarin (left) and Cantonese (right).**

appears that the robustness of the  $F_0$  range difference across intonations is language-dependent, as evidenced from the multiple-contour plots and standard deviation graphs (though further study is needed to confirm this). Finally, while we already answered “no” to Q3 (*Declarative-to-Echo-Question Mapping*) based on the Mandarin results, the Cantonese results provide us with more evidence that the declarative-echo question difference cannot be expressed as a single algorithm that simply takes the phonetic parameters of one as input to yield the phonetic parameters of the other. We have seen that the pitch range on the syllable is tone-dependent, and that neither the

final  $F_0$  height nor the total  $F_0$  excursion are kept constant (or even proportional), so the echo question function cannot be a simple interpolation to a target or an additive or multiplicative function. These language-specific differences and tone-specific melodic mechanisms all have implications for modeling, which will be discussed in subsequent chapters.

## **2.5 Production in North Kyeongsang Korean**

North Kyeongsang Korean (henceforth NKK) is a label for a dialect of Korean spoken in the Northern Kyeongsang Province in the southeastern region of the Korean peninsula (Lee 2008). Unlike Seoul Korean, NKK makes use of lexical tonal distinctions and is often analyzed as a “pitch accent language” (Kim 1997; Chang 2005; Jun, Kim et al. 2006).

### **2.5.1 Overview of NKK tonal classes**

It is generally accepted that words in NKK all fall into one of four tonal shapes—those with an accent on the initial syllable, those with an accent on the penultimate syllable, those with an accent on the final syllable, and a “special” class of words that bear a distinctive type of accent that is realized on the first two syllables. Jun (2006) considered the first three shapes to derive from one, singly-linked pitch accent type that can link to the first, penultimate, or final syllable of a word, respectively, and the fourth shape to derive from a second, doubly-linked accent type that only ever links to the first two syllables of a word. Lee (2008) notes that there is only a three-way contrast possible on disyllabic words given the system described above (since, in a disyllabic word, the penultimate syllable is also the initial syllable) and she uses the labels HL, LH, and HH to refer to the respective tonal classes on disyllables. Table 2.7 shows these three tonal categories with their equivalent Lee (2008)-style labels. On monosyllables, only a two-way distinction is possible; the two categories are considered to be initial-accented (HL) and double-accented (HH), respectively, because they pattern with those respective classes when converted into disyllables by the addition of particles. In the rest of this dissertation, the more descriptive labels from the left-hand column of Table 2.7 will be used in order to facilitate cross-

**Table 2.7: Tone classes and equivalent Lee (2008)-style labels on disyllables in NKK.**

tonal category	Lee (2008) label
initial-accented	HL
double-accented	HH
final-accented	LH

linguistic comparisons with the Kansai Japanese tonal categories, but no implicit theoretical analysis should be inferred from these labels—especially the *double-accent* label.

### 2.5.2 Subject

A female speaker in her thirties, born and raised in Andong, South Korea, was recorded.

### 2.5.3 Materials

All materials were recorded in a sound-attenuated booth with a Plantronics-500 DSP headset microphone connected to a laptop computer. The conditions that were included are given in (2.15):

(2.15) NKK production experiment conditions

20 words (representing 3 tonal categories: initial-, double-, and final-accented)

2 intonational categories (declarative statement vs. echo question)

2 contexts (isolation vs. frame sentence)

3 repetitions

The words, all nouns, were either monosyllabic or disyllabic. The specific words in the word list are given in Table 2.8. The frame sentence shown in (2.16) was used for the frame conditions:

(2.16) NKK frame sentence

*Eunhi-neun X* ‘Eunhi-TOP X’ (i.e. ‘As for Eunhi, X’)

This somewhat casual topicalized construction had to be used in order to get the target word (a noun) to be utterance-final, since Korean is an SOV language.

**Table 2.8: Categorized word list for NKK production experiment**

syllable count	tonal category	words
monosyllabic	initial-accented	<i>mal</i> ‘horse’ <i>nam</i> ‘south’
	double-accented	<i>mal</i> ‘end’ <i>nam</i> ‘third party’
disyllabic	initial-accented	<i>mole</i> ‘sand’ <i>mal-i</i> ‘horse-NOM’ <i>nam-i</i> ‘south-NOM’ <i>uli</i> ‘pigpen’ <i>mun-i</i> ‘door-NOM’
	double-accented	<i>nai</i> ‘age’ <i>kohyang</i> ‘birthplace’ <i>meil</i> ‘every day’ <i>mole</i> ‘day after tomorrow’ <i>mal-i</i> ‘end-NOM’ <i>nam-i</i> ‘third party-NOM’
	final-accented	<i>uli</i> ‘us’ <i>muni</i> ‘pattern’ <i>Nahi</i> ‘Nahi’ (feminine name) <i>Koyang</i> ‘Koyang’ (a city name) <i>meil</i> ‘e-mail’

#### 2.5.4 Differences from the Mandarin and Cantonese designs

There is one aspect of the NKK word list that made it quite different from the word lists in the Mandarin and Cantonese production experiments. In Mandarin and Cantonese it was easy to find four- and six-way minimal sets of words, respectively, that differed only in tone. While there are three contrastive tonal categories on disyllabic words in NKK (initial-accented, double-accented, and final-accented), it was not possible to find minimal triplets that satisfied the other criteria in the production experiment (nouns comprised of sonorant segments). Instead, multiple minimal pairs were included so that at least one comparison for every combination of two out of

the three tonal categories could be made (initial-accented vs. double-accented, initial-accented vs. final-accented, and double-accented vs. final-accented).

### 2.5.5 Results

Our single NKK speaker's duration results were, in general, comparable to those of the "B" subjects in Mandarin and Cantonese—that is, the echo question tokens were shorter on average than the declarative tokens. This was true both for the isolation condition, shown in Figure 2.25, and for the frame condition, shown in Figure 2.26.

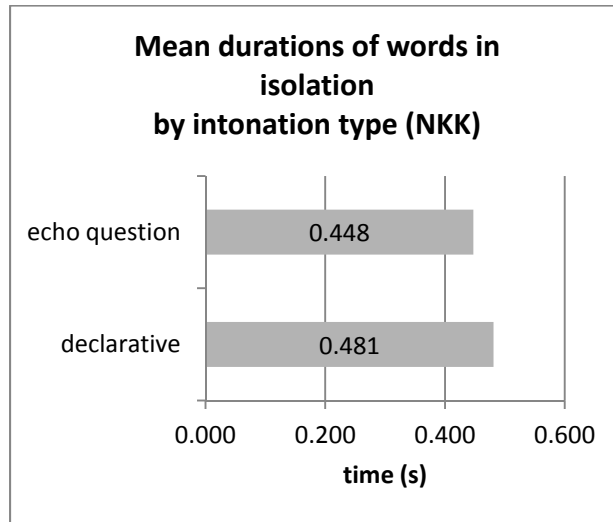


Figure 2.25: Mean durations for the isolation condition in NKK.

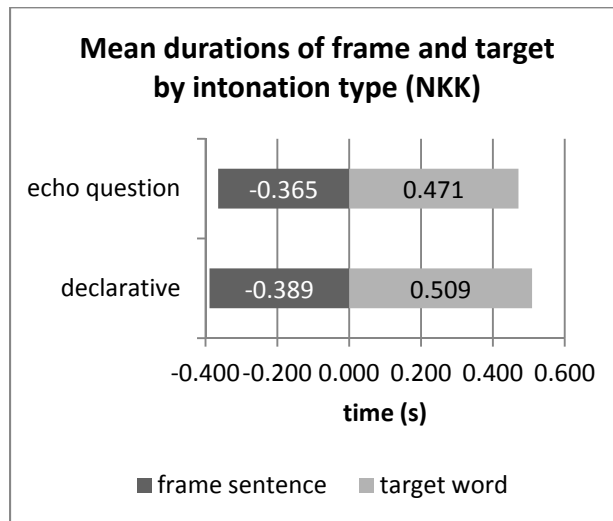


Figure 2.26: Mean durations for the frame condition in NKK.

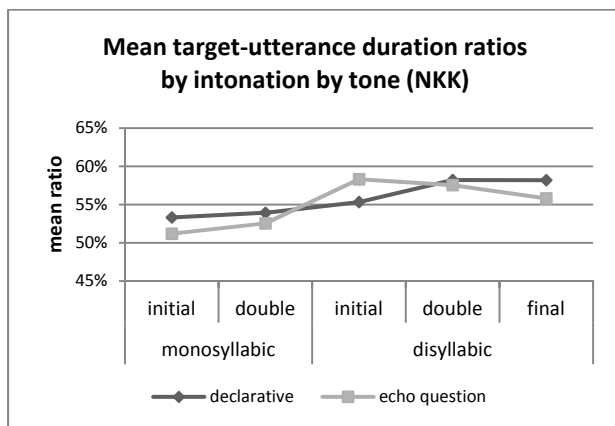


60 tokens were measured in each intonational category in each context. The target words in isolation as well as in a frame context were shorter on average in the echo question condition ( $p < .001$  in both cases). However, since—as with several of the other speakers—the mean duration of the frame was also shorter in the echo question context ( $p < .001$ ), it behooves us to look at the proportion results, shown in Table 2.9.

**Table 2.9: Target-utterance duration ratio by intonation for NKK.**

declarative %	echo question %	sig.
56.57	56.16	.462

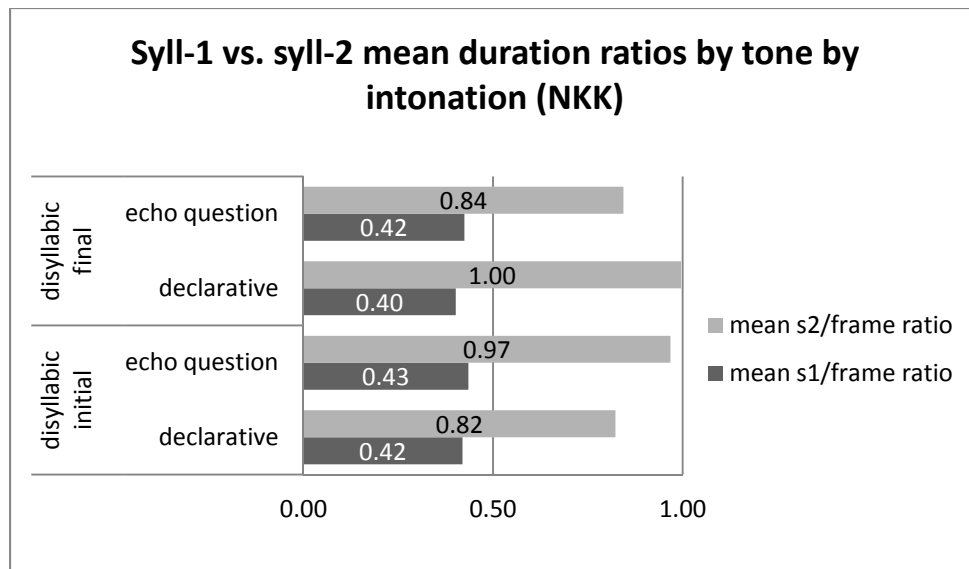
We see that the proportion of the utterance taken up by the target word was on average the same in the two intonational contexts. However, these numbers do not tell the whole story, as the trend was not the same across all tonal categories; this is apparent in Figure 2.27.



**Figure 2.27: Mean target-utterance duration ratios by intonation by tonal category (NKK).**

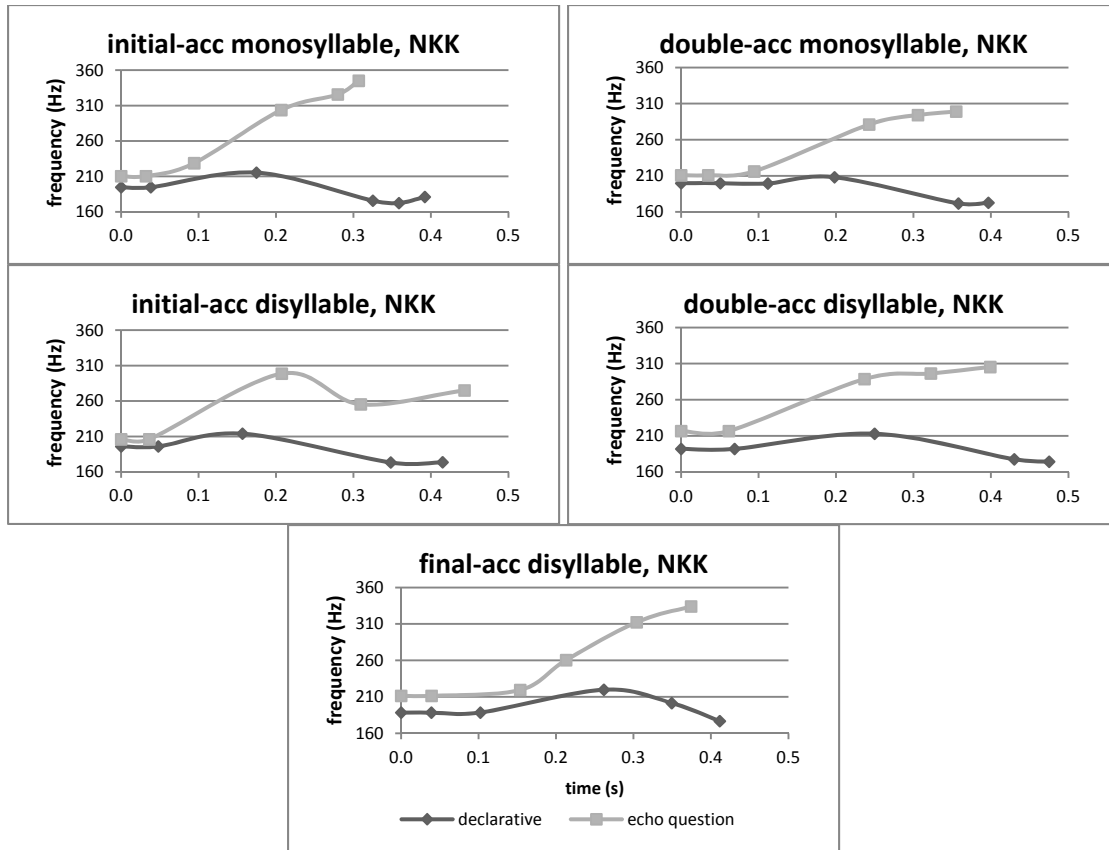
Only the differences observed for disyllabic initial-accented and disyllabic final-accented were significant ( $p = .073$  for monosyllabic initial-accented;  $p = .242$  for monosyllabic double-accented;  $p < .001$  for monosyllabic initial-accented;  $p = .329$  for disyllabic double-accented;  $p = .002$  for disyllabic final-accented). Since both of the tonal conditions that displayed an intonation-dependent duration difference were in the disyllabic category, it is worth taking a moment to break down the durational distribution by syllable. Figure 2.28 shows the mean ratios

of the durations of each of the syllables to the duration of the frame for each tonal/intonational condition. Note that, in this figure, different shaded bars represent different syllables in the target word. It is clear that virtually all of the durational differences between declarative and echo question renditions of disyllabic target words can be accounted for by differences in the second syllable. The syllable-1-to-frame ratio remains rather constant across intonational categories in both tonal categories ( $p = .610$  for initial-accented;  $p = .455$ ), while the syllable-2-to-frame ratio is greater in echo questions for initial-accented and greater in declaratives for final-accented ( $p < .001$  for both tonal categories) disyllables. Later on, a similar behavior will be observed in final syllables in Shiga Japanese disyllabic target words.



**Figure 2.28: Syll-1 vs. syll-2 mean duration ratios for disyllabic initial-accented and final-accented target words by intonation in NKK.**

Moving on now to the  $F_0$  results, the mean  $F_0$  contours for representative words from all five categories (two tones on monosyllables and three on disyllables) in isolation are shown in Figure 2.29. At first glance all of the declarative contours may look very similar. They all start in a middle range, reach a slight peak, and then fall. In this sense they are similar to T4 in Mandarin (the falling tone) and T1 for Cantonese Speaker A (the high level tone, which ended in a fall for that speaker). However, upon closer inspection we see that the different tonal



**Figure 2.29: Mean F<sub>0</sub> contours (isolation) for NKK.**

categories can be distinguished by the alignment of the peak in each case. This alignment difference has been noted previously by S.-E. Chang (2005) and Lee (2008). The peak is reached earliest in the initial-accented word, later in the double-accented word, and latest in the final-accented word. Lee (2008) argued that the peak in each tone is segmentally anchored. She finds that the peak of an initial or final accent is anchored to the boundary between the onset and nucleus of the post-accentual syllable (i.e. the C1-V1 boundary for initial-accented and the C3-V3 boundary for final-accented disyllabic words), while the double-accented peak is anchored to the interior of the nucleus of the post-accentual syllable (i.e. in the middle of V2). This is schematized in Figure 2.30, taken from Lee (2008). However, the exact generalization she made for where the peaks are anchored does not hold for the data collected here. Representative pitch contours for each of the tonal categories are shown on the left in Figure 2.31, with the boundaries of phones demarcated with vertical lines. On the right in Figure 2.31, contours from three

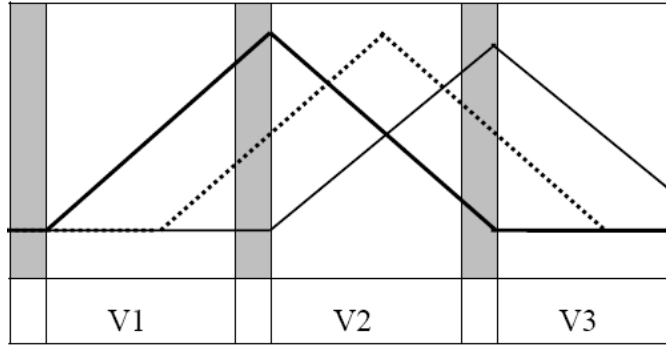


Figure 2.30: Segmental anchoring of initial-accented (thick solid), double-accented (dotted), and final-accented (thin solid) disyllables in an utterance-medial context.

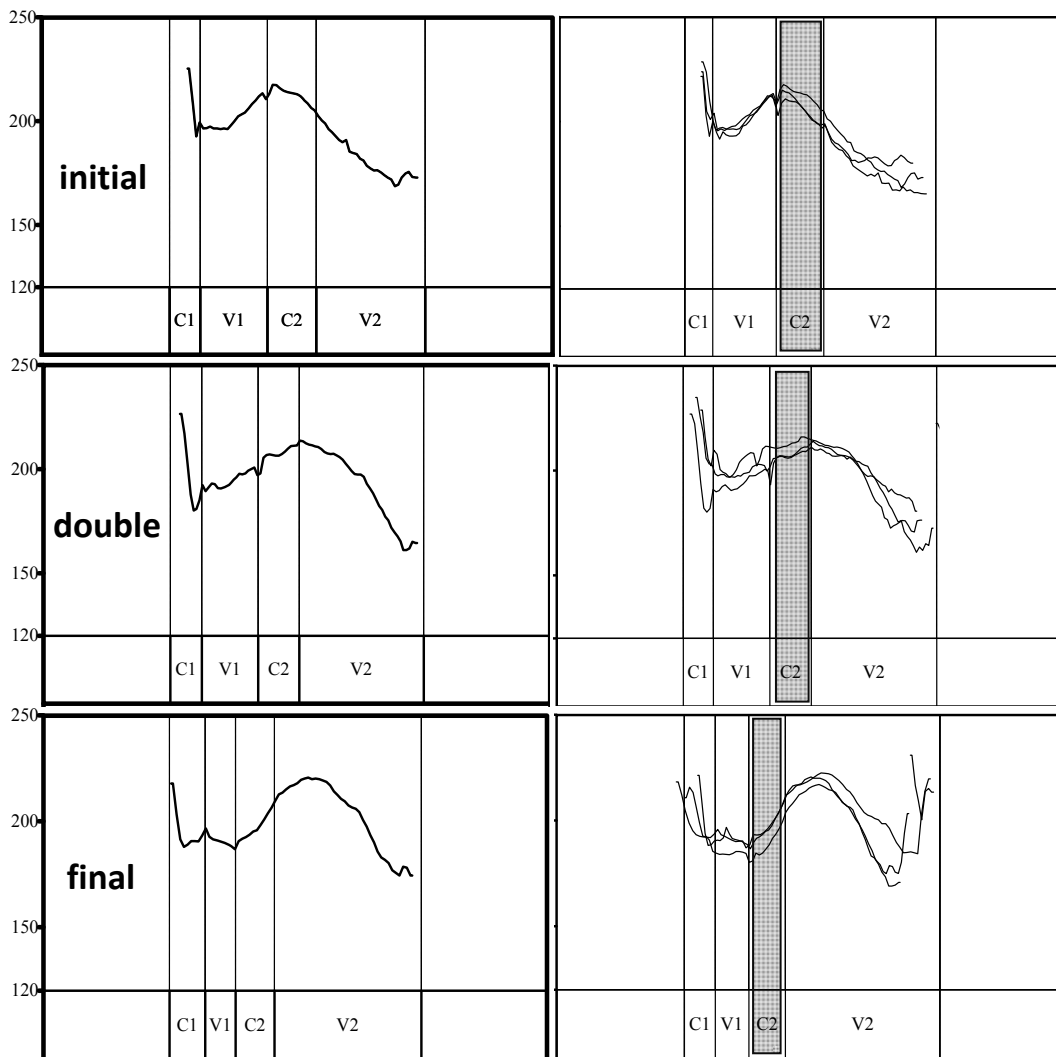
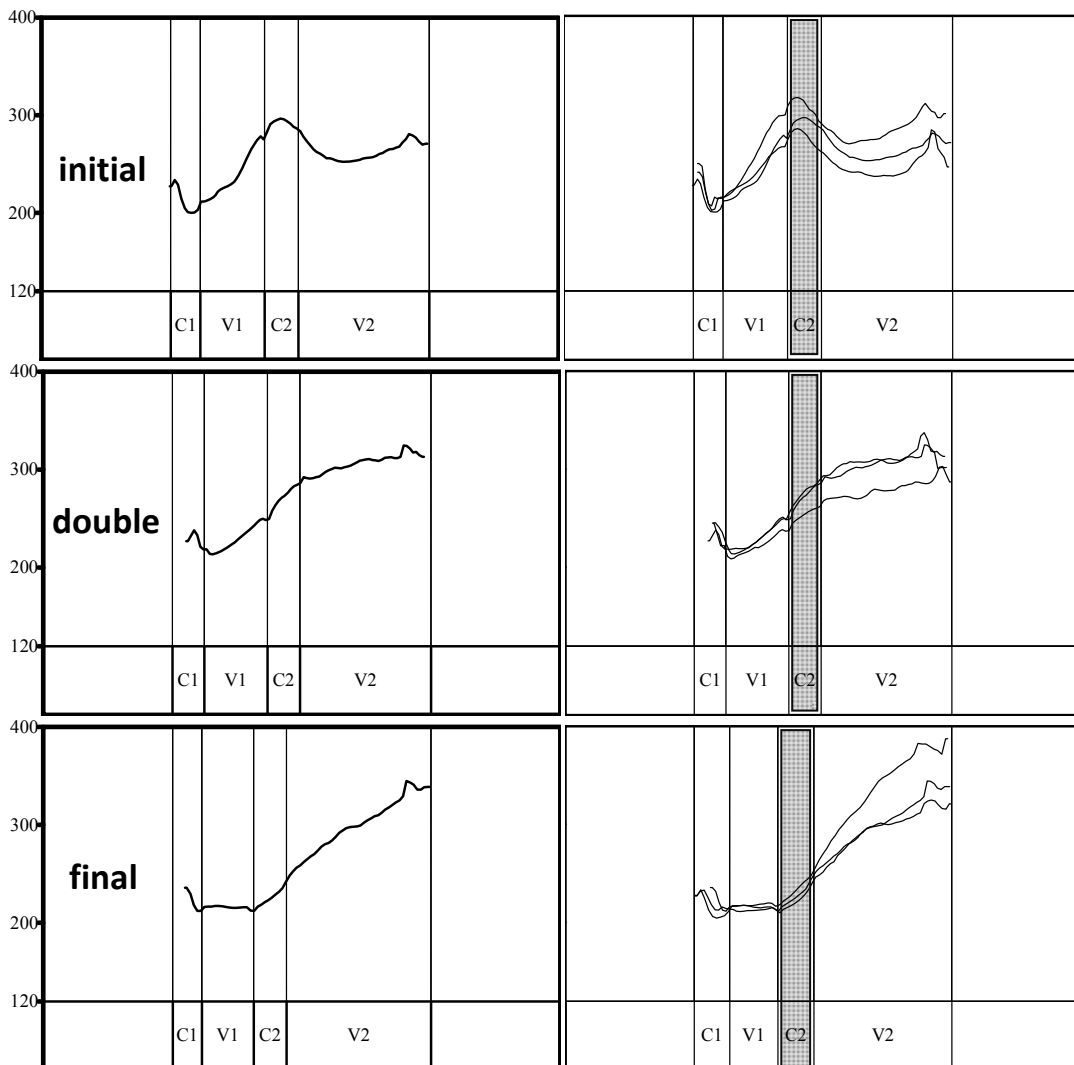


Figure 2.31: Representative F<sub>0</sub> contours for declarative renditions of initial- (*nam-i* ‘south-NOM’), double- (*nami* ‘third party-NOM’), and final-accented (*muni* ‘pattern’) disyllables in NKK.

repetitions of each word are shown, with the repetitions resized and aligned at the intervocalic consonant (C2). For the initial-accented word, the  $F_0$  rises through V1 and peaks at the boundary between V1 and C2, for the double-accented word it also rises through V1 but peaks later, at the boundary between C2 and V2. For the final-accented word the  $F_0$  stays flat through V1 and peaks partway through V2. It is also worth noting that the relative duration of V2 increases while that of V1 decreases as the peak moves from left to right. This three-way surface contrast on disyllables carries over to the echo question context; examples are shown in Figure 2.32.



**Figure 2.32: Representative  $F_0$  contours for echo question renditions of initial- (*nam-i* ‘south-NOM’), double- (*nam-i* ‘third party-NOM’), and final-accented (*muni* ‘pattern’) disyllables in NKK.**

The alignment behavior of monosyllables will be discussed in a moment, but first let us look at the same set of tonal categories in the frame context. These results, pooled and averaged, are given in Figure 2.33. The same overall patterns that were seen in the non-frame context can be seen here. The declarative contours differ with respect to alignment, and each of the echo question contours, with the exception of that for the disyllabic initial-accent, is entirely rising. Note that the question contour for the disyllabic initial-accent does not end in a rising trajectory here, as it did in the isolation condition, but rather it ends in a mid-level plateau. Lee (2008) showed that questions ending in a polysyllabic initial-accented word do not always end in a rise, but rather that they may end in a plateau or a fall depending on a variety of factors, including speaker age and sex. In addition, in a personal communication she expressed the intuition that, even for a single speaker, the level of emphasis on an echo question can be cued by the extent to

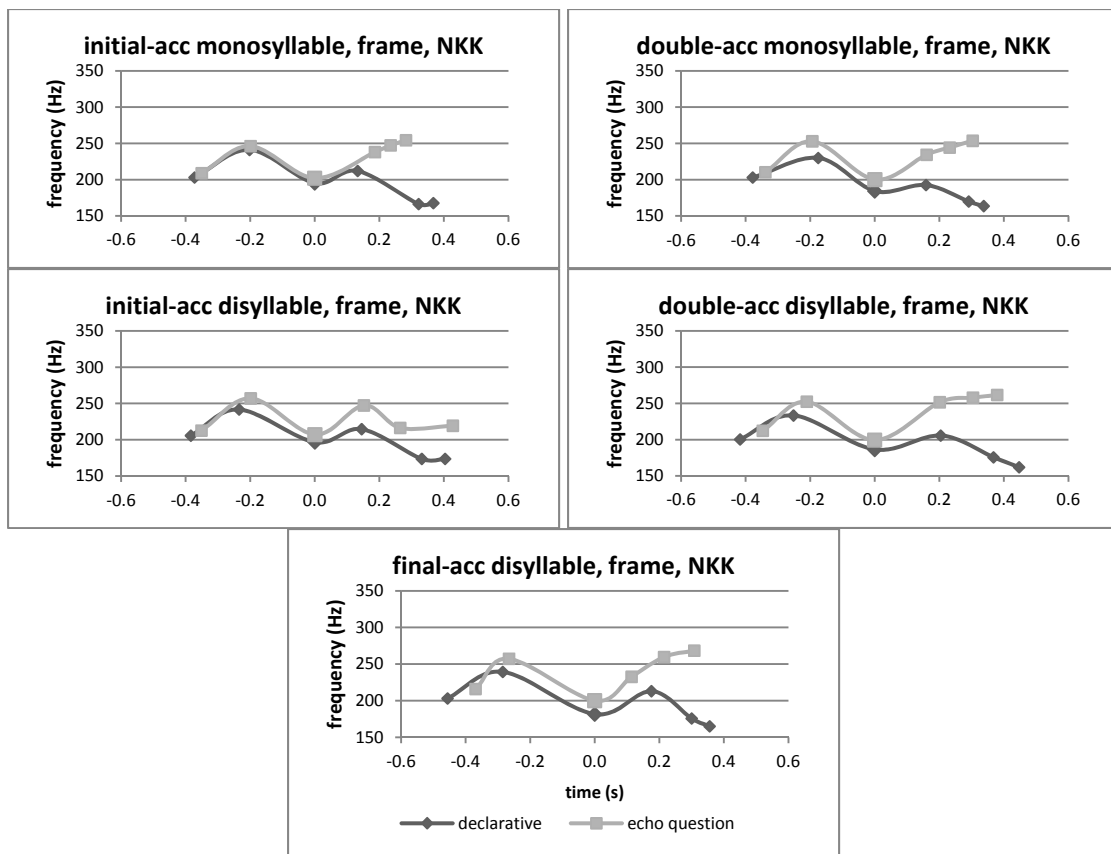
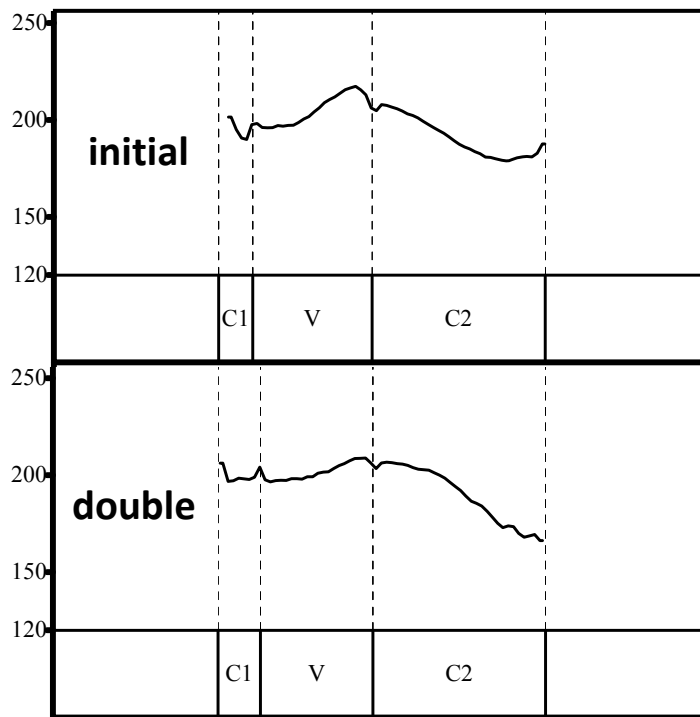


Figure 2.33: Mean  $F_0$  contours (frame) for NKK.

which the final  $F_0$  is actually rising. Crucially, though, the initial-accented disyllable in an echo question context always surfaces with a falling component. Finally, with the declarative and echo question contours superimposed, it is easier to see the general tendency for the global pitch range of echo questions to be slightly higher than that of declaratives.

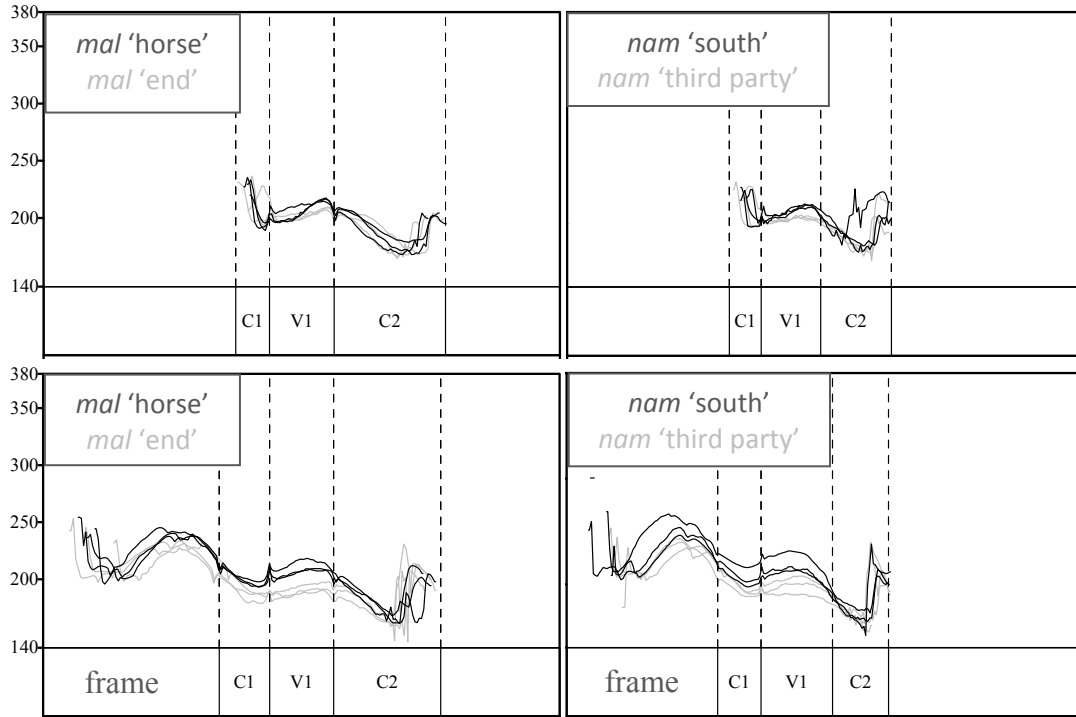
As promised, let us take some time to investigate the alignment properties of the monosyllables. Figure 2.34 shows the peak alignments of representative declarative initial-accented and double-accented monosyllables in NKK. As is apparent in the figure, there is not a



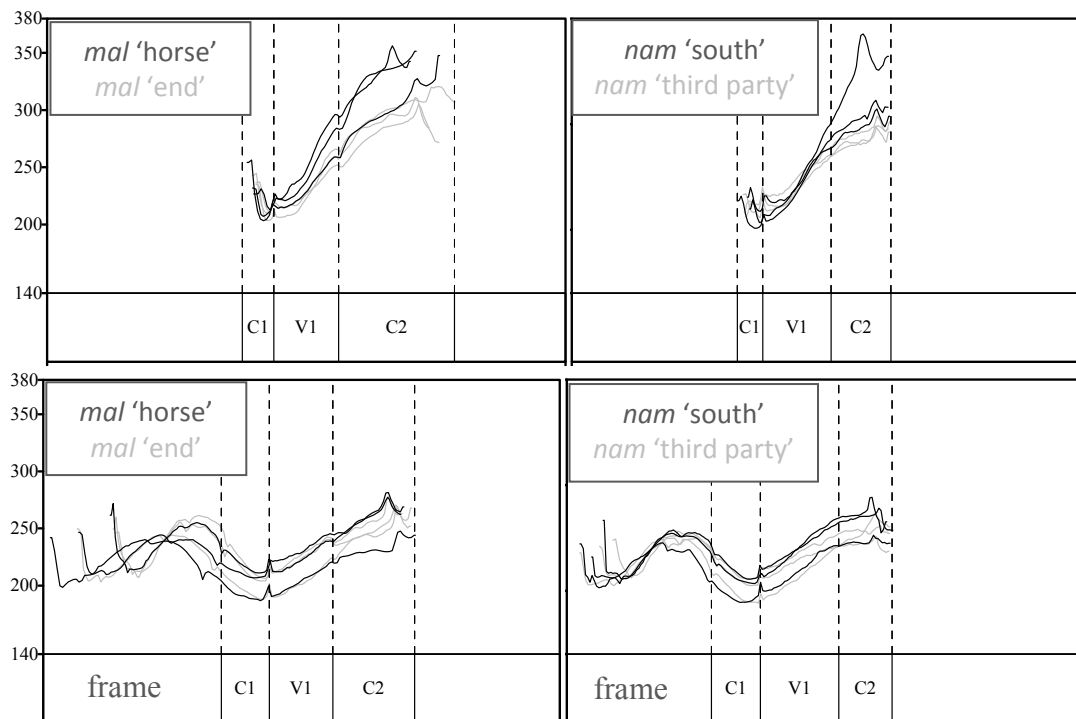
**Figure 2.34: Representative  $F_0$  contours for initial- (*nam* ‘south’) and double-accented (*nam* ‘third party’) monosyllables in NKK.**

clear-cut alignment difference when it comes to monosyllables<sup>15</sup>. In both tones the peak seems to occur at the V1-C2 boundary. Looking at these single tokens, it is possible to make the argument that the peak is higher for the initial- than for the double-accent, and that the fall after the peak is immediate for the initial- and delayed until partway through C2 for the double-accent.

<sup>15</sup> We know these words contrast in tone because of the alignment differences we see when particles are added to them, making them disyllabic.



**Figure 2.35: Multiple repetitions of initial- (dark gray) and double-accented (light gray) monosyllables in the declarative context, separated out by context and minimal pair.**

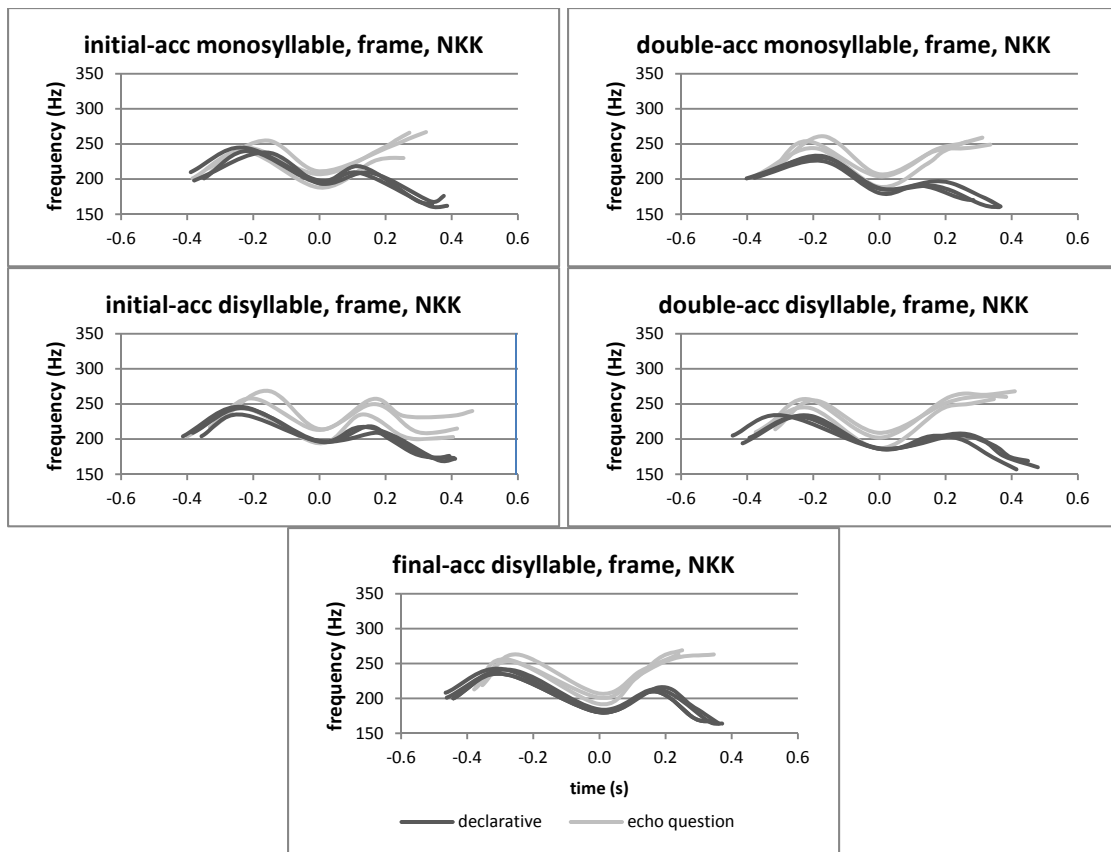


**Figure 2.36: Multiple repetitions of initial- (dark gray) and double-accented (light gray) monosyllables in the echo question context.**



To see if this is a general tendency, multiple repetitions of initial-accented and double-accented words in a declarative context and in an echo question context have been time-normalized and plotted on the same scale in Figures 2.35 and 2.36, respectively. It is difficult to say from these limited numbers of repetitions whether these generalizations can be made; as such, a more careful study regarding this issue is warranted.

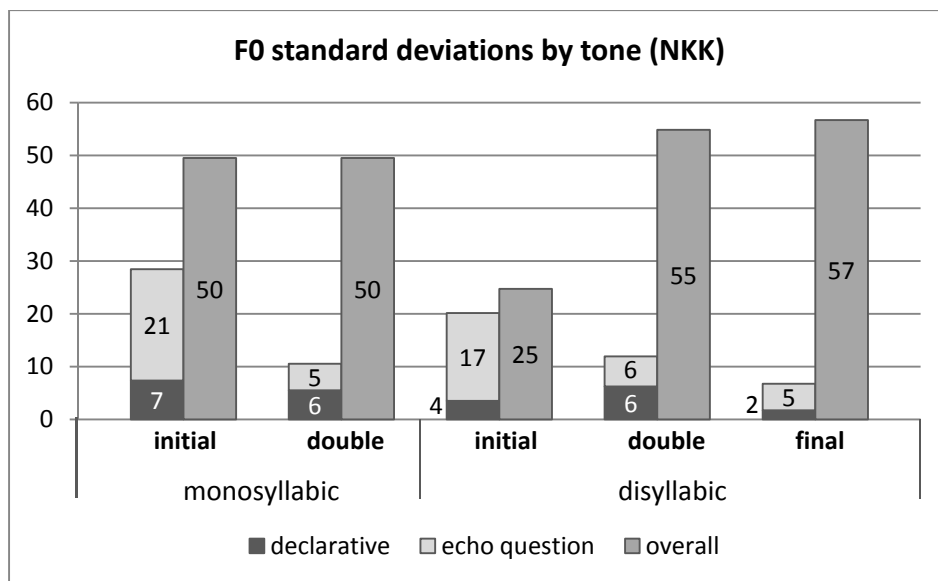
Finally, although only three repetitions for each tone-intonation combination were recorded, it is worth looking at the multiple-contour plots for each tonal category, with the declarative and echo question contours plotted on the same graphs. These are shown (for the frame conditions) in Figure 2.37. It is clear that the declarative-echo question difference is quite



**Figure 2.37: Multiple-contour plots for NKK. Declarative contours are in light gray and echo question contours are in dark gray.**

distinct in all cases, with the possible exception of the initial-accented disyllabic case. In that condition there was a wider range of final  $F_0$  values in the echo question context. An

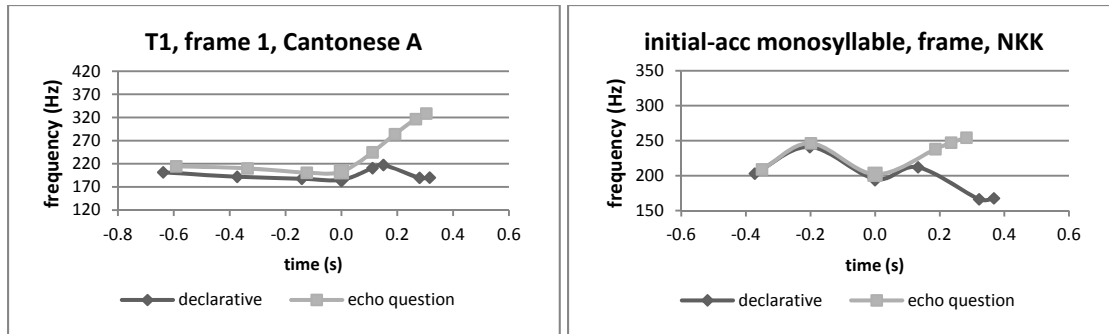
examination of the  $F_0$  standard deviation graph for NKK, shown in Figure 2.38, confirms that the initial-accented disyllabic category indeed has the lowest degree of separation among all of the tones.



**Figure 2.38: Standard deviations of declarative echo question, and overall  $F_0$ , by tone, in NKK.**

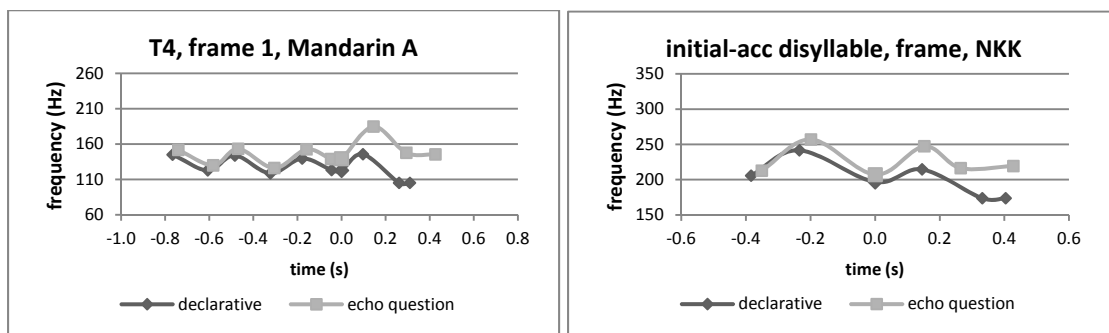
### 2.5.6 Discussion

The realization of echo question intonation relative to declarative intonation in NKK can be broadly characterized as a global upward shift in the pitch range as well as a raising of the final pitch target. In disyllables, the relative pitch height of this target increases the closer the peak in the corresponding declarative disyllable gets to the right edge of the second syllable. For monosyllabic words and for double-accented and final-accented disyllabic words, the resulting contour on the pitch is rising throughout the target word (though it tends to level out on the second syllable of double-accented disyllabic words). In this respect these tones behave similarly to T1 in Cantonese. A NKK monosyllabic initial-accent and a Cantonese T1 are shown side by side in Figure 2.39 for comparison. For initial-accented disyllables, the  $F_0$  falls but does



**Figure 2.39: Mean  $F_0$  contours for T1 in Cantonese (left) and initial-accented monosyllable in NKK (right).**

not fall as low as that of the declarative, and the contour flattens out at the end of the second syllable. In this respect they behave similarly to T4 in Mandarin A’s speech. They are shown side by side in Figure 2.40 for comparison.



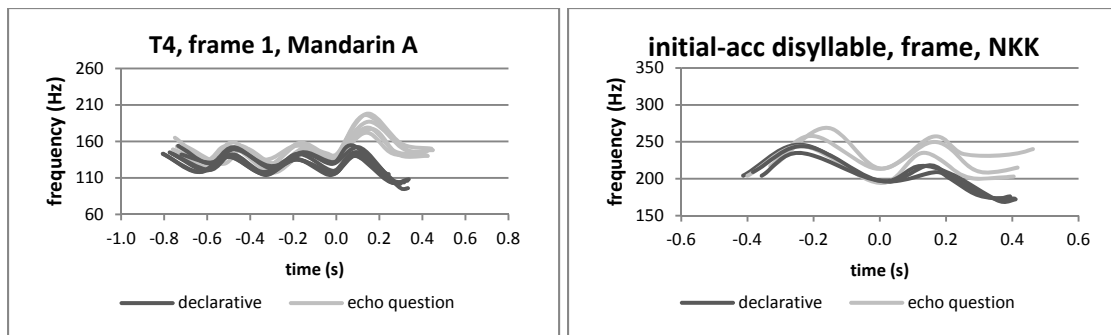
**Figure 2.40: Mean  $F_0$  contours for T4 in Mandarin (left) and initial-accented disyllable in NKK (right).**

As for durational effects of echo question intonation, the current results show that only the initial-accented disyllable is longer in the echo question context than in the declarative context. Although further study is needed to determine whether this apparent tone-specific exception is robust, it is interesting to note that this durational effect was observed in both the isolation and frame contexts.

Ostensibly, none of the echo question realizations in NKK resembles any of those for the two falling tones in Henanhua, despite the similarities between the double-final contrast in NKK and the T2-T4 contrast in Henanhua in the declarative context, contrasts which manifest themselves as peak alignment differences. With these results from NKK the overall typological

picture of melodic systems grows more complex, and we are further assured that the answer to Q1 (*Language-General Tone-Intonation Interaction*) at the beginning of this chapter is ‘no’; that is, utterance-level intonation does not interact with lexical tone in the same way in all languages.

Finally, it is worth noting that, once again, the degree of distinctiveness may not be consistent across all the tonal categories in NKK, as evidenced in Figure 2.37, which displayed multiple-contour plots for all the tones. In displaying less robust differences than the other conditions, the initial-accented disyllable seems to have the same status in the NKK system as T1 (high-level) in Mandarin and T2 (high-rising) in Cantonese. What we are starting to see, then, is that this dimension of “ $F_0$  range distinctiveness” seems to be independent of other factors, i.e. certain types of contour shapes do not seem to display inherent degrees of  $F_0$ -range distinctiveness. We saw this when Mandarin T1 and Cantonese T1 were compared in the last section, but to drive the point home we can compare the multiple-contour plots of Mandarin T4 and NKK initial-accented disyllables (both “falling” tones) to see how similarly shaped tones behave differently with respect to  $F_0$  range distinctiveness in different languages; this comparison is displayed in Figure 2.41. As a further reminder, the standard deviation



**Figure 2.41: Multiple-contour plots for T4 in Mandarin (left) and initial-accented disyllables in NKK (right).**

comparisons for Mandarin T4 (sum of intonational group standard deviations vs. overall standard deviations) were 12:23 and 46:90, respectively for Speaker A and Speaker B, while they were 21:25 for the NKK speaker. We will see in the next chapter that T4 in Mandarin elicited the highest rate of intonational accuracy among all the tones and the initial-accented disyllable

elicited one of the lowest in NKK. While these results need to be confirmed with a more comprehensive acoustic study, their implications are intriguing in that they suggest a dimension of language-specific phonetic implementation.

Having looked at analogous data from several different languages by now, we are becoming privy to the ways in which melodic systems may vary. In the next section we will examine data from one more language, Shiga Japanese (as well as related dialects), to help crystallize this emerging typological picture from the perspective of production.

## **2.6 Production in Shiga Japanese and Other Kansai Dialects**

This section presents results from a series of experiments investigating a family of dialects spoken in the *Kansai* region of Japan, in the western part of the main island of Honshu. This family of dialects is known collectively as *Kansai-ben*, and it includes dialects spoken in Osaka, Kyoto, Kobe, Nara, and other cities in the region, as well as in Shiga prefecture. Although non-standard, *Kansai-ben* enjoys a measure of prestige not afforded other non-standard Japanese dialects, since many comedians and other pop-culture celebrities who speak the dialect have gained exposure through the media. In the literature, Osaka Japanese is one of the most commonly discussed non-standard Japanese dialects (Kori 1987; Pierrehumbert and Beckman 1988; Haraguchi 1999). For this reason, the less-studied Shiga Japanese will be the main focus of the current study. Kyoto Japanese and Shiga Japanese are quite similar due to the proximity of Kyoto to Shiga, and for the purposes of the current study they will sometimes be lumped together as the Shiga-Kyoto dialect group.

### **2.6.1 Overview of the Kansai tonal system**

Like the standard Tokyo dialect, Kansai dialects are usually described as pitch-accent languages; all words are lexically marked as being “accented” or “unaccented”, and in all nouns the locus of the accent is also lexically specified. What makes the Kansai tonal system one step more complex is that words are also lexically specified as starting “high” or “low”, i.e. there is a

register contrast at the beginnings of words. Haraguchi (1999) gave the high-beginning unaccented class, the low-beginning unaccented class, the high-beginning accented class, and the low-beginning accented class the labels HH, LH, HL, and LHL, respectively. This follows the tradition in the Japanese phonology literature of using under- and over-lines to delineate accentual patterns—a four-mora word like *nokogiri* (‘saw’), which is low-beginning and bears an accent on its third mora, would be annotated for tone as shown in (2.17) (Kori 1987):

(2.17) no ko<sup>̄</sup>gi<sup>̄</sup>ri

Haraguchi (1999) considered words like this to be specified for a LLHL pattern. According to this tradition, high-unaccented words consist of all high-toned morae, low-unaccented words consist of all low-toned morae plus a final high-toned mora, high-accented words start high and stay high until the drop in pitch associated with the pitch accent, and low-accented words start low and stay low until the accented mora, which is high and then is followed by the drop in pitch associated with the accent. These labels are given in Table 2.10 with example tone specifications in Haraguchi-style notation.

**Table 2.10: Haraguchi-style labels and example sequences for the four tone classes in Osaka Japanese.**

tone class	class label	example tone sequence
high-unaccented	HH	HHHH
low-unaccented	LH	LLLH
high-accented	HL	HHLL
low-accented	LHL	LLHL

There is no contrast between high and low for initial-accented words, but Haraguchi (1999) considered them to be high-accented based on their morpho-phonological behavior<sup>16</sup>. There is also a gap whereby there are no high-accented words in which the accent falls on the last syllable.

<sup>16</sup> He reported that the genitive particle *-no* triggers deaccenting on high-accented nouns and that this deaccenting applies to initial-accented nouns.

In the current study, the results will mainly be limited to bimoraic words, so the possible tonal patterns will be *high-unaccented* (henceforth *H-unaccented*), *low-unaccented* (henceforth *L-unaccented*), *high-initial-accented* (henceforth *H-initial-accented*), and *low-final-accented* (henceforth *L-final-accented*).

### 2.6.2 Subjects

Twelve speakers of Kansai Japanese were recorded—five male and seven female, ranging in age from 22 to 72. Their geographical backgrounds were as listed in (2.18):

(2.18) Kansai Japanese speaker breakdown

4 (Subjects C, I, J, and K) were born and raised in Shiga Prefecture

1 (Subject G) was born in Nara and went to Kyoto for college before settling in Shiga for 24 years

1 (Subject H) was born and raised in Kyoto

6 (Subjects L, M, P, U, Y, and Z) were born and raised in the Osaka-Kobe area

The female subject from Nara (Subject G) seemed to have a melodic system typical of Shiga, despite the fact that she grew up only an hour away from Osaka. There were a handful of words that she put in a different tonal class from most of the other speakers from Shiga (but this was true of a couple of those other speakers as well; only words whose tonal category was unanimously agreed upon were included in the analysis of the pooled data for the Shiga group).

### 2.6.3 Materials

Recordings were conducted either in a sound-attenuated recording studio or in a quiet room, and they were made using a Plantronics DSP-500 headset microphone connected to a laptop computer. The four conditions given in (2.19) were included:

(2.19) Kansai Japanese production experiment conditions

59 target words

4 contexts

2 repetitions

The words were all nouns (consisting solely of sonorant segments when possible), but they varied in mora count (two to six morae) and tonal category, crucially including final-accented words. See the Appendix for the full word list. The four contexts in which the target words were recorded are given in (2.20):

(2.20) Kansai Japanese recording contexts

*X*. (isolation, declarative)

*Sore-wa X-yawa*. ‘That is X.’

*Sore-wa X-no hanasi-yawa*. ‘That is a discussion about X.’

*X?* (isolation, echo question)

The second frame sentence, in which the target word was followed by the genitive particle *-no*, was included because *-no* is reported in the Kansai Japanese literature as triggering tonal-category-dependent accent deletion—that is, it is said to change H-accented words into H-unaccented words (Okuda 1971; Haraguchi 1977). The order of the words was randomized for each speaker. A filler word was included at the beginning and at the end of the word list to avoid irregular intonation (although this turned out to be unnecessary as the filler words did not receive irregular intonation for any of the subjects). The four contexts were presented to the speakers prior to starting the recording so that they could memorize them and get comfortable with them, but the contexts were also printed out on a separate card that was kept off to the side as a reminder. Sessions lasted about 30 minutes.

Because the experiments done in Shiga were part of a larger field research study aimed at collecting data on lexical tone in this lesser-studied dialect, there was a much longer word list



and more contexts included in the experimental design than for the other languages. As such, the additional declarative and echo question frame sentence contexts in which the target word was utterance-final were not included. However, one subject (Subject C) was recorded saying both sides of the dialogue<sup>17</sup> given in (2.21), with two of the target words:

(2.21) Follow-up recording contexts for Shiga Speaker C

*Naoya-ga X-yade.* ‘Naoya is X (you know).’

*Naoya-ga X?* ‘Naoya is X?’

The sentence particle *-yade* was used in the declarative frame because the subject claimed he could not naturally end the declarative sentence with the object (Japanese is an SOV language)<sup>18</sup>. The target words plugged into this dialogue were the low-final-accented *ame* ‘rain’ and the high-unaccented *ame* ‘candy’. Three repetitions of the dialogue for each word were recorded. The reason for this brief follow-up recording was to determine if there was any indication of an effect of echo question intonation on the earlier part of the frame sentence for Speaker C.

#### **2.6.4 Comparison with other experiments**

Like in the NKK experiment, no full minimal sets (of words comprised of sonorant segments) cutting across all the tonal categories were available, so two-way comparisons of different combinations of tonal contrasts had to suffice. Also, target words were uttered in isolation and in a utterance-medial position, but not in an utterance-final one. Finally, because of the larger number of subjects, the longer word list, and the limited amount of time, only two repetitions of each condition were recorded, as opposed to six repetitions for Mandarin and Cantonese and three repetitions for NKK.

---

<sup>17</sup> The context given was that a game was being played in which players were assigned certain words, and the declarative utterance was to inform the listener what word Naoya had been assigned.

<sup>18</sup> This is in contrast with the NKK speaker, who felt comfortable with such constructions in a casual conversational context.

## 2.6.5 Results

Because the Shiga dialect is the least studied among the Japanese dialects that were examined, and because the Shiga melodic system provides the most relevant typological material for developing a model of speech melody, the results from that dialect will be the main focus of this section. Selected results from the other Japanese dialects will be presented only when relevant. When results from only Shiga Japanese are being discussed, the name of the dialect will be abbreviated to SJ. When other dialects are being discussed, they will be referred to by their geographical locations. Since the interaction of utterance intonation and tone is the focus of the current study, only results from the first and fourth recording contexts (isolation declarative and isolation echo question) will be discussed<sup>19</sup>. Starting once again with duration, the duration-by-intonation results for the six Shiga-Kyoto speakers are shown in Figure 2.42.

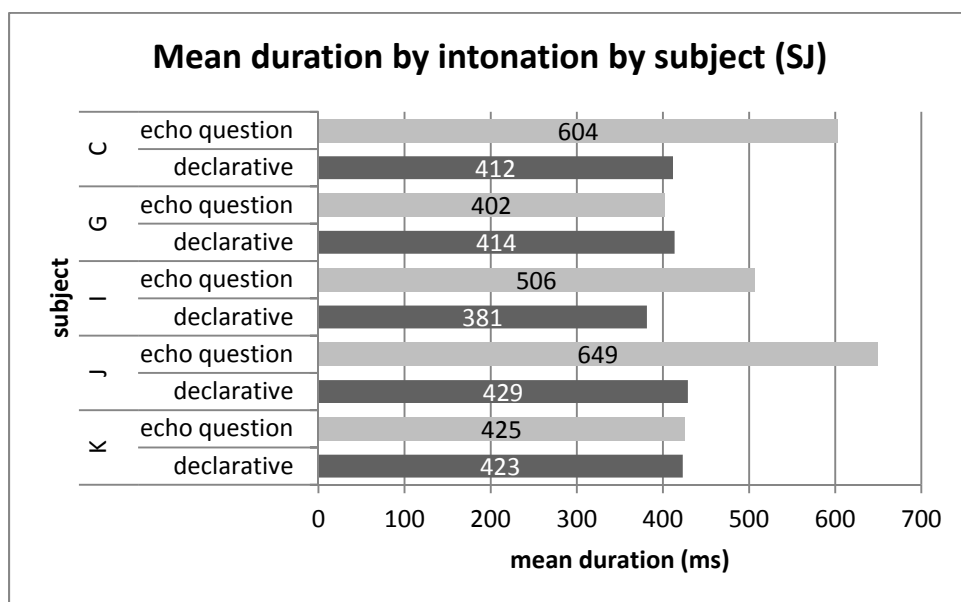


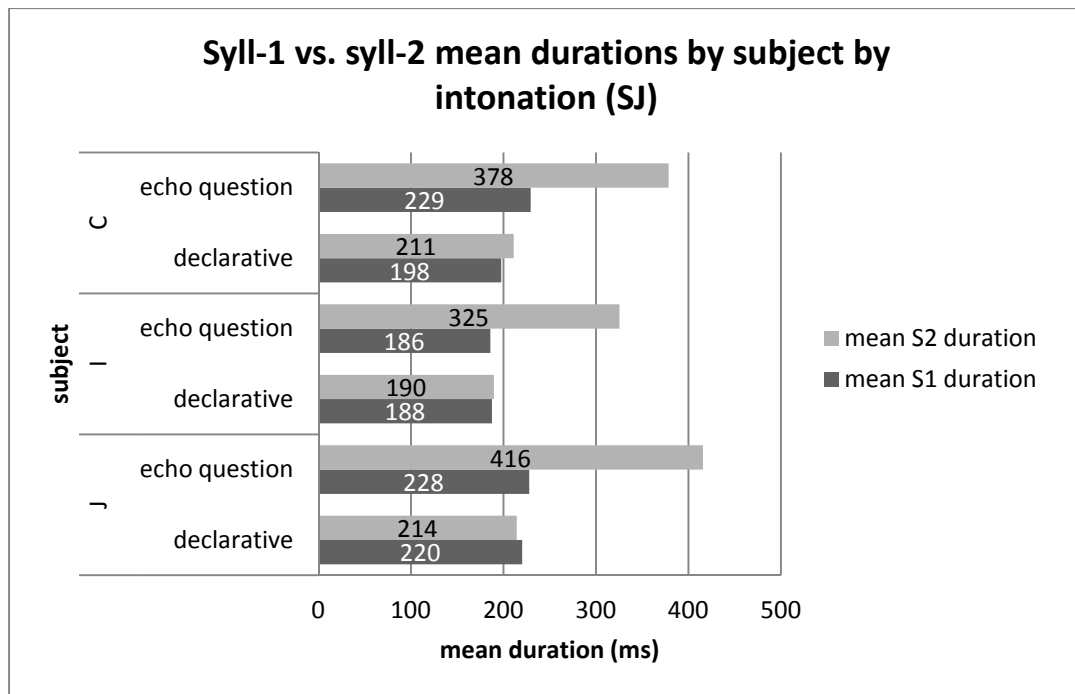
Figure 2.42: Mean duration of target word by intonation type for SJ speakers.

112 tokens were measured in each intonational category for each speaker. It is clear that there was a wide range of duration effects. The basic generalization that can be made is that echo

<sup>19</sup> As far as the second and third recording contexts are concerned, only a subset of speakers ever displayed the *no*-triggered accent deletion on H-accented words, and among those who displayed it, it appeared to be optional since no one displayed the accent deletion consistently with every H-accented word or even with every repetition of a given H-accented word.

question utterances were either about the same or longer in duration as compared to declarative utterances. The lengthening for echo questions was found to be significant for Subjects I and J, with Subject C's results just shy of being significant (Subject C:  $p = .061$ ; Subject G:  $p = .865$ ; Subject I:  $p = .000$ ; Subject J:  $p = .026$ ; Subject K:  $p = .749$ ).

For the subjects that showed consistent durational differences across intonational categories, the durational differences can all be accounted for by differences in the second syllable, just like in NKK. Figure 2.43 shows the durational difference between corresponding syllables in the two intonational contexts for the relevant subjects (Subject C is included because of the borderline significance results for that subject).



**Figure 2.43: Syll-1 vs. syll-2 mean durations for SJ Subjects C, I, and J by intonation.**

The mean duration of Syllable 1 remains relatively constant across intonational categories for all three speakers ( $p = .168$  for Subject C;  $p = .932$  for Subject I;  $p = .752$  for Subject J), while the mean duration of Syllable 2 is greater in the echo question context for all three speakers ( $p < .001$  across the board). This durational difference manifesting itself on the final syllable is

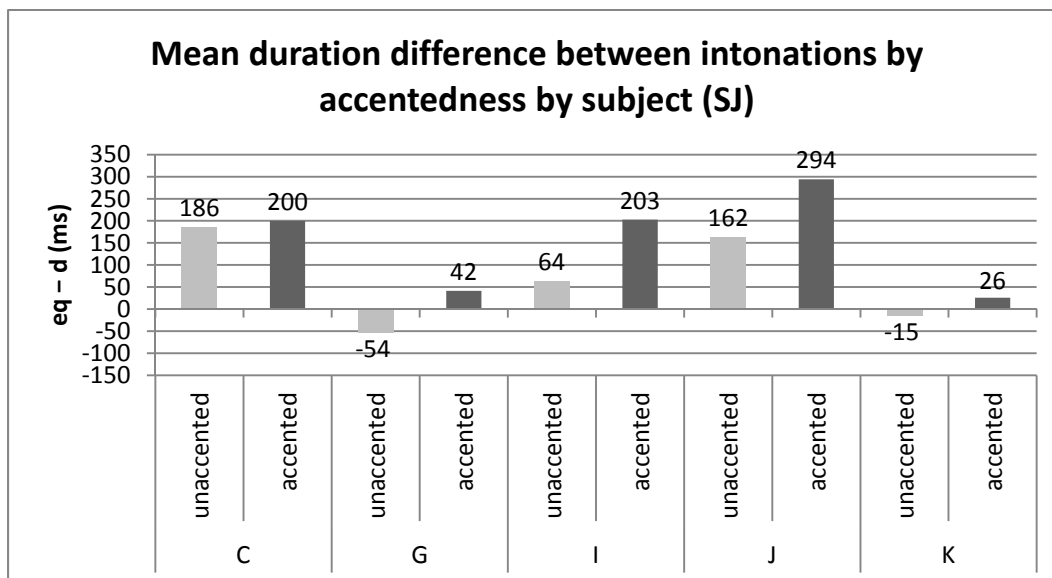
also apparent in the  $F_0$  contour plots that will be shown later, in which the  $F_0$  contours are time-aligned at the end of the onset of the second syllable.

When we break down the duration results along tonal dimensions, we can see that there was consistently a bigger difference between the two intonational categories when the target word was accented as opposed to unaccented. These pooled results are shown in Table 2.11.

**Table 2.11: Mean durational difference between intonations by accentedness for SJ.**

accentedness	mean $\text{dur}_{\text{EQ}} - \text{dur}_{\text{D}}$ (ms)
unaccented	69
accented	154

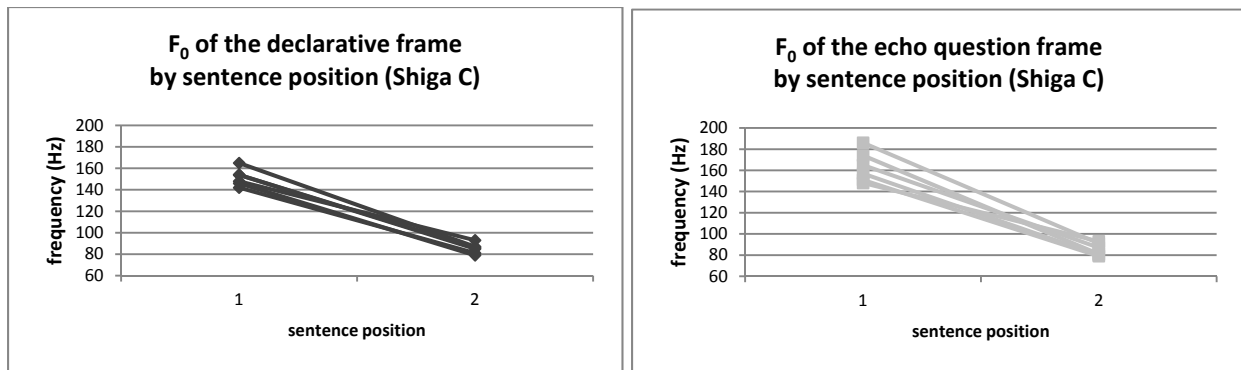
Even for the subjects for whom the absolute duration effect of longer echo questions did not hold (see Figure 2.42), this accentedness-dependent duration *difference* effect was fairly consistent. This can be seen in Figure 2.44, where negative bars for Subject G and Subject K indicate that echo questions were actually shorter on average than their declarative counterparts for those speakers. Note that in both cases the negative bars are in the *unaccented* category and that the positive bars are in the *accented* category.



**Figure 2.44: Duration difference between intonational types by accentedness by subject for SJ.**

The unaccented-vs.-accented difference was highly significant for all subjects other than Subject C ( $p = .503$  for Subject C;  $p < .001$  for all other subjects).

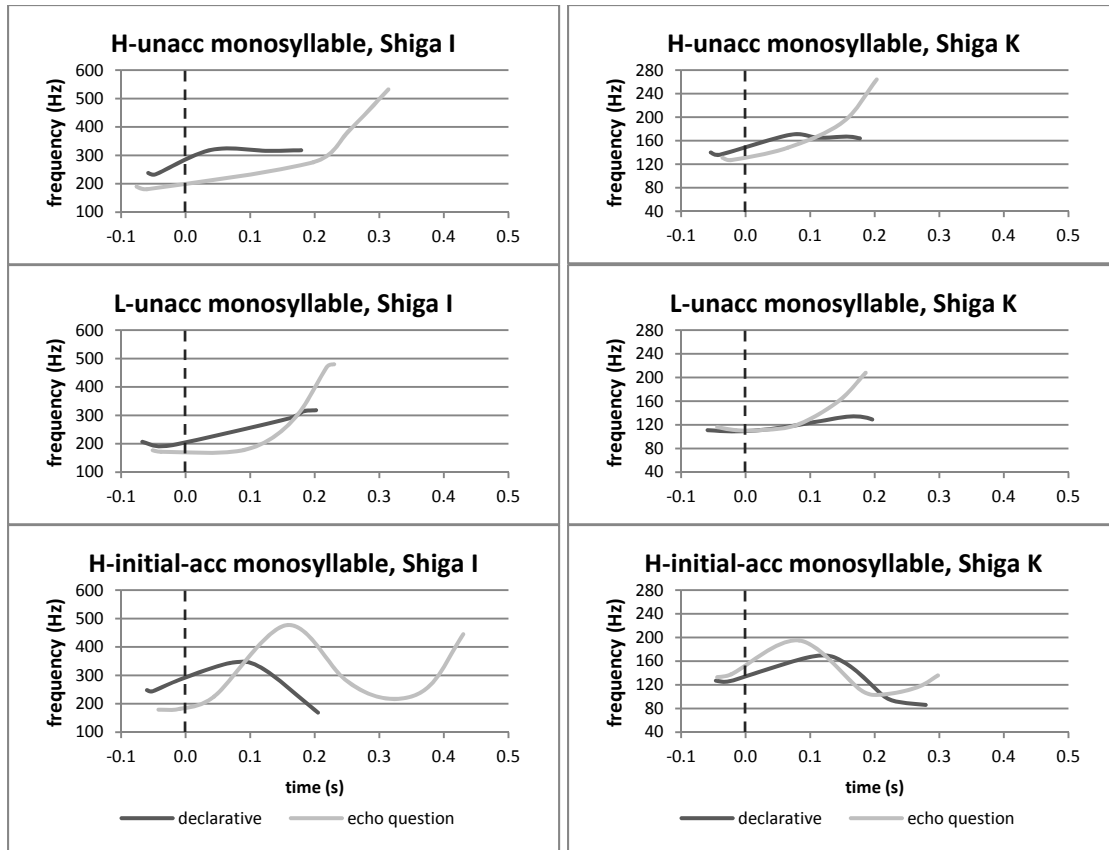
Before delving into the main  $F_0$  results for the target words, let us briefly examine the results for the declarative and echo question frame sentences obtained only for Speaker C (*Naoya-ga X-yade* and *Naoya-ga X?*). Measurements were taken at the first maximum (on the accented initial syllable of *Naoya*) and at the following local minimum (on the particle *-ga*) before the target word in each utterance. The  $F_0$  values at each of these measurement points are shown in Figure 2.45, with the results for each intonation type plotted separately. It is clear that



**Figure 2.45:  $F_0$  at two sentence positions in the frame for Shiga C (declarative on the left, echo question on the right).**

the minima on the second sentence position were unaffected by intonation type. The maxima on the first sentence position, on the other hand, displayed a higher mean and median and a wider range. Obviously a more comprehensive study is warranted, but it seems reasonable to draw the tentative conclusion that there is a tendency for local maxima in an echo question to be higher than those in a declarative utterance.

While there was some variation among the SJ speakers with respect to duration and  $F_0$  manipulation across the different tonal and intonational conditions, some general patterns are readily apparent. Stylized pitch contours for representative tokens of each of the tonal conditions on monosyllables are shown in Figure 2.46. Since words may contrast in their initial pitch height as well as accentedness (the presence or absence of a pitch accent), and since neither of these

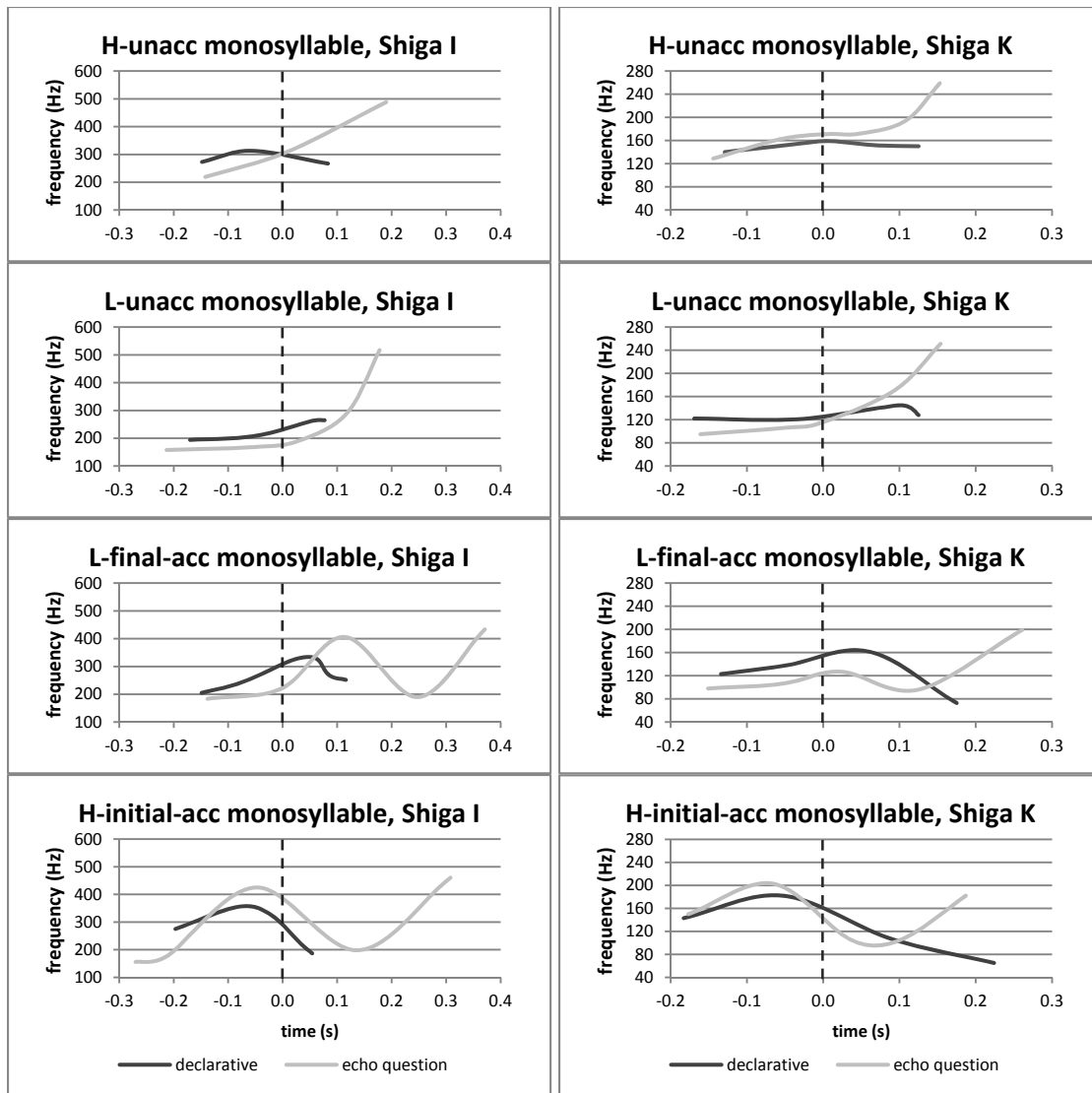


**Figure 2.46: Three-way tonal contrast on monosyllables in SJ.**

parameters is dependent on syllable alignment, there is the possibility for a three-way<sup>20</sup> contrast on monosyllables. From these initial results we can see that the echo question contours all end in a rising “tail”.

Representative contours on disyllables are shown in Figure 2.47 for Subject K (male) and Subject I (female). In each plot, the contours are time-aligned such that the end of the nasal onset is located at 0 seconds. At this point, based on the available results, several generalizations can be made. First of all, declarative contours may end in a salient fall or a quasi-level trajectory, depending on the accentedness of the tonal category. Among the accented categories, the alignment of the peak serves to distinguish between categories (as in NKK). As for the echo

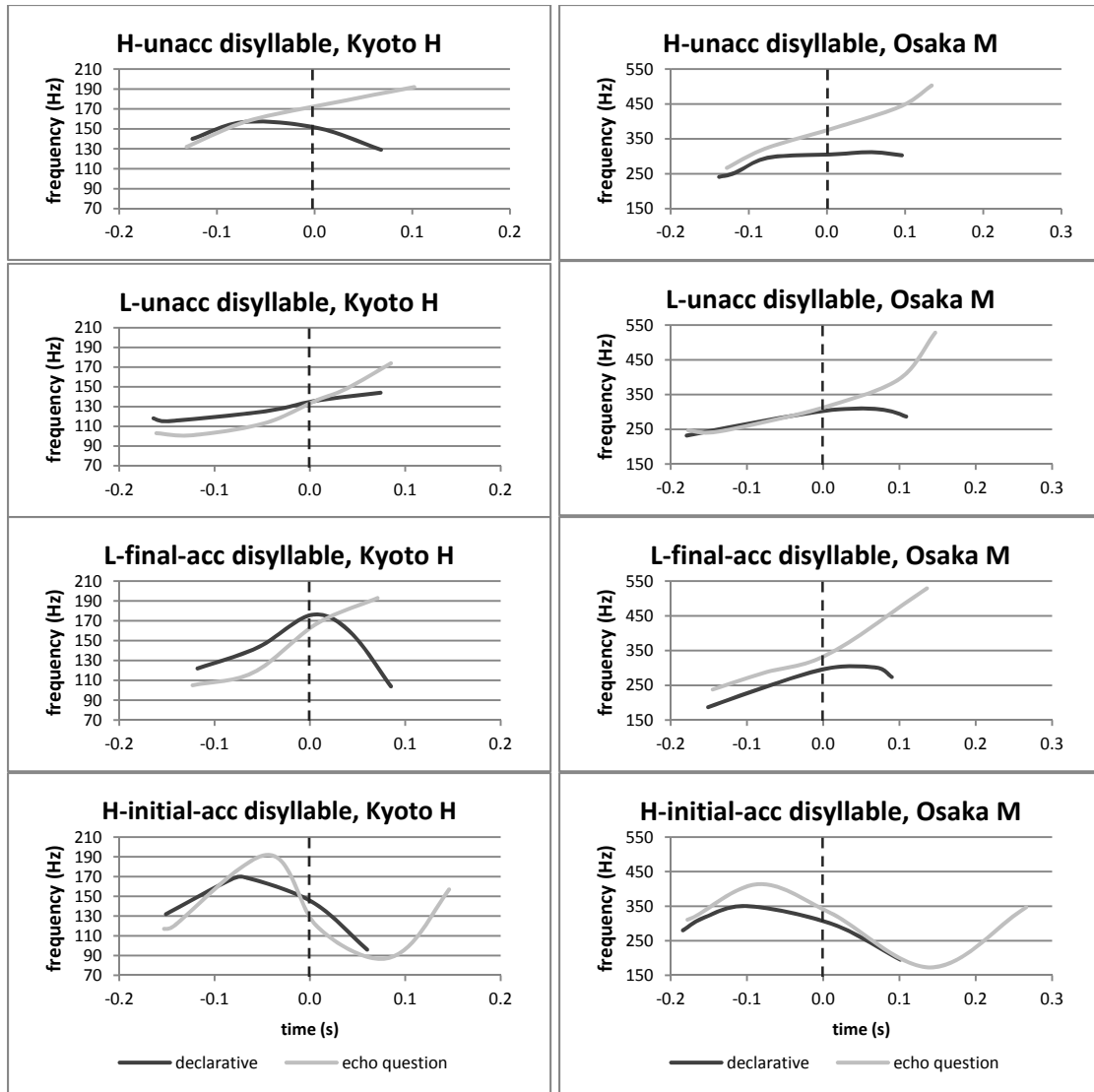
<sup>20</sup> Actually there are four logical possibilities on monosyllables, but—as discussed in Section 2.6.1—there is either a gap or a merger such that there are no low-beginning initial-accented words.



**Figure 2.47: F<sub>0</sub> contours for representative disyllabic tokens in SJ. Dotted vertical lines indicate the onset-nucleus boundary of the second syllable.**

questions, as mentioned above, most durational differences across intonational categories can be accounted for by differences in the second syllable. Secondly, in every case the echo question contour ends with a clearly rising trajectory. In the case of the unaccented tonal categories, in which the declarative contour is close to level by the end of the second syllable, the echo question contour rises throughout the second syllable. In the case of initial- and final-accented words, the echo question contour falls, hits a minimum, and then rises after that. These generalizations held across all four native Shiga speakers (C, I, J, K) as well as the subject born in Nara (G). Let us look briefly at the contours for the Kyoto speaker (H) and one of the Osaka

speakers (P) for comparison with each other and with the Shiga speakers. These results are shown in Figure 2.48. In general, the contours for these speakers look quite similar to those of



**Figure 2.48: F<sub>0</sub> contours for representative disyllabic tokens in Kyoto (left) and Osaka (right) Japanese. Dotted lines indicate the onset-nucleus boundary of the second syllable.**

the Shiga speakers. The exceptional tonal category that sets them apart is the L-final-accented category. For the Kyoto speaker, while there is a clear fall on the second syllable in the declarative context, this fall disappears in the echo question context and the contour has a rising trajectory throughout the second syllable in that context. For the Osaka speaker, the fall

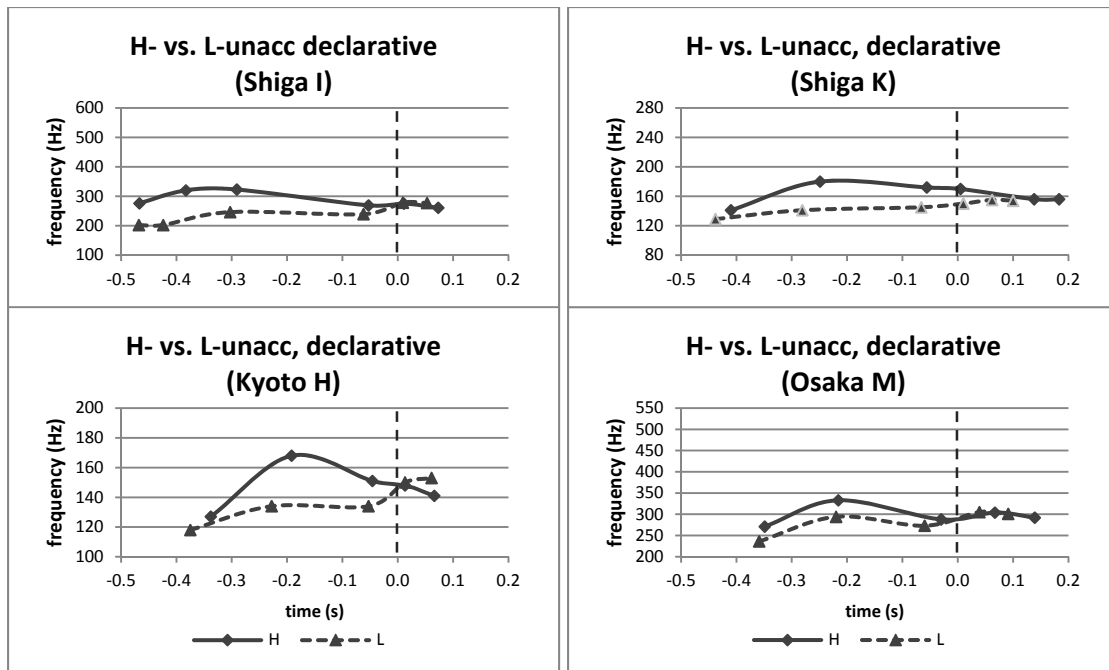


associated with that final accent is not realized in either intonational context<sup>21</sup>, and the L-final-accented and L-unaccented contours appear to be nearly neutralized. This small difference across the three dialects has important implications for our general model of speech melody.

One last result that will bear on the analysis for this dialect group is the behavior of H-unaccented vs. L-unaccented words in utterance-final position in the declarative context. A brief inspection of the declarative contours across these tonal categories in Figure 2.48 reveals that they tend to converge near the same  $F_0$  at the end of the utterance. To make this convergence more obvious, representative declarative contours for the four-syllable words *yamaimo* ('Japanese yam'; H-unaccented) and *nagaiimo* ('Chinese yam' L-unaccented) are plotted on the same graph for two Shiga speakers, the Kyoto speaker, and an Osaka speaker, respectively, in Figure 2.49. The contours in each graph are aligned such that the right boundary of the [m] in the last syllable [mo] is at time 0. It seems clear from these examples that there is a final target in declarative utterances that is lower than the initial high target of H-unaccented words and higher than the initial low target of L-unaccented words. In H-unaccented words, aside from aberrations due to segmental effects, the  $F_0$  appears to be more or less linearly interpolated from the initial high target (which itself is situated partway into the word) to the end. As for L-unaccented words, it appears that the  $F_0$  stays relatively low and then jumps up on the last mora/syllable. Shiga speaker K, shown in the upper right plot, may be the exception to this—in his L-unaccented word the  $F_0$  appears to creep up gradually leading up to the final mora/syllable. Kori (1987) noted that some scholars distinguish two speaker-specific realizations of L-beginning patterns in Osaka Japanese—"low flat" and "ascending from low". Assuming that this distinction applies across the Kansai dialects, presumably Shiga K would be considered an "ascending from low" speaker while the other three speakers represented in Figure 2.49 would be considered "low flat" speakers.

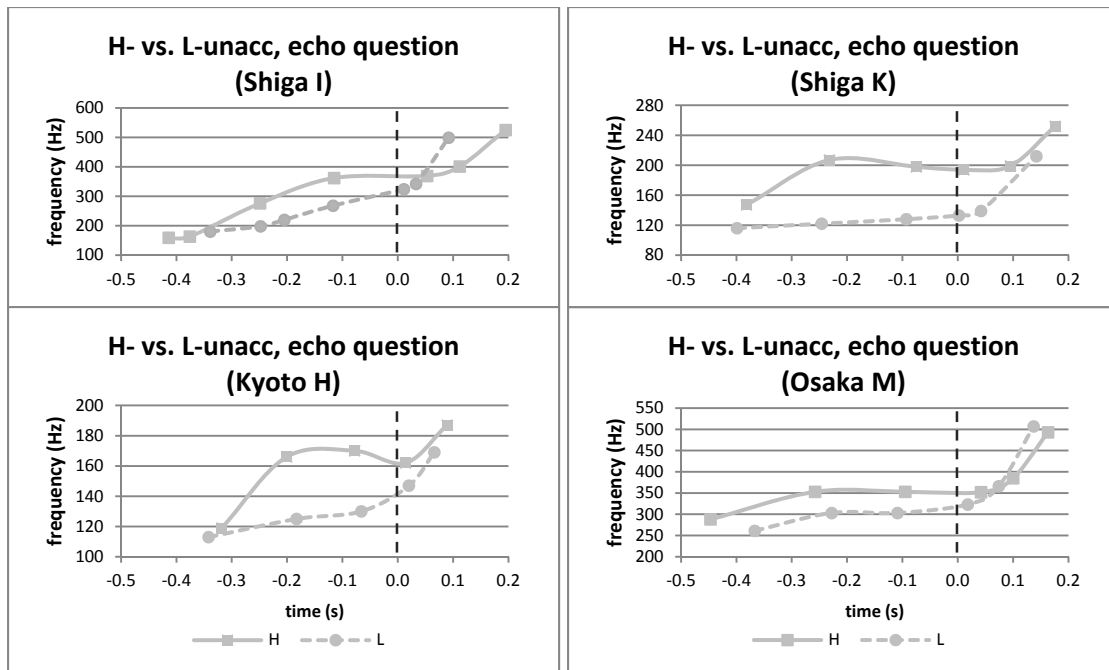
---

<sup>21</sup> It would be realized if the final-accented word were not phrase final, i.e. if there were a particle or some other material following the accented syllable. This loss of a post-accent fall in phrase-final position is akin to the well-known, analogous phenomenon in Tokyo Japanese.



**Figure 2.49: Representative  $F_0$  contours for the four-syllable tokens *yamaimo* (‘mountain yam’; H-unaccented) and *nagaimo* (‘long yam’; L-unaccented) in a declarative context for Shiga (top), Kyoto (bottom left), and Osaka (bottom right) Japanese. Contours on the same plot are zero-aligned at the onset of the last syllable.**

For comparison with the declarative renditions of these categories, the echo question renditions of the same two categories are shown in Figure 2.50. What the contours in Figure 2.50 all have in common is that they all end with a steeply rising trajectory on the last mora/syllable. While the tonal contrast is maintained by all the speakers in this context, the manner in which the echo question intonation affects the rest of the word leading up to the final mora/syllable is apparently speaker-dependent. For Shiga I and Osaka M, the  $F_0$  contours of the two respective tonal categories still seem to converge at a common target—one that is slightly higher than the target in the declarative context—and then there is a tail that rises from that point. For Shiga K and Kyoto H, the  $F_0$  difference between the tonal categories is exaggerated and maintained leading up to the final syllable (and in the case of Shiga K, through the onset of that syllable). Consequently, the final rising tail starts at a lower  $F_0$  for L-unaccented than for H-unaccented for those speakers. Incidentally, Shiga I and Osaka M are both female and Shiga K and Kyoto H are both male, but a more comprehensive investigation on the role of speaker sex in

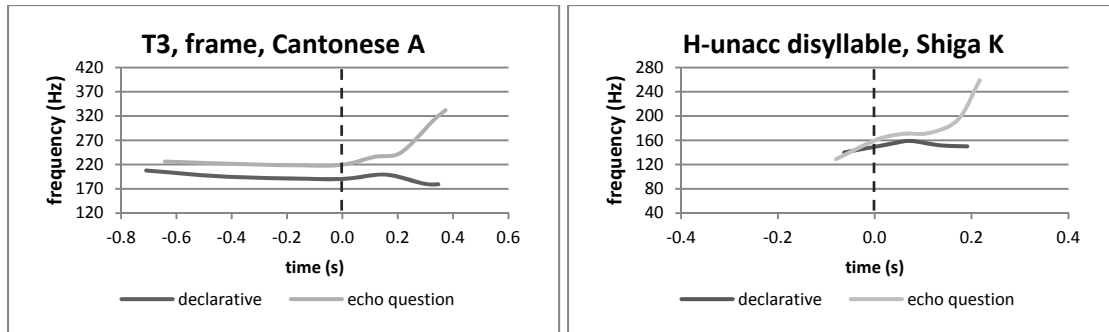


**Figure 2.50: Representative F<sub>0</sub> contours for yamaimo (‘mountain yam’; H-unaccented) and nagaimo (‘long yam’; L-unaccented) in an echo question context for Shiga (top), Kyoto (bottom left), and Osaka (bottom right) Japanese. Contours on the same plot are zero-aligned at the onset of the last syllable.**

phonetic implementation strategies is warranted. The implications of these results for the representation of melodic elements in these dialects are taken up in the following section.

### 2.6.6 Discussion

The realization of echo question relative to declarative intonation in SJ and the other Kansai dialects of Japanese can be broadly characterized as the serial addendum of a rising F<sub>0</sub> “tail” at the end of the utterance. If the declarative version of a word in a given tonal category ends at a relatively level trajectory on the last syllable, as in the unaccented categories, the echo question contour generally starts at that level at the beginning of the rightmost syllable and rises. In this respect the unaccented tonal categories in SJ behave similarly to the “level” tones in Cantonese (T1, T3, and T6). Examples from each language are shown in Figure 2.51 for comparison. Since the target word in Cantonese is monosyllabic and in a frame sentence and the one in SJ is disyllabic and in isolation, the dotted vertical lines have been placed at the left edge of the rightmost syllable in each case, to highlight the comparable portions of the respective contour

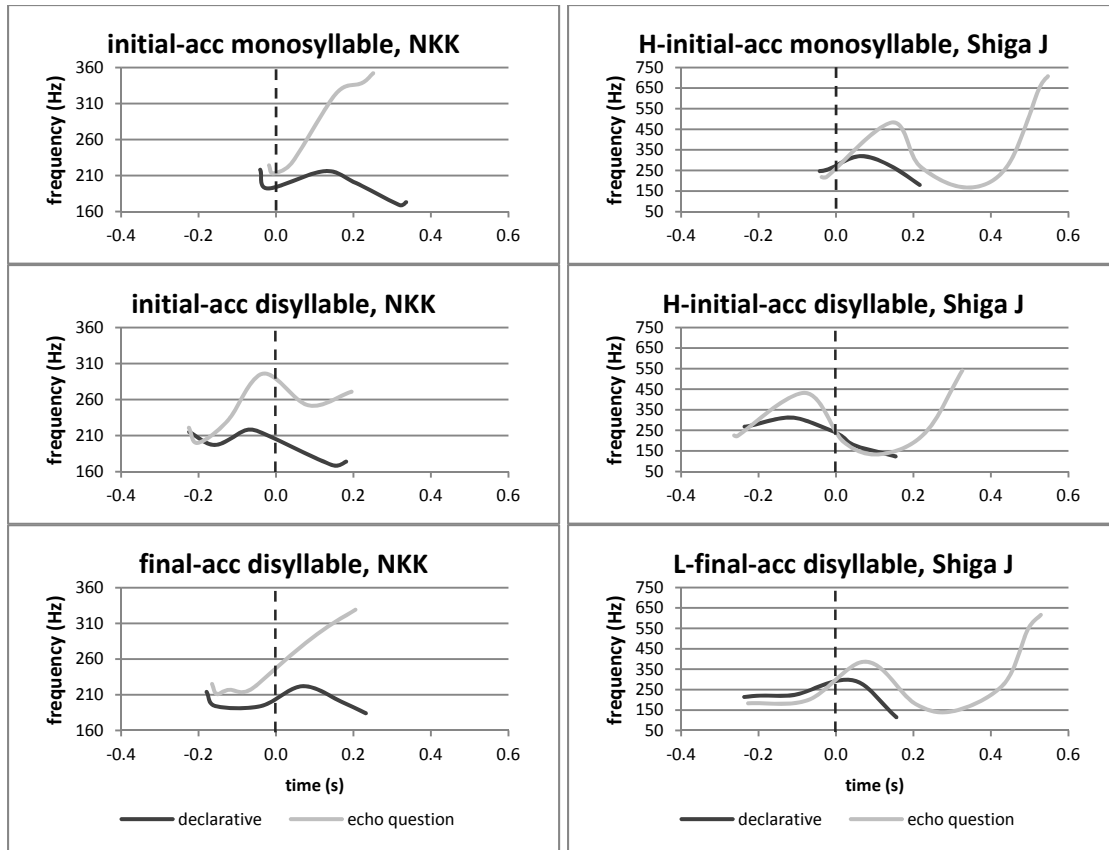


**Figure 2.51: Comparison of a level tone in Cantonese and an unaccented tone in Kansai. Dotted vertical lines indicate the left boundary of the rightmost syllable in each case.**

plots. It should also be kept in mind that the Cantonese plots represent mean contours, while the SJ plots are from single tokens. That being said, the similarity between the contours in the respective languages is striking, despite the differences in other parts of their respective tonal systems.

On the other side of the spectrum, since both NKK and the Kansai dialects have tonal systems that have been analyzed as utilizing pitch accents, it can be enlightening to compare them. A fundamental difference between the two is that words in isolation may be unaccented in Kansai but not in NKK. Therefore there is a very clear two-way contrast on monosyllabic words in the former—accented vs. unaccented—but there is a near-neutralization between initial- and double-accented in the latter. In both languages, pitch accents in a declarative context surface as a peak followed by a fall. It is interesting, therefore, to look at the comparable accented categories—both monosyllabic and disyllabic—side by side. This is shown in Figure 2.52, which compares NKK to SJ. It is clear that the analogous pitch accent categories behave very similarly in the two languages in a declarative context—there is a pitch peak in the nucleus of the accented syllable and a fall after the peak. There is a stark contrast between the languages, however, when it comes to how these tonal categories interact with echo question intonation. In NKK, there is no fall when the accent is in the final syllable—i.e. in the initial-accented monosyllabic or the final-accented disyllabic case<sup>22</sup>, but there is when the phrase is disyllabic

<sup>22</sup> And of course the double-accented monosyllabic case, not shown here.



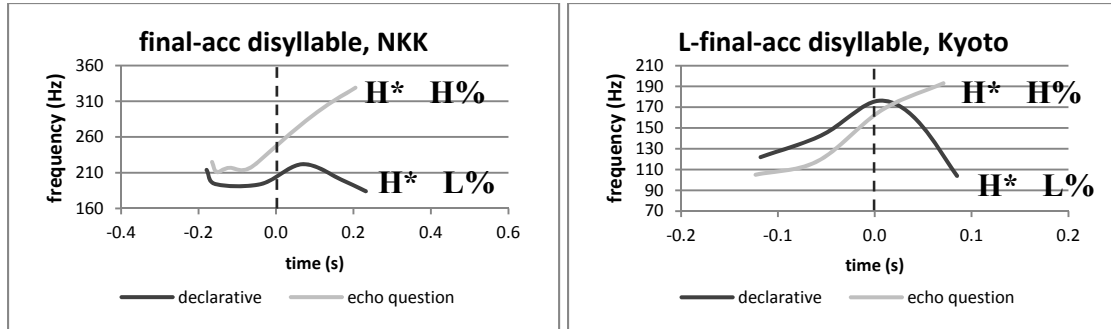
**Figure 2.52: Representative  $F_0$  contours for analogous accented tonal categories in NKK and SJ. Dotted vertical lines indicate the onset-nucleus boundary of the rightmost syllable.**

and the accent is on the first syllable—i.e. in the initial-accented disyllabic case. In all cases, the echo question intonation lifts up the entire pitch range of the rightmost syllable. In SJ, the fall is always present in the echo question context, and although a pre-final peak may be higher in that context than in the declarative context (we saw that this is not true for all speakers), the main robust effect appears to be a salient rise *after* a salient fall, no matter where that fall occurs in relation to the syllable<sup>23</sup>.

Superficially, Kyoto Japanese actually seems to be the Kansai dialect that is most comparable to NKK, since the fall is absent in the final-accented disyllabic case in the echo

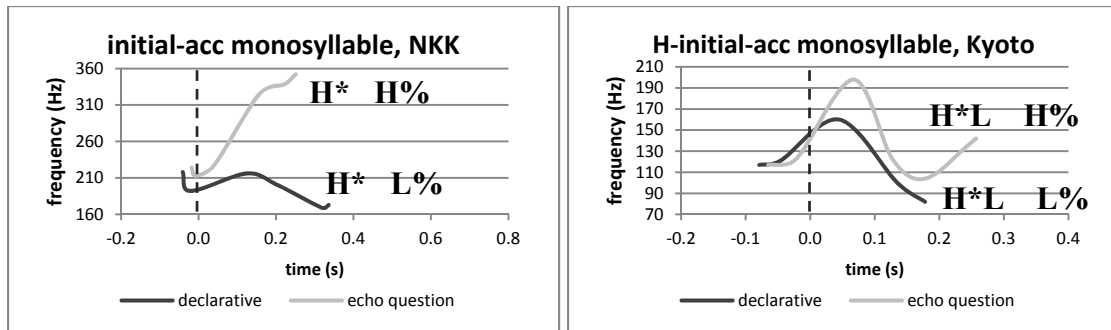
<sup>23</sup> It should be noted that the Shiga examples in Figure 2.52 were taken from Shiga speaker J, who tended to lengthen final syllables quite generously in the echo question context, presumably to accommodate the extra pitch movement. This strategy contrasts with that of Shiga speaker K (shown earlier in Figure 2.46 and in Figure 2.47), for whom the duration effects of echo question intonation were not as extreme or as consistent.

question context. Indeed, it is reasonable to posit similar surface-phonological representations for the two cases, as shown in Figure 2.53. However, if in both languages there is an underlying



**Figure 2.53: Preliminary surface-phonological representations for final-accented disyllables in NKK (left) and Kyoto (right).**

trailing L tone that gets deleted in certain phrasal environments, it is clear that the environments in question are not exactly the same, since the HL accent on a monosyllable in Kyoto Japanese does not behave in the same way as the HL accent on a monosyllable in NKK. This is evidenced in Figure 2.54<sup>24</sup>. More in-depth phonological analyses and their implications for a model of speech melody will be presented in Chapter 5.



**Figure 2.54: Preliminary surface-phonological representations for accented monosyllables in NKK and Kyoto Japanese.**

Considering Q2 (*Dimensions of Language-Specific Variation*) for a moment, although we have seen a wide range of durational effects of echo question intonation in SJ, with some

<sup>24</sup> The disyllable shown in Figure 2.53 for Kyoto Japanese is *mame* ('bean'), the second syllable of which is light (monomoraic). The monosyllable shown in Figure 3.48 for Kyoto is *nen* ('year'), which is a heavy (bimoraic) syllable. It is likely that the deletion of the L tone is sensitive to underlying mora structure.

speakers exhibiting greater intonation-dependent length differences than others, it is reasonable to posit that the phonetics in that dialect generally allows for considerable lengthening on that last syllable to accommodate the successive realizations of tonal and intonational targets, which we have not seen in any other language thus far. This is supported by the fact that accented words showed more lengthening than unaccented ones, and that most lengthening that was observed was observed on the second syllable in disyllabic tokens.

Within the Kansai dialect group, we saw a three-way split in how the L-final-accented category behaved. In Shiga, the post-accent fall was realized in both the declarative and echo question contexts. In Kyoto<sup>25</sup>, it was realized in the declarative context but not in the echo question context. In Osaka-Kobe, it was not realized in either intonational context. This is schematized in Table 2.12.

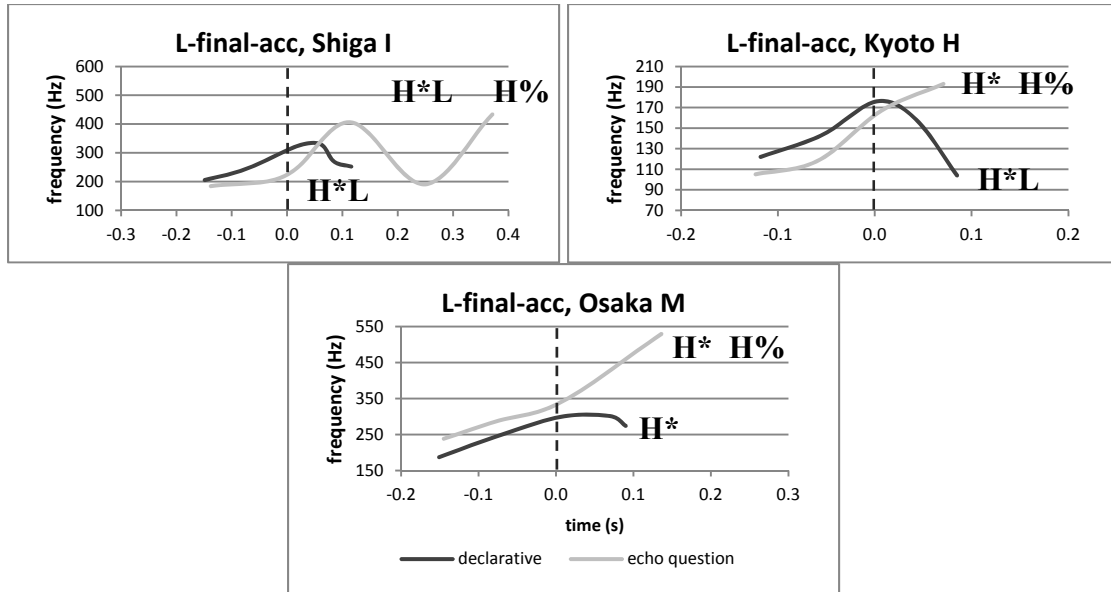
**Table 2.12: realization of post-accentual fall in a phrase-final context in the Kansai dialects**

Kansai dialect	fall realized in declarative	fall realized in echo question
Shiga	✓	✓
Kyoto	✓	✗
Osaka-Kobe	✗	✗

Preliminary surface-phonological analyses for this tonal category in each of the three dialects are shown in Figure 2.55. The implications of this typological breakdown will be discussed in greater depth in Chapter 4. Briefly, it suggests that the phonology needs to have access both to the lexical tonal category and to the intonational category (the latter traditionally being treated as postlexical in phonological analyses). At the very least, it provides further support to the notion that the answer to Q3 (*Declarative-to-Echo-Question Mapping*) is also “no”.

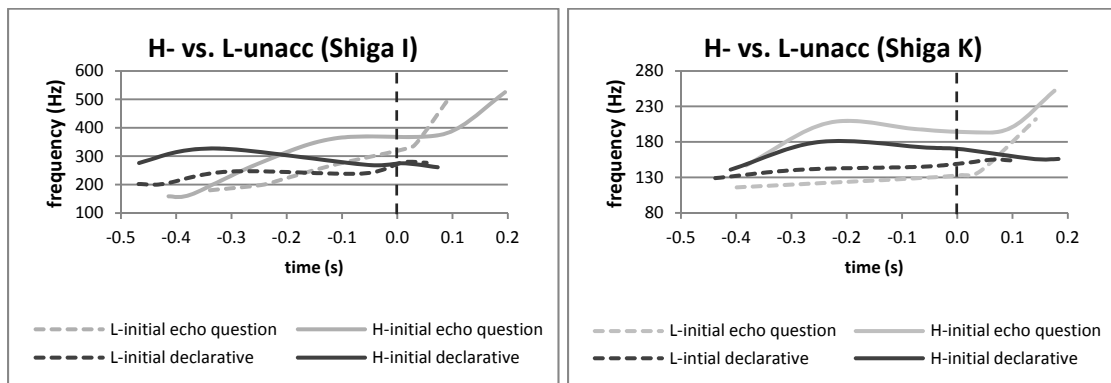
<sup>25</sup> Given that only one Kyoto speaker was recorded for the current study, it is possible that the results do not generalize to all Kyoto speakers. However, even if the typological variation observed here does not fall cleanly along geographic lines, it still must be accounted for. Hereinafter, the *Kyoto* label will continue to be applied to Speaker H’s dialect as it is a convenient way to differentiate it from the other Kansai dialects under discussion.

Lastly, the results from the polysyllabic unaccented words provide evidence that echo question intonation does more than just add a rising tail to the declarative contour. It was shown in the previous section that there are at least two different strategies when it comes to echo



**Figure 2.55: Preliminary surface-phonological representations for the L-final-accent in three Kansai dialects.**

question intonation on unaccented phrases. In Figure 2.56, the declarative and echo question contours for the two unaccented categories are superimposed on a single graph for each of the two SJ speakers that exemplify the different strategies. For Shiga I, the apparent common target



**Figure 2.56: Representative declarative and echo question contours for *yaimimo* (H-unaccented) and *nagaimo* (L-unaccented) for two SJ speakers.**



on the last syllable (whose onset is zero-aligned) gets shifted upward and the final rising tail starts from that point. The beginning of the word, meanwhile, gets shifted *downward*, as if to create a rising trajectory throughout the word leading up to the last syllable. Note that the interpolation from the starting point to the common target also changes in each tonal category, so that it is an oversimplification to say that the declarative contours are simply tilted to have an overall rising trend. For Shiga K, meanwhile, the  $F_0$  difference between the tonal categories is exaggerated and maintained leading through the onset of the final syllable. In essence, then, echo question intonation *raises* the  $F_0$  range of the H-initial-accented word and *lowers* that of the L-initial-accented word, and consequently the final rising tail starts at a lower  $F_0$  for L-unaccented than for H-unaccented. These somewhat complex interactions between the echo question intonation and the tonal contours of the words are not easily captured by a model that assumes that question intonation consists of a single H target or even a LH composite target that is added indiscriminately to the end of a phrase. The patterns observed in Shiga K's unaccented forms can loosely be characterized as a combination of Mandarin-like contour shape exaggeration and Cantonese-like addenda of rising tails.

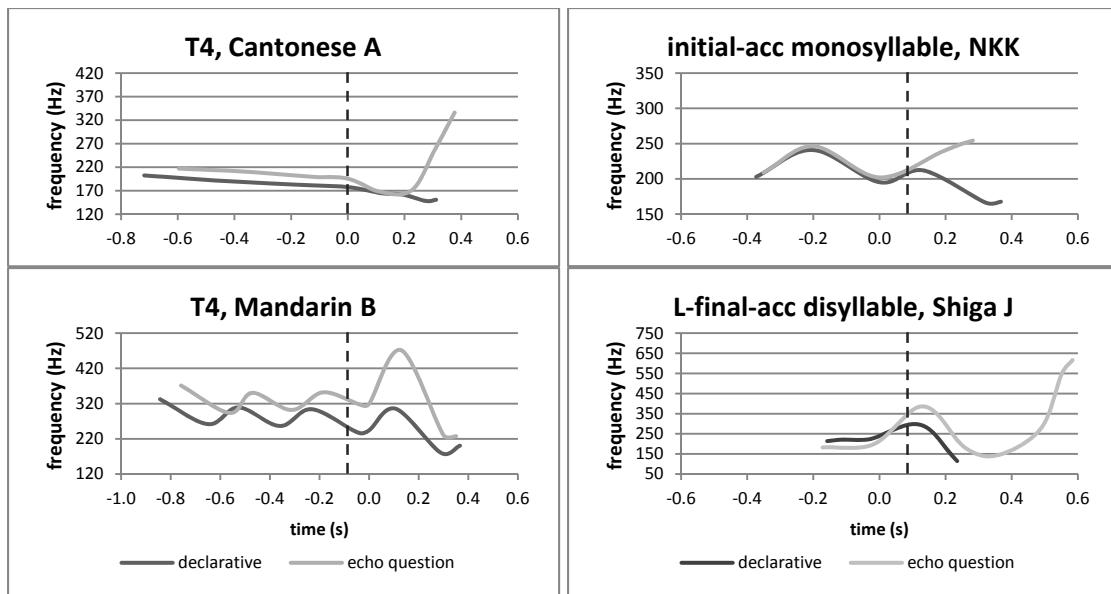
## **2.7 Summary and Conclusion**

In this chapter we have seen results from production experiments centered on the interaction of lexical tone and utterance intonation in four different languages. There were differences across the different languages as well as across tonal categories within each language with respect to the relative realizations of declarative and echo question intonation.

### **2.7.1 Cross-linguistic differences**

Focusing first on cross-linguistic differences, the realization of echo question intonation relative to declarative intonation in Mandarin is generally characterized by an upward shift in the peak  $F_0$  of the final syllable (whether the peak is at the beginning or end of the syllable) as well as an

increase in the  $F_0$  range. The “falling” tone, T4, is falling in both intonational contexts. In Cantonese, the  $F_0$  of the final syllable in an echo question starts at or slightly above where it would start in a declarative utterance (all else being equal), then steeply rises, in most cases starting partway through the syllable. All tones, including the “low falling” tone, T4, end with this steep rise in the echo question context. In NKK, all declarative utterances end with a peak and a fall, while in echo questions the contour hits a slightly higher point at the original peak time and then continues to rise, stays level, or falls a bit before leveling out, depending on the alignment of the original peak with respect to the last syllable. In SJ, the contour of the final syllable in the declarative context is preserved in the echo question context, with a steep rise added on to the end of it. For the sake of comparison, declarative and echo question renditions of a “falling tone” are shown for each language in Figure 2.57. We see then, that among languages that implement some sort of “rise” for echo question intonation, there are a few dimensions along which this implementation may vary.



**Figure 2.57:  $F_0$  contours for the “falling tone” category in Cantonese, NKK, Mandarin, and SJ. Dotted vertical lines indicate the beginning of the onset of the rightmost syllable.**

We can now start providing some answers to Q2 (*Dimensions of Language-Specific Variation*) at the beginning of the chapter. First, we see that in some languages the beginning of

the final syllable is relatively unaffected but the contour ends with a literally rising trajectory (i.e. its final slope is positive). Cantonese and SJ are such languages. There are other languages, such as Mandarin, in which the overall shape of the lexical tone is largely preserved, albeit distorted by the imposition of echo question intonation. In these languages the  $F_0$  contour may actually end with a falling trajectory if the lexical tone is falling. There are still other languages that seem to be hybrids of the above two types, sometimes imposing a final rising component and sometimes shifting the whole contour of the lexical tone upward. In NKK, which is such a language, the extent of “overlap” between the lexical tonal contour and the intonational contour is dependent on how far from the end of the utterance the last tonal specification is.

Among the languages that impose a final rise, we also see language-specific differences in how much duration may be added in the implementation of that rise. In Cantonese (and sometimes in NKK) it generally starts partway through the syllable or in some cases at the beginning of the syllable, partially or fully “wiping out” its lexical tonal contour as it would appear in a declarative context. In SJ it starts later, after the lexical tonal contour of the syllable has been established. We saw that some SJ speakers will elongate the final syllable to more than twice its declarative duration in order to accommodate the various “contour components” contributed by the lexical tone and the echo question intonation.

Among the languages that don’t allow any substantial lengthening but rather require the realization of the tone and the intonation to occur within the span of the syllable, we have seen some preliminary results that indicate that how robust (i.e. distinctive) the differences between the declarative and echo question versions of a given tonal contour are may be language-dependent. To examine this further in a methodical way, it would be necessary to elicit varying degrees of emphasis, incredulity, etc. from speakers in the echo question conditions. However, it is striking how the both of the Cantonese speakers maintained a more pronounced distinction in T1 than both of the Mandarin speakers in the equivalent T1, and it is most likely due to differences in how intonation is implemented in the two languages.

As for Q4 (*Structural Correlation of Tone-Intonation Interaction*), the prognosis does not look good! While both NKK and SJ have somewhat similar tonal systems involving pitch accents (i.e. word tone) and a lower density of tonal specification, and while Cantonese and Mandarin have more similar tonal systems involving syllable tone and a higher density of tonal specification, the intonational systems of these languages do not cluster along the same lines.

Before moving on from the topic of cross-linguistic differences, it is worth mentioning that the use of utterance-final particles is another dimension along which languages may vary, and although incorporating them into the discussion of utterance-level intonation complicates the overall typological picture, they are nonetheless a crucial part of that picture. As mentioned at the beginning of this chapter, all of the languages discussed in the current study optionally make use of utterance-final particles in questions. In some cases question particles are not given any special prosodic status compared to “lexical” material at the right edge of a question; Japanese (Poser 1984; Pierrehumbert and Beckman 1988; Venditti, Maekawa et al. 2008) and NKK (Lee 2008) appear to fall into this category. In other cases, while utterance-level intonation *may* manifest itself on final particles the way it does on lexical material, particles may also be realized with certain intonational properties not possible on lexical material; Cantonese is a classic example of such a language (Matthews and Yip 1994; Xu and Mok 2011), and to a more restricted extent Mandarin falls under this category as well (Lee 2005). This has partly to do with the particular range of semantic and pragmatic functions associated with certain syntactic constructions only possible with particles in those languages (Law 1990; Lee 2005). As important as question particles and utterance-final particles more generally are to the discussion of the typology of intonation and tone, they will be left aside for further research and not addressed in the rest of the current study, save a few passing remarks during theoretical discussions.

### 2.7.2 Tonal-category-dependent differences

Besides the various language-dependent differences, we have seen that even within a specific language, how intonational distinctions are realized can be dependent on the tonal category. In both Mandarin dialects, the nature in which each tone is shifted upward in the echo question context is not the same across all the tones. In Cantonese, where the rising trajectory starts in echo questions is tone dependent; even two relatively similar tones, T1 (high level) and T3 (mid-level) behave differently, with T1 rising throughout the syllable and T3 starting level and ramping upward starting partway through the syllable. Also, the overall  $F_0$  excursion in the echo question rendition of a tone is not independently derivable from the relative  $F_0$  height of the declarative rendition of that tone. In NKK, all of the tones in the declarative context end in a falling trajectory, but only the initial-accent, in a disyllabic environment, has a falling component on the last syllable in the echo question context. In Kyoto Japanese, all accented words surface with a falling then rising final contour, with the exception of final-accented words, whose contours lack a falling component in that environment. In Shiga Japanese, we have seen that echo question intonation affects the overall contour of polysyllabic H-unaccented words differently from that of polysyllabic L-unaccented words. All of these phenomena force us to conclude that the answer to Q3 (*Declarative-to-Echo-Question Mapping*) at the beginning of this chapter is “no”. This issue will be discussed further in Chapter 4.

### 2.7.3 Answers to questions about production

The questions from the beginning of this chapter are repeated in (2.22), along with some answers that have been revealed by the experiments presented in the chapter:

(2.22) Questions regarding melodic perception in tone languages

**Q1 – *Language-General Tone-Intonation Interaction*:** Does intonation interact with lexical tone in the same way in all tone languages? *No*.

**Q2 – Dimensions of Language-Specific Variation:** If not, what are the dimensions along which melodic systems may vary in a language-specific way? “*Translational*” (Mandarin) vs. “*overwriting*” (Cantonese) vs. “*serial*” (SJ)  $F_0$  functions for intonation; degree of accommodation for multiple melodic “events” on a single TBU (Cantonese vs. SJ); degree of robustness of  $F_0$  range differences across intonational categories.

**Q3 – Declarative-to-Echo-Question Mapping:** Is the  $F_0$  contour of a tone in the echo question context predictable purely from the phonetic attributes of the tone in the declarative context (i.e. is there an algorithm that can take just phonetic pitch parameters of the declarative form as input and successfully predict the  $F_0$  contour of the echo question form, or vice versa)? *No; we have seen evidence for many tone-dependent attributes of intonational realization that rule out the possibility of a successful mapping algorithm that does not take information about the tonal category into consideration. This phenomenon of tone-dependent intonation will be taken up in further detail in Chapter 4.*

**Q4 – Structural Correlation of Tone-Intonation Interaction:** Is there any way to predict certain aspects of tone-intonation interaction based on given characteristics of a language (the structure or inventory of its tonal system, e.g.?) *None has been revealed for the languages examined in this chapter (although this is of course merely the absence of evidence and not evidence of absence).*

#### **2.7.4 Conclusion**

The various language-specific and tone-specific phenomena revealed in this chapter will all need to be taken into account when working toward a viable model of speech melody. This task will be taken up in Chapters 4, 5, and 6. First, though, in Chapter 3, the melodic systems of several of the languages examined in this chapter will be approached from a perceptual point of view, as results from perceptual experiments in those languages are presented and analyzed.

## CHAPTER 3: PERCEPTUAL EXPERIMENTS

### 3.1 Questions about Perception

In the previous chapter it was shown that the interaction of the tonal and intonational systems in each of the four languages results in distinct patterns of relative pitch (and in some cases relative duration) in each language. In this chapter we will see how these differences affect the perception of lexical tone and utterance intonation. In particular, attempt will be made to answer the questions listed in (3.1):

(3.1) Questions about melodic perception in tone languages

**Q5 – *Functional Recoverability*:** Are different types of tonal systems (i.e. word tone vs. syllable tone; register vs. contour) equally successful in encoding multiple communicative functions (lexical distinctions vs. intonational meanings) in the speech melody in a recoverable way?

**Q6 – *Functional Prioritization*:** Is it the case that priority is given to the same communicative function (tone or intonation) in all languages when there is an overlap?

**Q7 – *Types of Confusion*:** What types of confusions, misperceptions, and “perceptual neutralizations” are possible and when do they occur?

A. Can the melodic expression of one communicative function be misperceived and reinterpreted as that of another (i.e. can a melodic cue for tone be misperceived as that for intonation or vice versa)?

B. Can we make any predictions about when such confusions occur based on the characteristics of a given tonal system?

A series of perceptual experiments was carried out with the above questions in mind. In order to be able to make meaningful comparisons among the languages in question, an attempt

was made to design the perceptual tests for the different languages in a relatively parallel way. The stimuli were taken from the production experiments presented in Chapter 2 and consisted of target words that were grouped as minimal pairs or sets uttered either in isolation or at the end of a short frame sentence. They were minimal sets in that they were segmentally identical and differed only in tonal category. If a frame sentence was used, care was taken to ensure that the speech melody would be the only source of cues to tonal and intonational categories. For example, while all of the languages in question employ utterance-final particles in some contexts, often providing morphological cues to utterance type and reducing the amount of overlap between intonational events at the right edge of the utterance and lexical tonal events in the rightmost “content word”, such constructions were not included in any of the stimulus sets. The basic task in each experiment was as follows: the subject listened to stimuli over a headset. Upon hearing each stimulus twice, the subject was asked to provide two judgments: what the target word was (i.e. what tonal category she perceived) and whether the utterance was a statement or a question. In addition, the subject was asked to rate her own confidence for each response on a scale of one to five, five being the highest. Detailed descriptions of the individual experiments and results are provided in the following sections, starting with Mandarin in Section 3.2. A summary of all of the results and a general discussion follow at the end of the chapter.

## **3.2 Perception in Mandarin**

### **3.2.1 Subjects**

There were 20 subjects who all identified themselves as native speakers of *Putonghua*, or “standard” Mandarin. Although age and sex were not evenly distributed, they ranged in age from 18 to 51 and included 7 males and 13 females. With one exception<sup>26</sup>, they all lived in Beijing for at least the first 17 years of their lives and they all had parents who spoke *Putonghua* at home.

---

<sup>26</sup> The exceptional subject lived in Urumqi for seven years before moving to Beijing.



### 3.2.2 Materials and procedure

All stimuli were presented as audio clips played over a high-quality headset connected to a laptop. Each subject was presented with a total of 96 utterances, satisfying the conditions given in (3.2):

(3.2) Mandarin perceptual experiment conditions

4 tonal categories (Tone 1, Tone 2, Tone 3, and Tone 4)

2 intonational categories (declarative statement vs. echo question)

6 repetitions

2 stimulus sets (i.e. two speakers who were recorded)

The stimuli were presented in two parts—48 stimuli produced by one speaker and 48 stimuli produced by another speaker. Within each part the order of the stimuli was randomized each time and the order in which the parts were presented was switched each time (i.e. half the subjects heard Speaker A’s utterances first and half the subjects heard Speaker B’s utterances first). In between the two parts, a short break was taken during which background information about the subject was obtained. The subjects indicated their responses on a multiple choice answer sheet, a portion of which is shown in Figure 3.1. The frame sentence, *na4liang4 nian4...* ‘Naliang reads...’ appears in the upper left-hand corner. The first four blank columns correspond to the four tonal categories, which are represented by the words *wan1* ‘bay’, *wan2*

娜觀念...						
	“湾”	“丸”	“碗”	“万”	。	?
1						
2						
3						
4						
5						
6						
7						

Figure 3.1: Sample answer sheet for the Mandarin perceptual test.

‘pill’, *wan3* ‘bowl’, and *wan4* ‘ten thousand’, respectively. The last two blank columns correspond to the two intonational categories, *statement* and *echo question*, respectively. Upon hearing a given stimulus twice, the subject was asked to indicate her choice of tonal category and intonational category by writing a number in the corresponding box for each, the number being an indicator of confidence on a scale of 1 to 5<sup>27</sup>. There was no time limit; once the subject was satisfied with her response for a given stimulus, the next stimulus was presented. On average, the entire test, including explanation time at the beginning, took about 30 minutes.

### 3.2.3 Results

The overall rate of accuracy for tone identification was nearly perfect, at 98.91%. The rate of accuracy for intonation identification was a bit more degraded but still very high, at 87.97%. These results are shown in Table 3.1. A generalized estimating equations (GEE) model<sup>28</sup> was

**Table 3.1: Overall rates of perceptual accuracy in judging tonal and intonational categories in Mandarin**

function	rate of accuracy
tone	98.91%
intonation	87.97%

used evaluate the effect of function (i.e. accuracy type—tone accuracy vs. intonation accuracy) on the rate of accuracy for Mandarin. The difference between the rate of accuracy for tone and that for intonation was highly significant ( $p < .001$ ).

<sup>27</sup> As the confidence results were not very illuminating—either the subjects’ levels of confidence very weakly correlated with their rates of accuracy but showed a much narrower overall range or the subjects were just overly confident across-the-board—these results are not reported on in this thesis. The same goes for the confidence level results in the other languages. Anecdotally, time and time again a subject would fret over her response, taking a lot of time to make up her mind and even scratching out one answer before settling on a different one, only to record her level of confidence for that response as a ‘5’. Based on these observations, response times probably would have been a more accurate (and precise) measure of level of confidence. Unfortunately, the test was not set up in such a way that response times were recordable.

<sup>28</sup> The results here and elsewhere in the chapter were analyzed using SPSS version 20; the GEE model was specified with a binary distributed outcome and a logit link function, where clustering of data at the subject level was accounted for by an exchangeable correlation structure.

Let us take a look at the tone identification results in more detail. Table 3.2 shows the rate of perceptual accuracy for tone plotted by the intonation type of the stimuli. Since the overall rate was so high, it is not surprising that there is little difference between intonation types. Unfortunately, the level of significance of this difference could not be computed within a logistic regression model due to the fact that a Hessian matrix singularity was caused by the parameters in the model for which one or more categories returned a rate of accuracy of 100%.

**Table 3.2: Rate of perceptual accuracy in judging tonal category, by intonation type of the stimulus, in Mandarin.**

intonation	tonal accuracy
declarative	99.90%
echo question	97.92%

If we plot the pooled responses in confusion matrices, we can see that this small rate of error was not evenly distributed across the tonal categories or stimulus sets. Figure 3.2 shows four confusion matrices for Mandarin. The two upper matrices show results from Stimulus Set A and the two lower matrices show those from Stimulus Set B. The left-hand matrices show results from the declarative condition, while the right-hand matrices show results from the echo question condition. The rows in the matrices correspond to the stimulus tonal category and the columns correspond to the response category. The numbers in the cells represent percentages (rounded to the nearest whole) and the cells are shaded accordingly. The matrix on the lower left, which shows results for the declarative condition for Stimulus Set B, is an example of the highest possible rate of accuracy, with no confusions in any category. Note that it is characterized by a pure black diagonal of 100s. From looking at Figure 3.2, we can see that there was some confusion between T2 and T3 in the echo question conditions. Specifically, T2 was misperceived as T3 8% of the time in the echo question condition for Stimulus Set A, while T3 was misperceived as T2 6% of the time in the echo question condition for Stimulus Set B. These rates of error, while small, were not confined to one or two individuals. The T2→T3 error was

stim	T1	T2	T3	T4
T1	100	0	0	0
T2	1	99	0	0
T3	0	0	100	0
T4	0	0	0	100

Stimulus Set A, declarative

stim	T1	T2	T3	T4
T1	98	2	0	1
T2	0	93	8	0
T3	0	0	100	0
T4	0	0	0	100

Stimulus Set A, echo question

stim	T1	T2	T3	T4
T1	100	0	0	0
T2	0	100	0	0
T3	0	0	100	0
T4	0	0	0	100

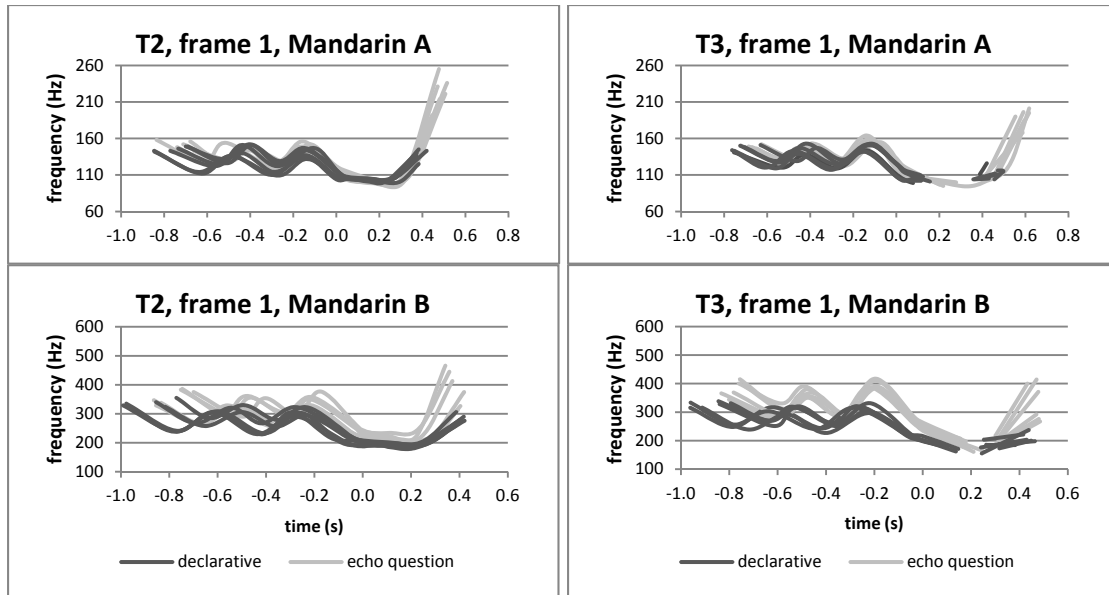
Stimulus Set B, declarative

stim	T1	T2	T3	T4
T1	100	0	0	0
T2	0	100	0	0
T3	0	6	93	1
T4	0	0	0	100

Stimulus Set B, echo question

**Figure 3.2: Tonal confusion matrices for Mandarin.**

made by 6 out of the 20 respondents. The T3→T2 error was made by 5 out of the 20. One individual made both errors, meaning that 10 out of the 20 respondents confused T2 for T3 or vice versa at least once. Indeed, although the level of significance here is indeterminate, previous studies have found that the T2-T3 confusion is the most common among all of the tonal pairs in Mandarin (Chuang, Hiki et al. 1972; Zue 1976; Shen and Lin 1991). These confusions are understandable if we go back and look at the actual pitch contours of the relevant categories. As we can see in Figure 3.3, since the rising tail of the dipping tone—T3—gets exaggerated in the echo question context, its overall shape becomes more similar to that of the rising tone—T2—in that context. In addition, the heavy glottalization that is present at the low points of the T3 contours (and that manifests itself as discontinuities in the plots), while still present in the echo question context, is less extreme in that context. Yang (2011) presented evidence that voice



**Figure 3.3: Multiple-contour plots for Mandarin T2 (left) and T3 (right), for speakers A (top) and B (bottom). Declarative contours are shown in dark gray and echo question contours in light gray.**

quality is used in the categorical perception of T3 by native listeners. It should be noted that all of the cases of T3 mistaken for T2 were for tokens that actually did contain some glottalization, suggesting that this weaker glottalization was not a strong enough cue to sway listeners' perceptions in those cases. Other than the T2-T3 confusion, which itself was minimal, it is striking how little the intonational melodies affected the perception of the tonal categories.

Unlike perceptual accuracy for tone, perceptual accuracy for intonation in Mandarin was highly category-dependent. The results for intonational accuracy, broken down by intonation type, are shown in Table 3.3. A GEE model used to analyze the effects of tonal category and intonational category on intonational perceptual accuracy revealed the effect of intonation type to be highly significant ( $p < .001$ ). These results are in line with those of Yuan (2004), who concluded that “statement intonation is easier to identify than question intonation” (p. 46).

**Table 3.3: Rate of perceptual accuracy in judging intonational category, by intonational category of the stimulus, in Mandarin**

intonational category	rate of accuracy
declarative	98.02%
echo question	77.92%

Yuan (2004) also noted:

That question intonation identification was less accurate means that many question intonation utterances were identified as statements. This suggests that statement intonation is a default or unmarked intonation type. That is, listeners fall back to this option when there is not enough information suggesting ‘question’, which is also supported by the fact that the tone of the last syllable does not affect statement identification. Question intonation is, however, a marked intonation type. It can only be identified if the listeners actually hear the ‘question’ features/mechanisms. (p. 64)

This apparent bias becomes clear if we reinterpret the data by displaying the rate of response for each intonational category—like a “bin test” to see how often listeners reported hearing declarative intonation and how often they reported hearing echo question intonation. This bin test is shown in Figure 3.4. We will see that this apparent “over-guessing” in favor of declarative intonation was observed for Cantonese and NKK as well.

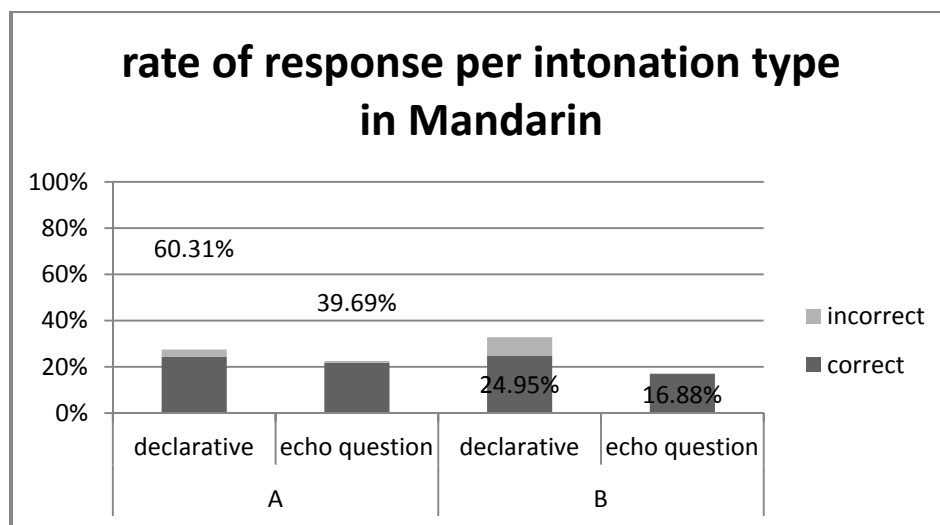
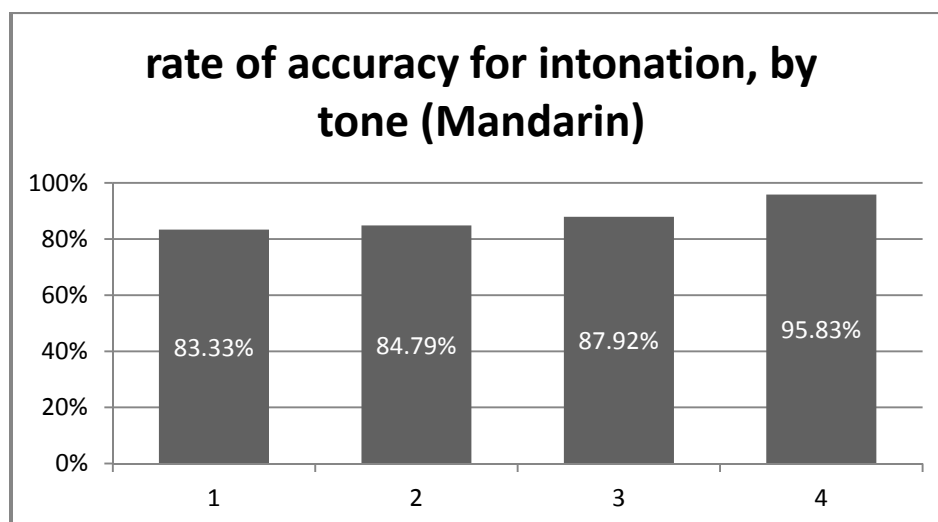


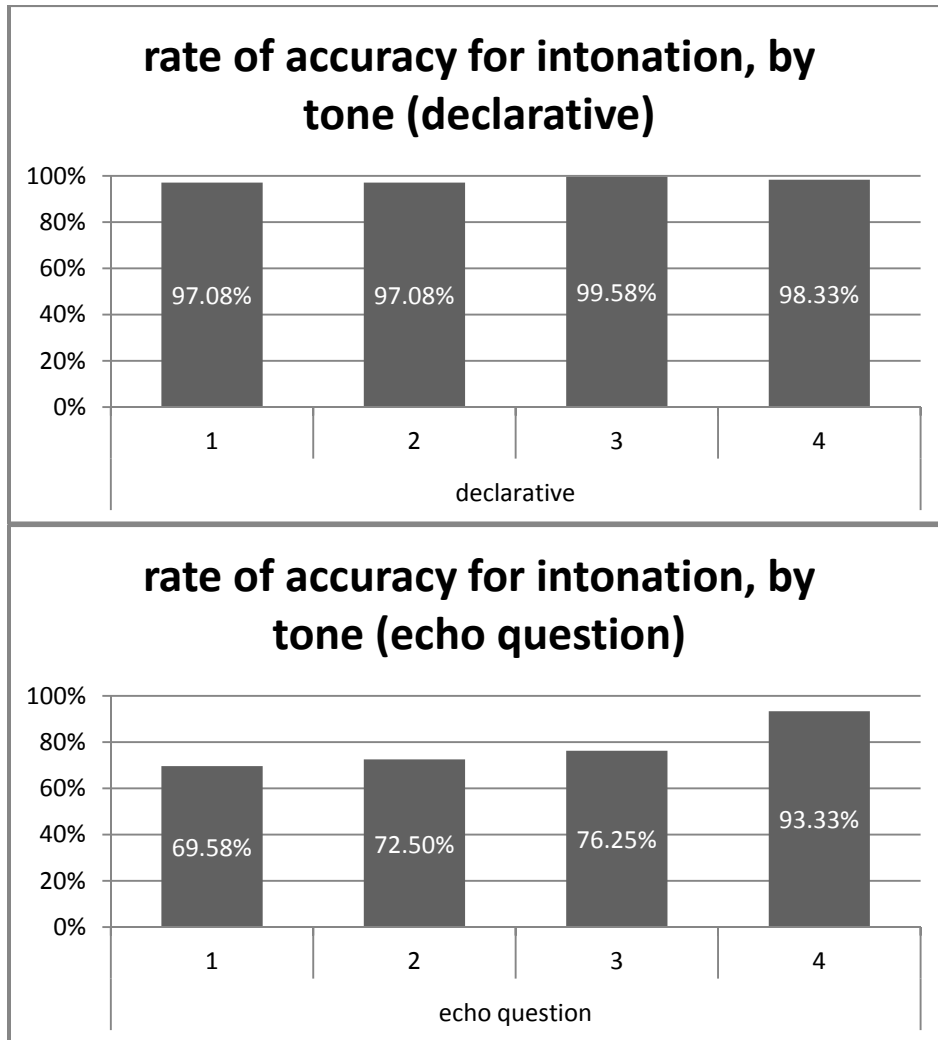
Figure 3.4: The rate of response—or “bin test”—for each intonation type in Mandarin

Let us now explore how tonal category affected the perception of intonation. These results are shown in Figure 3.5. We can see that the rate of accuracy was not the same across all



**Figure 3.5: The rate of accuracy in judging intonational category, by tonal category of the stimulus, in Mandarin.**

tonal categories. This uneven distribution is more pronounced if we separate out the intonation conditions. This is shown in Figure 3.6. It is apparent from Figure 3.6 that the effect of tonal category on the perception of the intonational category is amplified in the echo question condition but unapparent (most likely because it is masked by the over-guessing effect) in the declarative condition. The most striking result is that intonational perception remains the most accurate in T4, while it is more degraded in the other three tone conditions. According to a post-hoc analysis with a Bonferroni correction, the rates of accuracy were not different across tonal categories in the declarative context; in the echo question context only the T4 rate was significantly different from those of the other three tonal categories ( $p < .01$  in all three cases). Note that T4 is the falling tone, which we may characterize as the tone whose contour is the most “counter” to the intonational melody of the echo question. These results are mostly in line with those reported by Yuan (2004), who found that question intonation identification was best when the last syllable bore Tone 4 and worst when it bore Tone 2.



**Figure 3.6: The rate of accuracy in judging intonational category, by tonal category of the stimulus, in Mandarin—the declarative condition above and the echo question condition below.**

### 3.2.4 Discussion

Broadly speaking, we may characterize the perception of melody in Mandarin as follows: tone is highly recoverable across-the-board, while the perception of intonation is slightly degraded—the degree to which it is degraded being dependent on the tonal category of the rightmost syllable in the utterance. Independent of intonation, it is perhaps unsurprising that listeners fared well in distinguishing the four lexical tones on a purely melodic basis. In a declarative context, the four tones have very distinct contours in any pitch range—arguably maximally so. As for the echo question context, recall from Chapter 2 that we may characterize echo question intonation in



Mandarin as a more global phenomenon, mainly affecting the maximum point of the rightmost syllable's tonal contour as well as the peak in the penultimate syllable in some cases. It seems that this has at least two consequences for perception.

One, since in most cases this means simply increasing the overall pitch range of the melody on the target syllable, the signature contour of each tone is largely preserved, allowing tonal recognition to remain highly accurate. In the few cases where there is some tonal confusion, it is precisely in those cases where the  $F_0$  maximum for both tones (T2 and T3) is on the right edge, resulting in perceptually less distinct contours for the respective tonal categories involved. This type of confusion, while minimal among Mandarin listeners, is representative of a trend we will see in Cantonese as well, namely that confusions are likely to occur with rising tones (broadly characterized as lexical tones with a final rising component) in a rising intonational context. A second consequence of the more global nature of intonation in Mandarin is that the distinction between the two intonational melodies can be difficult to discern from a perceptual point of view. Unlike in Cantonese and SJ, the intonational melody associated with echo questions does not surface with a signature local contour of its own, and it also does not change the trajectory of the final tonal contour to any substantial degree; it certainly does not change the sign of the final slope from negative to positive (which, as we will see, happens in Cantonese, NKK, and SJ). Also, the nature of this type of intonational system is such that the degree to which the tonal contour is transformed is less pronounced in at least three out of the four categories—precisely those categories for which the perception of intonation was the most degraded.

Setting aside utterances that end in a purely level tone, we can crudely divide utterances into those that end with a rising contour and those that end with a falling contour. The results from the perceptual test in Mandarin suggest that it might be interesting to keep track of whether and how each type of utterance can be reinterpreted. Table 3.4 shows this schematically, in order to go part of the way toward answering Q7 (*Types of Confusion*) at the beginning of this chapter. We will see that, while a rising tone in an echo question context (in the top row) is

**Table 3.4: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations in Mandarin**

underlying combination	surfaces as	occasionally <sup>29</sup> reinterpreted as
<b>T↗ + I↗</b>	<b>rise</b>	<b>T↗ + I↘</b>
T↗ + I↘	rise	n/a
T↘ + I↗	fall	n/a
T↘ + I↘	fall	n/a

reinterpreted as a rising tone in a declarative context here and in Henanhua and Cantonese, a falling tone in a declarative context (in the bottom shaded row) does not usually get reinterpreted in any of the languages. We will return to a discussion of these results in a broader typological context at the end of the chapter. In the next section the perceptual results for Henanhua are presented.

### 3.3 Perception in Henanhua

#### 3.3.1 Subject

A full-fledged perceptual experiment with multiple listeners was not carried out for Henanhua. However, the speaker herself took a perceptual test in which she listened to randomized tokens of her own speech.

#### 3.3.2 Stimuli

All stimuli were presented as audio clips played over a Sennheiser PC 156 headset. They were presented in two sets—the isolation set followed by the frame set. Each set contained 24 randomized stimuli (4 tones x 2 intonations x 3 repetitions). The subject was given a response sheet very similar to the one used in the Mandarin experiment.

<sup>29</sup> More than 10% of the time

### 3.3.3 Results

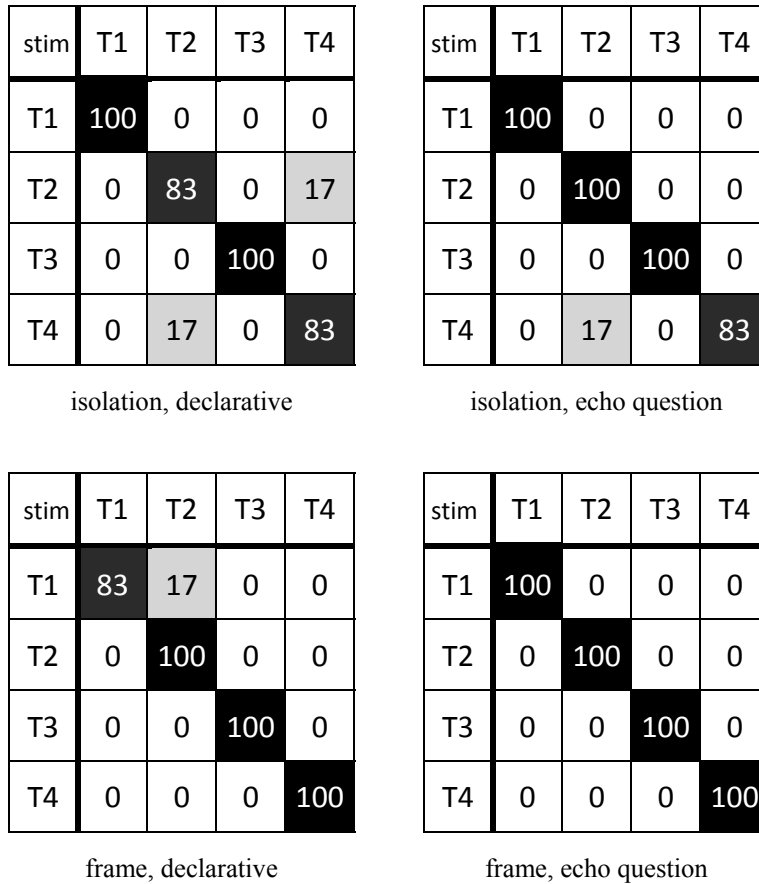
The robustness of the results needs to be confirmed with a full-fledged experiment; nevertheless, some selected results from this preliminary test are presented here as they raise some interesting questions. Also, they can be compared to the results of the perceptual test in NKK for which the original speaker also took the perceptual test.

The overall rates of perceptual accuracy were similar to those for Mandarin. These are shown in Table 3.5. The confusion matrices for each of the intonation conditions in each of the

**Table 3.5: Overall rates of perceptual accuracy in judging tonal and intonational categories in Henanhua**

function	rate of accuracy
tone	95.83%
intonation	86.46%

phrasal contexts are shown in Figure 3.7. The listener perceived her own T2 as a T4 twice and her own T4 as a T2 once (out of 24 T2 and 24 T4 stimuli). She also characterized a T1 as a T2 once in the frame context, but this was most likely a lapse in concentration rather than a misperception, since Henanhua T1 in no way resembles Henanhua T2 in the phrase-final context. The T2-T4 confusion is understandable, though, since those contours are quite similar. Notice that tone identification was just as good, if not better, in the echo question context. Unlike in Mandarin, echo question intonation did not elicit any confusion between the dipping tone and the rising tone. Perhaps the degree of similarity was not high enough, or perhaps some such confusions would occur in a larger sample of listeners (especially ones who are not the speaker).



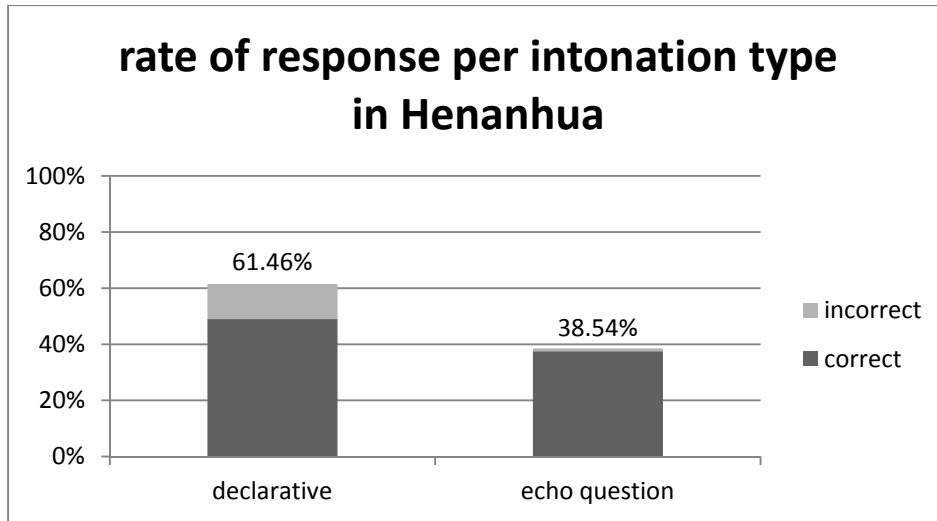
**Figure 3.7: Tonal confusion matrices for Henanhua, by phrasal context and intonation type. A 17% difference represents a one-stimulus difference.**

Moving on to the intonation results, we see in Table 3.6 that the breakdown of intonational accuracy by intonation type was very similar to the mean breakdown for Mandarin listeners.

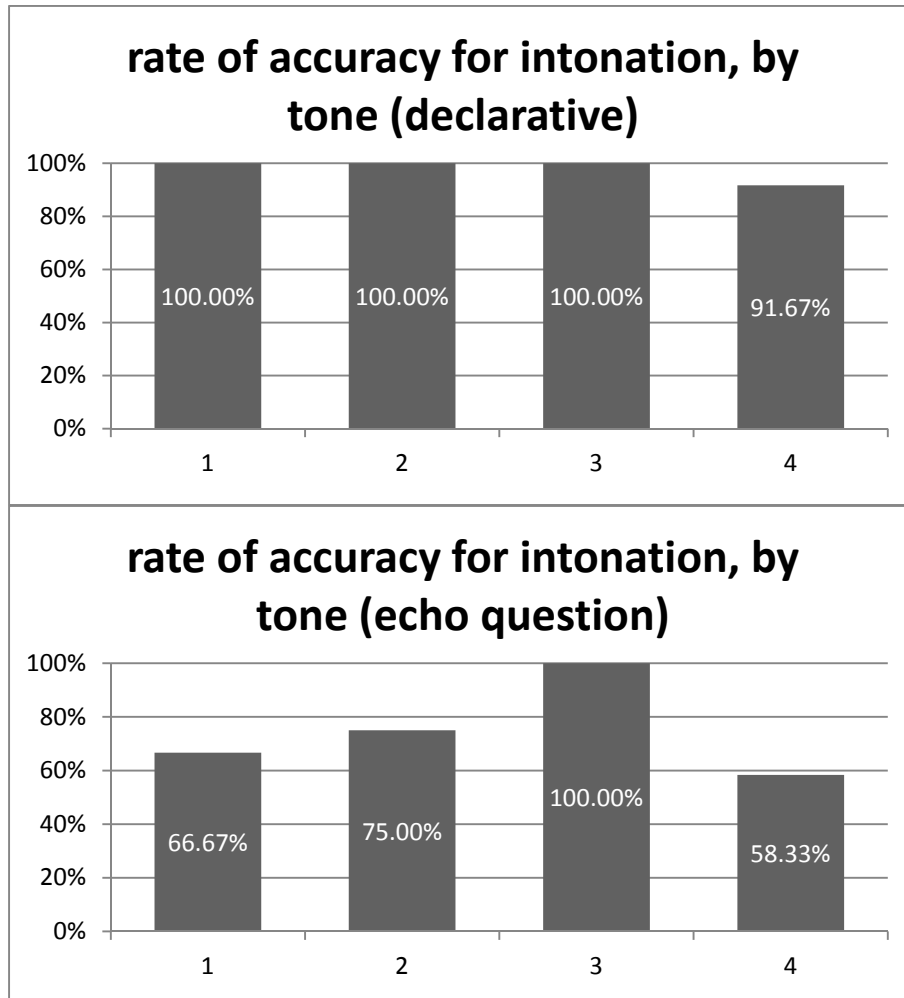
**Table 3.6: Rate of perceptual accuracy in judging intonational category, by intonational category of the stimulus, in Henanhua**

intonational category	rate of accuracy
declarative	97.92%
echo question	75.00%

Figure 3.8 shows the results in bin test form, again highlighting an apparent bias toward declarative intonation. Figure 3.9 shows the rate of perceptual accuracy for intonation type, by tone. The phrasal context has not been factored out, since the sample size is so small and the results were largely the same in both contexts. It's clear that T3, the level/“gently rising” tone,



**Figure 3.8: Bin test for intonation responses in Henanhua.**



**Figure 3.9: The rate of accuracy in judging intonation, by tonal category of the stimulus, in Henanhua—the declarative condition above and the echo question condition below.**

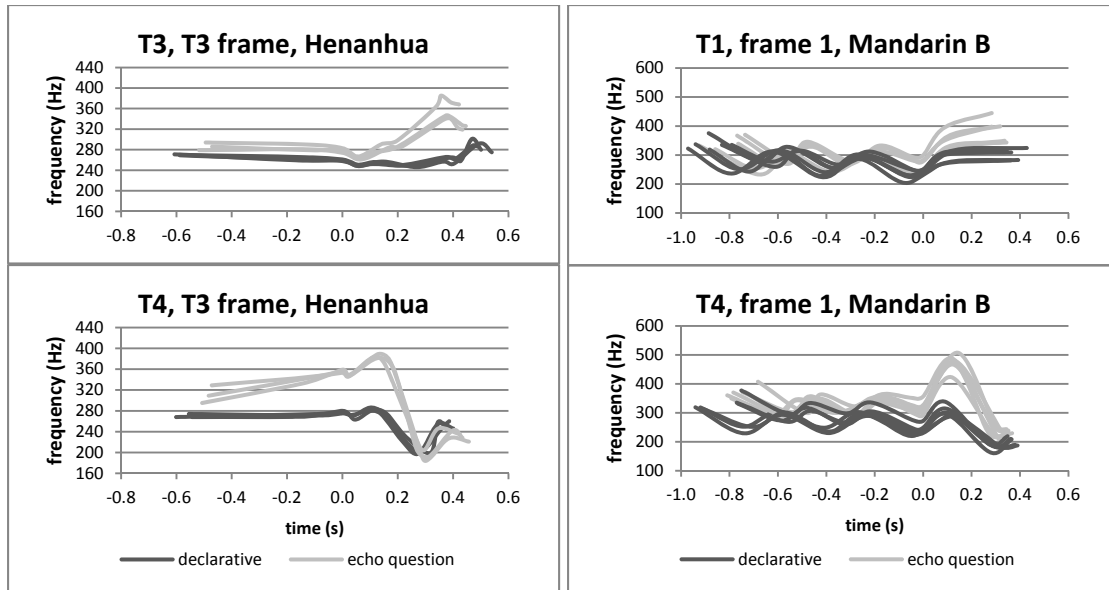
elicited the highest rate of accuracy; T4, one of the two falling tones, appears to have elicited the lowest rate of accuracy, but the robustness of this result needs to be checked in future studies with more data.

### 3.3.4 Discussion

These (admittedly preliminary) perceptual results indicate that tones in Henanhua are highly recoverable, with an occasional T2-T4 confusion when the stimuli are presented in isolation. The confusion is not surprising, given that both tones are falling (in both declarative and echo question contexts). In fact, the fact that only 3 out of 48 (6.25%) T2/T4 stimuli were miscategorized is quite impressive; we will see in Section 3.5 that the NKK speaker could only distinguish her own final-accented and double-accented monosyllables at a rate slightly greater than chance.

With just one listener, it is unclear whether the fact that the only T2-T4 confusions happened in the isolation set is meaningful. However, as noted in Chapter 2, the pitch contour of the frame sentence portion of the utterance leading up to T4 was different from that leading up to T2 in the echo question condition, and it would be worth finding out, with a broader study, if that intonational condition specifically resisted T2-T4 confusion the most. If so, it would provide an interesting contrast with the results for Mandarin, which indicated that the echo question renditions of T2 and T3 (in a frame sentence) yielded the *highest* rate of confusion, presumably because the echo question context rendered them more *similar* to one another.

As for the intonation results, the tone that yielded the lowest rate of accuracy was T4, one of the falling tones, and the tone that yielded the highest rate of accuracy was T2, the level/“gently rising” tone. Recall that in Mandarin the situation was reversed: the falling tone yielded the highest rate of accuracy while the level and rising tones yielded the lowest. We can attempt to reconcile this apparent discrepancy by revisiting the “difference-robustness” results for the two dialects from Chapter 2. They are shown here in Figure 3.10. It seems that these multiple-contour plots only provide us with half of a potential explanation for the discrepancy.



**Figure 3.10: Multiple-contour plots for Henanhua (left) and Mandarin (right), repeated from Chapter 2.**

T3 in Henanhua was much more distinctive across intonations than T1 was in Mandarin; this could explain why the Henanhua listener did so well on the former but Mandarin listeners struggled on the latter. However, T4 in Henanhua was just as distinct across intonations as T4 in Mandarin (if not more so), and yet the perceptual results for the two T4s were starkly different.

It should be noted that none of the Henanhua T4s that elicited intonational confusion were misidentified as T2s; that is, the listener accepted them all (correctly) as T4s. This discrepancy across dialects warrants further study.

Table 3.7 is a schematization of reinterpretations for Henanhua akin to the one shown in Section 3.2.4 for Mandarin. For the purposes of this schema, Henanhua T1 is considered a rising

**Table 3.7: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations in Henanhua**

underlying combination	surfaces as	occasionally <sup>30</sup> reinterpreted as
T↗ + I↗	rise	T↗ + I↘
T↗ + I↘	rise	n/a
T↘ + I↗	fall	T↘ + I↘
T↘ + I↘	fall	n/a

<sup>30</sup> More than 10% of the time.

tone and Henanhua T2 and T4 are considered falling tones. Note that both rising tones and falling tones in the echo question context (in the white rows) were reinterpreted as declarative. Like in Mandarin, the direction (i.e. the sign of the slope) of the contour uniquely reflects the direction of the tone, regardless of the intonation. As a result, the intonation type does not interfere with tone identification. In the next section, we will see that the situation is different in Cantonese.

### **3.4 Perception in Cantonese**

#### **3.4.1 Subjects**

There were 20 subjects who all identified themselves as native speakers of Hong Kong Cantonese. Although age and sex were not evenly distributed, they ranged in age from 19 to 63 and included both males and females. With three exceptions, they all lived in Hong Kong for at least the first 15 years of their lives (Subject C lived in Sichuan for three years before living in Hong Kong for 16 years, Subject G lived in Canton for two years before living in Hong Kong for 60 years, and Subject R lived in Canton for three years before living in Hong Kong for 52 years). They all had parents who spoke Cantonese at home.

#### **3.4.2 Stimuli**

All stimuli were presented as audio clips played over high-quality headsets. Each subject was presented with a total of 144 utterances, satisfying the conditions given in (3.3):

(3.3) Cantonese perceptual experiment conditions

6 tonal categories (Tone 1, Tone 2, Tone 3, Tone 4, Tone 5, and Tone 6)

2 intonational categories (declarative statement vs. echo question)

6 repetitions

2 speakers



The stimuli were presented in two parts—72 stimuli produced by one speaker and 72 stimuli produced by another speaker. Within each part the order of the stimuli was randomized each time and the order in which the parts were presented was switched each time (i.e. half the subjects heard Speaker A’s utterances first and half the subjects heard Speaker B’s utterances first). A short break was taken between the two parts during which background information about the subject was obtained. The subjects indicated their responses on a multiple choice answer sheet, a portion of which is shown in Figure 3.11. The frame sentence, *lei6lei6wa6...* ‘Leilei says...’ appears in the upper left-hand corner. The first six blank columns correspond to

莉莉話...								
	“欄”	“調”	“癩”	“蘭”	“懶”	“爛”	。	?
	豬欄 牛欄 果欄	調醒 調聰明 調有寶	蚊癩 風癩 生癩	蘭花 蘭桂坊 紫羅蘭	懶惰 懶蟲 偷懶	燦爛 腐爛 爛茶渣		
1								
2								
3								
4								
5								
6								
7								

**Figure 3.11: A sample answer sheet for the Cantonese perceptual test.**

the four tonal categories, which are represented by the words *laan1* ‘market’, *laan2* ‘claim’, *laan3* ‘rash’, and *laan4* ‘orchid’, *laan5* ‘lazy’, and *laan6* ‘decay’, respectively. The heading of each column includes the target word at the top and three example compounds utilizing the target word underneath. This was done because written Cantonese is not as commonly used as written Mandarin, and also because in some cases more than one tone may be associated with a given character depending on the compound it is in. The last two blank columns correspond to the two intonational categories, *statement* and *echo question*, respectively. Upon hearing a given stimulus twice, the subject was asked to indicate her choice of tonal category and intonational

category by writing a number in the corresponding box for each, the number being an indicator of confidence on a scale of 1 to 5.

### 3.4.3 Differences from Mandarin design

There are several aspects of the design worth noting, especially in comparison with that of the Mandarin experiment. First, since Cantonese has six contrastive tones compared to only four in Mandarin, the total number of utterances was greater for Cantonese, which made the whole test take considerably longer (about 35 minutes on average as opposed to 20 minutes for Mandarin). Also, as mentioned above, written Cantonese is less commonly used than written Mandarin, so care had to be taken to ensure that each subject was assigning the correct tone to each column. This was done by asking each subject to read the characters in the column headings aloud before starting the test. Finally, whereas the frame sentence for Mandarin consisted of three syllables bearing a falling tone (Tone 4 in Mandarin), the frame sentence for Cantonese consisted of three syllables bearing a low level tone (Tone 6 in Cantonese). As a result, the tonal context leading up to the target word was slightly different in the two experiments.

### 3.4.4 Results

The overall rate of accuracy for tone identification was 72.85%. The rate of accuracy for intonation identification was substantially higher, and in fact the same as that for Mandarin: 87.15%. These results are shown in Table 3.8. The same GEE model that was run on the

**Table 3.8: Overall rates of perceptual accuracy in judging tonal and intonational categories in Cantonese**

function	rate of accuracy
tone	72.85%
intonation	87.15%

Mandarin results was run on the Cantonese results. The difference between the rate of accuracy for tone and that for intonation was again found to be highly significant ( $p < .001$ ). Note that, while the rate of accuracy for intonation was the same as for Mandarin, the rate of accuracy for

tone was much worse, the result being that the relationship between the rates of accuracy for each of the melodic functions in Cantonese is the reverse of that in Mandarin.

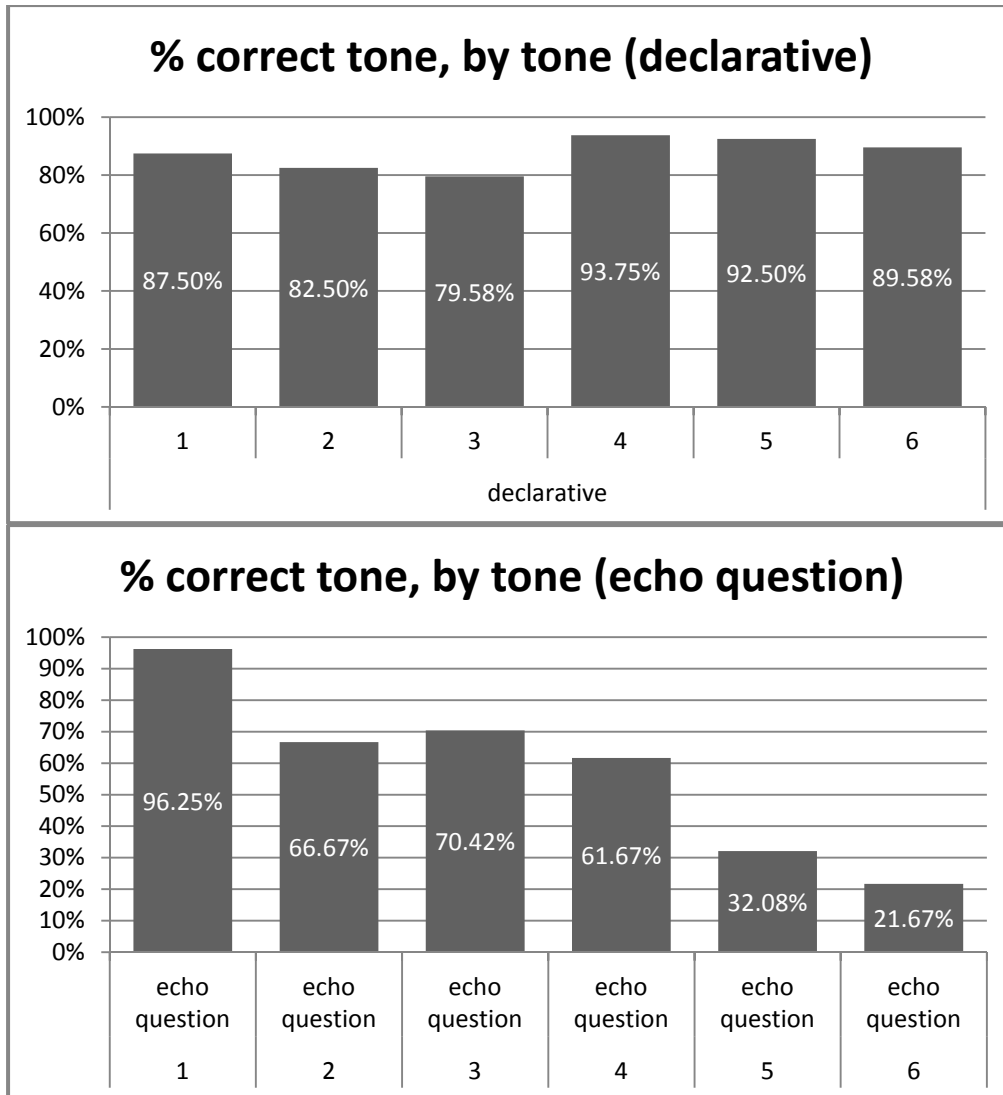
Splitting up tonal accuracy by intonation type, as shown in Table 3.9, we see a striking contrast between the intonation conditions—tone identification was much worse in the echo question condition than in the declarative condition. The difference between these rates was

**Table 3.9: The rate of accuracy in judging tonal category, by intonational category of the stimulus, in Cantonese.**

intonational category	rate of accuracy for tone
declarative	87.57%
echo question	58.12%

highly significant ( $p < .001$ ). This contrast presents a further difference from Mandarin, where the intonation condition hardly made any difference for tone identification. It is interesting to note that when Speaker A from the production experiment listened to her own speech, her rates of accuracy were even more separated, at 97.22% and 52.78%, respectively (her results were not pooled with the general results above). These perceptual results were not in line with her intuition during the production experiment that she was producing reliable cues to all tonal contrasts in both intonational environments.

Let us now break down the results further, by tonal category, in each intonation condition. This is shown in Figure 3.12. While there is some variability in the declarative condition, it is apparent that there was a much bigger effect of tonal category in the echo question condition. A post-hoc analysis of the results, with a Bonferroni correction, indicated that none of the differences in the declarative condition were significant, but that several of the differences in the echo question condition were significant. These post-hoc pairwise comparisons are shown in Table 3.10. In particular, the rate for T1 was significantly higher than those for all other tones. We then see T2, T3, and T4 clustering in the middle and T5 and T6 receiving the worst rates of accuracy.



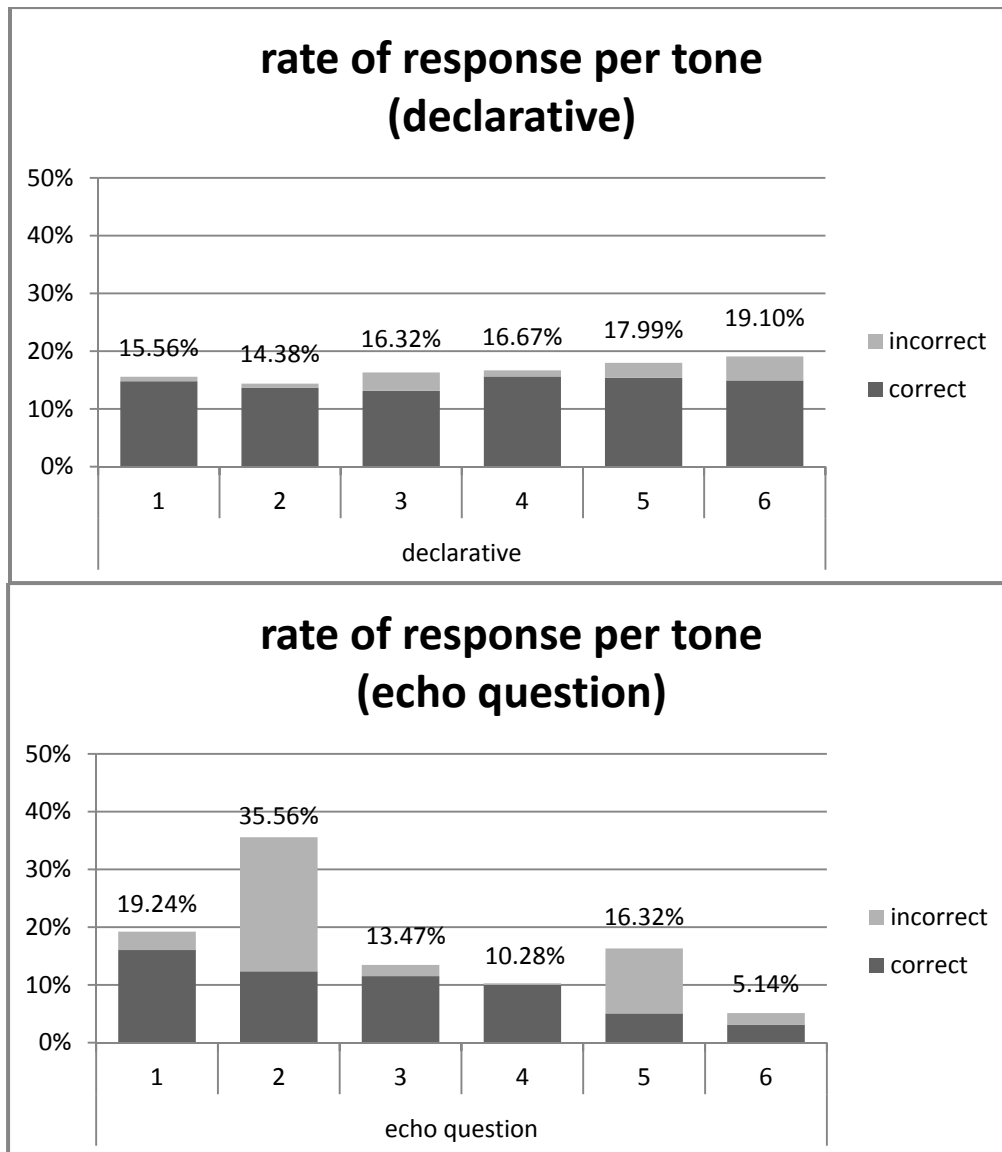
**Figure 3.12: Rate of perceptual accuracy for tone, by tone and intonation type of the stimulus, in Cantonese.**

The results thus far paint an interesting picture of tone perception in Cantonese, but it is not the whole picture. To see if there were any tendencies for listeners to respond with one or more of the tones more often than with the others, a bin test similar to the one performed for intonation responses in Mandarin was performed for tone responses in Cantonese. The results, broken down by tone in each intonation condition, are shown in Figure 3.13. It is clear that the rate of response, while being relatively even across the tonal categories in the declarative condition, was quite uneven in the echo question condition. The notable exceptions in the declarative condition are slight over-responding in the categories of T3, T5, and T6. We will see

**Table 3.10: Pairwise comparisons of tonal accuracy among tones in the echo question condition for Cantonese.**

(I) Tone	(J) Tone	Mean Difference (I-J)	Bonferroni Sig.
T1	T2	<b>.30*</b>	<b>.002</b>
	T3	<b>.26*</b>	<b>.018</b>
	T4	<b>.35*</b>	<b>.000</b>
	T5	<b>.64*</b>	<b>.000</b>
	T6	<b>.75*</b>	<b>.000</b>
T2	T3	-.04	1.000
	T4	.05	1.000
	T5	.35	.141
	T6	<b>.45*</b>	<b>.000</b>
T3	T4	.09	1.000
	T5	<b>.38*</b>	<b>.000</b>
	T6	<b>.49*</b>	<b>.000</b>
T4	T5	<b>.30*</b>	<b>.000</b>
	T6	<b>.40*</b>	<b>.000</b>
T5	T6	.10	1.000

this trend illuminated further in the confusion matrices below. As for the echo question condition, there was a substantial favoring of T2 and there was a secondary favoring of T5. As noted in the discussion of the production results for Cantonese in Chapter 2, the echo question renditions of all of the Cantonese lexical tones bear a resemblance to the rising contours of T2 and T5 (more so the former than the latter), and so these biases are not surprising. The primary bias toward T2 may partially explain the relative rates of accuracy for the six tones in the echo question context shown above in Figure 3.12. If we order the tones in terms of the proximity of their average minimum  $F_0$  (i.e. the  $F_0$  at the beginning of the final rise for each tone) to that of T2, we get the following order: T5 and T6 first, then T4 and T3, and finally T1. This order is reflected inversely in the relative rankings of the tones in terms of perceptual accuracy in the echo question context, and we will see in a moment that the order crops up again in terms of how



**Figure 3.13: The bin test for tone responses in a declarative context (above) and in an echo question context (below) in Cantonese.**

often each tone was confused with T2 (T5 and T6 were most often confused with T2; T1 was least often confused with T2).

As was done for Mandarin, confusion matrices for the different intonation conditions in each stimulus set condition are displayed in Figure 3.14. The two upper matrices show results from Stimulus Set A and the two lower matrices show those from Stimulus Set B. The left-hand matrices show results from the declarative condition, while the right-hand matrices show results from the echo question condition. Right away it is apparent that the patterns of confusion are

stim	T1	T2	T3	T4	T5	T6
T1	79	1	18	1	0	1
T2	0	84	0	0	14	2
T3	0	2	68	1	0	30
T4	7	0	0	92	0	2
T5	0	1	2	0	90	8
T6	0	0	5	6	0	89

Stimulus Set A, declarative

stim	T1	T2	T3	T4	T5	T6
T1	97	0	3	0	0	1
T2	2	68	1	2	28	1
T3	15	9	70	0	3	3
T4	3	31	0	54	12	0
T5	0	62	1	1	32	5
T6	1	49	2	1	33	15

Stimulus Set A, echo question

stim	T1	T2	T3	T4	T5	T6
T1	96	1	3	0	0	1
T2	1	80	1	3	15	1
T3	2	0	90	2	0	7
T4	3	0	1	96	1	0
T5	0	4	1	0	95	0
T6	0	0	8	1	1	90

Stimulus Set B, declarative

stim	T1	T2	T3	T4	T5	T6
T1	96	1	3	0	0	0
T2	0	66	0	0	32	3
T3	12	17	67	0	2	3
T4	5	20	0	66	8	1
T5	0	63	2	0	28	8
T6	1	43	14	0	20	23

Stimulus Set B, echo question

**Figure 3.14: Tonal confusion matrices for Cantonese.**

quite different from those in Mandarin; there was more confusion overall, but there was also a much bigger difference between the declarative and echo question conditions. Apart from the differences from Mandarin confusion, it is striking how consistent the confusion patterns were across the stimulus set conditions, i.e. not only were the rates of confusion for each tone very similar across stimulus set conditions, but the distribution in terms of which tones were interpreted as which other tones and how often was remarkably consistent in the two sets. This

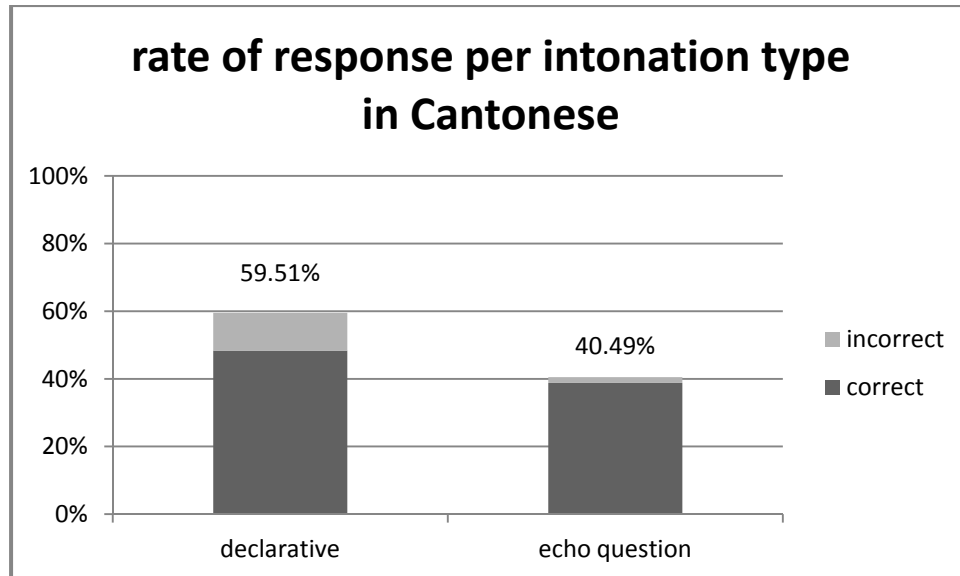
indicates that listeners were not merely guessing but rather picking up on certain cues they heard being produced by both speakers alike in the various tone-intonation combinations and interpreting those cues in a knowledge-based way. One notable exception to this cross-stimulus set consistency is a slightly higher rate of confusion in the declarative condition for Stimulus Set A, specifically in the case of T1, T2, and T3, which were more often confused for T3, T5, and T6, respectively, in that stimulus set. Recall that these were the tonal categories that received slightly higher rates of over-responding in the bin test. The “interfering” tones in these cases were those tones whose contours were the same but whose register was one category lower than that of the target tones (for example, T1, the high level tone, was mistaken for T3, the mid level tone). Note that the interfering effect of T3 on T1 was drastically reduced in the echo question condition for Stimulus Set A. The biases toward T2 and T5 in the echo question condition manifest themselves as vertical “smears” in the columns under “T2” and “T5”, respectively, in the echo question matrices of both stimulus sets. Focusing for a moment on the T2 columns in each of the echo question matrices, note that if we rank the tones other than T2 in terms of how often each one was interpreted as T2, we get  $T5 > T6 > T4 > T3 > T1$ , which is very nearly the inverse of the rankings of perceptual accuracy given above in Figure 3.12 (the relative positions of T5 and T6 being switched<sup>31</sup>), indicating that a large part of the degradation in positive identification for each tone is explained by how often it was reinterpreted as T2. These confusion patterns in the echo question condition are very much in line with the results reported by Ma, Ciocca et al. (2006), who performed a very similar perceptual test for Cantonese.

Let us now turn to intonation perception in Cantonese. Figure 3.15 shows the bin test for intonation responses in Cantonese. According to the bin test there is again an apparent bias toward perceiving declarative intonation, just as there was in both Mandarin dialects. In fact, the

---

<sup>31</sup> This reversal of T5 and T6 in the rankings makes sense, since reinterpretations for T6 were split between T2 and T5, whereas T5 was mainly perceived as itself or as T2.





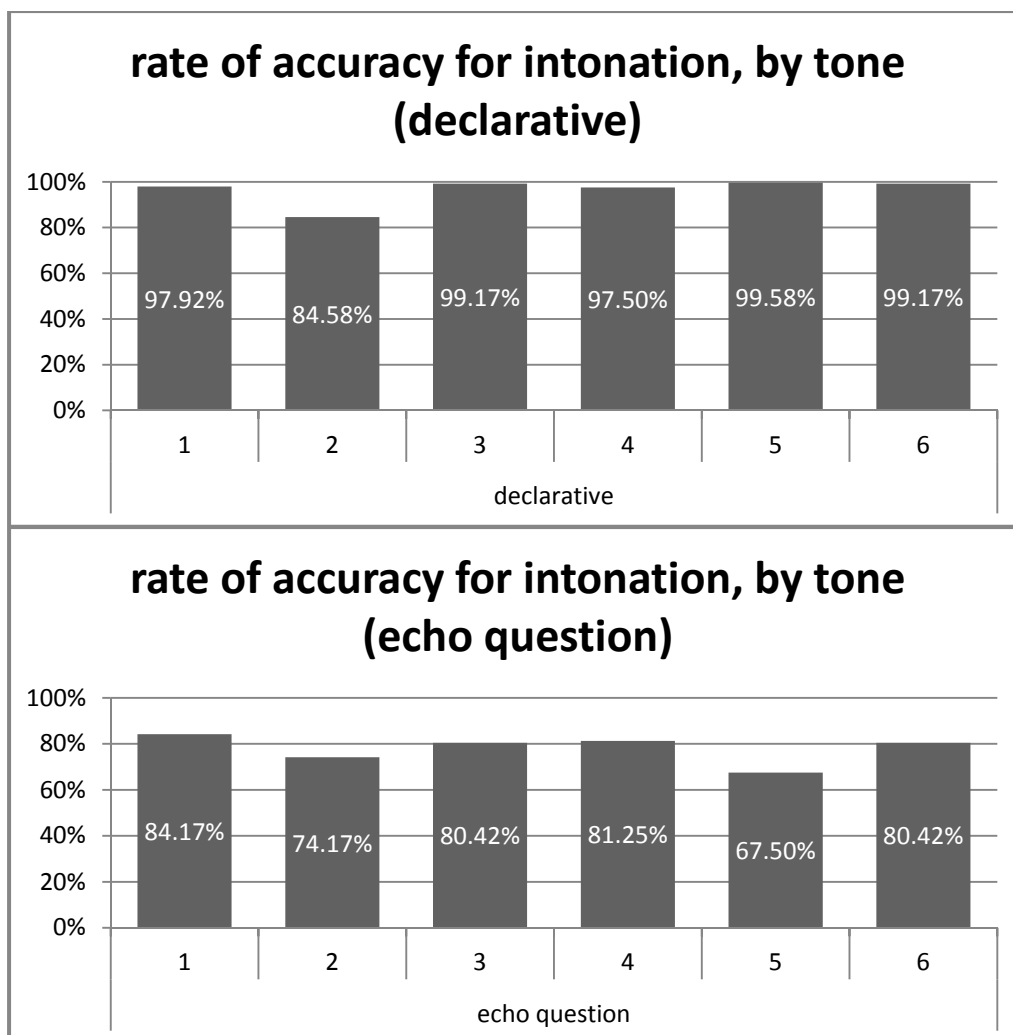
**Figure 3.15: Bin test for intonation responses in Cantonese.**

split is nearly identical, at 60-40 again. Table 3.11 shows the respective rates of perceptual accuracy for intonation. While the rate of accuracy for declarative intonation was nearly perfect at 96.32%, the rate of accuracy was much lower for echo question intonation, at 77.99%.

**Table 3.11: Rate of perceptual accuracy in judging intonational category, by intonational category, in Cantonese**

intonation type	rate of accuracy
declarative	96.32%
echo question	77.99%

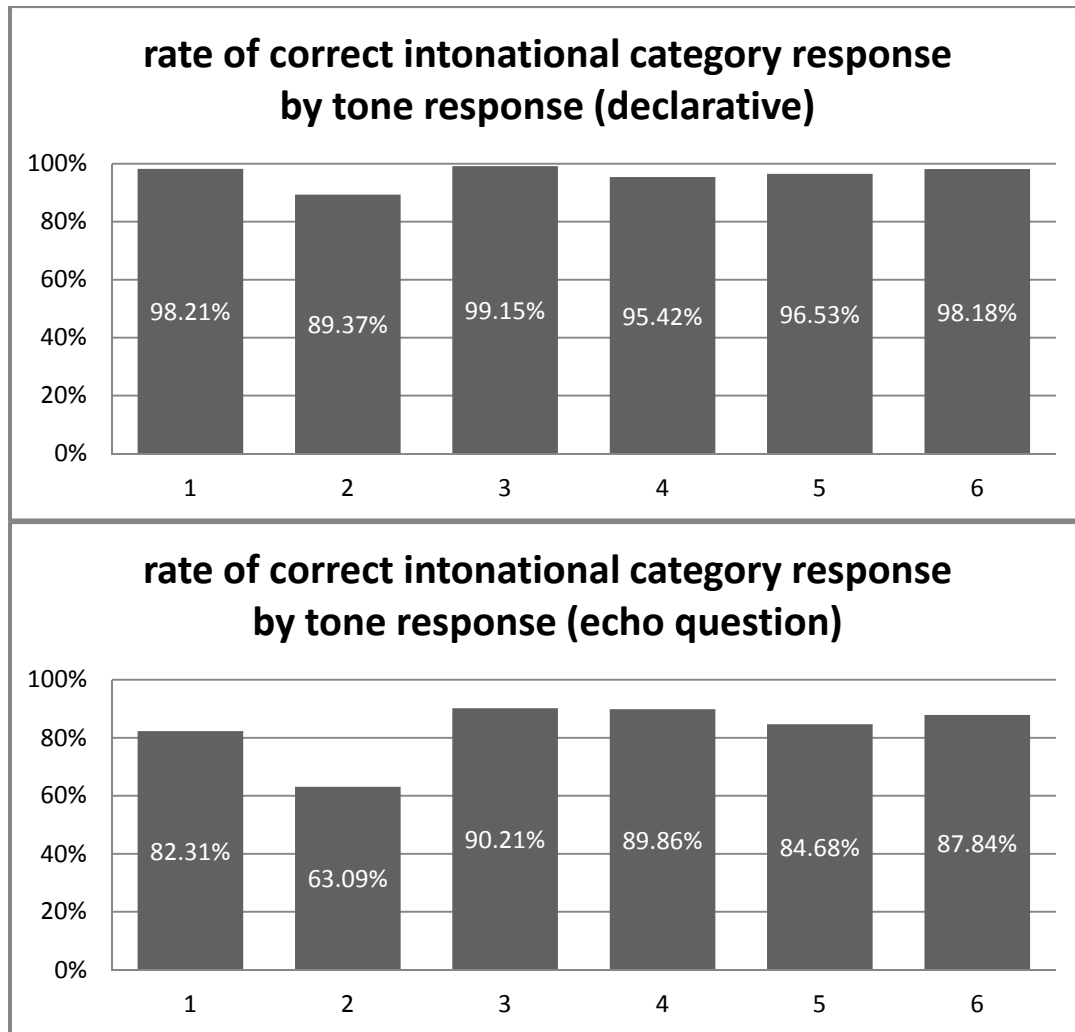
Breaking down the results by tonal category in Figure 3.16, we see that T2 had a negative effect on intonation perception in the declarative condition and T5 had a negative effect on intonation perception in the echo question condition. The post-hoc analysis with a Bonferroni correction indicated that the difference between T2 and all the other tones in the declarative condition was highly significant ( $p \leq .001$  for all pairs) and the difference between T5 and T6 in the echo question condition was significant ( $p = .032$ ). We might wonder if settling on a given tone response made listeners more or less likely to perceive the correct intonation for certain



**Figure 3.16: The rate of accuracy in judging intonational category, by tonal category of the stimulus, in Cantonese—the declarative condition above and the echo question condition below.**

tone-intonation combinations. This is the motivation behind Figure 3.17, which gives the percentage of ‘declarative’ responses for each tone response, in each of the intonation conditions. Figure 3.17 might be a bit opaque to the reader at first glance. What these graphs show us is, when listeners reported hearing a given tone in a given intonation condition, what percentage of the time they judged the intonational category correctly. In the declarative condition, the rates were close to 100%, although when listeners reported hearing T2 there was a slight drop in the rate of ‘declarative’ responses. On the other hand, we see an even more substantial effect of T2 in the echo question condition. The rate of correct intonational category responses dropped to

63.09% when the tone response was T2—in other words, about 37% of the time when listeners reported hearing T2 in the echo question condition, they reported (incorrectly) hearing declarative intonation. We will return to this finding in the next section.



**Figure 3.17: Rates of correct intonational category response by tone response, in the declarative condition (above) and the echo question condition (below), in Cantonese.**

### 3.4.5 Discussion

When it comes to pure tone identification, it is fair to say that Cantonese listeners are generally at a disadvantage compared to Mandarin listeners. With no pragmatic, semantic, or syntactic context clues, the listeners in this experiment were asked to differentiate six tones, as opposed to four in Mandarin. Perhaps more crucially, three out of the six (T1, T3, and T6) are level tones

that contrast among one another mainly in register and two out of the six (T2 and T5) are rising tones that contrast with each other mainly in register (Ma, Ciocca et al. 2006). It is thus expected that Cantonese listeners' rates of perceptual accuracy would be degraded in comparison with those of Mandarin listeners. However, we have seen that there is also an effect of intonation on tone identification, namely that the echo question melody on non-rising tones is often perceived as a tonal melody and reinterpreted as being one of the rising tones, T2 or T5<sup>32</sup>. This reinterpretation is perhaps spurred on by a general bias towards expecting and perceiving declarative intonation. In other words, when given a choice of interpreting a rising melody as echo question intonation or a rising lexical tone, the bias towards declarative intonation may have swayed listeners to favor the rising lexical tone interpretation. Indeed, the results in Figure 3.17 would seem to indicate that the identification of the lexical tone as T2 often gave listeners the "license" to perceive declarative intonation, fulfilling the bias. It was also noted that the ranking of the rates of positive identification for the five tones besides T2 largely correlated inversely (with the relative rankings of T5 and T6 being switched) with the ranking of how often each of those tones was reinterpreted as T2. Looking at the acoustic results for Cantonese in Chapter 2, this ranking order is perhaps not surprising, since it can be roughly derived by looking at the proximity of the mean minimum  $F_0$  for each tone in the echo question context (i.e. the starting pitch of its final rise) to that of T2: T5 and T6 were the closest, followed by T4, with T1 and T3 the farthest away. Although T1 and T3 had very similar minimum pitches, T3 was more often confused for T2 than T1. This could be due to the fact that the *shape* of the final rise for T1 was unique among all the tones; while the final rise for each of the other five tones had an exponential character, peeling away from the minimum, the rise for T1 was more linear. It

---

<sup>32</sup> Since the favoring of the T2 and T5 categories was limited to the echo question condition, we can surmise that, unlike the intonation bias, which presumably is due to the relative markedness of echo questions, uneven distribution in this case does not reflect a lexical frequency effect. Indeed, although no official Cantonese lexical frequency data was available to be consulted, the bin test results for the echo question condition did not correlate with one of the native speakers' intuitions on the relative frequencies of the words in the word list. Instead, it seems safe to surmise that the uneven response distribution was due to an interference effect from the intonational melody.

seems that the perceptual patterns reflect the tone-dependent nature of the echo question intonation that was revealed in Chapter 2!

Interestingly, the interference effect goes both ways, i.e. not only did the intonational category of the stimulus affect the perception of the tonal category, but the tonal category of the final syllable in the stimulus affected the perception of the intonational category. This was illustrated in Figure 3.16, where we saw that T2 stimuli caused a lower rate of accuracy for intonation in general and T5 stimuli caused a lower rate of positive identification of echo question intonation. Furthermore, the two forms of interference are somehow complementary in that the two tones that interfere the most with intonation perception are precisely the two tones that other tones are mistaken for in the presence of echo question intonation. This property of the Cantonese melodic system further sets it apart from that of both dialects of Mandarin, in which the direction of interference is mainly one-way (tonal category affects intonation perception). We can see in Table 3.12 how this difference is reflected in the reinterpretation schematization (only T2 is considered as a “rising” lexical tone for the purposes of this schema). In contrast with Mandarin and Henanhua, any melodic combination that has a rising component from either the tone or the intonation surfaces as a rising contour and leaves room for reinterpretation. Like Mandarin and Henanhua, however, the falling tone-declarative intonation combination (in the shaded row) resists reinterpretation.

**Table 3.12: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations in Cantonese**

underlying combination	surfaces as	occasionally <sup>33</sup> reinterpreted as
T↗ + I↗	rise	T↗ + I↘
T↗ + I↘ <sup>34</sup>	rise	T↘ + I↗
T↘ + I↗	rise	T↗ + I↘ or T↗ + I↗
T↘ + I↘	fall	n/a

<sup>33</sup> More than 10% of the time.

<sup>34</sup> This type of reinterpretation only applied in the case of underlying declarative T2, which was interpreted as bearing echo question intonation 13.04% of the time. Declarative T5 got reinterpreted less than 1% of the time.

Aside from the differences between Cantonese and Mandarin that have been brought to light by this and other studies such as Ma, Ciocca et al. (2006), there are a couple of generalizations that can be made based on certain common aspects of the perceptual results for the two studies. First, as was noted above, both groups of listeners displayed the same rate of bias (60-40) toward perceiving declarative intonation (see Figure 3.4 and Figure 3.15), which skewed their responses in such a way that the resulting rates of accuracy of intonation perception for the respective intonational conditions was virtually the same in the two languages, about 96-98% for the declarative conditions and about 78% for the echo question conditions (see Table 3.3 and Table 3.8). Further, the handful of cases where echo question intonation *did* interfere with tone identification in Mandarin (the T2-T3 confusion), it was precisely in those cases where the local melodic effects were such that the final trajectory of the  $F_0$  curve was steepened in a positive direction, effectively making the net melodic effect of echo question intonation on those tonal contours (T2 and T3) more like the general melodic effect of Cantonese echo question intonation. In the next section we will see how perception is affected in NKK, whose lexical tonal system is quite different from those of Mandarin and Cantonese.

### **3.5 Perception in North Kyeongsang Korean**

#### **3.5.1 Subjects**

There were 24 subjects who all identified themselves as native speakers of North Kyeongsang Korean. Although age and sex were not evenly distributed, they ranged in age from 20 to 62 and included both males and females. With one exception, they all lived in Daegu in the North Kyeongsang Province for their entire lives (one subject lived in nearby Yeongju, for one year at age 6). They all had parents who spoke NKK at home.

#### **3.5.2 Stimuli**

All stimuli were presented as audio clips played over high-quality headsets. Each subject was presented with a total of 192 stimuli, satisfying the conditions given in (3.4):

(3.4) NKK perceptual experiment conditions

16 words (representing 3 tonal categories: initial-, double-, and final-accented)

2 intonational categories (declarative statement vs. echo question)

2 contexts (isolation vs. frame sentence)

3 repetitions

The specific words in the word list are given in Table 3.13. The frame sentence, when used, was *Eunhi-neun...* ‘Eunhi-TOP...’ (i.e. ‘As for Eunhi...’). There were only stimuli from one speaker.

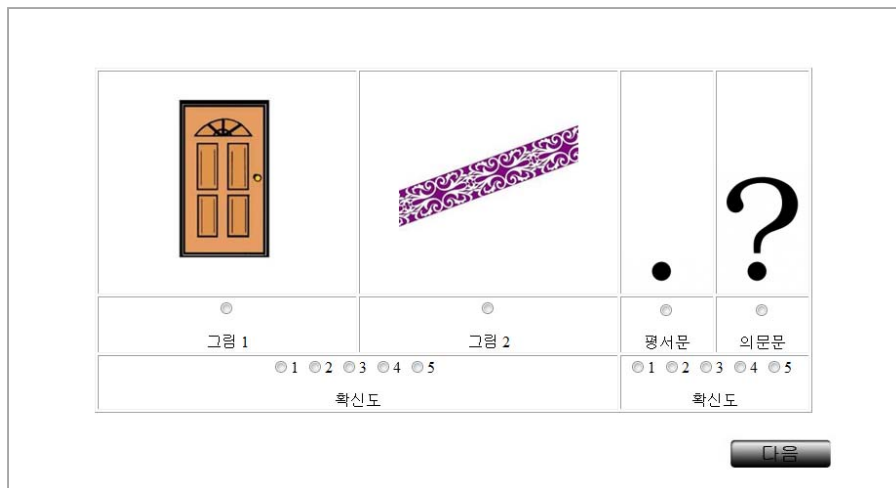
**Table 3.13: Word list for NKK perceptual test**

syllable count	tonal category	words
monosyllabic <sup>35</sup>	initial-accented	<i>mal</i> ‘horse’ <i>nam</i> ‘south’
	double-accented	<i>mal</i> ‘end’ <i>nam</i> ‘third party’
disyllabic	initial-accented	<i>mal-i</i> ‘horse-NOM’ <i>nam-i</i> ‘south-NOM’ <i>uli</i> ‘pigpen’ <i>mun-i</i> ‘door-NOM’
	double-accented	<i>mal-i</i> ‘end-NOM’ <i>nam-i</i> ‘third party-NOM’ <i>nai</i> ‘age’ <i>meil</i> ‘every day’
	final-accented	<i>uli</i> ‘us’ <i>muni</i> ‘pattern’ <i>Nahi</i> ‘Nahi’ (feminine name) <i>meil</i> ‘e-mail’

The stimuli were presented in two parts—96 single-word stimuli followed by 96 frame sentence stimuli. Within each part the order of the stimuli was randomized each time. A short break was taken between the two parts. Unlike for Mandarin and Cantonese, the test was administered on a computer, and instead of indicating their responses on a paper answer sheet, subjects were presented with cartoon representations of the words on the computer monitor and asked to indicate their choices by clicking on bubbles on the screen. As each new screen loaded, the

<sup>35</sup> Recall that final-accented monosyllabic words do not exist in NKK.

corresponding audio stimulus was automatically played twice with a two-second gap in between. An example screenshot is shown in Figure 3.18. Here we can see that the first row of cells



**Figure 3.18: A sample screenshot from the NKK perceptual test.**

correspond to two lexical choices (initial-accented *mun-i* ‘door-NOM’ vs. final-accented *muni* ‘pattern’) and two intonational choices (*statement* vs. *question*). The row below that provides the subject with bubbles on which to click to choose one of the lexical choices and one of the intonational choices. The bottom row provides the subject with bubbles corresponding to five levels of confidence for each choice. The button at the bottom right is a ‘next’ button, to be clicked on by the subject when she is satisfied with her responses for the current stimulus. The order of the cartoon figures on the page was randomized (e.g. the picture of the door did not always appear on the left).

### 3.5.3 Differences from Mandarin and Cantonese designs

There are several aspects of the design worth noting, especially in comparison with that of the Mandarin and Cantonese experiments. Perhaps the two most significant differences are that cartoon illustrations were used instead of orthography and that no more than two tonal categories were being presented as choices at any given time (as opposed to four in Mandarin and six in Cantonese). The former was due to the fact that Korean orthography does not disambiguate



segmentally identical minimal pairs. As for the latter, while there are three contrastive tonal categories on disyllabic words in NKK (initial-, double-, and final-accented), it was not possible to find minimal triplets that satisfied the other criteria in the production experiment (nouns comprised of sonorant segments). Therefore the word list was filled out with two-way comparisons of the three tonal categories (initial- vs. double-, initial- vs. final-, and double- vs. final-). Finally, whereas in the Mandarin and Cantonese experiments all stimuli included frame sentences, the first part of the NKK test included only words in isolation. The context condition was introduced here in lieu of a speaker condition. While a similar context condition would have been interesting to include in the other experiments, it would have doubled the total number of stimuli, making the test unwieldy and impractical to complete in a single session for each subject.

### **3.5.4 Results**

In assessing the results for tone identification, it is important to keep in mind that the task for NKK listeners was not exactly equivalent to that of Mandarin and Cantonese listeners in that they only ever had to choose from two tonal categories, as opposed to four or six, respectively, for the other two languages. This means that pure guessing would result in rates close to 50% (compared with 25% for Mandarin and 17% for Cantonese). It is true that the comparison between Mandarin and Cantonese was also not perfect since the total number of choices was different in each of those languages, but in a sense the comparison there was fair because the number of choices was a direct reflection of each language's natural tonal inventory. While there are indeed only two tonal categories on monosyllables in NKK, there are actually three categories on disyllables, but we artificially limited the number of possible choices to two on disyllables as well. Keeping all this in mind, let us look at the overall rates of accuracy for tone and intonation in NKK, shown in Table 3.14. The rate for intonation was strikingly similar to that for Mandarin and Cantonese, at 86.55%. Although the languages up until now have indeed displayed similar rates of perceptual accuracy for intonation, we will see in Section 3.6 that SJ will break this trend with a much higher rate. Furthermore, we have been seeing that the

**Table 3.14: Overall rates of perceptual accuracy in judging tonal and intonational categories in NKK**

function	rate of accuracy
tone	76.95%
intonation	86.55%

*relationship* between intonational accuracy and tonal accuracy is not the same across languages. The perceptual rate for tone in NKK was surprisingly low at 76.95%, making it appear similar to Cantonese, for which the tonal accuracy rate was 72.85%. If we break down the tone results by syllable count of the target word in the stimulus, however, we see that this overall rate was pulled down quite a bit by the rate for the monosyllabic condition. This breakdown is shown in Table 3.15. It is clear that listeners were basically guessing when the target word was

**Table 3.15: Rate of perceptual accuracy in judging tonal category, by syllable count of the stimulus, in NKK**

syllable count of target	perceptual accuracy for tone
monosyllabic	53.65%
disyllabic	84.72%

monosyllabic (remember there were only two tonal categories to choose from). This is not surprising considering how similar the contours of the stimuli from the different tone conditions were in the monosyllabic condition, as we saw in Chapter 2. Even the original speaker had trouble distinguishing her own tones on monosyllables, despite the fact that she had had the intuition that she was producing them distinctly. Her individual results (excluded from the pooled results above and subsequent results in this section) are shown in Table 3.16. Although

**Table 3.16: Rate of perceptual accuracy in judging tonal category, by syllable count of the stimulus, for NKK speaker A (listening to her own speech)**

syllable count of target	tonal accuracy
monosyllabic	62.50%
disyllabic	100.00%

she did slightly better than chance on her own monosyllables (62.5%), it was quite a low rate of accuracy compared to her perfect score (100%) on disyllables. The phrasal context (isolation vs. frame) did not have any effect; her rate of accuracy was exactly 62.5% in both contexts.

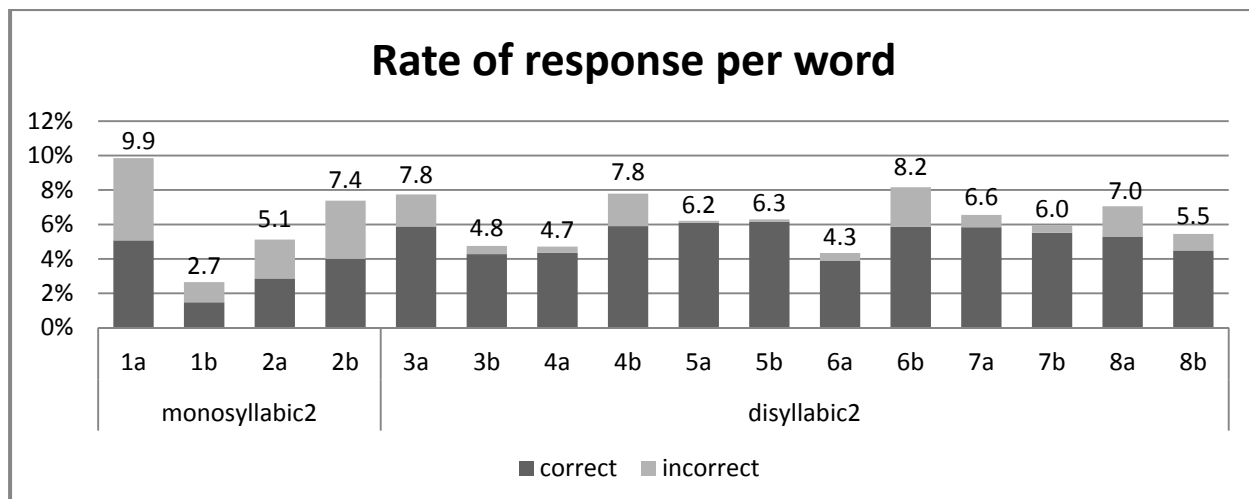
If we exclude the monosyllables from the pooled results, the revised overall rate of accuracy for tone is 84.72%. The revised rate for intonation, on the other hand, remains virtually unchanged, at 86.05%. These revised numbers are shown in Table 3.17. These adjusted rates

**Table 3.17: Revised rates of perceptual accuracy in judging tonal and intonational categories in NKK (excluding monosyllables)**

function	rate of accuracy
tone	84.72%
intonation	86.05%

were found not to be significantly different from one another ( $p = .510$ ). These revised numbers set NKK apart from the previously discussed languages, in that they suggest that one function is not more successfully encoded than the other in this language.

This revised rate for tonal accuracy is still remarkably low considering there were only ever two choices and the pitch cues in disyllables were quite distinct (as a non-native speaker the author was able to distinguish them easily). It turns out that there may have been a frequency effect depending on the minimal pair. Figure 3.19 shows a bin test by word for NKK. The words are given number and letter designations, where segmentally identical pairs share the same number (i.e. 1a and 1b are a minimal pair differing only in tone). The rate for each word has been rounded to one decimal place and % symbols have been omitted for ease of reading. Equal distribution would yield a rate of about 6% for each word. It is clear that the distribution was not equal among some of the pairs, namely the 1s, 2s, 3s, 4s, and 6s. These disparities are likely due to lexical frequency effects. Words 1b and 2a were *nam* ‘south’ and *mal* ‘end’, respectively. These are Sino-Korean words that are less likely to appear as stand-alone words than native Korean words, which likely led to the biases in those pairs (H.-S. Lee, personal communication).



**Figure 3.19: Bin test for word responses in the NKK perceptual test.**

Words 3b and 4a are those same words with a particle added. It is interesting to note that, while the biases in the 3s and 4s go in the same direction as those in the 1s and 2s, the biases are not as strong. It looks as if, when faced with a lack of pitch cues in the monosyllabic cases, listeners were swayed even more strongly by the lexical biases. As for the 6s, 6a was *uli* ‘pigpen’ and 6b was *uli* ‘we/us’. The word *uli* ‘pigpen’ appeared only 3 times in a corpus presented in Cho (2002), while *uli* ‘we/us’ appeared 6,583 times<sup>36</sup>. It is not unreasonable to assume that this large disparity between the levels of frequency of these words contributed to the unequal distribution in this case.

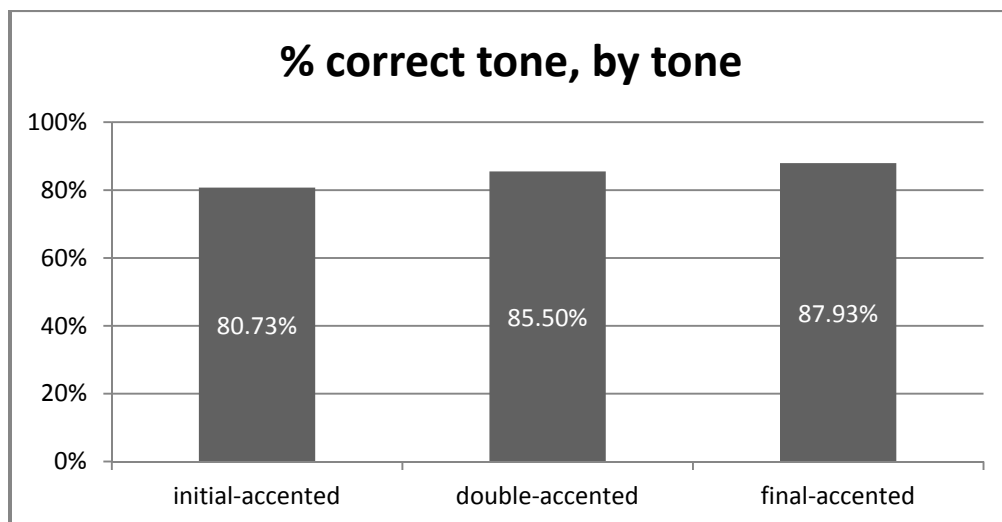
Keeping these lexical effects in mind, let us turn to the effect of intonation on tone identification. Table 3.18 shows the overall rates of accuracy for tone (excluding the monosyllabic condition) by intonation type: There was no significant effect of intonation type

**Table 3.18: The rate of perceptual accuracy in judging tonal category, by intonational category of the stimulus (excluding the monosyllabic condition), in NKK.**

intonational category	rate of accuracy for tone
declarative	85.13%
echo question	84.32%

<sup>36</sup> Thanks to Jiwon Yun for looking up the words and obtaining these frequency figures.

on tone identification ( $p = .531$ ). Just to check if there was any effect of tonal category on tone perception, the results were broken down by tonal category of the stimulus. This is shown in Figure 3.20. The final-accented condition yielded the highest overall rate of accuracy by a small



**Figure 3.20: Rate of perceptual accuracy for tone (excluding the monosyllabic condition), by tone of the stimulus, in NKK.**

margin, followed by double-accented, followed by initial-accented. The difference between initial- and final-accented was highly significant ( $p < .001$ ), but double-accented was not significantly different from either of the other two tonal categories. If we revisit the bin test results for individual words, we see that, of the three minimal pairs that seemed to be affected most by lexical biases, the “loser” was double-accented in one of the pairs (the 3s) and initial-accented in two of the pairs (the 4s and 6s). It seems safe to conclude, then, that any small effect seen here is ultimately traceable to the lexical biases. Since neither the intonation condition nor the tone condition seems to have had much effect on tone identification, it is not expected for the confusion matrices to yield any inherently interesting insights for NKK. However, it is perhaps worth showing them here (Figure 3.21) so that they can be compared with the confusion matrices for Mandarin and Cantonese. Like in Mandarin, the matrices seem pretty similar across the intonation conditions. The rate of accuracy was not as high as in Mandarin, however, leading to a lot of moderately shaded cells outside of the ideal diagonals.

	initial	double	final
stim			
initial	80	10	10
double	8	86	6
final	1	9	90

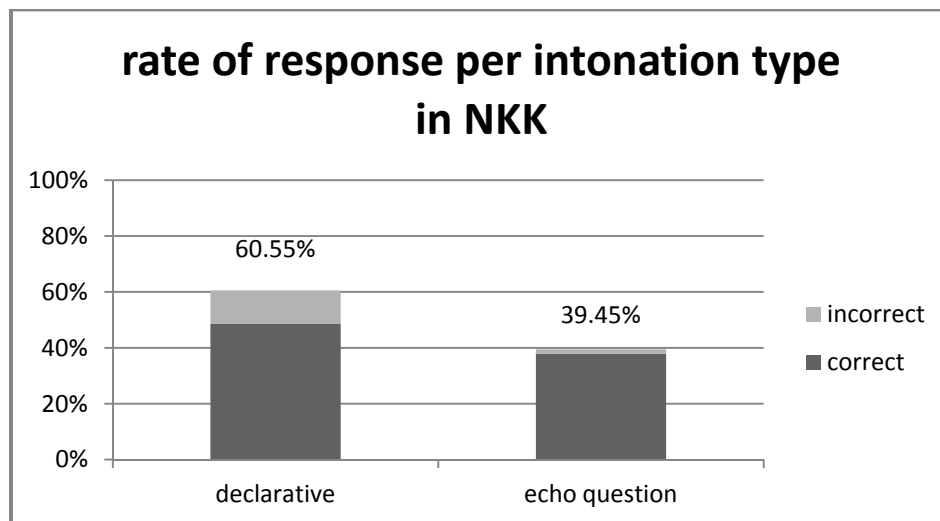
declarative

	initial	double	final
stim			
initial	82	9	9
double	10	85	5
final	3	11	86

echo question

**Figure 3.21: Tonal confusion matrices for the disyllabic condition in NKK.**

Let us turn now to intonation identification in this language. Figure 3.22 shows the bin test for intonation type in NKK. Once again, there is a 60-40 bias toward declarative intonation,



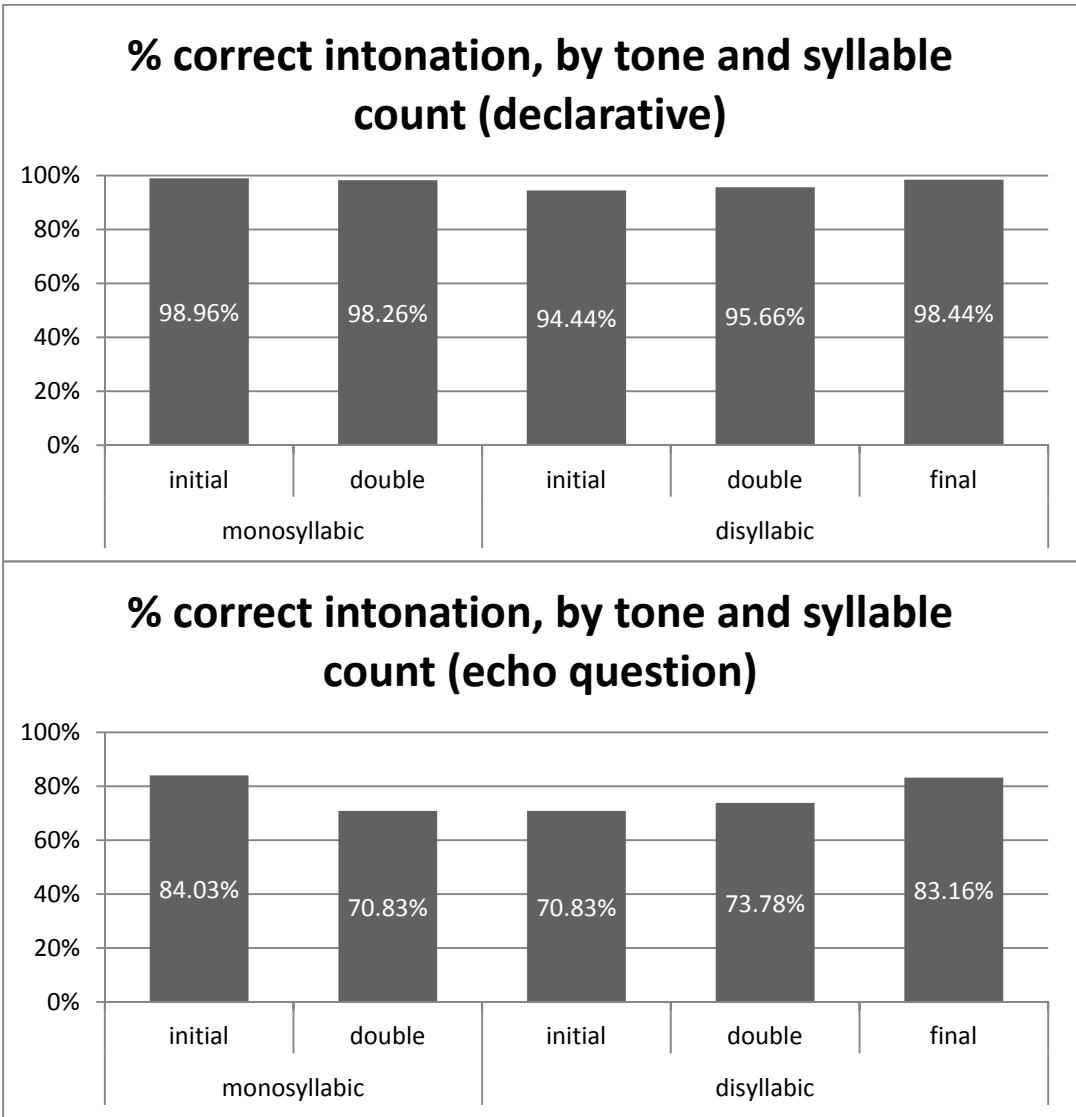
**Figure 3.22: Bin test for intonation responses in NKK.**

the same bias ratio as in Mandarin and Cantonese. As shown in Table 3.19, the rate of accuracy in the declarative condition was quite high at 96.18%, while the rate of accuracy in the echo question condition was much worse, at 75.93%. This difference was highly significant ( $p < .001$ ). Finally, the rate of intonational accuracy broken down by intonation, syllable count,

**Table 3.19: Rate of perceptual accuracy in judging tonal category, by intonational category of the stimulus, in NKK**

intonational category	rate of accuracy
declarative	96.18%
echo question	75.93%

and tonal category is shown in Figure 3.23. As we have seen for the other languages, the effect of tone appears to be exaggerated in the echo question condition. A post-hoc test with a Bonferroni correction indicated that, while not all of the differences among the rates were



**Figure 3.23: Rate of intonational accuracy by intonation, syllable count, and tonal category for NKK.**

significant, the rates for double-accented monosyllables and initial-accented disyllables were significantly lower than those for initial-accented monosyllables and final-accented disyllables ( $p < .05$ ) in the echo question condition.

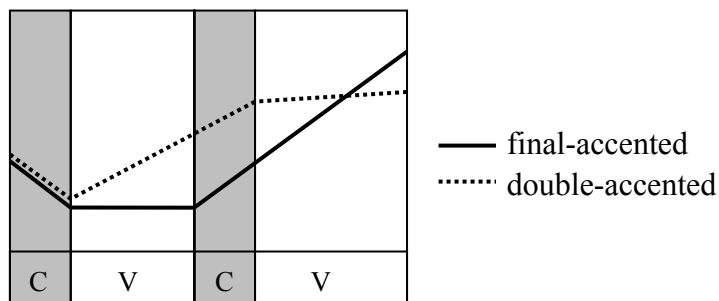
### 3.5.5 Discussion

The extent to which tonal identification was not perfect was unexpected, and no doubt lexical bias effects played a much stronger role than intended. Why did lexical frequency seem to have a much stronger effect in the NKK experiment than in the Mandarin and Cantonese experiments? There are two factors that are probably most relevant. First, while the intention was for the words in the test to be interpreted as words rather than the denotations of those words, this was not made explicit in the instructions. The frame sentence, when it was used, also did not force this interpretation the way the frame sentences in Mandarin and Cantonese did. The frame for NKK was *Eunhi-neun...* ‘Eunhi-TOP...’ (i.e. ‘As for Eunhi...’). This construction was used as opposed to a more conventional ‘X said...’ or ‘X read...’ because Korean is generally a verb-final language and the topic-object construction is one of the few in which the sentence can naturally end with a noun. At the same time, the rates of accuracy in the disyllabic condition ranged from 68% on the low end to 98% (100% for the speaker herself) on the high end, indicating that the  $F_0$  information is indeed available and the tonal category is recoverable but that some speakers are more sensitive to it than others or that the association of the tonal contrast with the lexical contrast is more strongly ingrained for some people than for others. This would make sense considering that the functional load played by tone is much lighter in NKK than in Cantonese and Mandarin. Also, while the “standard” dialects of Cantonese and Mandarin used in the media and in educational settings have strictly-defined tonal systems, the “standard” Seoul variety of Korean is not tonal at all.

Lexical frequency effects aside, it can be stated that NKK behaves like Mandarin in the sense that intonation has a minimal effect on tone identification. This may be a surprise at first glance,



given that the effect of echo question intonation on the right edge of the  $F_0$  curve is much more similar to Cantonese than to Mandarin in all cases except for disyllabic initial-accented words. However, this result needs to be considered in light of the following facts: 1) There are three tonal categories, and the fact that one of them, initial-accented, behaves in an echo question context like Mandarin and the other two, double-accented and final-accented, behave in that context like Cantonese actually provides an unambiguous contrast that makes initial-accented words unmistakable in that context. So, the tonal identification task then reduces to a 50-50 choice between double-accented and final-accented. 2) Although the overall trajectory of the contours for double-accented and final-accented appear superficially very similar in an echo question context, there are alignment, slope, and pitch range differences that all potentially cue the distinction in an unambiguous way, as was shown in Chapter 2. These differences are shown schematically in Figure 3.24. All things considered, the echo question melody does not interfere with the tonal melodies to the extent that it degrades their level of contrast.



**Figure 3.24: Schematization of slope and alignment of double-accented and final-accented contours on disyllables in an echo question context in NKK.**

Also unexpected was the extent to which listeners had trouble identifying echo question intonation in NKK. One of the lowest rates of perceptual accuracy was observed on initial-accented disyllables, which makes sense given that it displayed the lowest degree of intonation-dependent  $F_0$  range separation (see Figures 2.37 and 2.38 in Chapter 2) and the least salient final rise in the echo question context. Initial-accented disyllables aside, though, the right edge of every echo question token was marked with a salient rise (one was not followed by a fall), much

as in Cantonese but minus the danger of the rising contour being misconstrued for any lexical rising tone in a declarative context. It is true that there was a wide range of scores for intonation identification; omitting the initial-accented disyllabic conditions, the rates of accuracy ranged from 30% to 100%. Age was not evenly distributed among the listeners, but three out of the four older listeners (age 40 and above) scored below 60%. As far as sex goes, the females (excluding the older listeners) scored an average of 86.46% and the males 73.96%. These results suggest that the age and sex demographics may be relevant, but even the listeners in the same demographic as the speaker did not do as well as expected. Since the confusions were not systematic and not predictable by the surface contour direction (i.e. melodic combinations that are normally realized as rising were reportedly perceived as those that are normally realized as falling, and vice versa), it is unenlightening to fill out the rise vs. fall reinterpretation table for NKK. Further study is needed to shed light on these confusions.

Finally, recall that the contrast between initial-accented and double-accented polysyllabic words is one of alignment. The fact that the speaker herself only correctly identified her own tones 62.5% of the time on monosyllables (compared with 100% on disyllables) indicates that a perceptual neutralization (near, if not complete) is taking place in that environment. One might imagine that hearing the tones in the context of a frame sentence would increase the rate of perceptual accuracy (e.g. because the frame could provide the listener with information about the speaker's pitch range), but that did not turn out to be the case. These results are in stark contrast with the results for Henanhua, in which the speaker was able to distinguish her two falling tones—T2 and T4—on monosyllables 93.75% of the time. In that language, the contrast between T2 and T4 was also mainly one of alignment, especially in the frame context (where, incidentally, the rate of accuracy was 100%). This perceptual difference between the two languages indicates that the nature of the representations of tone—and in particular these falling tones—in the two languages must be different, despite the similarities in surface contrasts. In the next section we turn to the results from a pilot perceptual study in SJ, which might be expected

to yield perceptual results closer to those of NKK, given that the tonal representations in NKK are usually presumed to be similar to those in Japanese.

### **3.6 Perception in Shiga Japanese**

Due to constraints of time and resources, a full-fledged perceptual study was not carried out for SJ. However, a pilot study with the same basic design was performed. Since the author did not have first-hand contact with the handful of subjects that participated, the stimulus files were delivered electronically to the subjects and the perceptual tests were self-administered.

#### **3.6.1 Subjects**

There were six subjects that participated, including two of the speakers that provided a subset of the stimulus files (i.e. they were hearing their own speech some of the time). They all lived and worked in Shiga at the time of the test, and they all identified themselves as native speakers of SJ, but it was not possible to glean specific information about their backgrounds (except for the two listeners who had participated in the production experiment) due to privacy concerns.

#### **3.6.2 Stimuli**

There were 56 stimuli in total. The stimuli consisted of speech produced by three different speakers. All stimuli were words in isolation. All words were disyllabic and there were four tonal categories represented, although they were not evenly distributed, nor was every two-way comparison represented. The source word list is given in (3.5):

(3.5) SJ perceptual experiment target word list

2 H-unaccented words: *hana* ‘nose’, *ame* ‘candy’

1 L-unaccented word: *asa* ‘hemp’

1 H-initial-accented word: *hana* ‘flower’

2 L-final-accented words: *ame* ‘rain’, *asa* ‘morning’

Only one of the three speakers' sets contained all of the above words. The other two sets lacked the *asa* 'hemp' / *asa* 'morning' minimal pair; consequently, the L-unaccented tonal category was only represented for one speaker. The additional conditions listed in (3.6) were controlled for:

(3.6) SJ perceptual experiment additional conditions

2 intonational contexts (declarative statement vs. echo question)

2 repetitions

Each subject heard three groups of stimuli, each group comprised of utterances from one of the three speakers. The order within each group was randomized differently for each subject. The subjects were instructed to play each sound file twice and then indicate their responses on a multiple-choice answer sheet, a portion of which is shown in Figure 3.25. The first six empty

	鼻	飴	麻	雨	朝	花	。	?
例				4				5
1								
2								
3								
4								
5								
6								
7								

Figure 3.25: A sample of the answer sheet in the SJ perceptual test.

columns correspond to the six words on the word list, and the last two, as in the other experiments, correspond to two intonational categories (*statement* vs. *question*). As in the other experiments, subjects were asked to choose a word and an intonational category and to indicate their levels of confidence on a scale of 1 to 5.

### 3.6.3 Comparison with other experiments

As shown above, the design of the SJ experiment shared some elements with those of the Mandarin and Cantonese experiments and some elements with that of the NKK experiment. Namely, like in the Mandarin and Cantonese experiments, orthography was sufficient to

disambiguate lexical contrasts, but like in the NKK experiment, no full minimal sets were available so two-way comparisons of different combinations of tonal contrasts had to suffice. Also, because the test contained fewer total stimuli and they were all words in isolation, the SJ test took less than ten minutes to complete and, overall, was much less taxing than any of the other tests.

### 3.6.4 Results

The SJ listeners performed very well on both tone and intonation identification, with overall rates of accuracy at 97.02% and 96.73%, respectively, as shown in Table 3.20. The rates shown in Table 3.21 suggest that intonation had little effect on the perception of tone. The imperfect perceptual scores were mostly due to a mysterious handful of L-unaccented-vs.-L-final-accented confusions, as evidenced in the confusion matrices in Figure 3.26. It is not clear what caused these L-unaccented-vs.-L-final-accented confusions, which were shared among three out of the six listeners.

**Table 3.20: Overall rates of perceptual accuracy in judging tonal and intonational categories in SJ**

function	rate of accuracy
tone	97.02%
intonation	96.73%

**Table 3.21: The rate of perceptual accuracy in judging tonal category, by intonational category of the stimulus, for SJ.**

intonational category	rate of accuracy for tone
declarative	97.62%
echo question	96.43%

	H-unacc	L-unacc	H-initial-acc	L-final-acc
stimulus				
H-unacc	1	0	0	0
L-unacc	0	75	0	25
H-initial-acc	0	0	100	0
L-final-acc	0	2	0	98

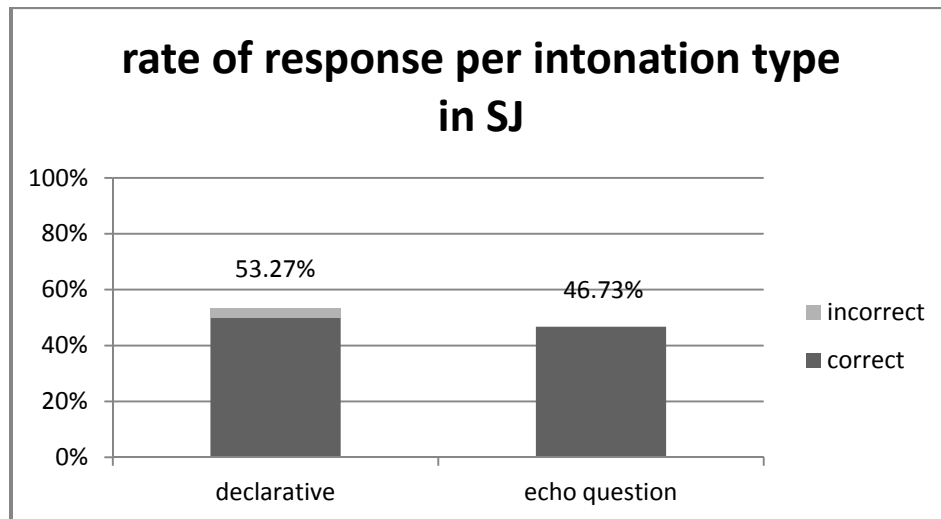
declarative

	H-unacc	L-unacc	H-initial-acc	L-final-acc
stimulus				
H-unacc	97	0	0	3
L-unacc	0	67	0	33
H-initial-acc	0	0	100	0
L-final-acc	0	0	0	100

echo question

**Figure 3.26: Tonal confusion matrices for the disyllabic condition in SJ.**

Turning now to intonation perception, let us look at the bin test for intonation in SJ, shown in Figure 3.27. This is the first time we do not see such a strong bias toward declarative utterances. At 53-47, the bias is still there, but it is weaker than the 60-40 ratio we saw for the



**Figure 3.27: Bin test for intonation responses in SJ.**

other three languages. This bias does seem to affect the perception of intonation by intonation condition, however. As shown in Table 3.22, the declarative condition yielded a perfect rate of accuracy, 100%, while that in the echo question condition was slightly degraded, at 93.45%.

**Table 3.22: Rate of perceptual accuracy in judging intonational category, by intonational category of the stimulus, in SJ.**

intonational category	rate of accuracy
declarative	100%
echo question	93.45%

Since the number of stimuli was not the same for every tonal category, it is probably not useful to see the accuracy results for intonation broken down by tonal category. Suffice it to say that at least one incorrect response for intonation was given in each of the four tonal categories, and that these errors were shared among three of the six listeners. At this level of investigation, there are probably not enough data points to decide if any one of the tonal categories interferes with intonation perception more than the others.

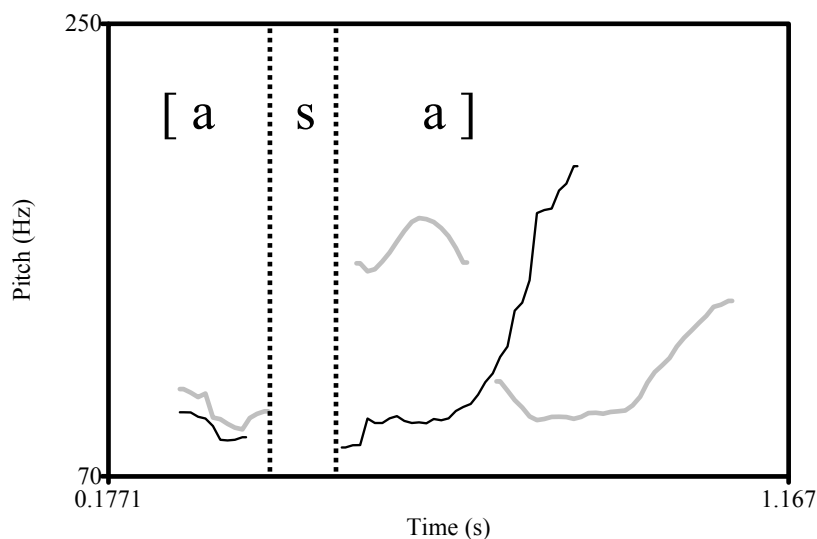
### **3.6.5 Discussion**

While it must be reiterated that this perceptual experiment in SJ was merely a pilot study, the results provide some interesting insights into the melodic system of that language and its place in the typology among the other languages under investigation. The overall rates of accuracy for tone and intonation perception were very good. The rate for tone was on par with that for Mandarin and much better than that for NKK or Cantonese. The rate of accuracy for intonation was much better than that for any of the other languages.

It makes sense that SJ listeners outperformed Cantonese listeners in tone identification. SJ has fewer tonal contrasts and the tone is a “word tone” as opposed to a “syllable tone”, the consequence being that any local melodic interference on the last syllable is only going to affect part of the “word tone”. Furthermore, whereas the echo question melody in Cantonese (at least in the tokens recorded for these perceptual tests) tends to have an “overwriting” effect, encroaching backwards on the latter portion of the tonal melody on the final syllable, the echo question melody in SJ is generally kept more delineated from the tonal melody on that last syllable, often invoking durational adjustments to preserve the necessary melodic contours.

It also makes sense that SJ listeners would outperform Cantonese and Mandarin listeners in intonation identification. Unlike in Mandarin, the melodic cues for echo question intonation are localized and well-delineated. Unlike in Cantonese, there is no danger of a listener reinterpreting a final echo question rise as a rising lexical tone, since none of the lexical tones ends in a rise in SJ. In fact, the few misperceptions that occurred for both tone and intonation are somewhat puzzling.

As for the tonal confusions, it seems safe to assume that the confusions did not occur at the level of pitch perception, since the  $F_0$  contours in the different conditions were so distinct. Example contours for the two most often confused tone classes in an echo question context are shown in Figure 3.28. The L-final-accented contour (displayed in gray) shows a very salient peak



**Figure 3.28: Pitch tracks for L-unaccented *asa* ‘hemp’ (thin black line) and L-final-accented *asa* ‘morning’ (thick gray line) in isolation in an echo question context, in SJ.**

on the nucleus of the second syllable followed by a dip and another rise, while the L-unaccented contour (in black) stays low at the beginning of the nucleus and then rises sharply. Moreover, since none of the three listeners consistently made the same mistake, it is unlikely that any of them categorized the words in the opposite lexical classes. The only factor that seems plausibly relevant is lexical frequency, where presumably *asa* ‘morning’ is much more common than *asa* ‘hemp’ (although this would not explain the one confusion that went in the other direction). As



for the intonation confusions, the bias toward declarative intonation seems to be playing a role. Both types of confusion in SJ warrant further study, but for now a tentative version of the rise vs. fall perceptual reinterpretation schema that was presented for Mandarin and Cantonese is shown in Table 3.23 for SJ. Note that SJ has no rising tones, so the first two melodic combinations do not occur. For the combinations shown, no reinterpretations are observed in SJ; it goes without

**Table 3.23: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations in SJ**

underlying combination	surfaces as	occasionally interpreted as
T↗ + I↗	n/a	n/a
T↗ + I↘	n/a	n/a
T↘ + I↗	fall-rise	n/a
T↘ + I↘	fall	n/a

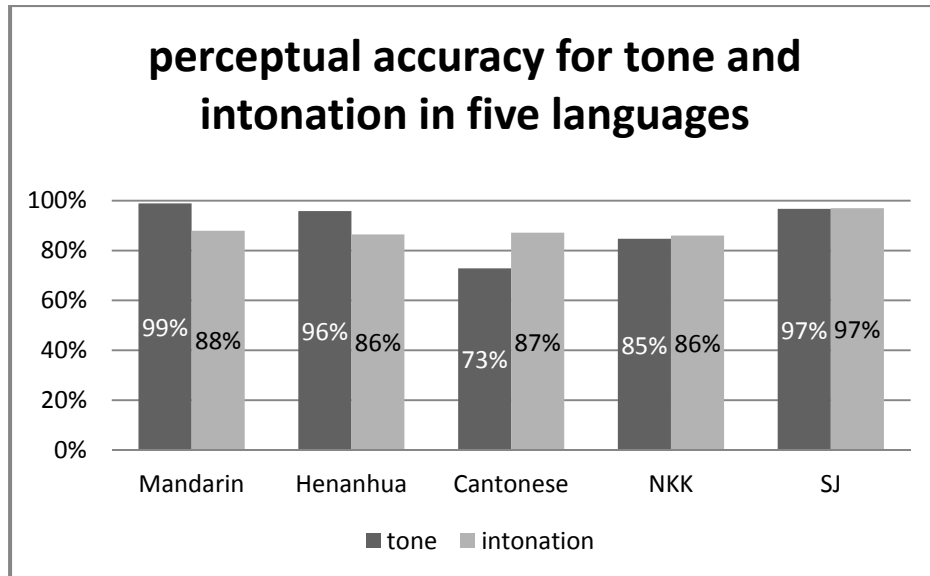
saying that the falling-tone-declarative-intonation combination does not get reinterpreted in SJ; this result is consistent with the reinterpretation results for Mandarin and Cantonese.

Another question worth asking is why SJ listeners did so much better all around than NKK listeners, when at first blush their tonal systems seem quite similar. Indeed, those not familiar with the tonal systems of the two languages might lump them together as “pitch accent” languages. There are, however, some crucial differences that are most likely relevant. First, in a disyllabic domain, SJ maximizes its use of pitch space in differentiating among the different lexical tones to a greater extent than NKK. While NKK mainly varies the horizontal alignment of its pitch peak, necessitating a three-way distinction in horizontal alignment, SJ manipulates a few different dimensions—whether the first syllable starts in a high register or a low register and whether there is a pitch peak on the first or the second syllable (no pitch peak anywhere being one of the possibilities). Second, SJ speakers may be one step more in tune with the tonal nature of their language due to certain sociolinguistic factors. While SJ is not the national standard, the national standard that is prescribed and used in the media is indeed tonal, and from a young age children in school are made aware of well-known tonal minimal pairs. Speakers of the local

dialect may therefore be more aware that dialects can differ along tonal dimensions, which highlights the very existence of a tonal system in their own dialect. The situation is different in South Korea—that is, the national standard is not tonal, so speakers of NKK may not be as fully conscious of the tonal nature of their own dialect. Finally, some of the differences may be due to differences in the experimental conditions. The words used in the SJ experiment were all native Japanese words and did not include any Sino-Japanese words, while the mixture of native Korean and Sino-Korean words in the NKK experiment may have led to certain lexical biases in that experiment. Also, the two demographics that participated in the respective studies just may not have been comparable. A majority of the 24 listeners in the NKK experiment were university students while the six listeners in the SJ pilot study were high school English teachers who may have been more sensitive to language phenomena. A broader study in SJ with a larger pool of participants would help test this hypothesis.

### **3.7 Summary and Conclusion**

In this chapter, results from perceptual experiments in Mandarin, Cantonese, and NKK, as well as results from a pilot perceptual study in SJ, were presented. Despite necessary language-specific differences in the experimental designs in each study, an attempt was made to design the experiments in such a way that the results were easy to compare across languages. Figure 3.29 summarizes the perceptual accuracy results for each language, rounded to the nearest percent. Note that the percentages given for NKK exclude the monosyllabic condition, which brings the average for tonal accuracy down to 77% when included. When the data were pooled for the three languages for which the sample sizes were sufficiently large—Mandarin, Cantonese, and NKK—there was a highly significant effect of language on tonal accuracy ( $p < .001$ ) but not on intonational accuracy ( $p = .435$ ). Perhaps more importantly, we can see that the relationship between the rate of tonal accuracy and the rate of intonational accuracy is different in each of these languages (recall that the difference between these rates was positive and highly significant



**Figure 3.29: Cross-linguistic comparison of perceptual accuracy for tone and intonation.**

for Mandarin, negative and highly significant for Cantonese, and not significant for NKK). We are now equipped to answer some of the questions posed at the beginning of this chapter. As for Q5 (*Functional Recoverability*), in a context stripped of all pragmatic, semantic, and syntactic cues, different languages are *not* equally successful in encoding multiple communicative functions in the speech melody in a perceivable way. In the case of SJ, both tone and utterance-type intonation appear to be quite recoverable, whereas the rate of intonation recovery is lower in Mandarin and the overall rate of recovery is lower in NKK and Cantonese. In the case of NKK, the difference may be due to a tonal system in that language that makes less use of the tonal pitch space as well as to a lower degree of tonal sensitivity on the part of native speakers. In the case of Cantonese the difference is likely due to its larger inventory of contrasting tones as well as denser tonal specification (or, rather, more localized tone-domain).

As for Q6 (*Functional Prioritization*), there indeed seems to be a trade-off in some languages whereby one melodic function is more perceptually recoverable than the other. The trade-off goes in one direction for Cantonese, where intonation is recovered at a higher rate than tone, whereas it goes in the other direction for Mandarin and Henanhua. In the case of Cantonese, there seems to be an “overwriting” effect by the echo question melody, which

encroaches on the tonal melody of the final syllable in the utterance, in some cases reversing the slope and obscuring register contrasts. This is not a problem in Mandarin or Henanhua, where echo question intonation for the most part enhances the differences among the tonal contours, with the possible exception of T2 and T3 in Mandarin. (Note that this is also not a problem in SJ, where the echo question melody is implemented locally but kept delineated from the tonal melody at the right edge of the utterance, and where there is no danger of the echo question melody being mistaken for one of the tonal melodies.) Meanwhile, it seems to be the case that this enhancing nature of echo question intonation in Mandarin is precisely what causes it to be less salient in the presence of some of the lexical tones. This tendency for the overall shapes of the Mandarin and Henanhua tonal contours to be preserved, along with an apparent bias toward expecting or perceiving declarative intonation causes a slight degradation in the rate of echo question perception in those languages.

As for Q7 (*Types of Confusion*), we have certainly seen that confusions may cause reinterpretations in several directions. As Cantonese has shown us, it is indeed possible for the melodic expression of one communicative function to be misperceived as that of another. We saw that echo question intonation in Cantonese was often reinterpreted as T2 or T5. We also saw that T2 in that language could degrade overall utterance-type intonation perception. This was also true in Mandarin for T1, T2, and T3. As for when these types of interferences and confusions are more likely to occur, it seems reasonable to surmise that a language will have more trouble if its tonal inventory makes use of both register and contour, since any intonation system is likely to affect either the register (if it is parallel or translational in its implementation) or the contour (if it is partially serial). Cantonese is such a language. We may also say that, within a melodic system, confusions are more likely to occur involving lexical tones that have a rising component at their right edge. This was of course true for Cantonese T2 and T5, the two rising tones in that language, but also to a small degree for T2 and T3 in Mandarin, the rising and dipping tones, respectively, in that language. Note that *none* of the lexical tonal melodies in SJ ends with a rising component, and echo question intonation did not cause any substantial

interference in that language. Table 3.24 summarizes the types of reinterpretations we observed in Mandarin, Cantonese, and SJ. Note that there is an apparent asymmetry when it comes to tone-intonation combinations that “agree”—i.e. rising tones in an echo question context vs. falling tones in a declarative context. While the former, which surfaces as a rise in the relevant languages (Mandarin, Henanhua, and Cantonese), can be reinterpreted as being declarative, the latter is hardly ever<sup>37</sup> reinterpreted as being interrogative. In other words, a rising surface contour can be incorrectly attributed solely to an underlying rising tone but a falling contour “resists” being incorrectly attributed solely to an underlying falling tone.

**Table 3.24: Schematic breakdown of melodic combinations, their surface realizations, and their possible reinterpretations across languages**

underlying combination	occasionally <sup>38</sup> reinterpreted as	reinterpreted in
T↗ + I↗	T↗ + I↘	Mandarin, Henanhua, Cantonese
T↗ + I↘	T↗ + I↗	Cantonese
T↘ + I↗	T↗ + I↗ or T↗ + I↘	Henanhua, Cantonese
T↘ + I↘	n/a	n/a

At this point it is worth noting a difference between the types of perceptual *interference* we observe in the realm of speech melody and the types of perceptual *neutralization* we see in other realms of the phonology and phonetics. A useful comparison is with the phenomenon of word-final devoicing that occurs in some languages. We can think of word-final devoicing as a case of context-triggered neutralization. It might be tempting to seek an analogous type of neutralization in the melodic systems of tone languages such as the ones investigated here, for example a case where a certain tonal contrast gets neutralized in a certain intonational context or vice versa. However, this turns out to be a futile pursuit, at least for the languages tested in these studies. First, how do we define “perceptual neutralization” given the types of results we have

<sup>37</sup> Less than 3% of the time in any of the languages, except for Henanhua, where this type of reinterpretation occurred once in twelve stimuli, i.e. 8.3% of the time. A larger sample size is needed to gauge the robustness of that result in Henanhua.

<sup>38</sup> More than 10% of the time.

seen? Is it something that happens on a case-by-case basis or is it defined by a large sample of perceptual responses? In the case of Mandarin, there were a handful of instances in which the steep final rise due to the echo question intonation obscured the contrast between T2 (the rising tone) and T3 (the dipping tone) enough that about half of the listeners got confused about what tone they were hearing some of the time. Would we call this subset of cases an instance of perceptual neutralization despite the fact that most of the time listeners recovered the T2-T3 contrast with no problem in the echo question context? Another confounding issue is exemplified by the situation in Cantonese. It seems fair to say that in many cases the steep rise in the last syllable of echo questions in that language completely obscures tonal contrasts on that syllable. However, rather than the intonational melody being easily identified at the expense of tone identification, often the intonational melody is misperceived as being a *tonal* melody, the result being that both the tone *and* the intonation are perceived incorrectly! This situation is in sharp relief with more “conventional” types of perceptual neutralization exemplified by word-final devoicing, where the context is in one dimension (word position) and the relevant contrast is in another (voicing). A language that would provide a true positive example of a context-sensitive melodic neutralization is a language we can hypothetically call *Cantonese'*, in which the intonational system is the same as that in Cantonese but there are no rising tones in the lexical inventory. In the absence of such a language, however, it is more useful to talk about perceptual *interference* when it comes to the interaction of different components of speech melody.

### **3.7.1 Answers to questions about perception**

The questions from the beginning of this chapter are repeated in (3.7), along with some answers that have been revealed by the experiments described in the chapter:

(3.7) Questions about melodic perception in tone languages

**Q5 – Functional Recoverability:** Are different types of tonal systems (i.e. word tone vs. syllable tone; register vs. contour) equally successful in encoding multiple communicative functions (lexical distinctions vs. intonational meanings) in the speech melody in a recoverable way? *No. Some systems are more successful overall (SJ), some are less successful overall (Cantonese, NKK), some are better with one function than with the other (Mandarin, Cantonese).*

**Q6 – Functional Prioritization:** Is it the case that priority is given to the same communicative function (tone or intonation) in all languages when there is an overlap? *No. Mandarin does better at preserving tonal contrasts, while Cantonese does better at preserving intonational contrasts.*

**Q7 – Types of Confusion:** What types of confusions, misperceptions, and “perceptual neutralizations” are possible and when do they occur?

A. Can the melodic expression of one communicative function be misperceived and reinterpreted as that of another (i.e. can a melodic cue for tone be misperceived as that for intonation or vice versa)? *Yes. A positive pitch slope that is the realization of tone can be construed as an intonational cue, and a positive pitch slope that is the realization of intonation can be construed as a tonal cue.*

B. Can we make any predictions about when such confusions occur based on the characteristics of a given tonal system? *Possibly. Confusions may be more likely to occur in languages with a tonal system that makes use of register and contour differences; within a tonal system, if the inventory of lexical tones includes one or more “rising” tones, melodies involving those tones may be more likely to elicit confusions than those involving other tones.*

In this chapter and the chapter before it, results from cross-linguistic production and perception studies were presented and discussed mainly from a typological perspective. In the next chapter,

the melodic system in each language is evaluated internally and various points of unpredictability within each one are highlighted.



## CHAPTER 4: TONE-DEPENDENT INTONATION AND ITS THEORETICAL CONSEQUENCES

### 4.1 Introduction

In Chapter 1 of this dissertation, during the discussion of current and past models of speech melody that can be found in the literature, Ladd's (1996; 2008) distinction between *overlay* models and *sequential* models was adopted. It was noted that one of the fundamental differences between overlay models and sequential models is that the former do not include any phonological interactions among various melodic functions, while the latter do. Ladd (2008), in his discussion of Kochanski and Shi's Stem-ML model for Mandarin (Shih and Kochanski 2000; Yuan, Shih et al. 2002; Kochanski, Shih et al. 2003) and Xu's PENTA model (Xu 2005), noted:

[T]he issue of sequential and overlay models discussed in this section may have already turned into a more tractable question of how  $F_0$  realisation parameters interact. However, I have included these models in the discussion here because they are still, to a considerable extent, based on the assumption that intonational meanings or functions (such as focus and interrogativity) are directly signaled by the variability of acoustic parameters, and not mediated by the occurrence of phonological events. (p. 30)

While the need for this kind of phonological "mediation" between grammatical function and acoustic output has been established in other areas of phonology—and therefore taken for granted in the realm of speech melody by proponents of the AM approach—the notion that speech melody is somehow special in this regard and does *not* share this need nevertheless persists in the field. Ladd (2008) continued:

In other words, the more basic question of how intonation conveys meaning [...] remains a fundamental point of disagreement. Xu's work, in particular, is clearly motivated by a purely parametric approach to intonational meaning (see especially Xu and Xu 2005). It seems unlikely that those who find the purely parametric approach attractive will take a

few pieces of evidence for phonological structure as signaling the end of the debate. (pp. 30-31)

One of the goals of this dissertation—and the aim of this chapter in particular—is to bring data from languages with lexical tone to bear on this issue; by investigating whether the acoustic realization of utterance-type intonation is dependent on the phonological categories of lexical tone in a way that is not predictable (i.e. it does not just fall out from the phonetic properties inherent to each lexical tonal unit), we can test whether or not purely parallel encoding of the two melodic functions is tenable. In this regard, one of the questions raised at the beginning of Chapter 2, Q3 (*Declarative-to-Echo-Question Mapping*), is relevant. It is repeated in (4.1):

(4.1) Question about production from Chapter 2

**Q3 – *Declarative-to-Echo-Question Mapping***: Is the  $F_0$  contour of a tone in the echo question context predictable purely from the phonetic attributes of the tone in the declarative context (i.e. is there an algorithm that can take just phonetic pitch parameters of the declarative form as input and successfully predict the  $F_0$  contour of the echo question form)?

Note that the term *mapping* is used here in a non-directional sense. This question is not about whether echo question intonation is somehow *derived* from declarative intonation; it is a much more superficial question about whether the effect of utterance-type intonation can be factored out of the surface contour in a predictable way without using any information about the lexical tones involved. The question is relevant whether declarative intonation is somehow “neutral” and echo question intonation is derived from it or both declarative and echo question intonation involve the manipulation of some other neutral, baseline contour. In the latter scenario, if there is one tone-independent function that yields declarative intonation and another tone-independent function that yields echo question intonation (given some neutral “underlying”

contour), it should be possible to describe a mapping from declarative to echo question intonation (as well as the reverse mapping) by combining the two functions into a single function.

Various results presented and discussed in Chapter 2 strongly indicate that the answer to Q3 is indeed “no”. In the rest of this chapter, the extent to which the implementation of utterance-type intonation is lexical-tone-dependent is made explicit and characterized in more detail. Data from the syllable-tone languages (Mandarin, Henanhua, and Cantonese) are discussed in Section 4.2 and data from word-tone languages (NKK and the Kansai Japanese dialects) are discussed in Section 4.3. Section 4.4 summarizes and concludes the chapter.

## **4.2 Accounting for Tone-Dependent Intonation in Syllable-Tone Languages**

In this section, the need for tone-dependent mechanisms in our model is highlighted by revisiting relevant results from the three syllable-tone languages (Mandarin, Henanhua, and Cantonese). In each case we start with the null hypothesis, namely that *no such mechanisms are necessary and the function or mechanism that underlies echo question intonation in general should be inferable from the behavior of any given tone in the tonal inventory*. For each of the syllable-tone language we see that the null hypothesis is not supported, and consequently a purely phonetic model with a built-in assumption of parallel encoding breaks down.

### **4.2.1 Mandarin**

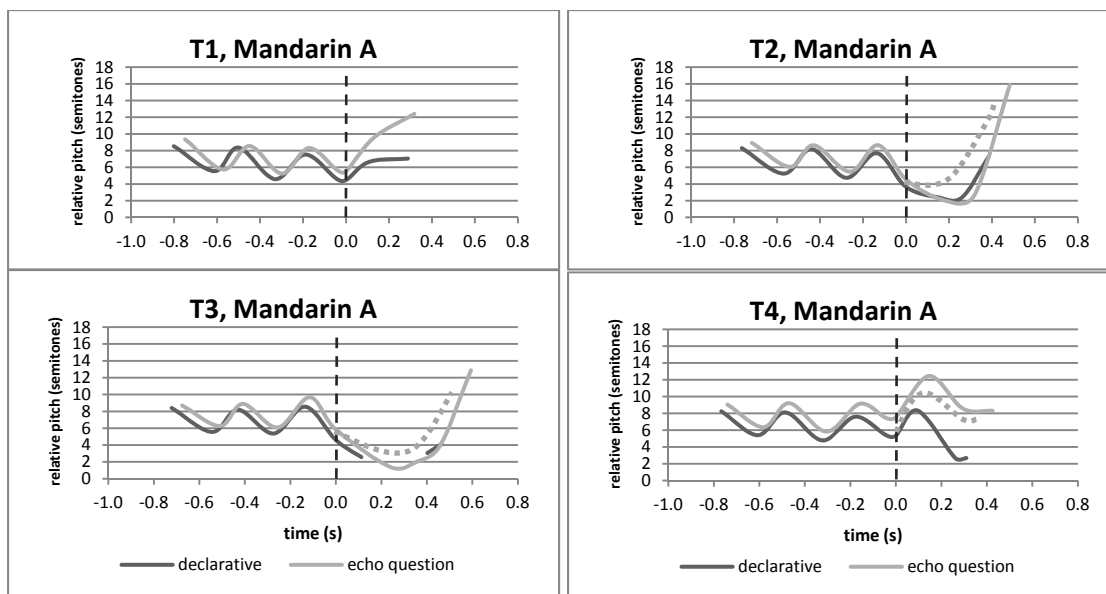
All four tones of Mandarin are distinguishable by their contours in a declarative utterance-final context—level, rising, dipping, and falling. We saw in Chapter 2 that the effect of echo question intonation on the tonal targets appears mainly to be one of register—all or part of the tonal contour is shifted upwards. This is apparent, for example, in the case of T4, which is falling in both intonational contexts but falls from a higher point in the echo question context. In this way the *implementation* of tone and intonation can be considered to be parallel<sup>39</sup>. However, there are

---

<sup>39</sup> Although, see Chao (1968) for examples of tonal particles in Mandarin that might constitute examples of serial implementation of lexical tone and intonational tone.

category-specific effects that are apparent in the output, ruling out a model in which the *encoding* of the two is strictly parallel. In this section, these effects are examined from the point of view of modeling the utterance-type intonation in the phonetics.

Speaker A’s mean contours for all four tones are shown in Figure 4.1, where the unit along the *y*-axis is semitones (as opposed to Hz)<sup>40</sup>. Using one of the tones as a starting point, we can attempt to characterize the realizational differences between declarative and echo question contours as a generalizable function and see if the predicted contours for the other tones match the actual contours. For example, given the behavior of Mandarin Speaker A’s T1, shown on the upper-left in Figure 4.1, a reasonable mechanism to invoke would be an intonation-dependent



**Figure 4.1: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for each of the four lexical tones in Frame 1, for Mandarin Speaker A, with predicted echo question contours (dotted light gray) for T2, T3, and T4 based on the hypothetical mechanism of phrase curves that diverge starting at the beginning of the last syllable (modeled after T1).**

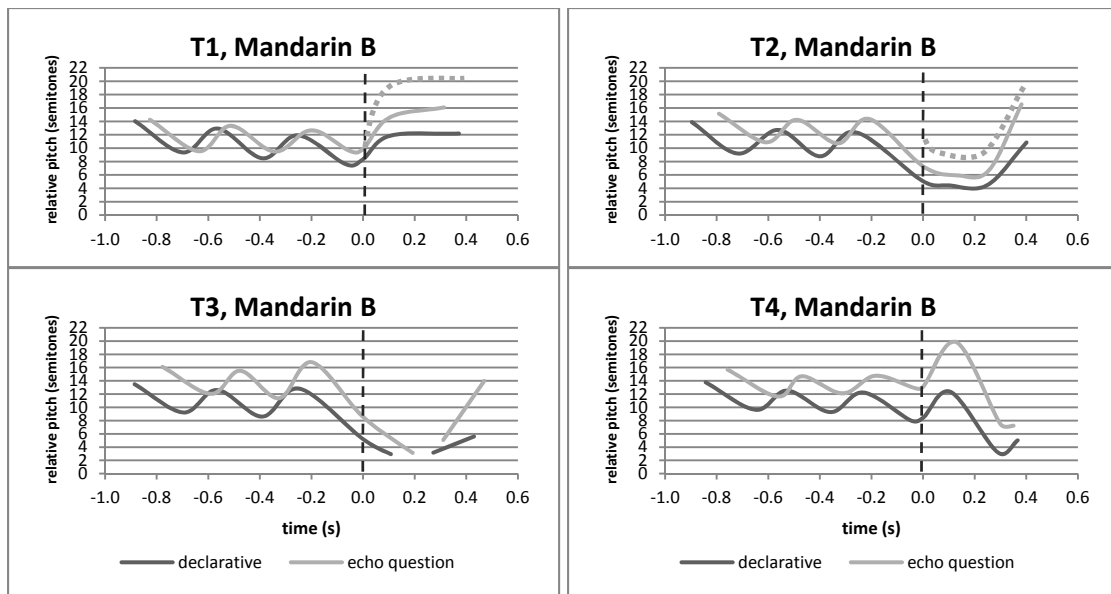
<sup>40</sup> This scale was chosen so that the shapes of contours relative to one another might more accurately reflect how speakers perceive them, and so that any possible transformational functions invoked to describe their relationship might be expressed more simply (for example, an additive function only makes sense if the units are in semitones). The semitone values displayed in the plots were calculated using the following formula:  $\text{semitone value} = 12 \log_2 r$ , where  $r$  was the ratio of the measured  $F_0$  value to some sensible reference value near the bottom of the speaker’s speaking range (for Mandarin Speaker A it was 90 Hz). The choice of 2 for the base follows from the assumption that the perceived pitch changes by an octave when the  $F_0$  of one pitch is twice that of another pitch in Hz, and since there are twelve semitones per octave, the log value is multiplied by 12 to yield the semitone value.

phrase curve or baseline mechanism. The echo question phrase curve would diverge from the declarative phrase curve starting at the beginning of the final syllable, resulting in a much larger vertical displacement at the end of the syllable than at the beginning of the syllable. A look at the other tones in Figure 4.1 reveals that this mechanism does not predict the observed echo question contours for them, however (the predicted contours, appearing in the figure as light gray dotted curves, do not line up with the observed contours). The divergence for T2 starts much later, basically at the end of the syllable, while for T3 it appears to start midway through the syllable (evidenced by the gap in the declarative contour caused by glottalization that is missing in the echo question contour, which suggests a lower target pitch for the former than for the latter in that portion of the syllable). For T4 the divergence actually starts a bit before the final syllable. In addition, the duration of the last syllable is extended in the case of T2 and T3, resulting in rising “tails” that further increase the vertical displacement between the ends of the declarative and echo question contours; this effect is also seen to a lesser extent for T4, which receives a small, flat tail at the end of it in the echo question condition.

This notion of diverging phrase curves is compatible with the mechanism that Xu (2005) proposed for handling echo question intonation. He suggested that the phrase curve for echo questions might rise exponentially, resulting in a divergence that is mostly observed on the last syllable of the utterance. We have seen from the results for Speaker A that such a mechanism is promising, but it would only work if the settings that determine the shape of the rise for echo questions (how early it starts, how steep it is, etc.) can be made dependent on the lexical tonal category. Also, some mechanism for modeling the “tails” observed at the ends of T2, T3, and T4 would be necessary. The null hypothesis is not supported for Speaker A.

What about Mandarin Speaker B? Recall from Chapter 2 that the two Mandarin speakers were observed to employ slightly different strategies when it came to T4. The slope of T4 was slightly flattened in the echo question context for Speaker A, but it was steepened in that context for Speaker B. In other words, the vertical displacement was smaller at the peak than at the end of the fall for Speaker A, while the reverse was true for Speaker B. Obviously a continually

diverging phrase curve mechanism cannot produce the right effect for Speaker B, whose mean contours are given again in Figure 4.2. Instead, a multiplicative function might be more appropriate, since there appears to be a correlation between the pitch level of the declarative contour and the degree of vertical displacement between the contours (i.e. greater vertical displacement at higher pitch levels). The multiplicative function that works for T4 appears to predict more or less the correct echo question contour for T3. At first glance, it might appear to make the right prediction for the other tones as well, but note that if the same baseline and multiplication factor are used for all four tones, the predicted echo question contours are much higher than the actual contours for T1 and T2. This is obvious if we consider that the lowest point of T4 is lower than the lowest point of T2 and yet the vertical displacement between the declarative and echo question contours is greater at that point for T4 than it is for T2; similarly, despite the fact that the high point of T4 has the same mean pitch as T1, the vertical displacement at that point for T4 is much greater than that for T1.



**Figure 4.2: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for each of the four lexical tones in Frame 1, for Mandarin Speaker B, with predicted echo question contours (dotted light gray) for T1 and T2 based on the hypothetical mechanism of a multiplicative function, modeled after T3 and T4.**

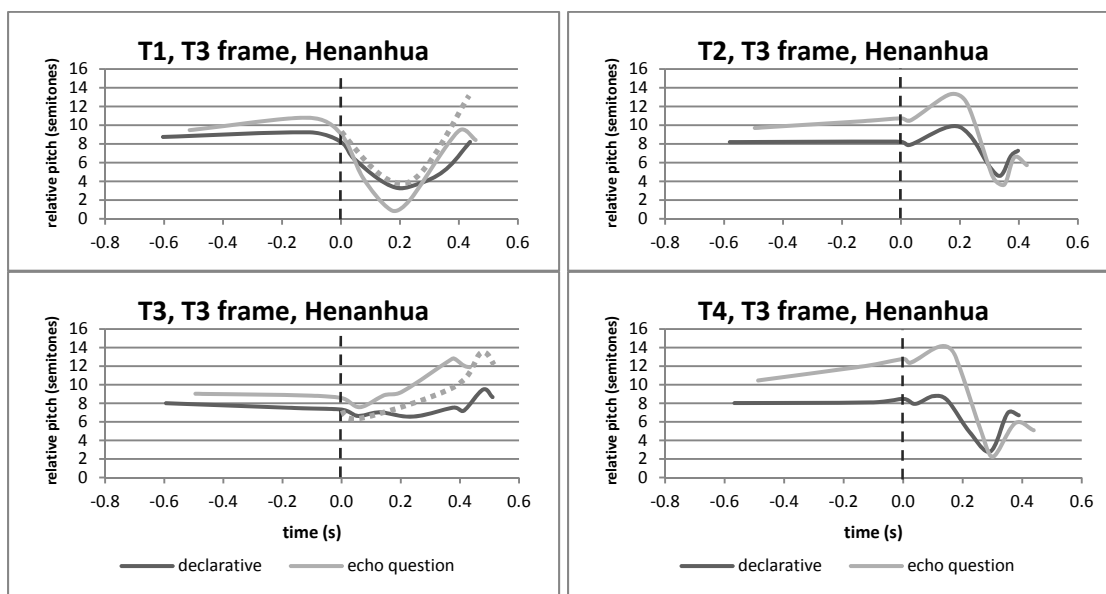
The results for Speaker B do not lend themselves as well to an analysis that depends solely on a rising phrase curve mechanism. While assigning intonation-dependent phrase curves might still be useful for modeling the global upward shift we see sometimes over the entire frame portion of the utterance, an additional mechanism that involves a multiplicative function is probably needed to account for the effects on the final syllable. This is exactly the recipe adopted by Yuan, Shih et al (Yuan, Shih et al. 2002), who worked within the Stem-ML framework in order to model the differences between the pitch contours for eight declarative utterances and those of their respective interrogative counterparts (spoken by a single Mandarin speaker). They claimed that a single phrase curve setting (to model the global upward shift in the frame) and a single strength setting were sufficient for modeling question utterances ending in each of the four tones. However, when Yuan (2004) reported on a more comprehensive acoustic and perceptual study involving multiple speakers and many more renditions of interrogative utterances, he noted that some tone-dependent mechanisms were indeed necessary to account for certain asymmetrical results in both his production and his perception experiments. Likewise, the results obtained here for Speaker B suggest that, either the strength parameter values set in the model would need to be different for different tones, or some additional tone-specific adjustments would be needed. Again, the null hypothesis is not supported.

Before moving on to other languages, it should be noted that tone-dependent intonation effects in Mandarin are not limited to the domain of utterance-type intonation. Chen and Gussenhoven (2008) present results from an experiment in which they elicited two degrees of emphasis on target words in declarative frames in Mandarin. Their results indicate that there appears to be a ceiling effect when it comes to the degree of  $F_0$  expansion associated with emphasis, but the authors argue that this ceiling effect cannot be due to physiological constraints, since the value of the  $F_0$  ceiling appears to be tone-dependent. They further point out that the Stem-ML model and the PENTA model (from Xu 2005) have no direct way to account for any such tone-dependent ceiling effect.

## 4.2.2 Henanhua

In many respects, the Henanhua melodic system is very similar to that of standard Mandarin. The relevant similarities, discussed in more detail in Chapter 2, crucially include a global upward shift in the overall  $F_0$  range and a greater pitch excursion on the last syllable in the echo question context. Just as in Mandarin, however, we saw some tone-specific effects that preclude a model with parallel encoding.

For ease of reference, the mean contours for the four tones placed at the end of the all-T3 frame are repeated here in Figure 4.3; once again the Hz increments have been replaced with semitones. Right away we see that any kind of additive function or a mechanism involving a

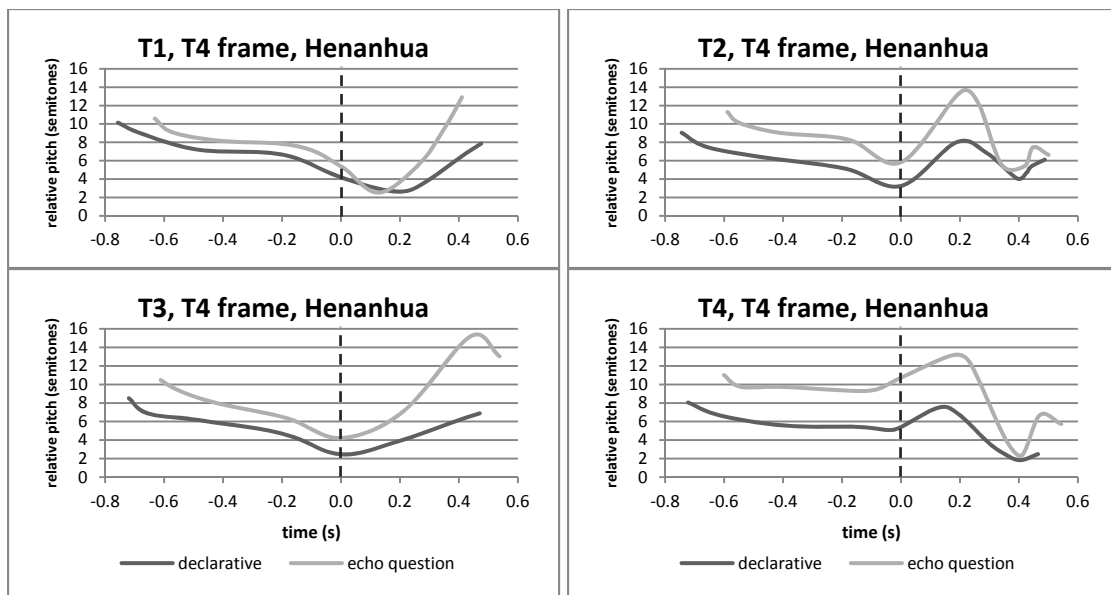


**Figure 4.3: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for each of the four lexical tones in an all-T3 frame in Henanhua, with predicted echo question contours (dotted light gray) for T1 and T3 based on the hypothetical mechanism of a multiplicative function with a movable baseline, modeled after T2 and T4.**

phrase curve with a positive slope throughout the final syllable is not appropriate, since the two falling tones show a greater vertical displacement at the beginning of the final syllable than at the end (much like T4 for Mandarin Speaker B). Some kind of multiplicative function seems more promising, but it is obvious that all of the tones cannot share a common baseline for multiplication, since the declarative and echo question versions of the two falling tones, T2 and



T4, converge at their respective low points, which are at different pitch levels. The only hope, then, is to allow for the baseline for multiplication to be set according to the lowest point reached in the declarative rendition of each tone. This predicts more or less the correct baseline in the case of T3, although the multiplication factor appears to be a bit off for that tone. Truly problematic, though, is T1. Instead of the low points of the declarative and echo question contours converging on roughly the same pitch target, the echo question contour dips way *below* the declarative contour, implying a multiplicative baseline floating somewhere higher than the lowest point in the declarative contour. Whether this sub-declarative dip for T1 is a reliably consistent phenomenon is unclear. The results from the all-T4 frame, shown in Figure 4.4, indicate that, at least in some environments, T1 behaves as predicted by the moveable



**Figure 4.4: Mean declarative (dark gray) and echo question (light gray) pitch contours for each of the four lexical tones in an all-T3 frame in Henanhua.**

multiplicative baseline model. Further study is needed to determine if this cross-frame difference is indicative of a wide range of variation for pitch scaling when it comes to T1 (note how consistent the low points for the two falling tones—T2 and T4—are across frame types) or if there is an articulatory interaction between the preceding T3 frame and the target T1 that causes the dip in the echo question context. Whatever the scenario, it appears likely that T1

interacts with echo question intonation in a way that is distinct from how T2 and T4 interact with it.<sup>41</sup> The behavior of T1 notwithstanding, a worthwhile question to ask is whether a conventional overlay model can accommodate a mechanism that sets a multiplicative baseline according to a local minimum that is determined by a lexical tone's inherent minimum pitch. It is not clear how the models of Kochanski, Shih et al. (2003) or Xu (2005) would handle this task, since it entails the specification of two constants (a multiplicative factor and a  $y$ -intercept constant), one of which is inherent to the lexical tonal category. The null hypothesis is not supported for Henanhua.

### 4.2.3 Cantonese

In Cantonese, the realization of echo question intonation relative to declarative intonation involves imposing a sharp rising trajectory on the right edge of the utterance. This rising contour starts partway through the final syllable and “overwrites” the tonal contour for that part of the syllable. This is readily apparent in the case of T4, where the trajectory is falling for the first half of the syllable and then the  $F_0$  takes a sharp turn and rises throughout the second half. Thus, the expression of the lexical tone on the last syllable precedes the expression of the echo question intonation in time<sup>42</sup>. In this way the implementation can be described as serial, although the expression of both functions occurs within the duration of the final syllable and there is very little durational difference between a syllable in the declarative context and one in the echo question context<sup>43</sup>. The mere fact that the implementation is serial does not preclude an

---

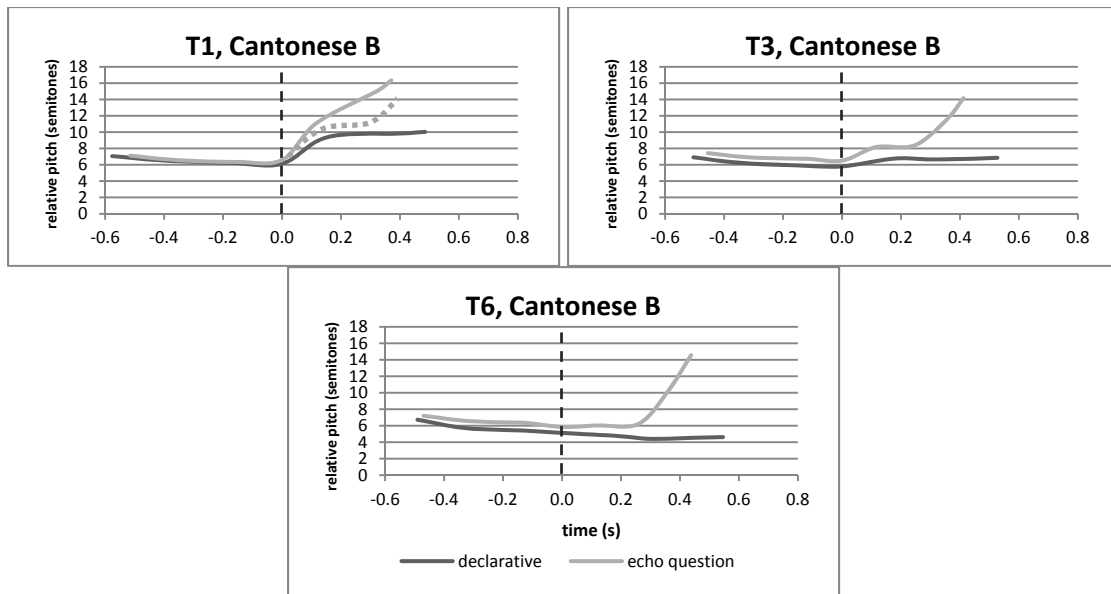
<sup>41</sup> If the relevant difference between T1 on the one hand and T2 and T4 on the other hand turned out to be that T1 simply shows a much wider range of scaling effects in an echo question context, it is possible this tone-dependent behavior could have a perceptual explanation: It could be that pitch scaling is more precise for T2 and T4 in Henanhua because more precision is necessary to maintain a contrast between those tones. This explanation would be similar to the explanation provided by Chen and Gussenhoven (2008) for the tone-dependent ceiling effects they observed for focus manifestation in Mandarin.

<sup>42</sup> This does not take into account the more global effect that is sometimes observed on Cantonese echo questions, in which the overall  $F_0$  range of the sentence is slightly higher than that of the equivalent declarative utterance. If present, this global  $F_0$  shift can be considered to be implemented in parallel with all of the lexical tones in the sentence, including the one in the last syllable.

<sup>43</sup> This is in contrast with the behavior of intonation on final particles in Cantonese, which may elongate substantially to accommodate various pitch movements. Their behavior is akin to general sentence-final intonation in Japanese dialects.

encoding model in which the two melodic functions are independently encoded. Once again, however, we observe category-specific effects that cannot be captured by a parallel encoding model. Once again, these category-specific effects will be approached from the point of view of modeling the intonation in the phonetics.

A good place to start looking for possible generalizations that can be made regarding intonation is the behavior of the level tones, which are distinguished by register in a declarative context. As seen in Figure 4.5, the echo question renditions of T3 and T6 would seem to indicate that there is a static (all else being equal<sup>44</sup>) high target associated with echo questions and that the first half of the lexical tone is expressed on the first half of the syllable, after which the contour deviates from the steady-state declarative trajectory in order to reach the high target at the end. Since the contour starts from a lower place in T6 than in T3, the resulting slope is

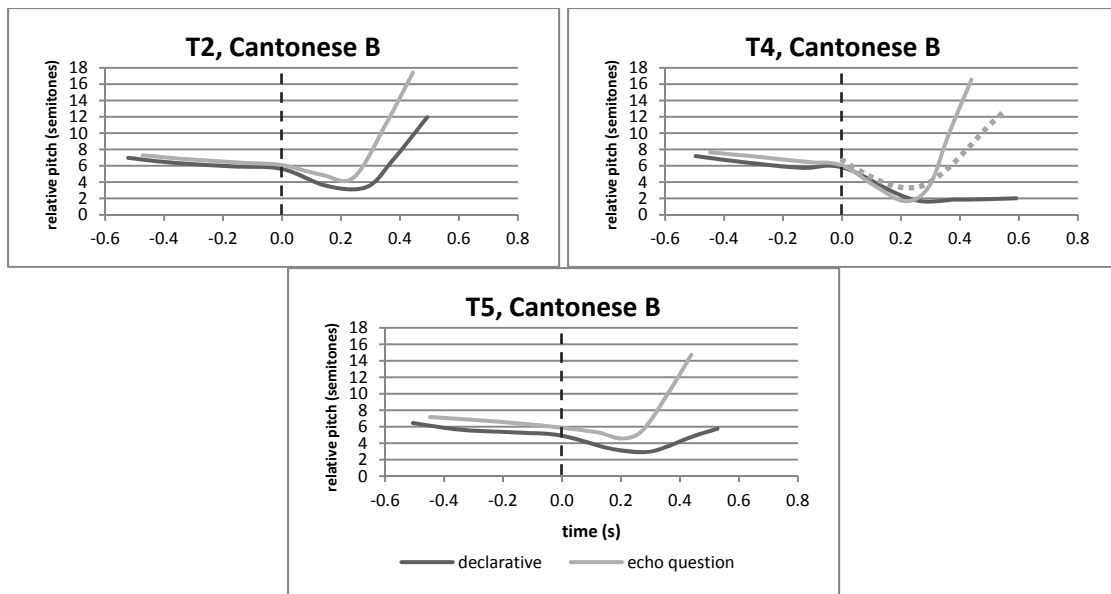


**Figure 4.5: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for Cantonese Speaker B's T1, T3, and T6, with a predicted echo question contour (dotted light gray) for T1 based on the hypothetical mechanism of a static high target approached after an initial steady-state pitch, modeled after T3 and T6.**

<sup>44</sup> i.e., relative to the lexical tonal pitch space, which itself might change depending on many factors other than sentence-type intonation.

steeper for T6 than for T3. Extrapolating from these realizations, we would expect T1, the high level tone, to follow the same pattern—the contour should start in a higher register than for T3 and then peel away partway through the syllable to reach the high target (about 14 semitones from the bottom of the speaker’s range). Since T1 is the highest-registered level tone, we expect the rising portion of the syllable in the echo question context to cover the shortest  $F_0$  range. In reality, the realization of echo question T1 does not match the predicted contour. For one thing, the echo question contour does not cleave to the declarative contour and then peel away partway through the syllable; rather, it starts to deviate right from the onset of the syllable and plots a direct (and for all practical purposes *linear*) course for the final high target. In addition, the high target is higher for this tone than for the other two tones, which makes it difficult to maintain a static-high-target analysis.

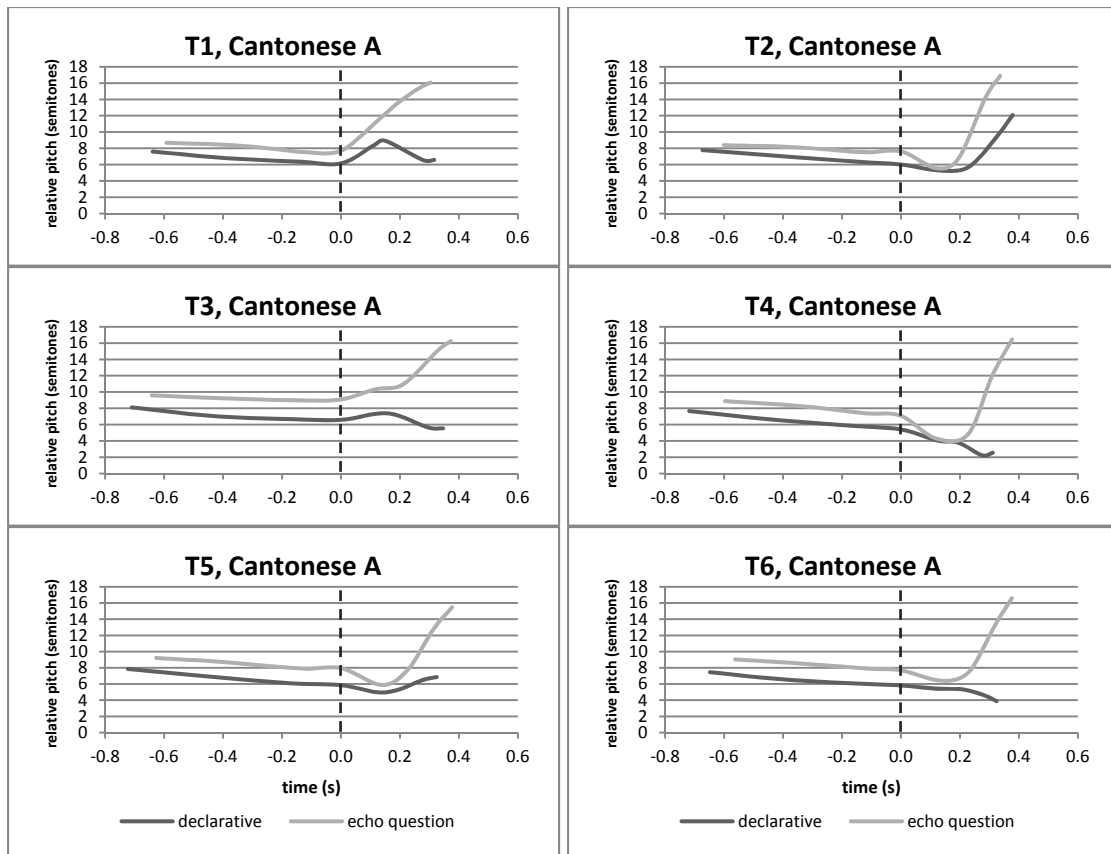
What if we had started our analysis with the two rising tones—T2 and T5? These are shown in Figure 4.6 alongside T4. Here, the echo question contours appear to behave similarly



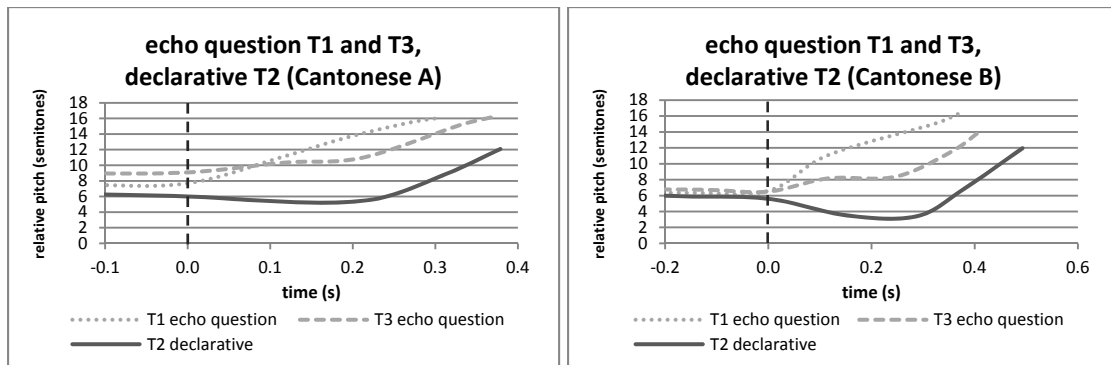
**Figure 4.6: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for Cantonese Speaker B’s T2, T4, and T5, with a predicted echo question contour (dotted light gray) for T4 based on the hypothetical mechanism of a relative high target dependent on relative declarative pitch height, modeled after T2 and T5.**

to those for T3 and T6 in that we see the same “peeling away” effect. However, now it looks as though the final high target is higher for T2 than that for T5. This preserves the contrast between the two tones, which both start at around the same pitch but end at different pitch levels in the declarative context. If we were to restrict our analysis to just this pair of tones, we might conclude that the relative height of the final high target is dependent on the relative height of the final high target of the lexical tone in question (i.e. the target is determined by an additive function). However, the echo question contour predicted by this mechanism for T4, the lowest tone in the inventory, clearly does not match reality. In fact, for Speaker B, echo question T4 reaches a higher pitch than three of the other tones (T3, T5, and T6), and exhibits the greatest pitch excursion of all the tones. Clearly, a single analysis that neatly accounts for the types of realizational differences observed with all six tones is not possible. Once again, our null hypothesis is not supported and the encoding mechanism cannot be parallel.

For the sake of thoroughness, let us briefly examine all six tones for Cantonese Speaker A. These are shown in Figure 4.7. It appears that a mechanism involving a static high target may actually be appropriate for Speaker A, who rather consistently reaches a final pitch level that is about 16 semitones from the bottom of her range. However, T1 is still exceptional for this speaker in terms of the shape of its echo question contour. All of the other tones exhibit an “elbow” in their echo question trajectories; in fact, a comparison of T1 and T3 reveals that T3 actually starts a bit higher than T1 in the echo question context, so that the main salient difference between the two tones in that context is in fact the presence vs. absence of the elbow. To show that this is the case for both speakers, mean T1 and T3 echo question contours are superimposed along with the mean T2 declarative contour on the same plot for each speaker in Figure 4.8. Recall that, in the perceptual test reported on in Chapter 3 for Cantonese, T1 was the least often misperceived out of all the tones in the echo question context, and in particular it was the least often mistaken for T2. Of course further experimental study is warranted, but it seems likely that the perceptual results are affected by the tone-dependent nature of the shape of the



**Figure 4.7: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for all six tones for Cantonese Speaker B.**



**Figure 4.8: T1 and T3 echo question contours compared to T2 declarative contours for Cantonese speakers A (left) and B (right).**

final rise for T1. In the future it would be interesting to run a controlled perceptual test in which an inflection point has been added to the echo question T1 contour to make it concave, in order to see if such a manipulation increases the rate at which it is perceived as a T2.

The tone-dependent results of the current production study for Cantonese are largely in line with previous studies. Ma, Ciocca et al. (2006) reported on results from a production study with multiple Cantonese speakers producing three-syllable utterances with declarative and interrogative intonation. Although they did not make many observations about the contours of the various final tones in the interrogative context beyond noting that T2, T4, T5, and T6 in that context all resembled T2 in the declarative context, the mean contour plots they presented spoke for themselves. They showed that interrogative T1 was clearly set apart from the other level tones by rising linearly instead of peeling away, the final high pitch levels reached were neither quite static nor consistent relative to the declarative contours. Meanwhile, Gu, Hirose et al. (2006) performed an analysis-by-synthesis study on Cantonese declarative and interrogative utterances, just as Yuan, Shih et al. (2002) did for Mandarin. Gu, Hirose et al. (2006) used a command-response model based in the same framework as the one proposed by Fujisaki and Nagashima (1969) and Fujisaki and Hirose (1984) for Japanese intonation. Where the Stem-ML approach used by Yuan, Shih et al. (2002) employs phrase curves and strength parameters to model global and local  $F_0$  excursions, respectively, the command-response approach uses phrase commands and tone commands. One of the fundamental differences between the two approaches is that the parameters in Stem-ML act on inherent target contours for each of the lexical tones, while positive and negative tone commands are the primitives in the command-response model, the result being that inherent tonal contours are described in terms of combinations of these commands, while the contours of tones in “non-neutral” settings are described in terms of different combinations of these commands. For this reason, it is impossible to compare the tone-independent strength parameters used by Yuan, Shih et al. (2002) directly to the *highly* tone-dependent tone commands used by Gu, Hirose et al. (2006). In the command-response framework, if the mechanisms were *tone-independent*, we would expect one of two results: either the amplitude of the tone commands for interrogative intonation would be the same for all tone conditions (for a static high target mechanism) or the difference between the amplitude of each tone command in the declarative context and that of its corresponding tone

command in the interrogative context should be the same (for a relative high target mechanism). In fact, neither turns out to be the case. Gu, Hirose et al. (2006) report a unique tone command amplitude for each of the six tones in the interrogative context, and their values are not derivable from the corresponding tone command amplitude values in the declarative context. Specifically, they find the greatest amplitude differences for T2, T4, T5, and T6, a lesser difference for T3, and the smallest difference for T1.

What about the unique shape of the rising trajectory for T1 as opposed to that of the other tones that both speakers in the current study exhibited? Since the command-response model allows for two unique tone commands for each tone in the echo question context as well as a variable distance between them, the more linear rise for T1 is achieved through its unique combination of commands<sup>45</sup>.

Note that the Stem-ML strength parameter would not be useful for Cantonese, since the inherent shapes of the tones are not preserved in the echo question context. The Stem-ML model for Mandarin would need to be expanded to allow for an additional “intonational target” in the echo question context (abandoning the parallel implementation), and the interpolation from the rightmost lexical tone to this target would need to be different in the case where the rightmost lexical tone is T1 (abandoning parallel encoding). It is unclear how Xu’s (2005) PENTA model would handle Cantonese, short of employing phrase curves that rise up sharply starting partway through the final syllable (abandoning syllable-synchronicity<sup>46</sup>); of course, in that case the shape of the phrase curve would have to be tone-dependent (abandoning parallel encoding). Regardless of the mechanisms employed, it is clear that the null hypothesis is not tenable for Cantonese.

---

<sup>45</sup> Specifically, they set the intrinsic starting pitch of T3 as the baseline, so T1 receives two positive tone commands placed adjacent to one another, T3 receives no command in its first half followed by a positive command in its second half, and all the other tones receive a negative command followed by a positive command, with space in between the two.

<sup>46</sup> Recall that Xu’s (2005) assumes parameter assignment to be synchronized with the syllable (the temporal domain of the syllable corresponding to an articulatory domain of jaw movement).



#### **4.2.4 Interim summary and conclusion**

Thus far we have addressed Q3 (*Declarative-to-Echo-Question Mapping*) with respect to syllable-tone languages, including two dialects of Mandarin and a dialect of Cantonese. It has been clearly demonstrated that modeling the melodic systems of these languages entails incorporating tone-dependent intonational mechanisms to account for the tone-dependent phenomena that pervade those systems. It is reasonable to ask whether this is a property unique to syllable-tone languages. It is conceivable, for example, that the higher tonal density of syllable-tone languages entails in some way a more complex interaction of tonal events at the right edge of an utterance, giving rise to tone-dependent intonational phenomena. In the following section, therefore, the melodic systems of word-tone languages are critically assessed from the point of view of modeling their intonation in the phonetics; it will be shown that the null hypothesis is not supported for word-tone languages, either.

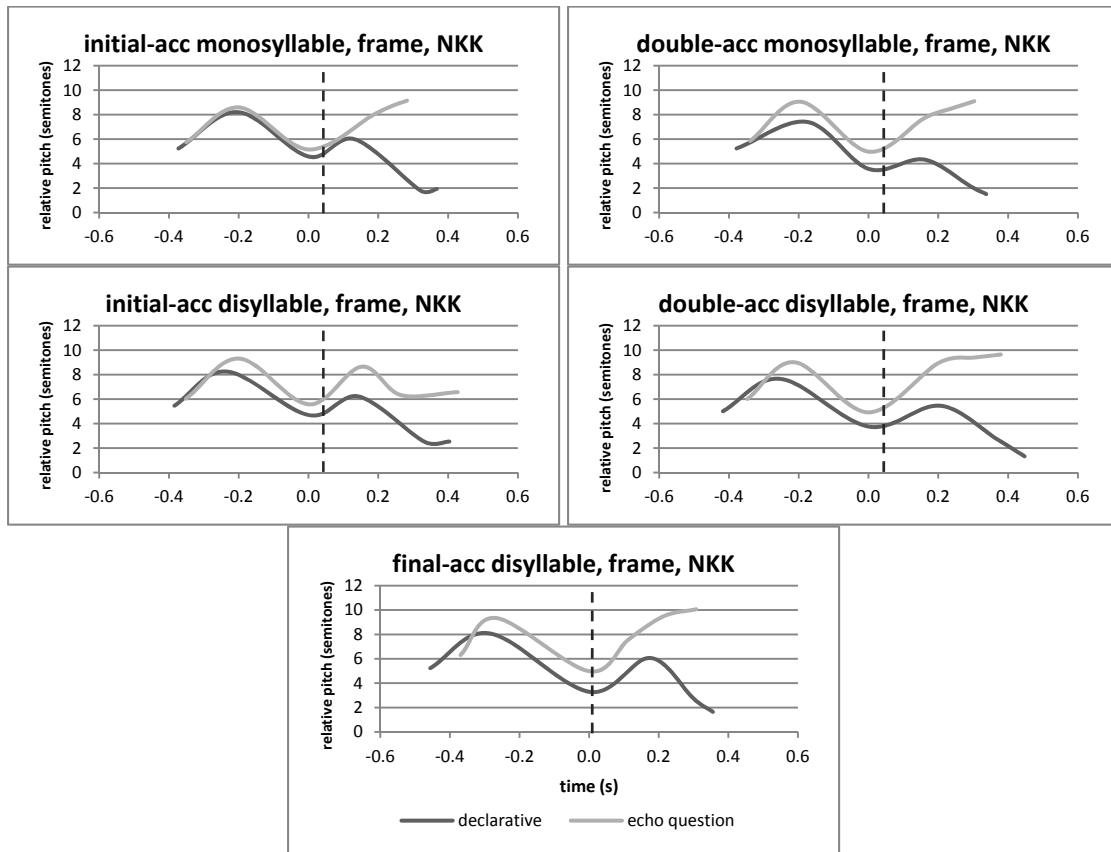
#### **4.3 Accounting for Tone-Dependent Intonation in Word-Tone Languages**

In this section, the results from NKK and the Kansai Japanese dialects that bear on the issue of tone-dependent intonation will be discussed. We saw in Section 4.2 that there is no getting around incorporating tone-dependent intonational mechanisms into our model when it comes to syllable-tone languages. However, syllable-tone languages such as Mandarin and word-tone languages such as Japanese traditionally receive radically different treatments when it comes to their melodic systems, and as such it is worth pursuing the possibility that we might be able to get away with maintaining a parallel encoding model for the latter (which is favorable inasmuch as such a model is simpler and less powerful). In Section 4.3.1 the apparently exceptional behavior of the initial-accented disyllable is given a tentative articulatory explanation, but this articulatory explanation is shown to be ruled out empirically. In Section 4.3.2, tone-dependent intonational behavior in Kyoto Japanese is highlighted, with the equivalent tone-intonation combinations in Shiga Japanese provided as a reference. These phenomena in NKK and Kyoto

Japanese extend the need for tone-dependent encoding in our intonational model to word-tone languages.

### 4.3.1 NKK

It was shown in Chapter 2 that the primary dimension of phonetic contrast among the three tonal categories on disyllables in NKK appears to be horizontal peak alignment. The initial-accent has the earliest pitch peak, final-accent has the latest pitch peak, and that of the double-accent falls somewhere in between. Echo question intonation can generally be described as the global raising of the overall pitch range and the realization of a high target at the end of the utterance. Declarative and echo question contours are shown in Figure 4.9, in where pitch values are plotted on a semitone scale. Upon first glance, it appears that the initial-accented disyllable

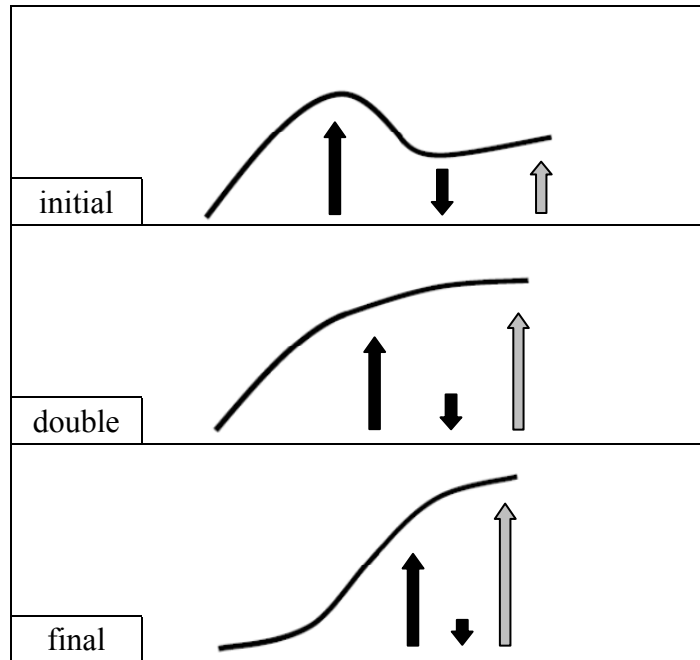


**Figure 4.9: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for all tonal categories on monosyllabic (top two) and disyllabic (bottom three) words in NKK.**

behaves exceptionally in that it is the only tonal-category-syllable-count combination for which the contour has a falling component in the echo question context (despite the fact that all of the tonal contours are characterized by a falling component in the declarative context). As such it appears that we have a case of tone-dependent intonation in this melodic system. However, if we assume that the phonetics is implementing three targets—a high target, a low target, and another high target—for every tonal category in the echo question context, the seemingly exceptional behavior of the initial-accented category can be seen as falling out from phonetic co-articulatory effects on the other two tonal categories. The target configurations corresponding to the three tonal categories on disyllables can be characterized according to the relative distance of the peak of the lexical tone from the right edge of the final syllable, and therefore from the high target contributed by echo question intonation. As the initial high target gets closer to the final high target, the former “boosts” the latter and the latter gets realized with a higher surface pitch. Meanwhile, the intervening low target gets more and more undershot<sup>47</sup>, as articulatory priority is given to the realization of the high targets. This interaction is schematized in Figure 4.10. The assumption of such an analysis is that the low target in the initial-accented case is realized as lower than the first high target because there is sufficient time for it to be at least partially realized. In the double-accented and final-accented cases, the low target is undershot to the point where there is no actual dip in the contour corresponding to it. Two points regarding this analysis should be noted. For one thing, as the three targets get closer to one another, the realization of each target is affected in a unique way—the realization of the initial high target remains relatively consistent and unaffected, the low target gets more and more undershot, and the pitch realization of the final high target gets more and more boosted. Second, the analysis hinges on the premise that the realizations of the three targets can be predicted solely on the basis of temporal proximity.

---

<sup>47</sup> Lee (2008) also interpreted the shape of the initial-accented echo question contour as a case of undershot low and high targets.



**Figure 4.10: Schematization of a purely phonetic interaction between high and low targets in NKK.**

The above premise can be tested empirically by manipulating the temporal distance between the initial high target and the final high target via slower and faster speech rates. If the temporal-distance-based phonetic model is correct, we should observe the same effects across the same temporal distances regardless of the tonal categories or the number of syllables involved. An informal follow-up experiment was carried out with the original NKK speaker in order to test this hypothesis. The NKK speaker was instructed to produce slower and faster renditions of each of the sentences shown in (4.2):

(4.2) NKK sentences for testing the temporal-distance hypothesis

- a. monosyllabic initial-accented, declarative

*Eunhi-neun nam.*  
Eunhi-TOP south

- b. monosyllabic initial accented, echo question

*Eunhi-neun nam?*  
Eunhi-TOP south

c. disyllabic initial-accented, declarative

*Eunhi-neun nam-i.*

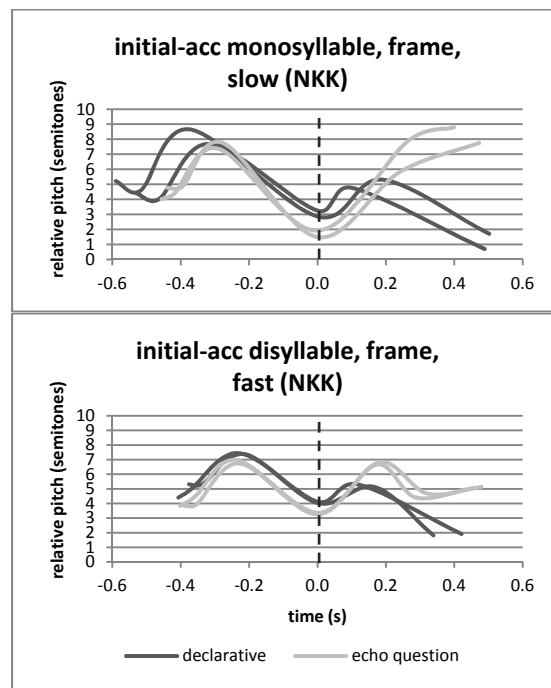
Eunhi-TOP south-NOM

d. disyllabic initial-accented, echo question

*Eunhi-neun nam-i?*

Eunhi-TOP south-NOM

The crucial comparison—the slower rendition of the monosyllabic case vs. the faster rendition of the disyllabic case—is shown in Figure 4.11. Note that the temporal distance between the



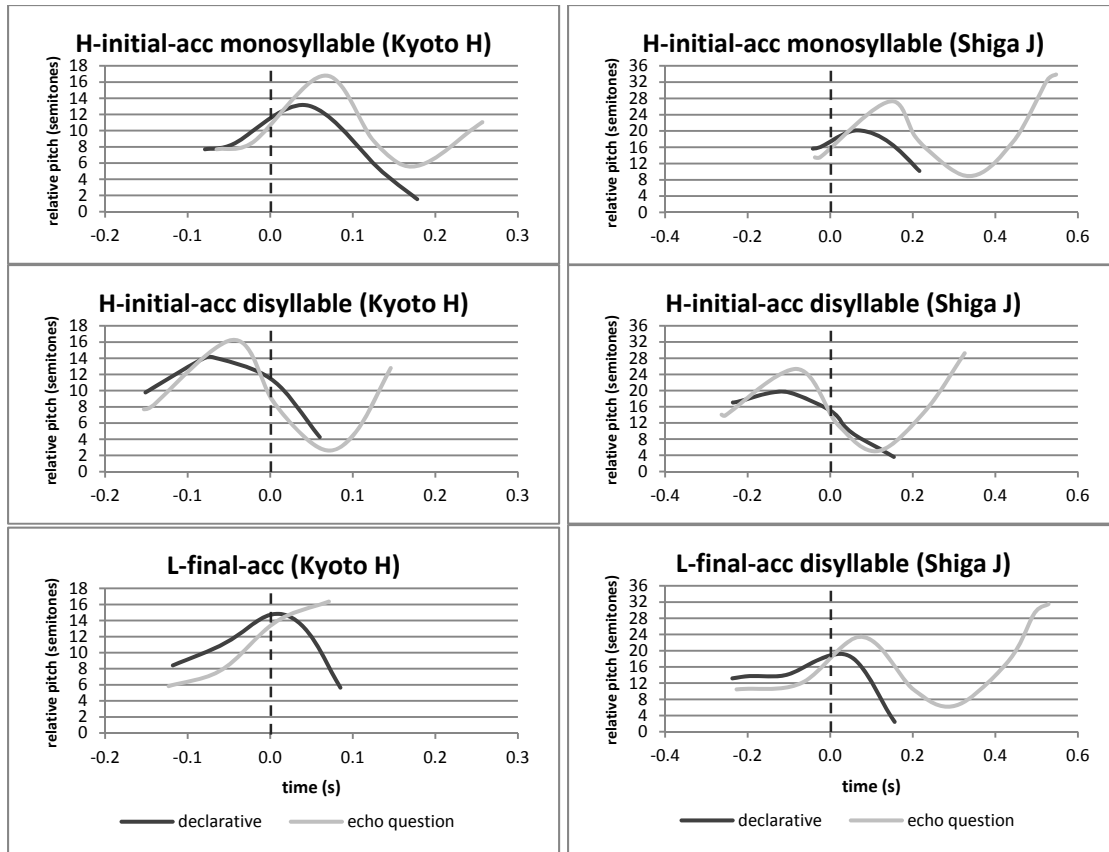
**Figure 4.11: Individual declarative (dark gray) and echo question (solid light gray) pitch contours for an initial-accented monosyllable at a slow speech rate (above) and an initial-accented disyllable (below) in NKK.**

accentual peak and the end of the utterance is clearly shorter in the (fast) disyllabic declarative case than in the (slow) monosyllabic declarative case. In both cases, the echo question utterances were produced at speech rates comparable to those for the corresponding declarative utterances (as evidenced by the similarly aligned frame contours in both cases); yet the disyllabic echo question contours include a falling component while the monosyllabic contours lack one.

Thus, it is clear that the temporal distance hypothesis is not supported—within the initial-accented lexical tonal category, the shape of the echo question contour is dependent on syllable count and not duration. Since this syllable-count-dependent alternation is only observed for the initial-accented category, it must indeed be treated as a tone-dependent intonational phenomenon. Most likely, the overlay models discussed thus far would handle this alternation by including or excluding a low target in the encoding scheme, but it is unclear how its inclusion or exclusion would be conditioned other than by making reference both to the lexical tonal category and to the syllable count of the final word. The null hypothesis (that tone-specific mechanisms do not need to be incorporated into a phonetic model of speech melody) is not supported for NKK.

### **4.3.2 Kansai Japanese**

The contours of the accented categories in Kansai Japanese, like those in NKK, are characterized by a rise to a peak and then a subsequent fall in a declarative context. For the most part, the echo question contours include peak and a fall followed by a subsequent rising “tail” (the peak is slightly raised relative to the declarative contour). In Kyoto Japanese, however, one tonal category is realized differently in the echo question context. Figure 4.12 shows mean declarative and echo question pitch contours for all accented categories on monosyllables and disyllables for Kyoto, with the equivalent categories in Shiga alongside it for comparison. Both the H-initial-accented monosyllable and the H-initial-accented disyllable in Kyoto Japanese follow the general pattern in the echo question context, but the L-final-accented disyllable in that dialect shows a very different interaction with echo question intonation. Specifically, there is no falling component or subsequent rising tail in the echo question contour; the pitch rises through the end with no apparent addendum of any kind. As is apparent in Figure 4.12, the equivalent contour in Shiga Japanese behaves as expected.



**Figure 4.12: Mean declarative (dark gray) and echo question (solid light gray) pitch contours for accented monosyllables (top) and disyllables (middle and bottom) in Kyoto Japanese (left) and Shiga Japanese (right).**

Accounting for this tone-dependent intonation in Kyoto Japanese in a purely phonetic model is a tall order. On the other hand, if a decompositional phonological representation is incorporated into the model, the behavior of the L-final-accented contour can be explained by a phonological deletion of part of the melody, conditioned by the intonational context. A representation of this kind is proposed in Chapter 6.

#### 4.4 Summary and Conclusion

In this chapter, we have seen evidence from both syllable-tone languages (Mandarin, Henanhua, and Cantonese) and word-tone languages (NKK, Kyoto Japanese) that a comprehensive model of speech melody must allow for tone-dependent intonation. In purely phonetic overlay models, the only way to handle such tonal-category-specific effects is to modify them to allow for a unique

phonetic implementation for every tone-intonation combination; this is the solution adopted by Gu, Hirose et al. (2006) in their command-response model of Cantonese speech melody; in assigning unique amplitude values to every tone-intonation combination, they were abandoning any notion of parallel encoding of tone and intonation and giving their model (and by extension the phonetics) enough power to generate virtually any unique contour shape for any tone-intonation combination. While this level of power indeed allows the model to produce the correct contours for echo questions in Cantonese, it is not clear what the constraints on such a model would be. The main appeal of overlay models that maintain parallel encoding and parallel implementation is that various melodic functions can be encoded independently and then the final pitch contour can be derived by combining the functions additively, multiplicatively, or in some other general algorithmic way. Since the findings in the current study preclude such models, the question becomes whether all of the power to generate category-specific differences within the melodic system of a given language (as well as differences across languages and dialects such as those discussed in Chapters 2 and 3) should be bestowed on the language-specific phonetics, or whether some or all of it should be handled by phonology.

In Chapter 5, it is argued that shifting at least some of the burden of explanation to the phonology allows some tone-dependent intonational phenomena, as well as some cross-linguistic similarities and differences, to be explained—and constrained—structurally. Further, incorporating phonological representations in our model of speech melody allows for more useful characterizations of language-specific phonetics, because cross-linguistic or cross-dialectal variation can be captured in terms of structurally similar representations being realized with different phonetics. A model that attempts to strike the right balance of power between phonology and phonetics is presented in Chapter 6.



## CHAPTER 5: A PHONOLOGICAL APPROACH TO MELODIC INTERACTIONS

### 5.1 Introduction

In Chapter 4, it was shown that phonetics-only overlay models in their “purest” form are not powerful enough to handle the tone-dependent intonational phenomena that are apparent in the melodic systems of the various languages that were investigated in Chapters 2 and 3. It was suggested that one way to “rescue” the overlay models so that they are able to handle those phenomena is to abandon the tenet of parallel encoding and to allow the phonetic mechanism that encodes utterance-type intonation to be sensitive to the lexical tonal categories involved in the utterance. It was noted that Gu, Hirose et al. (2006) resorted to this approach by making the parameter values for an echo question rise at the end of an utterance dependent on the identity of the last lexical tone in the utterance. This approach centers on the implementation of *tone-dependent intonation algorithms* in the phonetics. An alternative approach to this problem is to include an active phonological component in the model—one in which the various melodic functions can interact at a structural level such that the actual phonological representations being sent to the phonetics can alternate depending on the melodic (or metrical or segmental) context. This second approach can be characterized as a kind of *intonation-dependent allotony*.

In Section 5.2 of this chapter, the concept of intonation-dependent allotony is briefly explored and its effectiveness in handling at least a subset of the melodic irregularities discussed in Chapter 4 is illustrated. Then, in Section 5.3, analyses couched in an established formalized framework that allows for such phonological interactions—the *autosegmental-metrical* framework—are evaluated for their ability to account for the results obtained in Chapter 2 for the word-tone languages. The merits and shortcomings of the traditional autosegmental-metrical accounts that are highlighted in this chapter are taken into account in Chapter 6, where a more powerful autosegmental-metrical model is proposed.

## 5.2 Tone-Dependent Algorithms vs. Intonation-Dependent Allotones

If we look at traditional treatments of certain tonal alternations outside of the domain of utterance-type intonation, such as those dependent on phrase or word position, we see that the accounts given are often structural in nature. For example, traditional autosegmental treatments of Tokyo Japanese (Haraguchi 1977, e.g.) account for the lack of a falling trajectory on final-accented words by invoking a deletion rule (or, more precisely, a “failure-to-associate” rule) that keeps the L-tone component of the accentual tune from being realized when there is no post-accent mora on which to realize it. In essence, then, the contour that surfaces without the fall is a kind of “allotone” of the accentual tune. Compare this to the treatment of T3 in standard Mandarin, which surfaces with a rising tail (optional for some speakers) when the tone is phrase-final, as a rising tone before other T3s, and as a low tone everywhere else. These three realizations are given allotone status in many traditional analyses (e.g., see Chen 2000), so that they are represented with different sets of Chao-letter numerals—**214** for phrase-final, **35** for pre-T3, and **21** elsewhere. Note the similarity between a rule that says **HL**→**H** and one that says **214**→**21**.

Let us briefly revisit the Kansai Japanese example discussed in Section 4.3.2. Recall that the three dialects—Osaka Japanese, Kyoto Japanese, and Shiga Japanese—are quite similar in many respects. But Osaka Japanese behaves like Tokyo Japanese when it comes to word-final accents; the falling component of the accent is absent when the accent falls on a word-final<sup>48</sup> syllable. So for Osaka Japanese as well as Tokyo Japanese we could posit a rule<sup>49</sup> such as the one shown in (5.1) for Osaka:

(5.1) Word-final L-deletion

**HL**→**H** / \_#

---

<sup>48</sup> It may be more precise to formulate this as accentual-phrase-final L-deletion, but the distinction is not crucial here.

<sup>49</sup> This can be thought of as shorthand for several rules in a strictly autosegmental framework: the H is associated with the accented mora, the L fails to associate for lack of a post-accentual TBU, and unassociated tones get deleted.

We saw that Shiga Japanese would have no such rule, since the fall gets realized even when the accent is word-final. In a constraint-based framework, one could characterize the dialect difference thus: markedness outranks faithfulness in Osaka Japanese, and faithfulness outranks markedness in Shiga Japanese.

So, if tone sandhi (i.e. an allotonic alternation) can be dependent on phrase or word position, or by the tonal identities of surrounding TBUs, why can it not be dependent on intonational context? Such a rule, presented in (5.2), would be perfect to account for the distribution of the HL surface tune in Kyoto, where the fall appears to surface on final-accented words in the declarative context but not in the echo question context:

(5.2) Utterance-type-conditioned word-final L-deletion in Kyoto Japanese

**HL**→**H/**\_#?

Formulated thus, this rule seems suspect, since it appears to be making reference directly to semantics. But if we put it in more formal terms by assuming that the implementation of echo question intonation involves the insertion of a H tone, it becomes more acceptable. This strictly phonological formulation of the rule is given in (5.3):

(5.3) Intonation-conditioned L-deletion in Kyoto Japanese

**HL**→**H/**\_H%

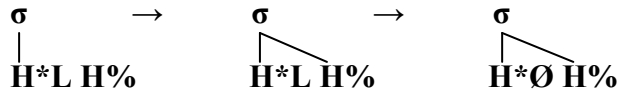
In autosegmental terms, the rule would appear as shown in (5.4):

(5.4) Autosegmental intonation-conditioned L-deletion in Kyoto Japanese

$\begin{array}{c} \mu \\ | \\ \mathbf{H^*L H\%} \end{array} \rightarrow \begin{array}{c} \mu \\ | \quad \backslash \\ \mathbf{H^*L H\%} \end{array} \rightarrow \begin{array}{c} \mu \\ | \quad \backslash \\ \mathbf{H^*\emptyset H\%} \end{array}$

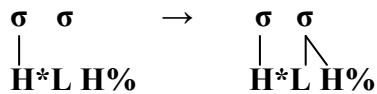
A similar rule, with a different TBU, could be formulated to handle the alternations observed in the NKK melodic system. This is given in (5.5):

(5.5) Intonation-conditioned L-deletion in NKK



Note that this rule would affect both final-accented polysyllabic forms and monosyllabic initial-accented forms at the right edge of an utterance as long as both are represented with a **H\*** associating with the rightmost syllable in the utterance. Initial-accented polysyllabic (including disyllabic) words would be unaffected, though, since the **H\*** and the **L** would each have their own TBUs with which to associate. This case is shown in (5.6):

(5.6) Lack of L-deletion for polysyllabic initial-accented forms in NKK



How the double-accented forms fit into the picture is not obvious, mostly because it is not obvious how they should be represented structurally; this issue will be taken up in Section 5.3.1 and then again in Chapter 6. For now it suffices to say that we would want the L-deletion rule in NKK to affect double-accented monosyllabic and disyllabic forms.

There is precedent for rules like these in other languages. Consider the following example from Shingazidja, taken from Patin (2008). Shingazidja is a Bantu language that can be analyzed as having a /H,  $\emptyset$ / tonal contrast underlyingly. In Shingazidja, polar questions are usually analyzed as triggering the insertion of a “superhigh” tone on the penultimate syllable of the utterance. The syllables that separate the superhigh-toned syllable from the syllable bearing the preceding tone get realized as high. Some representative declarative/interrogative pairs given by Patin (2008) are shown in (5.7):

(5.7) Shingazidja statements and polar questions

- a. ha-níká        ze    ɲ-uŋgu        m-6íli  
3SG.PST.give At<sub>10</sub> 10-cook.pot 10-two  
'he gave the two cooking pots'
- b. ha-níká        zé    ɲ-úŋgú        m-6íli  
'did he give the two cooking pots?'
- c. ha-níka        ɲ-úŋgu        n-dziro  
3SG.PST.give 10-cook.pot 10-heavy  
'he gave (some) heavy cooking pots'
- d. ha-níka        ɲ-úŋgú        n-dzíro  
'did he give (some) heavy cooking pots?'

Note that the superhigh tone on the penultimate syllable in (5.7b) and that in (5.7d) are realized at the same relative pitch height, suggesting a true “overwriting” of the underlying tonal specification, which is /H/ in (5.7a) and /Ø/ in (5.7c). Interestingly, when the final syllable of a declarative utterance bears a high tone, the superhigh tone in the polar question version of the utterance gets placed on the *antepenultimate* syllable, and the high tone on the final syllable is deleted. The relevant alternation is shown in (5.8):

(5.8) Shingazidja high-tone-final statement and question

- a. ha-ono        m̄-pirá  
SG.PST-see 3-balloon  
'he saw a balloon'
- b. ha-ono        m̄-pira  
'did he see a balloon?'

In this analysis of the Shingazidja melodic facts, we see instances of both insertions and deletions of tonal units triggered by the insertion of an intonational melodic unit.

This type of analysis for Kyoto, NKK, and Shingazidja might still be comfortable territory for phonologists since the intonational trigger has a somewhat discreet, tangible surface realization (in the form of an inserted H intonational unit), and its effect on surrounding TBUs

can be characterized by insertions or deletions of other lexical tonal units. But could such rules be called upon to handle the following typological facts, reported by Köhnlein (2011) for Arzbach, a Franconian tone dialect? What makes Arzbach unique as a dialect is that its Class 1 and Class 2 tonal contours correspond respectively to Class 2 and Class 1 tonal contours in surrounding dialects such as Cologne. But even more interesting is that in an interrogative context, there is no reverse-correspondence from one dialect to the other. This tune breakdown is schematized in Figure 5.1. Since the interrogative contours maintain the lexical tone contrast

Condition	Arzbach		Cologne	
	Class 1	Class 2	Class 1	Class 2
Declaration, non-final position				
Declaration, final position				
Interrogation, non-final position				
Interrogation, final position				

**Figure 5.1: Idealized contours associated with Class 1 words and Class 2 words in declarative and interrogative contexts in Arzbach and Cologne (from Köhnlein 2011).**

but cannot be easily derived from a combination of the declarative version of the tone contour and some single interrogative intonational element, the most straightforward analysis is that an actual “tune change” takes place in an interrogative context. This view makes Arzbach look a lot like a hypothetical version of Mandarin in which T3 sandhi is triggered by the interrogative context instead of lexical tonal context. It also becomes tempting to apply a similar type of analysis to the unique contours observed for each of the tonal categories in Cantonese echo questions. Under this allotonic alternation view, the Cantonese echo question rendition of T1, which starts high and rises linearly throughout, would simply be the echo question *allotone* of T1,

whose declarative allotone starts high and stays level. The echo question allotone of T3, meanwhile, would start at a mid-register and peel upward starting partway through the syllable.

The danger in allowing for such power in the phonology is that it is unclear where to draw the line in terms of how many intonation-sensitive allotones a given toneme may have. Would this mean, for example, that every tone has various allotones for various other speech acts? Also, if allotones can be thought of as the tonal equivalents of allophones, can we take the concept one step further and think of them as equivalent to *allomorphs*? In other words, could both the declarative and echo question versions of a given lexical tone just be stored in the lexicon? While this seems like a radical position, consider the analogy espoused by Myers and Tsay (2003) regarding the differences between T3 sandhi in Taiwan Mandarin and T3 sandhi in Beijing Mandarin, respectively. Citing results that indicate that the sandhi process is more “categorical” in Taiwan Mandarin (i.e. the sandhi form of T3 is acoustically the same as the citation form of T2) and more “gradient” in Beijing Mandarin, they compared the former to the *a/an* allomorphic alternation in English and the latter to flapping in English. Their results and concomitant discussion make the notion of putting different allotones in the lexicon seem more reasonable, but the same question asked above would just get displaced to here—how many allotones can possibly be stored in the lexicon?

On the other hand, if we were to shift the *entire* burden of explanation to the phonetic implementation, it would without a doubt require that the phonetic implementation be privy to lexical tonal categories, and therefore it would preclude parallel encoding. Furthermore, algorithms for implementation would have to be quite powerful, determining interpolation strategies, direction of register shift, degree of distortion, etc., all on a tone-by-tone basis. Some implementation strategies—such as the flattening of the slope of T4 in Mandarin echo questions—could be viewed as being coarticulatory in nature, while others—such as the exaggeration of the pre-T3 peak in Mandarin—could be viewed as dissimilatory. Still others—like the linear rise of T1 in Cantonese echo questions—would not be easily explained by either syntagmatic principle but could perhaps be understood in paradigmatic terms, e.g. maximizing

tonal contrast. Indeed, Chen and Gussenhoven (2008) offered such a notion as the driving force behind the tone-dependent effects they observed in connection with the realization of multiple levels of emphasis in Mandarin. The types of alternations seen in Kyoto Japanese and NKK would be the most difficult to capture in purely phonetic terms, since they seem best characterized in terms of the inclusion or exclusion of tonal targets.

Ultimately, there is no empirical way to decide between tone-dependent phonetic algorithms and intonation-dependent allotony (or a model that incorporates both) given the results available in the current study, although some of the results seem to lend themselves to one approach over the other. Until such conclusive experimental evidence<sup>50</sup> becomes available, the theoretical choice remains largely an aesthetic one; however, if we adopt a particular phonological model—say, one that is tailored toward handling cases like the ones in Kyoto Japanese and NKK, it then becomes easier to talk about what the phonetics must handle. To that end, in the rest of this chapter the autosegmental-metrical framework is assessed for its applicability to the Kansai and NKK melodic systems.

### 5.3 The Autosegmental-Metrical Framework

The notion of representing lexical tones as autosegments was adopted early on for African tone languages by Leben (1973), Goldsmith (1975b; 1976), and Williams (1976), and for Japanese by Goldsmith (1975a) and Haraguchi (1977). Meanwhile, Liberman (1975), Bruce (1977), and Pierrehumbert (1980) proposed autosegmental frameworks for representing intonation in the phonology. With the widespread adoption of ToBI (discussed in Chapter 1), the representational framework introduced in Pierrehumbert (1980) came to be widely adopted and was later coined the *autosegmental-metrical* (henceforth AM) theory of intonation by Ladd (1996). While the

---

<sup>50</sup> The kind of evidence that would be convincing in this regard would be the kind referred to by Cohn (1998; 2006) as *phonetic and phonological doublets*—cases where “similar but distinct effects of both a categorical and gradient nature are observed in the same language” (Cohn 2006, p. 29). One example discussed by Cohn (1998) is from Sundanese, where nasal airflow data suggest the co-existence of both phonological and phonetic nasalization in that language.



version of the AM theory presented in Pierrehumbert (1980) was originally geared toward languages like English that lacked lexical tone, Poser (1984) and Pierrehumbert and Beckman (1988) brought lexical tonal units and intonational units onto the same dimension of representation in their treatments of the Japanese melodic system. In its current instantiation, the standard AM theory includes the assumptions listed in (5.9):

(5.9) Assumptions of standard AM theory (adapted from Ladd 2008)

- a. At every level of the prosodic hierarchy there is at most a two-level pitch contrast.
- b. Pitch accents are representable as complexes of autosegmental H and L tones.
- c. The phonetics is blind to the source of a tonal autosegment (i.e. whether it is lexically or postlexically specified).

One advantage of the AM approach to speech melody over purely phonetic approaches is that it enables us to place structurally based constraints on the types of interactions that are permitted among various melodic functions. In addition, the AM framework has the potential to make cross-linguistic comparisons more fruitful because analogous structural elements can be thought of as being subject to different phonological processes or, in some cases, because analogous elements are being manipulated in the same way across languages (or dialects), so any observed differences in the surface contours can then be attributed to a more restricted language-specific (or dialect-specific) phonetics. Of course, the widespread adoption and “splintering” of ToBI—i.e. the development of various, subtly incompatible ToBI-based transcription systems in different languages—has obscured the overall typological picture somewhat (for reasons given in 1.3 of Chapter 1; also see Wightman 2002 for a critical discussion of ToBI), but this reality does not lessen the potential usefulness of the AM framework or AM-derived frameworks.

How can the advantages of the AM framework be exploited and/or augmented to explain the cross-linguistic differences among the melodic systems presented in this dissertation? For one thing, it is necessary to acknowledge that, in some melodic systems, melodic functions

appear to be implemented serially, and in some melodic systems they appear to be implemented in parallel. The need for a parallel implementation mechanism for Mandarin, for example, was highlighted by the use of multiple tiers for melodic specification in the M\_ToBI system proposed by Peng, Chan et al. (2005). A structural mechanism that reconciles parallel and serial implementation within one framework is proposed in Chapter 6; as such, this issue will be set aside for the rest of the current chapter. An additional challenge for the traditional AM framework is how to represent the lexical tones of syllable-tone languages (like Mandarin, Henanhuà, and Cantonese) in ways that foster comparison with those of word-tone languages (like Japanese dialects and NKK); this is also taken up in Chapter 6.

Before a theory or a framework can be shown to be effective from a typological perspective, however, it of course must be deemed capable of handling a system within a single language. As such, the focus of this section will be not on the typological issues mentioned above but on how the traditional AM framework fares in handling data internal to each of two word-tone languages—languages whose melodic systems closely resemble that of standard Japanese and therefore should be quite amenable to traditional AM accounts. For each language, (NKK in 5.3.1 and Kansai Japanese in 5.3.2), the AM analyses are followed through to their logical conclusions. In the end it becomes clear that, either the AM representations proposed in the literature for these languages are insufficient, or we need to rethink how the phonetics interacts with the phonology when it comes to tonal realization, or both.

### **5.3.1 AM analysis for NKK**

A first attempt at a set of rules governing NKK melody is given in (5.10); the rules given here are unordered, other than the postlexical rules following the lexical rules:

(5.10) Melodic rules for NKK

I. Lexical rules

- a. Associate H\* with every accented syllable<sup>51</sup> in the word.
- b. Associate trailing L with the syllable following the rightmost accented syllable, if there is one.

II. Postlexical rules

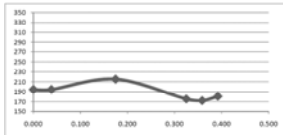
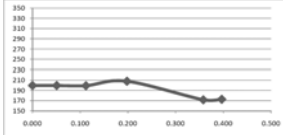
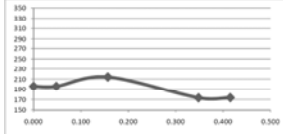
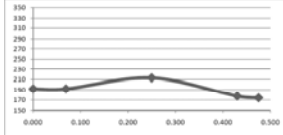
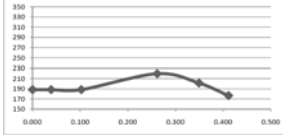
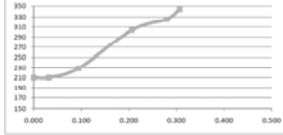
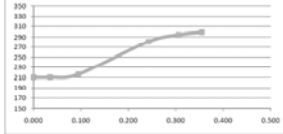
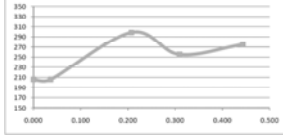
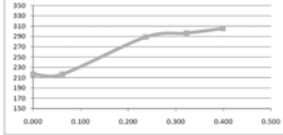
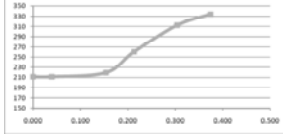
- a. Insert initial L% and associate it with the initial syllable.
- b. Insert final L% (for declaratives) or H% (for echo questions) and associate it with the final syllable.
- c. Delete unassociated melodic units.

Derivations for the various combinations of lexical tones and utterance-type intonational units are shown in Table 5.1. In this first attempt, the double-accented class is analyzed as a doubly-associated H in the underlying representation, after Jun, Kim et al. (2006). The assumption for double-accented monosyllables (not covered by Jun, Kim et al. 2006) is that the association of the accent with the second syllable is contingent upon there being a second syllable. In the “postlexical” column, dashed lines represent “default” associations and Ø represents a deleted tone.

---

<sup>51</sup> This assumes an underlying representation with the accented syllable specified with a diacritic; an alternative analysis would be to start with H\* associated with the relevant TBU(s).

**Table 5.1: First attempt at derivations for NKK melodies**

name	lexical	postlexical	output	surface
monosyllabic initial declarative	$\acute{\sigma}$   H*L L%	$\sigma$ / \ \ L% H*Ø L%	$\sigma$ / \ \ L%H*L%	
monosyllabic double declarative	$\acute{\sigma}$ (')   H*L L%	$\sigma$ / \ \ L% H*Ø L%	$\sigma$ / \ \ L%H*L%	
disyllabic initial declarative	$\acute{\sigma}$ $\sigma$     H*L L%	$\sigma$ $\sigma$ / \ / \ \ L% H*L L%	$\sigma$ $\sigma$ / \ / \ \ L%H*L L%	
disyllabic double declarative	$\acute{\sigma}$ $\acute{\sigma}$ / \ H*L L%	$\sigma$ $\sigma$ / \ / \ \ L% H*Ø L%	$\sigma$ $\sigma$ / \ / \ \ L% H* L%	
disyllabic final declarative	$\sigma$ $\acute{\sigma}$     H*L L%	$\sigma$ $\sigma$ / \ / \ \ L% H*Ø L%	$\sigma$ $\sigma$     L%H*L%	
monosyllabic initial echo question	$\acute{\sigma}$   H*L H%	$\sigma$ / \ \ L% H*Ø H%	$\sigma$ / \ \ L%H*H%	
monosyllabic double echo question	$\acute{\sigma}$ (')   H*L H%	$\sigma$ / \ \ L% H*Ø H%	$\sigma$ / \ \ L%H*H%	
disyllabic initial echo question	$\acute{\sigma}$ $\sigma$     H*L H%	$\sigma$ $\sigma$ / \ / \ \ L% H*L H%	$\sigma$ $\sigma$ / \ / \ \ L%H*L H%	
disyllabic double echo question	$\acute{\sigma}$ $\acute{\sigma}$ / \ H*L H%	$\sigma$ $\sigma$ / \ / \ \ L% H*Ø H%	$\sigma$ $\sigma$ / \ / \ \ L% H* H%	
disyllabic final echo question	$\sigma$ $\acute{\sigma}$     H*L H%	$\sigma$ $\sigma$ / \ / \ \ L% H*Ø H%	$\sigma$ $\sigma$     L%H*H%	

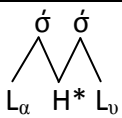
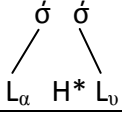
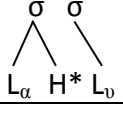
This analysis is also expressible in terms of constraints, a la Gussenhoven (2004). The relevant constraints for such an analysis are given in (5.11):

(5.11) Relevant constraints for NKK melodies

- a. **L<sub>α</sub>**: APs have initial L phrase tones.
- b. **MAXIO(H\*)**: H\* in the input has a correspondent in the output.
- c. **MAXIO(T%)**: Boundary tones in the input have correspondents in the output.
- d. **ALIGN(L<sub>α</sub>, Left)**: Align L<sub>α</sub> left in the AP.
- e. **ALIGN(T%, Right)**: Align boundary tones right in the utterance.
- f. **ALIGN(H\*, Lex)**: Align H\* with the accented syllable.
- g. **CONCATENATE**: Trailing L is left-aligned with the right edge of H\* (i.e. no intervening tones).
- h. **H\*→TBU**: H\* is associated with a TBU.
- i. **NOSPREAD**: One TBU per lexical tone.
- j. **NOCONTOURLEX**: One lexical tone per TBU.
- k. **T→TBU**: A tone must be associated with a TBU.
- l. **MAXIO(L)**: Trailing L in the input has a correspondent in the output.
- m. **NOCONTOUR**: One tone per TBU.

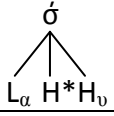
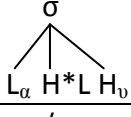
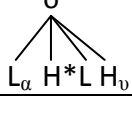
Constraints (a-g) are highly ranked and collectively ensure that every AP has an initial L<sub>α</sub> that aligns left, that H\* is preserved and aligns with the accented syllable, that final boundary tones are preserved and align right, H\* associates with accented syllables, and that trailing L associates with the post-accentual syllable (when it associates). **ALIGN(H\*, Lex)** and **H\*→TBU** both dominating **NOSPREAD** ensures that the double-accent associates with both accented syllables, as shown in (5.12):

(5.12) **ALIGN(H\*, Lex), H\*→TBU >> NOSPREAD**

	<b>ALIGN(H*, Lex)</b>	<b>H*→TBU</b>	<b>NOSPREAD</b>	<b>NOCONTOURLEX</b>	<b>T→TBU</b>	<b>MAX-IO(L)</b>	<b>NOCONTOUR</b>
$\acute{o}$ $\acute{o}$ $H^*L$ $L_v$							
a. 			*			*	*
b. 		*!			*	*	
c. 	*!						*

The constraints **NOCONTOURLEX** and **T→TBU** crucially dominate **MAXIO(L)**, resulting in the deletion of the trailing L when there is no post-accentual syllable, as seen in (5.13):

(5.13) **NOCONTOURLEX, T→TBU >> MAXIO(L)**

	<b>ALIGN(H*, Lex)</b>	<b>H*→TBU</b>	<b>NOSPREAD</b>	<b>NOCONTOURLEX</b>	<b>T→TBU</b>	<b>MAX-IO(L)</b>	<b>NOCONTOUR</b>
$\acute{o}$ $H^*L$ $H_v$							
a. 						*	**
b. 					*!		**
c. 				*!			***

**MAXIO(L)** dominating **NOCONTOUR** means that the trailing L is realized on the post-accentual syllable even if that syllable is utterance-final (and therefore has a boundary tone associated with it). This is shown in (5.14):

(5.14) **MAXIO(L) >> NOCONTOUR**

	<b>ALIGN(H*, Lex)</b>	<b>H*→TBU</b>	<b>NOSPREAD</b>	<b>NOCONTOURLEX</b>	<b>T→TBU</b>	<b>MAX-IO(L)</b>	<b>NOCONTOUR</b>
$\acute{\sigma}$ $\sigma$ $H^*L$ $H_v$							
a.							**
b.						*!	*
c.					*!		
d.				*!			**

To summarize, in addition to the highly ranked constraints mentioned in (5.11), the crucial rankings mentioned thus far give us the overall hierarchy given in (5.15):

(5.15) Constraint hierarchy for NKK melody

**L-**, **MAXIO(H\*)**, **MAXIO(T<sub>v</sub>)**, **ALIGN(L-, Left)**, **ALIGN(T<sub>v</sub>, Right)**, **ALIGN(H\*, Lex)**, and **CONCATENATE** not crucially dominated;

**H\*→TBU >> NOSPREAD**

**NOCONTOURLEX**, **T→TBU >> MAXIO(L) >> NOCONTOUR**

Ranked thus, these constraints would yield the same outputs as in the “output” column of Table 5.1.

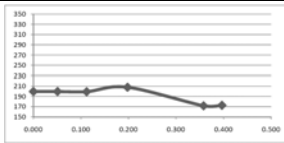
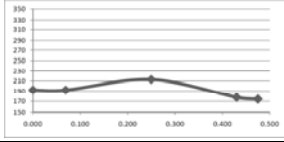
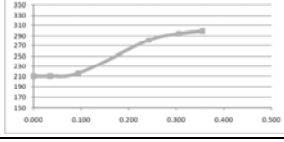
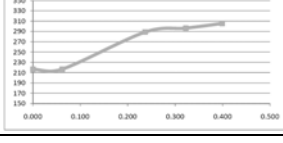
Superficially, both the rule-based and the constraint-based analyses presented above seem reasonable. At least based on the sequences of L and H in the output forms, the resulting contours seem to go up when they’re supposed to and down when they’re supposed to. However, a closer inspection reveals that these surface representations may not provide the most satisfying explanation of the facts. If the phonetic implementation of a syllable linked to a H is the same across-the-board, we do not expect the rise of the double-accent to start any later with respect to the initial syllable than it is for the initial-accent, and we do not expect the fall of the double-accent to be any earlier in the syllable than in the final-accent (consider, as a point of reference, how doubly linked H tones are generally expected to be realized in languages that display H tone spreading from an initial syllable to a second syllable), but we have seen clear evidence in the disyllabic contours that these expectations are not met in either case. In other words, the phonetic implementation is apparently sensitive to the tonal category. Furthermore, the derivations imply that, in monosyllables, there is a phonological neutralization between the initial-accent and double-accent categories at the level of the output, both in declarative and echo question contexts. The results of the perceptual test reported on in Chapter 3 suggest that they are indeed neutralized from a perceptual point of view. However, as shown in Chapter 2, the production results are not as definitive—there may still be a subtle contrast preserved in the monosyllables, albeit one that is not perceptually significant. Indeed, the native speaker who produced them had the intuition that she was preserving the contrast even in the monosyllabic context. So, it may not actually be correct to analyze the neutralization as a phonological one; rather, it is possible that the phonetic implementations of two distinct phonological representations are different, but similar enough that they are *perceived* as neutralized.

One seemingly plausible explanation for the fact that the phonetic implementation seems to be sensitive to the lexical tonal category is that the association of the double-accent happens at a different stage from that of the initial accent and the final-accent, respectively. Indeed, this is



the analysis proposed by Lee (2008), who posited that the double-accent gets associated with the first two syllables of the word *postlexically*, and that the phonetics realizes lexical association lines differently from postlexical ones. Such an analysis would enable us to maintain the traditional, generalized representation of the underlying lexical accent as H\*L for all tonal categories and to stay within the bounds of a traditional autosegmental framework while accounting for the unique phonetic implementation of the double-accented category. It would also provide a nice explanation for the fact that the double-accent only ever manifests itself on the first two syllables of a word, even among words with more than two syllables. The revised derivations for the double-accented conditions are shown in Table 5.2. Since this analysis

**Table 5.2: Revised derivations for double-accented melodies in NKK**

name	lexical	postlexical	output	surface
monosyllabic double declarative	$\sigma$ H*L L%	$\sigma$ L% H* $\emptyset$ L%	$\sigma$ L%H*L%	
disyllabic double declarative	$\sigma \sigma$ H*L L%	$\sigma \sigma$ L% H* $\emptyset$ L%	$\sigma \sigma$ L% H* L%	
monosyllabic double echo question	$\sigma$ H*L H%	$\sigma$ L% H* $\emptyset$ H%	$\sigma$ L%H*H%	
disyllabic double echo question	$\sigma \sigma$ H*L H%	$\sigma \sigma$ L% H* $\emptyset$ H%	$\sigma \sigma$ L% H* H%	

depends on a lexical/postlexical distinction when it comes to tone-to-TBU association, it cannot be translated into a flat (i.e. monostratal) constraint-based analysis. This issue aside, there is a potential problem with this revised derivational analysis. If there are no underlyingly accented syllables in double-accented words, the association of the post-accentual L tone cannot be with the syllable “following the rightmost accented syllable”. If it were left unassociated going into the postlexical stage, the post-accentual L would get deleted across-the-board for double-

accented words, even those with more than two syllables. However, Lee (2008) reported that a post-H pitch fall is indeed observed in double-accented words with more than two syllables (which is why she analyzed them as accented as opposed to unaccented). In order to express the L-deletion as a postlexical rule that only applies when appropriate, it would have to be rather precisely formulated, formulated as something like, “Associate the post-accentual L with the third syllable in the word if there is one; otherwise delete it.” At any rate, given such an analysis, and given the alignment behavior of the surface contour of double-accented words, we are forced into one of three conclusions, none of which is otherwise immediately obvious; these are given in (5.16):

(5.16) Conclusions regarding NKK tonal representation

- I. Double associations are interpreted differently by the phonetics from single associations, *or...*
- II. Postlexical associations are interpreted differently by the phonetics from lexical associations, *or...*
- III. The lexical tones of NKK are not representable purely in terms of H\*L accents and association lines

Conclusion I would render the double-linking representation an arbitrary, category specific diacritic that gets treated differently by the phonetics, and it would force us to represent the tonal neutralization of monosyllables as a phonological neutralization, which may not be correct. Conclusion II, while explaining the exceptional realization of the double-accented tune on disyllables and monosyllables, would preclude the notion that the phonetics is blind to whether a tonal association is lexical or postlexical. Conclusion III would make NKK look more like languages with a complex lexical tonal inventory such as Henanhua and less like typical word-tone languages such as Japanese.

### 5.3.2 AM analysis for Kansai Japanese

Pierrehumbert and Beckman (1988) devoted a large part of a chapter (Chapter 8) to providing an AM analysis for Osaka Japanese based on data from Kori (1987). Kori (1987) himself had already laid out, in an appendix, a very brief sketch of the melodic system in an AM-style framework. Let us first examine his proposal.

Kori (1987) argued that six tones are necessary to account for all the possible lexical contrasts in Osaka Japanese. He posited the inventory given in (5.17):

(5.17) Kori's (1987) Osaka Japanese tonal inventory

Boundary tones:	High boundary tone (H%)
	Low boundary tone (L%)
Phrase tones:	High phrase tone (pH)
	Low phrase tone (pL)
Nuclear tones:	Nuclear high tone (nH)
	Nuclear low tone (nL)

He also specified the intrinsic-pitch-height hierarchies given in (5.18):

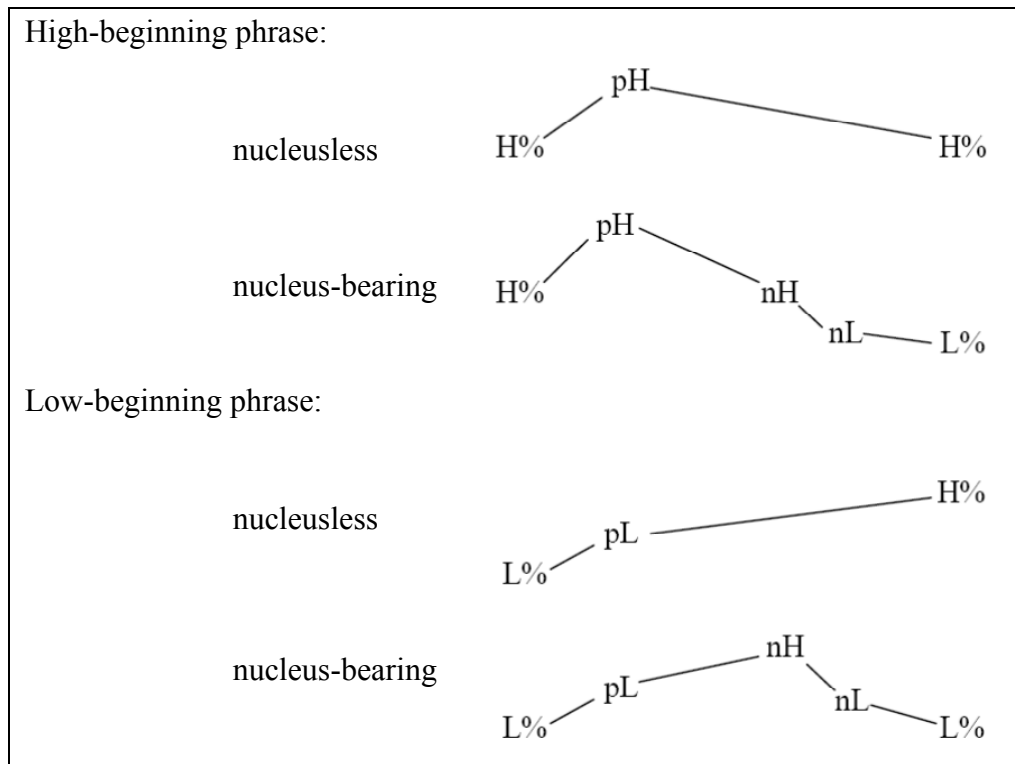
(5.18) Hierarchies of intrinsic pitch heights proposed by Kori (1987)

$H\% < pH > nH > nL > L\%$

$L\% < pL < nH$

(Kori 1987 did not explicitly mention it, but presumably H% should also be specified as higher than L%). He then sketched the four possible melodies on a phrase in Osaka Japanese using the above melodic inventory. This inventory is shown in Figure 5.2. Kori used the term “nucleus” in lieu of the word “accent”. The relative heights of the tonal targets in the schematization reflect the intrinsic pitch heights of those targets relative to one another. Note that certain combinations of tonal targets are predictable. A nL only ever follows nH, an accented phrase (i.e.

one that contains a nH) always ends in L%, an unaccented phrase (i.e. one that lacks a nH) always ends in H%, pH is always preceded by an initial H%, and a pL is always preceded by an initial L%. This analysis (henceforth the *Kori analysis*), while containing some redundancies, has the advantage of capturing the observed facts.



**Figure 5.2: Kori's schematization of the four phrase melodies in Osaka Japanese.**

Pierrehumbert and Beckman (1988) gave a slightly modified version of Kori's (1987) analysis. They dispensed with initial boundary tones and only specified whether a phrase<sup>52</sup> starts with a phrasal L or a phrasal H (the equivalent of Kori's pL and pH, respectively). Also they combined Kori's (1987) nH and nL into one bitonal target HL that associates directly with an accented mora. Finally, they dispensed with the notion of a final L%, positing that accented

<sup>52</sup> Actually, they refuted the notion that there is a prosodic constituent larger than the word but smaller than the intermediate phrase, and therefore they described everything in terms of the word (although they still called the word-initial tones 'phrasal'). Kori (1987) acknowledged that there is not an exact equivalent in Osaka Japanese to the well-established 'accentual phrase' in Tokyo Japanese, but he made a case for a constituent he called the 'OJ-phrase'. This theoretical distinction is not relevant for the discussion here.

phrases are unspecified following the HL tone and that only unaccented phrases end in a final H%. To summarize, the four possible melodic sequences in their analysis are as shown in (5.19):

(5.19) Pierrehumbert and Beckman's (1988) analysis of Osaka melodies

- a. pH... HL...
- b. pL... HL...
- c. pH...       ...H%
- d. pL...       ...H%

Note that this analysis, while more parsimonious than that of Kori (1987), still contains an element of predictability—unaccented phrases always end in H% and accented ones end unspecified. Also, the short rise observed at the beginning of all phrases, regardless of whether they are high-beginning or low-beginning, would have to be accounted for in the phonetic implementation. This exclusion seems strange in the context of their analysis for Tokyo Japanese, which involved the postlexical insertion of a L% at the beginning of an accentual phrase to account for a similar rising contour observed on the initial mora of every phrase. Likewise, the exclusion of a final boundary tone for accented phrases seems strange given how they went through the trouble of invoking a process of catathesis to explain why accented phrases in Tokyo Japanese end much lower than unaccented ones, despite both being specified for a final L% tone. A slightly modified version of their analysis, one that better captures the facts while remaining true to the spirit of parsimony in the melodic inventory, is shown in (5.20):

(5.20) A modified version of Pierrehumbert and Beckman's (1988) analysis for Osaka Japanese

(henceforth the *P&B' analysis*):

- a. L% pH... HL... H%
- b. L% pL... HL... H%
- c. L% pH...       ...H%
- d. L% pL...       ...H%

To capture the phonetic facts, they would have to posit that L% before pH is realized at a higher F<sub>0</sub> than L% before pL, and that H% undergoes a drastic lowering after HL, due to catathesis. This would account for the fact that H% is realized lower than the trailing L.

Neither the Kori analysis nor the P&B' analysis takes echo questions into account, although Pierrehumbert and Beckman (1988) mentioned at the end of their discussion that they would most likely render echo questions with an utterance-final H%. Of course, to explain the facts presented in 2.6.5, namely that all echo questions end in a sharp local F<sub>0</sub> rise, utterance-level H% would have to have an intrinsic pitch height that is higher than that of phrase-final H%. The expanded version of P&B', dubbed P&B'', is shown in (5.21).

(5.21) An expanded version of P&B' (henceforth P&B''):

- a. pL% pH... HL... pH% (high-beginning, accented, declarative)
- b. pL% pL... HL... pH% (low-beginning, accented, declarative)
- c. pL% pH... ...pH% (high-beginning, unaccented, declarative)
- d. pL% pL... ...pH% (low-beginning, unaccented, declarative)
- e. pL% pH... HL... pH% uH% (high-beginning, accented, echo question)
- f. pL% pL... HL... pH% uH% (low-beginning, accented, echo question)
- g. pL% pH... ...pH% uH% (high-beginning, unaccented, echo question)
- h. pL% pL... ...pH% uH% (low-beginning, unaccented, echo question)

Just to be explicit, an expanded version of the Kori analysis, one that takes echo questions into account, is shown in (5.22).

(5.22) An expanded version of the Kori analysis (henceforth Kori'):

- a. pH% pH... HL... pH% (high-beginning, accented, declarative)
- b. pL% pL... HL... pL% (low-beginning, accented, declarative)
- c. pH% pH... ...pH% (high-beginning, unaccented, declarative)
- d. pL% pL... ...pL% (low-beginning, unaccented, declarative)

- e. pH% pH... HL... pH% uH% (high-beginning, accented, echo question)
- f. pL% pL... HL... pL% uH% (low-beginning, accented, echo question)
- g. pH% pH... ...pH% uH% (high-beginning, unaccented, echo question)
- h. pL% pL... ...pL% uH% (low-beginning, unaccented, echo question)

Note that here there is no need to invoke a catathesis rule but we do need to place uH% above pH% in the intrinsic pitch height hierarchy.

The above representations of course correspond to the surface phonological representations for the various melodies in Osaka Japanese. They should also be sufficient to cover all the contrasts seen in the other Kansai dialects, so we are now ready to make an initial attempt at some derivations for Shiga Japanese. This is shown in Table 5.3, where the P&B'' surface representations are used (HL has been relabeled H\*L to make these representations compatible with those used in Section 5.3.1 for NKK). At the lexical level, the H\*L tone is associated with the initial mora of the accented syllable. Everything else is taken care of at the postlexical level. The postlexical rules are delineated in (5.23):

(5.23) Postlexical rules for P&B'':

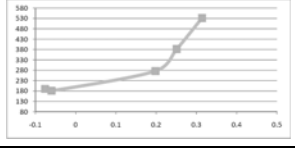
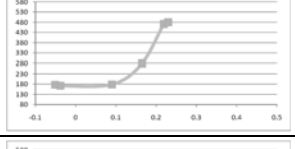
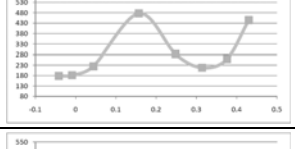
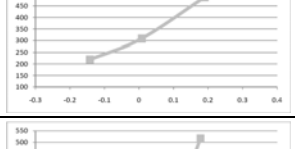
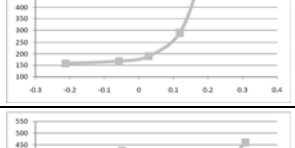

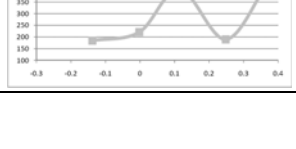
- a. pL, if present, is associated with the first mora.
- b. pH is inserted and associated with the first mora if there is no pL and the first mora is unassociated.
- c. A trailing L tone is associated with the mora following the H\*-associated mora if one is available. Otherwise it is associated with the same mora as the H\*.
- d. Initial pL% and final pH% are inserted and associated with initial and final morae, respectively.
- e. pH% is associated with the final mora if the H\* is not associated with the penultimate mora.
- f. uH%, if present, is associated with the final mora.

**Table 5.3: First attempt at derivations for Kansai melodies**

name	lexical	postlexical	output	surface
monosyllabic H-unaccented declarative	$\begin{array}{c} \sigma \\ \mu \mu \end{array}$	$\begin{array}{c} \sigma \\ \mu \mu \\ \text{pL}\% \text{pH} \quad \text{pH}\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \mu \\ \text{pL}\% \text{pH} \text{pH}\% \end{array}$	
monosyllabic L-unaccented declarative	$\begin{array}{c} \sigma \\ \mu \mu \\ \text{pL} \end{array}$	$\begin{array}{c} \sigma \\ \mu \mu \\ \text{pL}\% \text{pL} \quad \text{pH}\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \mu \\ \text{pL}\% \text{pL} \quad \text{pH}\% \end{array}$	
monosyllabic H-initial-accented declarative	$\begin{array}{c} \acute{\sigma} \\ \mu \mu \\ \text{H}^* \text{L} \end{array}$	$\begin{array}{c} \sigma \\ \mu \mu \\ \text{pL}\% \text{H}^* \text{L} \quad \text{pH}\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \mu \\ \text{pL}\% \text{H}^* \text{L} \quad \text{pH}\% \end{array}$	
disyllabic H-unaccented declarative	$\begin{array}{c} \sigma \sigma \\ \mu \mu \end{array}$	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL}\% \text{pH} \quad \text{pH}\% \end{array}$	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL}\% \quad \text{pH} \text{pH}\% \end{array}$	
disyllabic L-unaccented declarative	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL} \end{array}$	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL}\% \text{pL} \quad \text{pH}\% \end{array}$	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL}\% \quad \text{pL} \quad \text{pH}\% \end{array}$	
disyllabic H-initial-accented declarative	$\begin{array}{c} \acute{\sigma} \sigma \\ \mu \mu \\ \text{H}^* \text{L} \end{array}$	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL}\% \text{H}^* \text{L} \quad \text{pH}\% \end{array}$	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL}\% \quad \text{H}^* \text{L} \quad \text{pH}\% \end{array}$	
disyllabic L-final-accented declarative	$\begin{array}{c} \sigma \acute{\sigma} \\ \mu \mu \\ \text{pL} \quad \text{H}^* \text{L} \end{array}$	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL}\% \text{pL} \quad \text{H}^* \text{L} \quad \text{pH}\% \end{array}$	$\begin{array}{c} \sigma \sigma \\ \mu \mu \\ \text{pL}\% \quad \text{pL} \quad \text{H}^* \text{L} \quad \text{pH}\% \end{array}$	



**Table 5.3 (cont'd): First attempt at derivations for Kansai melodies.**

monosyllabic H-unaccented echo question	$\begin{array}{c} \sigma \\ \mu \ \mu \\ uH\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \ \mu \\ pL\%pH \quad pH\%uH\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \ \mu \\ pL\%pH \quad pH\%uH\% \end{array}$	
monosyllabic L-unaccented echo question	$\begin{array}{c} \sigma \\ \mu \ \mu \\ pL \quad uH\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \ \mu \\ pL\%pL \quad pH\%uH\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \ \mu \\ pL\%pL \quad pH\%uH\% \end{array}$	
monosyllabic H-initial-accented echo question	$\begin{array}{c} \acute{\sigma} \\ \mu \ \mu \\ H^*L \ uH\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \ \mu \\ pL\% \ H^*L \ pH\%uH\% \end{array}$	$\begin{array}{c} \sigma \\ \mu \ \mu \\ pL\% \ H^*L \ pH\%uH\% \end{array}$	
disyllabic H-unaccented echo question	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ uH\% \end{array}$	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL\%pH \quad pH\%uH\% \end{array}$	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL\% \quad pHpH\%uH\% \end{array}$	
disyllabic L-unaccented echo question	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL \quad uH\% \end{array}$	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL\%pL \quad pH\%uH\% \end{array}$	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL\% \quad pL \quad pH\%uH\% \end{array}$	
disyllabic H-initial-accented echo question	$\begin{array}{c} \acute{\sigma} \ \sigma \\ \mu \ \mu \\ H^*L \ uH\% \end{array}$	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL\% \ H^*L \ pH\%uH\% \end{array}$	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL\% \quad H^*L \ pH\%uH\% \end{array}$	
disyllabic L-final-accented echo question	$\begin{array}{c} \sigma \ \acute{\sigma} \\ \mu \ \mu \\ pL \quad H^*L \ uH\% \end{array}$	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL\%pL \quad H^*L \ pH\%uH\% \end{array}$	$\begin{array}{c} \sigma \ \sigma \\ \mu \ \mu \\ pL\% \quad pL \quad H^*L \ pH\%uH\% \end{array}$	

As was the case with NKK, these derivations seem reasonable (albeit cumbersome) given the broad strokes of how the corresponding  $F_0$  contours look. Of course, as mentioned above, certain implementation “rules” are still necessary to precisely predict those contours from the surface phonological representations (the catathesis that lowers  $pH\%$  drastically after  $H^*L$ , or the fact that  $pL\%$  gets raised slightly before  $uH\%$ , e.g.). However, there is actually a larger problem with both the P&B’ and the Kori’ surface representations. Note how, in accounting for the facts, we were forced into differentiating between  $pL\%$  and  $pL$ . This means that there are tones with three different intrinsic pitch heights associated to the phrase— $pL$ ,  $pH$ , and  $pL\%$ . In the original Kori (1987) analysis, this state of affairs is obscured by the fact that the first two are called

“phrasal tones” and the last one is a “boundary tone”, but it is a *phrase* boundary tone! Likewise, Pierrehumbert and Beckman (1988) posited for Tokyo Japanese accentual phrases an initial L% as well as a “phrasal H”. Although this deviates from the convention established in Pierrehumbert (1980) that boundary tones are only associated with a specific level in the prosodic hierarchy (the intonational phrase in the case of English), it does not seem too problematic because they can still be represented simply as a L and a H, respectively, that are associated postlexically to the accentual phrase node of the prosodic tree. Gussenhoven (2004) recognized their “kinship” in this regard and renamed them  $L_\alpha$  and  $H_\alpha$ , respectively in his OT analysis. However, these three tones in the Kansai system—pL, pH, and pL% would pose a problem for such a convention because pL and pL% would get conflated into something like  $L_\phi$  (or  $L_\omega$  for Pierrehumbert and Beckman), even though they clearly have different intrinsic relative pitch heights.

Pierrehumbert and Beckman (1988) mentioned the fact that a crucial difference between phrasal H in Tokyo Japanese on the one hand and phrasal H and L in Osaka Japanese on the other hand is that the latter are lexically specified while the former is inserted postlexically. We might extend this notion to the pL% in Kansai, i.e. the crucial difference between pL and pL% is that pL is lexically specified while pL% is inserted postlexically, and the phonetic implementation is sensitive to this difference. This attempt at “rescuing” the AM analysis is analogous to the proposal for NKK by Lee (2008).

In the end, then, it seems that one of two conclusions must be drawn; these are presented in (5.24):

(5.24) Conclusions regarding Kansai tonal representation

- I. Phrasal tones can either be lexically specified or postlexically inserted, and the phonetic implementation is sensitive to this difference *or...*
- II. We need to posit more than two relative pitch levels for tones in the phrase (or word) domain in Kansai.

Note that Gussenhoven (2004) would have had no choice but to submit to Conclusion II above unless he had adopted a Stratal OT model. Pierrehumbert and Beckman (1988) would probably not have wanted to submit to such a conclusion, though, because it deviates from a fundamental assumption of their model (and all conventional AM models)—that all melodies can be expressed in terms of a two-way contrast at every level of the prosodic hierarchy. Conclusion I, however, also requires a departure from a basic assumption built into the standard AM theory—that the phonetics is blind to the source of a melodic unit (i.e. whether it is lexical or postlexical).

#### **5.4 Conclusion**

In this chapter, the notion of intonation-dependent allotony was introduced, and it was suggested that AM approaches to speech melody are appropriate for handling the type of allotony that would need to be invoked to explain the types of alternations observed in Kyoto Japanese and NKK. With this potential advantage of the AM approach in mind, AM analyses of the NKK and Kansai Japanese melodic systems were critically analyzed in light of the phonetic facts we have seen in earlier chapters for those respective languages. It was shown that, in order to maintain the types of representations assumed in the AM analyses (complexes of Hs and Ls associated with syllables or morae), fundamental assumptions about how the phonology and phonetics interact (the phonetics treats any given TBU that is associated with a certain kind of tonal unit the same) must be violated; alternatively, in order to maintain those conventional assumptions, the traditional AM representations must be abandoned. While there is no hard-and-fast evidence that one tack is favorable over the other, it is worth noting the sheer number of postlexical rules that are needed to explain the facts within the AM framework, especially for Kansai Japanese. Indeed, Kori's (1987) sketch of the different combinations of accentedness and starting register in Osaka comes across as a thinly veiled attempt to capture four different classes of "tunes" in a decompositional way. It begs the question of whether it might be easier just to put these tunes directly in the lexicon. We have seen, though, that the decompositional analysis lends itself to

accounting for certain alternations and typological patterns (the special status of disyllabic initial-accented echo questions in NKK; the lack of a post-accentual fall in monosyllabic echo questions in Kyoto Japanese), so rather than throwing out the baby with the bathwater it is worth pursuing a model of tone that incorporates this structure while being powerful enough to accommodate the various dimensions along which categorical contrasts are made in different languages. Such a model is sketched out in the next chapter.

## CHAPTER 6: TOWARD A UNIFIED SCOPAL MODEL OF SPEECH MELODY

### 6.1 Introduction

It is clear from the results and discussion presented in Chapters 2 through 4 that, although phonetics-only overlay models of speech melody are well-suited for capturing the apparently parallel nature of melodic implementation at the right edge of utterances in some languages (e.g., Mandarin and Henanhua) and perhaps also for modeling the global, utterance-wide upward shift in the pitch range that occurs in echo question intonation in many languages, they cannot account for many of the facts—in particular the tone-dependent intonational phenomena that are apparent in the languages under investigation. Meanwhile, sequential models couched in the widely-used AM framework are able to handle some such phenomena by virtue of the fact that they include an active phonological component where tone and intonation may interact, but these models are not without their drawbacks. For one thing, they may be better suited for languages in which lexical tone and utterance-type intonation are realized serially than for languages in which the implementation appears to be parallel. Secondly, they do not offer a representational framework for lexical tone that allows for fruitful comparison across different types of tone languages (e.g. comparisons between word-tone languages and syllable-tone languages). Third, as seen in Chapter 5, the traditional assumptions of the model are too restrictive even for word-tone languages like NKK and Kansai Japanese.

In this chapter, a cross-linguistic melodic model is sketched out, with certain “design choices” (vis-à-vis what is taken care of in the phonology vs. what is left for the phonetics) made based on what seems reasonable given the amount of theoretical baggage each one entails as well as on what allows for the most fruitful cross-linguistic comparisons and typological characterizations. The representational framework proposed for lexical tones in Section 6.2 is a more flexible version of the autosegmental feature geometry framework used extensively in the

literature of the 1980s and 1990s for Chinese tonal systems; by allowing the sub-parts of the geometry to be “unpacked” according to secondary association conventions, the framework provides the means for a unified representational template to be used for lexical tone in languages like Mandarin as well as in languages like Japanese. The melodic framework proposed in Section 6.3 for handling the phonological interactions between tone and intonation is an extension of the AM framework established in Pierrehumbert and Beckman (1988), with association lines used to associate various melodic units with various constituents in the prosodic hierarchy. While more recent AM-based analyses in the literature continue to employ association lines between melodic units and terminal nodes in the prosodic tree (either morae or syllables), they tend to eschew association lines that connect melodic units to higher level prosodic nodes in favor of subscripts or indices on the tonal units (e.g. Gussenhoven 2004; Zsiga and Zec to appear). While the two mechanisms are basically equivalent—they both allow for a structural distinction between, say, a “high” tonal unit that is associated with a syllable and one that is associated with a phrase (or a phrase boundary)—the use of higher-level association lines lends itself to a seamless integration of the tonal geometry proposed in Section 6.2 into the prosodic tree, which in turn provides us with a way to represent both lexical and intonational tunes with the same geometrical template. This integration of lexical and non-lexical melodic units is mapped out in Section 6.3.

The other power afforded to us by the use of association lines—and in particular the option to associate a melodic unit with multiple levels in the tree—is the power to encode phonologically the *scope* of a given tune or melodic unit and whether it is interpreted by the phonetics in parallel with nearby melodic units or in series with them. This mechanism, introduced in Section 6.3, enables the model to capture the typological difference between a Mandarin-type melodic system and a SJ-type melodic system. Given all of this novel power, a stipulation is suggested in Section 6.4 to constrain precisely what aspects of the phonological structure the phonetics can access. The chapter is summarized and concluded in Section 6.5.

## 6.2 Autosegmental *Tune* Geometry

Ever since Leben (1973), Goldsmith (1976), Williams (1976), and Haraguchi (1977), the advantages of using autosegmental tonal representations for explaining observable patterns in tone languages as well as typological generalizations have been illustrated over and over again. However, in Chapter 5 we saw the ways in which the more restrictive AM instantiation of such a representation might be inadequate for the specific languages of NKK and Kansai Japanese, respectively. In this section, an attempt is made to build an autosegmental template that addresses these shortcomings and indeed handles the range of languages we have examined while maintaining the fundamental advantages of the AM approach.

A reasonable initial strategy is to try to find an autosegmental geometry that is able to handle the languages with larger tonal inventories and higher tonal densities, the assumption being that such a geometry should be powerful enough to handle other languages with simpler inventories. With this in mind, geometries proposed for Chinese tonal systems could be a good place to start. The literature on Chinese tonal systems has shown that contour tones in those systems behave at times like single units and at other times like complexes of constituent features (Clements 1985; Yip 1989; Bao 1990; Duanmu 1990). Chen (2000) compared and contrasted various autosegmental models and showed that only Bao's (1990) representation was able to predict the complete typology of the attested patterns involving the apparent spreading and shifting of various constituent features (contour only, register only, terminal node only, and whole contour tone)<sup>53</sup>. Accordingly, a version of Bao's (1990) representation is adopted for the current model; the template for the representation is shown in Figure 6.1.

---

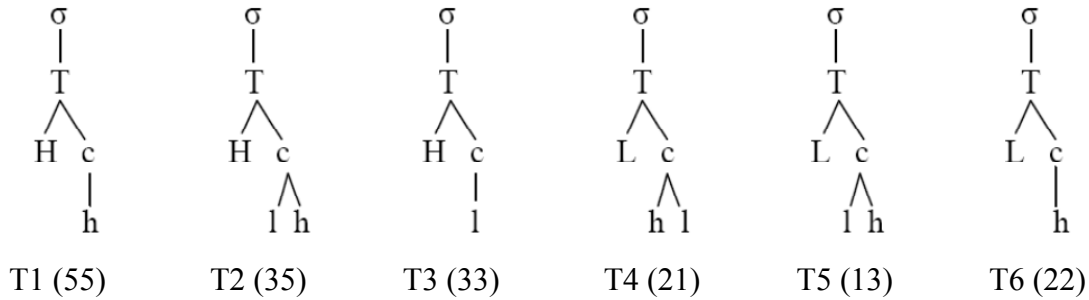
<sup>53</sup> Bao (1990) cited data from Zhenjiang and Wenzhou as examples of contour spreading and data from Wuyi and Pingyao as examples of register spreading in order to justify a model that allowed for both types of spreading. Yip (1995) argued that the level of power offered by Bao's (1990) model was not warranted based on the data that Bao (1990) himself provided. However, Chen (2000) brought new data to bear on the issue, namely those from Zhenhai for contour spreading and Chaozhou for register spreading (the latter from Bao (1996)). It was based on these new findings that Chen (2000) concluded that the power offered by Bao's (1990) was indeed necessary, and Yip (2002) agreed.



**Figure 6.1: Template for a lexical tune based on Bao's (1990) autosegmental representation of lexical tone. T = *tune*, R = *register*, c = *contour*, and t = *contour endpoints***

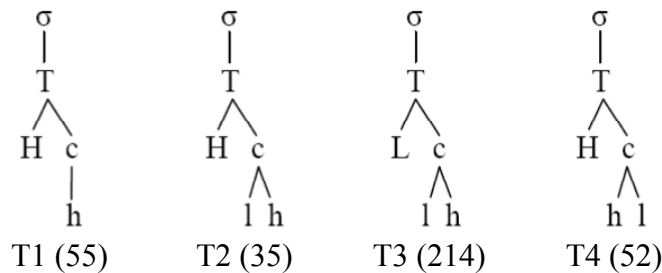
In Figure 6.1, the *R* node is for *register* (possible values are H and L), *c* is for *contour*, and the terminal *t* nodes are for *contour endpoints* (possible values are h and l). The branchingness of the *c* node is optional. One terminological departure from the tradition in the literature is the label of *tune* instead of *tone* for the whole constituent that is associated with the TBU (which now becomes the *tune-bearing unit*). This terminology is not likely to catch on within Chinese tonal phonology circles, since the term *tone* for referring to this type of melodic unit is so entrenched, but *tune* seems a more appropriate label for a melodic constituent that is decomposable, and it frees up the term *tone* so that the latter can be used to refer to the subparts of this melodic constituent when necessary. In addition, it will become apparent that using this terminology allows us to formulate a more cohesive model when it comes to cross-linguistic comparisons as well as handling melodic interactions between lexical and non-lexical melodic units. To minimize terminological shock, the term *tone* will still be used as a mass noun when referring to the melodic phenomenon of lexical melodic specification as opposed to the (countable) lexical melodic constituent. So, what will be referred to from now on as the six Cantonese *lexical tunes* would be represented as shown in Figure 6.2. Note that, for Bao (1990), the labels on terminal nodes are actually shorthand for feature values ([ $\alpha$ stiff] and [ $\alpha$ slack]), but it is the structural properties of his model that are being adopted here, not the featural ones. The general assumption that the features on these nodes are abstract phonological features is retained, however, so they need not capture the precise differences in relative pitch height across tonal categories. As it was for Bao (1990), the goal here is to have enough distinctive features to define a tonal inventory and to account for spreading and shifting phenomena that suggest certain





**Figure 6.2: Bao (1990)-style representations for the six Cantonese lexical tones. The corresponding tune labels and Chao tone-letter-numerals are shown underneath.**

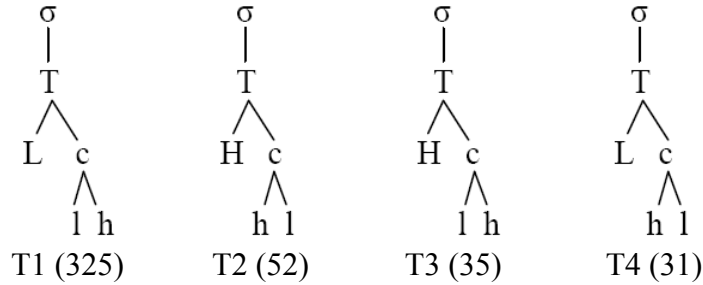
levels of constituency. Thus, branching in the geometry is limited to being binary; convex and concave tunes are represented underlyingly with only two contour points, and the phonetic implementation is responsible for realizing the tunes as convex or concave. The four lexical tones in Mandarin can be represented as shown in Figure 6.3. Note that T3, which is a dipping



**Figure 6.3: Representations for the four lexical tones of Mandarin.**

tune in isolation, is represented in the geometry as a low rising tone. The phonetic implementation is responsible for realizing it as a dipping tone in the appropriate context. When the T3 syllable is not utterance-final and not followed by a T3 syllable, a phonological rule or constraint would cause the **h** terminal node in T3 to be delinked from the **c** node, resulting in an *allotune* without a rising tail. When it is followed by a T3 syllable, the value on the register node is changed from L to H, making it equivalent to a T2 syllable<sup>54</sup>. The four lexical tones of Henanhua can be represented as shown in Figure 6.4. Once again, the dipping tune, T1, is

<sup>54</sup> Myers and Tsay (2003) showed compelling evidence that Mandarin speakers from Beijing do not *completely* neutralize T3 and T2 before T3, whereas speakers from other areas of China, such as Taiwan, do so “categorically”. If this distinction is correct, the above account would apply to the second category of speakers, since it results in a



**Figure 6.4: Representations for the four lexical tones of Henanhua**<sup>55</sup>.

represented as a low rising tune. Note also that the contrast between T2 and T4 is captured as one of register, even though in many cases the surface distinction is one of alignment (the fall of T2 is aligned earlier than that of T4). This alignment distinction would be taken care of by the phonetic implementation<sup>56</sup>.

The proposed application of the above geometry to languages like Japanese and NKK—languages traditionally treated as having lexical “pitch accents”—deviates from its original intended use in the Chinese literature, but the proposed deviations involve mechanisms that have all been employed elsewhere in the autosegmental literature. First, since tone is lexically distinctive at the level of the word and not the syllable in Japanese, it makes sense to have the word be the TBU. In other words, the relevant typological distinction is that Mandarin and Cantonese are *syllable-tune* languages and Japanese and NKK are *word-tune* languages. From a featural perspective, this concept has precedents in Rowlands (1959) and Edmonson and Bendor-Samuel (1966), who treated tone as a feature on phonological words, and in Welmers (1962), McCawley (1970), who treated it as a feature on morphemes. Leben (1973) cited the above examples and settled on the morpheme as the tonal-feature-bearing unit, positing phonological processes that treated tonal features as independent suprasegmental units in and of themselves. As for the structural concept of melodic units being associated with nodes higher in the prosodic

---

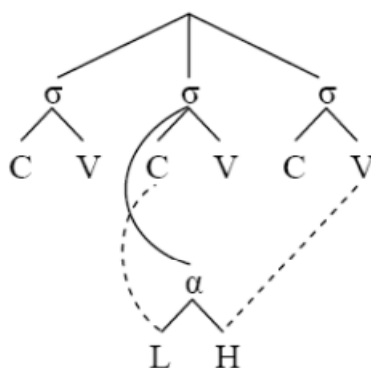
representational neutralization. For Beijing speakers, the two tones would retain their structural differences and the *near*-neutralization before T3 would be accounted for at a more superficial level, either postlexically or phonetically.

<sup>55</sup> The tone-letter numerals shown here are approximations based on the results from the individual speaker in the current study, so they do not match those given by Zhang, Chen et al. (1993).

<sup>56</sup> If a representation that captures horizontal alignment differences with a mechanism separate from that for vertical register differences is desired, perhaps the tones could anchor themselves to different subparts of the syllable, like morae or segments, as will be proposed below for other languages like SJ and NKK.

tree than the mora or the syllable nodes, such high-level associations are well-established for postlexical units like “phrase accents” and “boundary tones” in AM models; Pierrehumbert (1980) and Poser (1984) talked about such units “aligning with” or “associating with” the boundaries of higher level prosodic constituents, while Pierrehumbert and Beckman (1988) explicitly drew association lines connecting such postlexical melodic units with nodes at various levels in the prosodic tree. The current proposal extends this mechanism to allow lexical melodic units to be associated with nodes above the syllable node as well.

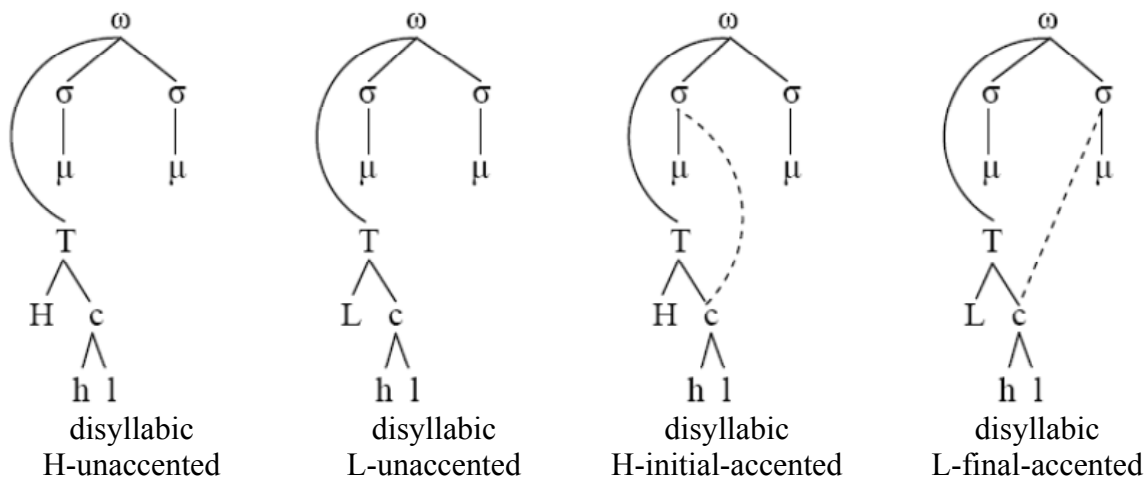
The second deviation from the original conception of the tonal geometry is the use of secondary association lines to “anchor” certain sub-constituents of the melodic unit to certain sub-constituents of the TBU. This anchoring can be specified underlyingly or invoked according to rules or constraints, and how the phonetics interprets the tonal geometry depends on whether or not there are any secondary associations. If there are, then the tonal features get decoupled and realized in series and aligned with segmental material according to the secondary association lines. The concept of phonologically anchoring components of decomposable tunes to sub-syllabic units such as segments has been explored in the past for postlexical tunes by Ladd and others in languages like Greek and Dutch (Arvaniti, Ladd et al. 1998; Ladd, Mennen et al. 2000). Ladd (2008) schematized this concept with the phonological representation shown in Figure 6.5.



**Figure 6.5: A phonological representation of the segmental anchoring of tones utilizing secondary association lines (adapted from Ladd 2008 Figure 5.2).**

However, Xu and Sun (2002) and Prieto and Torreira (2007) argued—and Ladd (2008) agreed—that such segmental anchoring accounts of language-specific melodic alignment phenomena are not warranted. The arguments against this phonological account were based in part on results from various studies that showed that subtle alignment differences within a language can arise that do not lend themselves to such representations, and also on facts about dialect-dependent alignment differences that seem best analyzed as dialect-specific phonetic implementations of equivalent phonological representations. These arguments are valid, and as such the idea of phonologically anchoring melodic units to segments is not pursued here. However, the idea of sub-constituent-anchoring, as a *structural* concept, will be exploited.

With the adoption of the above mechanisms, contrastive tonal patterns on disyllabic words in SJ can be represented as shown in Figure 6.6. Whether a word is high-beginning or

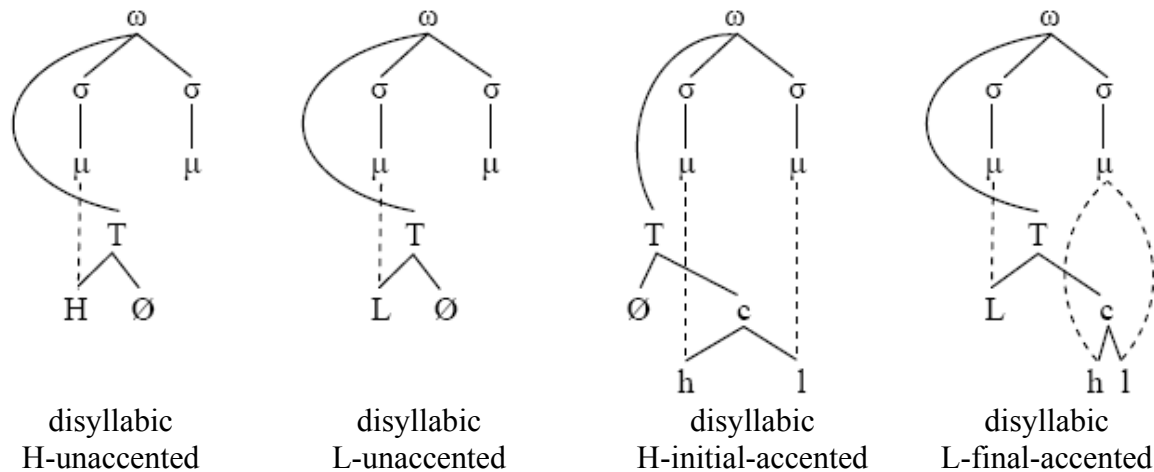


**Figure 6.6: Underlying forms for lexical tunes on disyllables in SJ<sup>57</sup>.**

low-beginning is treated as a register contrast, and what is traditionally referred to as the “accent” is captured on the contour node side of the tune, where the contour node dominates a **h-l** tone sequence. The tune-bearing unit is the word; since the “accent-bearing” unit is the syllable, the contour node is lexically associated with the accented syllable, if there is one. Note that the

<sup>57</sup> Here the assumption is that the syllable is the accent-bearing unit, as it is in Tokyo. Kori (1987) claimed that the accent-bearing unit (he called it the “nucleus-bearing unit”) was the mora in Osaka, but he acknowledged that this was rapidly changing under influence from Tokyo. None of the Kansai speakers recorded in the present study showed any evidence of accents falling on second halves of bimoraic syllables.

contour node here is equivalent to the  $\alpha$  node in Ladd's representation, but the contour node is dominated by another node that is associated with the entire word, so the association of the contour node with one of the syllables is actually a secondary association. Given these underlying forms, we can posit certain rules or constraints that would result in the proposed surface forms shown in Figure 6.7 (deleted constituents are indicated with the symbol  $\emptyset$ ). For



**Figure 6.7: Surface forms for lexical tunes on disyllables in SJ.**

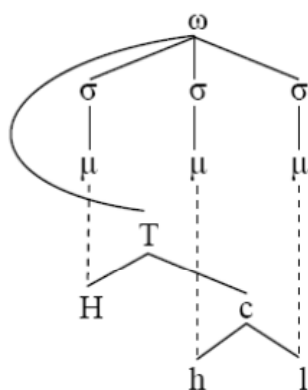
simplicity's sake, the phonological processes involved are expressed in (6.1) in terms of rules:

(6.1) Lexical tune association rules for SJ

1. If a **c** node is associated with a syllable, associate the terminal **h** node under it with the initial mora of that syllable and the terminal **l** node with the following mora, if there is one; otherwise associate the **l** node with the same mora as the **h** node.
2. If the first mora of the word is still unassociated, associate the register node with it.
3. Delete unassociated tonal constituents.
4. Remove any association lines connecting contour nodes to syllables<sup>58</sup>.

<sup>58</sup> Whether these “intermediate” secondary association lines are needed in the surface representation in addition to the ones on terminal **h** and **l** nodes may depend in part on the type of generative model at play; for example, with a non-derivational constraint-based approach it might be trickier to omit these intermediate association lines unless correspondence constraints of the flavor “for every associated contour node in the input there is one associated h

Regarding Rules 2 and 3, the **l** tone under the contour node does not get deleted when the contour node is associated with a phrase-final monomoraic syllable, because associations of multiple contour terminals (**ls** and **hs**) with a single mora are allowed in SJ; note that it would indeed get deleted in Osaka Japanese, and in Kyoto Japanese its fate would depend on the presence or absence of a tone from echo question intonation. Also, note that there is a neutralization of register that occurs when the contour association falls on the first syllable. This is a desirable result given that there are no register contrasts on initial-accented words in SJ. As mentioned in Section 2.6 of Chapter 2, there also happens to be a gap whereby there are no high-beginning final-accented words; there are longer words that are high-beginning and accented, though, and just for the sake of illustration the surface form of such a word is shown in Figure 6.8. Here, each terminal tonal node is associated with a different mora.



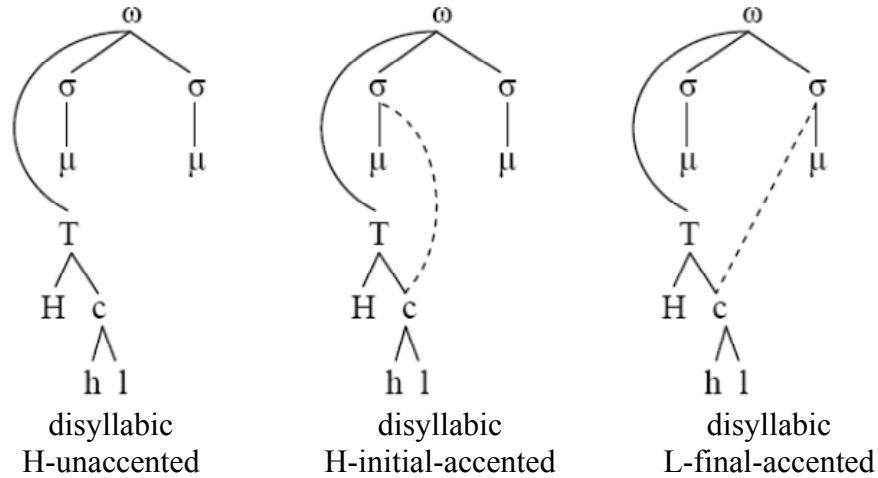
**Figure 6.8: Surface representation for a high-beginning, second-accented trisyllable in SJ.**

Before moving on to the representations for NKK, it is worth thinking about how Tokyo Japanese would be handled. The underlying representations would largely be the same as for SJ, except that there would only be H-registered tunes. The underlying representations for tunes on disyllables are shown in Figure 6.9 and the corresponding surface forms are shown in Figure 6.10. Since it is well-established that there is an “accentual phrase” in Tokyo Japanese between

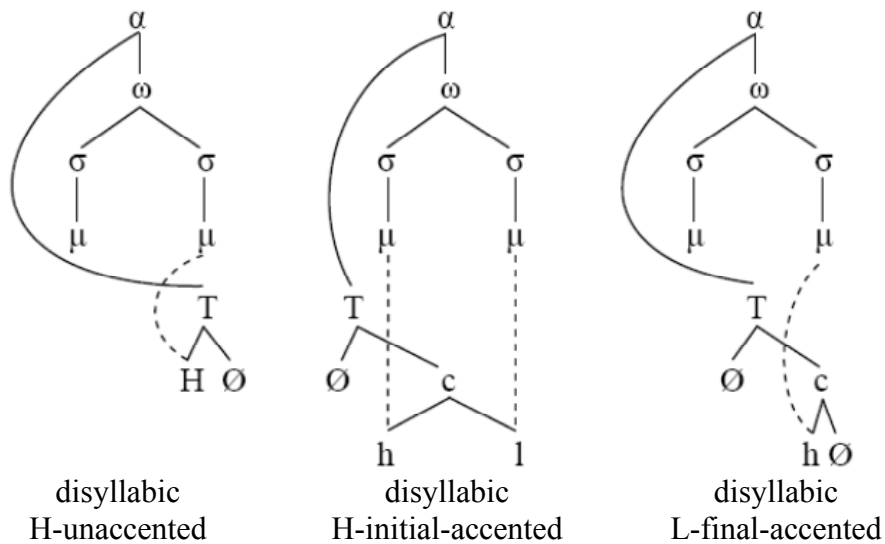
---

node in the output” and “associated contour nodes in the input have corresponding contour nodes in the output” are allowed. It could also be the case that the terminal **h** node is associated with a mora underlyingly.

the level of the word and the level of the utterance, in the surface representation the T node is assumed to be re-associated with the accentual node dominating the word node (if the accentual



**Figure 6.9: Underlying representations for lexical tunes on disyllables in Tokyo Japanese.**



**Figure 6.10: Surface representations for lexical tunes on disyllables in Tokyo Japanese.**

phrase contained two words, the tune of one of the words “wins out” over that of the other one according to certain lexical and morpho-syntactic rules). For our present purposes the tune

association rules for Tokyo Japanese, given in (6.2), assume that the “dominant” **T** node is marked as such in the lexicon:

(6.2) Lexical tune association rules for Tokyo Japanese

1. Re-associate the **T** node with the accentual phrase node (if there is more than one word in word in the accentual phrase, re-associate the dominant **T** node only).
2. If a **c** node is associated with a syllable, associate the terminal **h** node under it with the initial mora of that syllable and the terminal **I** node with the following mora, if there is one.
3. Associate the register tone with the second mora of the accentual phrase, if it is free.
4. Delete all unassociated tonal constituents.
5. Remove contour association lines.

Note that multiple contour terminals may not be associated with a single mora in Tokyo Japanese, and so **I** tones stranded at the right edge are deleted<sup>59</sup>. Also, the **H** tone on the register node—which would correspond to the **H** “phrase tone” of Pierrehumbert and Beckman (1988)—is secondarily associated with the *second* mora of the accentual phrase if that mora is free. It is possible that a postlexical tune containing a **L** tone is inserted and associated with the first mora of the accentual phrase in Tokyo Japanese—this would correspond to the initial **L** boundary tone of Pierrehumbert and Beckman (1988)—but such postlexical tunes are left out of Figure 6.10 for simplicity’s sake.

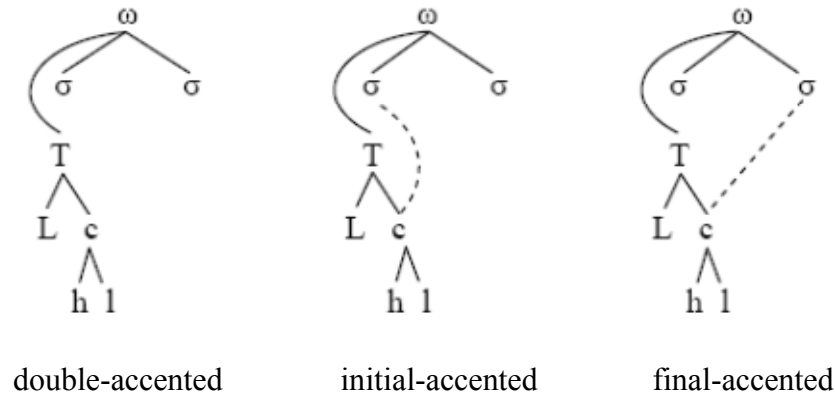
The underlying forms for NKK would be very similar to those for Tokyo Japanese, as suggested by previous analyses (e.g. Lee 2008). The underlying representations for all three tonal categories on disyllables are shown in Figure 6.11. Here, the register node bears a **L** tone, since all of the syllables leading up to the accented syllable in NKK are relatively low in pitch (like low-beginning words in Kansai). The corresponding surface forms of these three categories

---

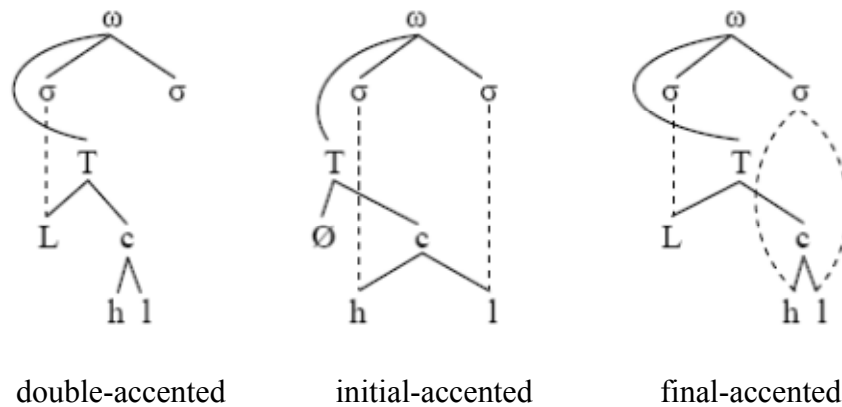
<sup>59</sup> Note that this terminal **h** tone deletion is the same mechanism that would be invoked for T3 sandhi in Mandarin; it is just the triggering environment that would be different in the two cases.



are shown in Figure 6.12. Crucially, unlike that of either of the Japanese dialects, the phonology of NKK would *not* delete the contour node when it is left unassociated, thus accounting for the



**Figure 6.11: Underlying representations for lexical tones on double-, initial-, and final-accented disyllables in NKK.**



**Figure 6.12: Surface representations for lexical tones on disyllables in NKK.**

difference in behavior between the Japanese unaccented word and the NKK “double-accented” word (recall that Lee 2008 rejected the analysis of these words as unaccented since there is still a fall in pitch realized on them). The fact that the contour node is left intact but unassociated captures the phonetic facts, namely that a pitch fall is present but its locus is predictable in this tonal class (i.e. it is not lexically “anchored”). Any segmental anchoring that is observed is then attributed to phonetic realization rules. As discussed in Chapter 4, one could stipulate that the

contour node associates with the first two syllables postlexically, as Lee (2008) did, but then one would have to further stipulate that the phonetics is sensitive to the difference between lexical and postlexical association lines, resulting in a different realization strategy for this tune<sup>60</sup>. (This alternative representation will be considered again in Section 6.3, where it is suggested that the context-sensitive deletion of the **I** tone may be simpler to explain if the **h** tone associates with the second syllable at some stage; for now the current analysis is pursued.) The association rules for NKK may be rather obvious at this point, but they are given in (6.3) for the sake of completeness:

(6.3) Lexical tune association rules for NKK:

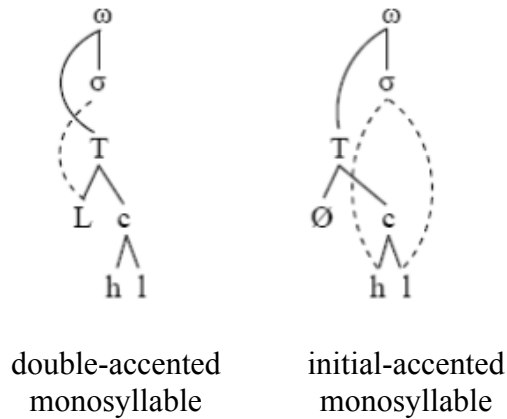
1. If a **c** node is associated with a syllable, associate the terminal **h** node under it with that syllable and the terminal **I** node with the following syllable, if there is one; otherwise associate the **I** node with the same syllable as the **h** node.
2. If the first syllable of the word is still unassociated, associate the register node with it.
3. Delete unassociated register tones.
4. Remove any association lines connecting contour nodes to syllables.

It is worth fleshing out the representations for monosyllables in NKK, as well, since the production results showed that the respective realizations of the initial-accent and the double-accent were very similar, and since perceptual results indicated that the two-way tonal contrast was largely perceptually neutralized on monosyllables. It is possible that a structural difference between the two categories is captured by underlyingly associating the contour node with the syllable node for the initial-accent but leaving it unassociated for the double-accent. Consequently the register node is deleted in the first case and associated with the syllable in the other case. The proposed surface representations for the two lexical tonal categories on monosyllables are presented in Figure 6.13. The structural difference between these two tunes

---

<sup>60</sup> Lee (2008) herself mentioned in a footnote that the phonetic realization of a “floating” H tone may be preferable to the postlexical association of the tone, since the former more transparently captures the phonetic facts.

on monosyllables reflects the intuition of the native speaker that there is a contrast between these categories on monosyllables; the extent to which this contrast is neutralized on the surface (as



**Figure 6.13: Surface representations for lexical tunes on double- and initial-accented monosyllables in NKK.**

discussed in Section 2.5.5 of Chapter 2 and Section 3.5.4 of Chapter 3) is down to the phonetic implementation. Although the contour node is not associated with the syllable in the double-accented case, there is only that one syllable on which to realize it, so the syllable surfaces with a peak and a fall that largely overwrite the effect of the L register tone being present as opposed to deleted<sup>61</sup>, and the two tunes end up surfacing with very similar pitch contours.

To summarize, a multi-tier autosegmental tune representation based on the geometry proposed by Bao (1990) provides us with a good template on which to base lexical tonal representations for various languages, including non-Chinese ones. The T node may associate with what is traditionally thought of as the TBU in some languages, like the syllable in Mandarin and Cantonese, or with a higher-level lexical constituent, like the word in Japanese and NKK, which becomes the TBU (tune-bearing unit) in the current framework. In the cases where it associates with a constituent higher than the terminal prosodic node, the lower-level nodes may

---

<sup>61</sup> As was mentioned for Tokyo Japanese, it is possible that a postlexical tune containing a L tone is associated with the leftmost—or in the case of monosyllabic words, only—syllable in NKK, making the representations on monosyllables even more similar with one another and explaining the rise observed at the left edge of the surface contours for both forms.

be anchored to its sub-constituents. These anchors can be lexically determined, as in the initial- and final-accented tonal categories in NKK and Kansai, or they can be left unspecified, in which case the phonology may delete the unanchored lower-level nodes, as in the Japanese dialects, or allow them to surface with some default alignment, as in NKK. When the register node is left intact but unassociated, it gets interpreted in parallel with the contour node, as in Mandarin and Cantonese, but if it is associated with a prosodic unit it gets “unpacked” from the tune and realized serially as a tone in its own right. This dichotomy between parallel and serial realization of melodic units is important and will come into play at higher levels of the prosodic hierarchy as well.

Note that the orientation of the register node relative to the contour node is significant for word tune languages that make use of register as well as “tonal prominence” within a word, since association lines originating from sisters are prohibited from crossing. The placement of the register node to the left of the contour node predicts that we will not find a language whose word tunes are the mirror image of those in Kansai Japanese dialects, i.e. a language in which register contrasts are realized *after* the “accentual” portion of the word tune. If such a language were shown to exist, the tune template would have to be relaxed to accommodate either orientation.

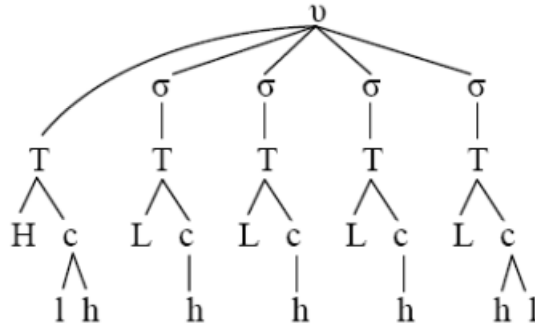
### **6.3 Where Tone and Intonation Meet**

In Section 5.3.2 of the current chapter, the point was made that committing to AM-based analyses forces us into certain theoretical conclusions when it comes to tonal contrasts and tone-dependent intonation, namely that those in Kansai Japanese would need to be handled via stipulative tone-specific restrictions (e.g. in Kori’s 1987 analysis of Kansai melodies) and category-specific phonetic implementation. The more powerful autosegmental tune geometry proposed in Section 6.2 tips the scales back a bit the other way in that it allows us to account for a wider range of contrasts than the traditional AM analyses do with structural explanations alone—the stipulative co-occurrence restrictions of the sort found in the Kori analysis for Kansai

are a non-issue in this type of geometry because word tunes are simply hard-coded in the lexical inventory as units. But what about the various ways in which echo question intonation seems to manifest itself in different languages and, in particular, the different ways in which it appears to interact with lexical tunes in each language and for each lexical tonal category within each language?

Of course, if we are assuming that a phonological representation, and not semantic or syntactic information, is what is sent to the phonetics, there is no reason to expect that echo questions feed the same input to the phonetics in every language. However, at least for the languages examined here, there are certain tendencies that are difficult to ignore. In every case, most of the “action” associated with echo question intonation appears to take place toward the right edge of the utterance. Furthermore, in many cases there is a tendency for the overall pitch range of the utterance to be affected, either in the sense that the pitch is slightly higher overall or in the sense that the pitch ceiling is slightly higher and local maxima are higher as a result. If we think in terms of the tune geometry we have established for lexical tunes, it is almost as if there is an utterance-level tune invoked by echo questions that has a **H** register and some sort of contour involving a **h** terminal node that gets associated with a constituent at or near the right edge. With that in mind, let us look at how tone and intonation might interact within the phonology of each of these languages to produce the various phonetic effects we have observed.

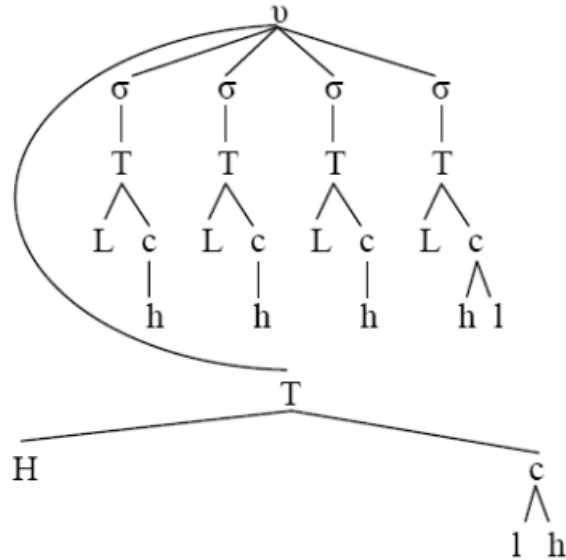
A possible phonological representation for the melody of a Cantonese echo question (*Lei6lei6wa6laan3?*) is shown in Figure 6.14. The current theoretical analysis makes no claims about levels of the prosodic hierarchy between the syllable and the utterance in Cantonese, and as such the immediate domination of the former by the latter should be taken with a grain of salt (although Wong, Chan et al. 2005 did note that there was little evidence for prosodic constituents in Cantonese between the syllable and what they called the intonational phrase). Crucially, though, the lexical tunes are associated with syllable nodes and the intonational tune affiliated with echo questions is associated with the utterance node (represented in the figure by the



**Figure 6.14: Underlying representation for the melody of the Cantonese utterance *Lei6lei6wa6laan4?* ('Leilei says 'orchid'?').**

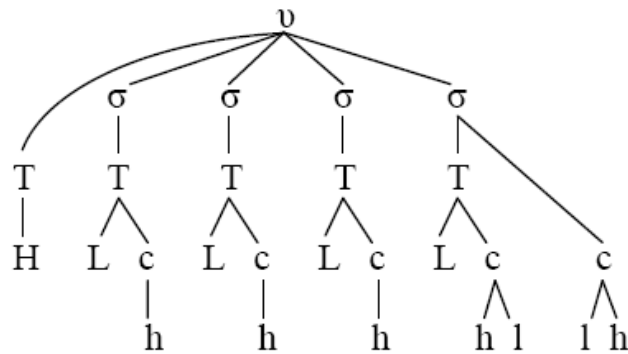
symbol **v**). In this case, the geometry selected for the echo question tune consists of a **H** register, which will translate into a slight raising of the overall pitch of the utterance, and a rising (**l-h**) contour, which will account for the consistent rise seen at the right edge of echo questions in Cantonese. The fact that this tune has the exact same geometry as one of the lexical tunes (that for T2) is a coincidence; the lexical tune inventory and the intonational tune inventory of a language should be considered to be independent of one another<sup>62</sup>. At the same time, though, the phonetics is considered to be blind to the respective sources of these tunes, and the extent to which they are interpreted differently by the phonetics is down to structural differences (which are themselves due to functional differences). This assumption is carried over from the traditional AM framework. Before moving on to what the surface representation might look like, it is worth rearranging the elements of the tree in Figure 6.14 to drive home the point that, at this level of representation, the intonational tune is on a different plane from the lexical tunes. This less orthodox presentation of the tree, shown in Figure 6.15, makes it clear that the register node of the intonational tune has access to (i.e. the capacity to secondarily associate with) any of the syllables in the utterance, as does the contour node and its daughters. It also shows that the choice to represent the tune as sitting to the left of the lexical tunes in the utterance in Figure 6.14 is arbitrary. The only restriction on secondary association (or re-association) in this context is that association lines coming from nodes dominated by the **T** node are prohibited from

<sup>62</sup> For Cantonese, the inventory of intonational tunes would have to be sufficient to account for the wide range of utterance-final intonational particles the language boasts (see Law 1990 for a comprehensive discussion of these particles).



**Figure 6.15: Rearranged underlying representation for the melody of the Cantonese utterance *Leilei6wa6laan4?* ('Leilei says 'orchid'?').**

crossing one another, so the register node would not be allowed to be associated with a prosodic node to the right of a prosodic node with which the contour node or one of its daughters is associated. Let us now imagine what the surface structure derived from such a configuration might look like. The proposed surface representation is shown in Figure 6.16. The idea here is



**Figure 6.16: Surface representation for the melody of the Cantonese utterance *Leilei6wa6laan4?* ('Leilei says 'orchid'?').**

that the **T** node, along with its **H** tone daughter, has remained associated with the utterance node, but the contour node has been “delinked” from the **T** node and has been re-associated with the rightmost syllable node. The phonetics then interprets the tune consisting of a single **H** tone in parallel with the entire utterance—i.e. the scope of the tune spans the entire utterance. On the

other hand, the now registerless contour unit is primarily associated with the rightmost syllable, to the right of the lexical tune that is associated with that syllable. Because these two melodic constituents are now sisters under the same syllable node, the phonetics interprets them in series. The language-specific phonetics determines how the combination of lexical and intonational tunes associated with the final syllable are realized, but the fact that the rightmost contour constituent is registerless is reflected in the fact that the pitch range of the final rise does not appear to be fixed in the same relative pitch space as the lexical tune (crucially, even when the lexical tune is T1, the high level tune, the pitch rises to the end<sup>63</sup>) and the fact that the phonetics is interpreting the melodic units on the last syllable in series is reflected in the fact that the left part of the syllable is realized according to the identity of the lexical tune and the right part of the syllable is realized according to the identity of the intonational tune. In this proposed model, the phonetics is not only language-specific but also sensitive to the particular combination of lexical and intonational tunes at the right edge and therefore the phonetic implementation at that right edge can be tonal-category-dependent in the ways discussed in Chapter 4.

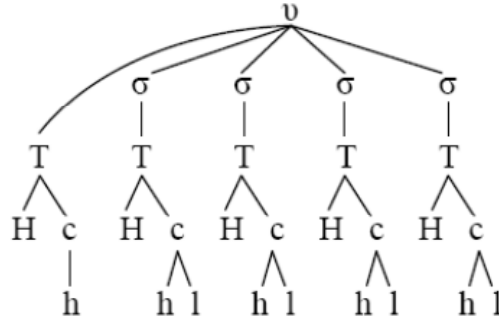
Let us turn to an example from Mandarin to see how this interaction works in another language. Figure 6.17 shows the underlying structure for the melody of the utterance *Na4liang4nian4wan4?* Apart from the makeup of the individual tunes, this structure is identical to that for the Cantonese echo question. The intonational tune is represented with a unary-branching contour node, with a single **h** tone under it. The structure of the surface representation, shown in Figure 6.18<sup>64</sup>, is slightly different from that for Cantonese. The crucial difference in Mandarin is that the contour node on the right-hand side of the utterance is simultaneously associated with the utterance node and the right-most syllable node. The phonetics then

---

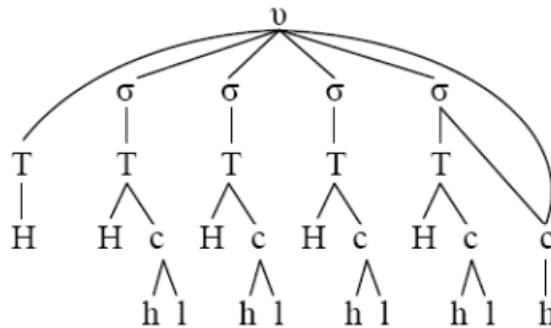
<sup>63</sup> An alternative analysis could be that the intonational tune places a single high target at the very end of the utterance and that the target is simply higher than even the highest lexical target, resulting in a rise regardless of the identity of the last lexical tune. However, this would require us either to come up with a way of representing more than four levels of relative pitch in our geometry or to stipulate that the phonetics is actually privy to the functional sources of melodic units.

<sup>64</sup> Note that any phonological sandhi processes, such as T3 sandhi in Putonghua, would be reflected as a restructuring of the affected lexical tunes on the surface.





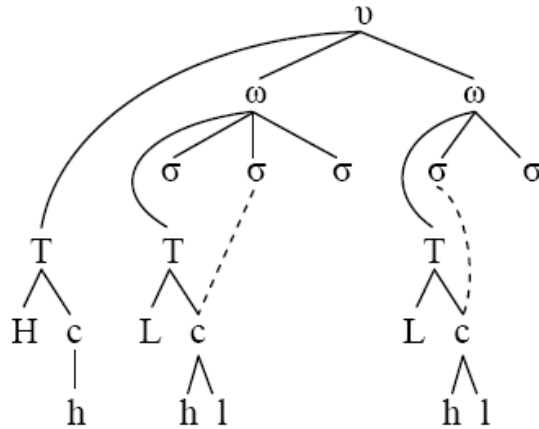
**Figure 6.17: Underlying representation for the melody of the Mandarin utterance Na4liang4nian4wan4? ('Naliang reads 'ten thousand'?').**



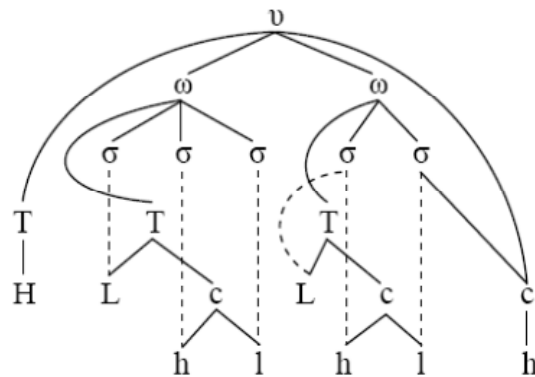
**Figure 6.18: Surface representation for the melody of the Mandarin utterance Na4liang4nian4wan4? ('Naliang reads 'ten thousand'?').**

interprets this structural configuration such that the lexical tune and the intonational tone are implemented in parallel (possibly via an increased strength value, as well as some other tonal-category-dependent mechanisms) by the association of the right-hand intonational tune (which is still associated with the utterance node) to that syllable.

Let us turn now to NKK. The proposed underlying representation for *Eunhi-neun nam-i?* ('Eunhi-TOP horse-NOM?') is shown in Figure 6.19. Again, no theoretical claims are being made about the exact structure of the prosodic hierarchy between the level of the prosodic word and the utterance (Jun, Kim et al. 2006 suggested that accentual phrases in NKK are grouped into intermediate phrases, which are in turn grouped into intonational phrases). Here we see that the intonational tune is given the same representation as that for Mandarin. The surface representation for the NKK melody is shown in Figure 6.20. Like in Mandarin, the contour node on the right-hand side is doubly associated with the utterance node and the last syllable node.

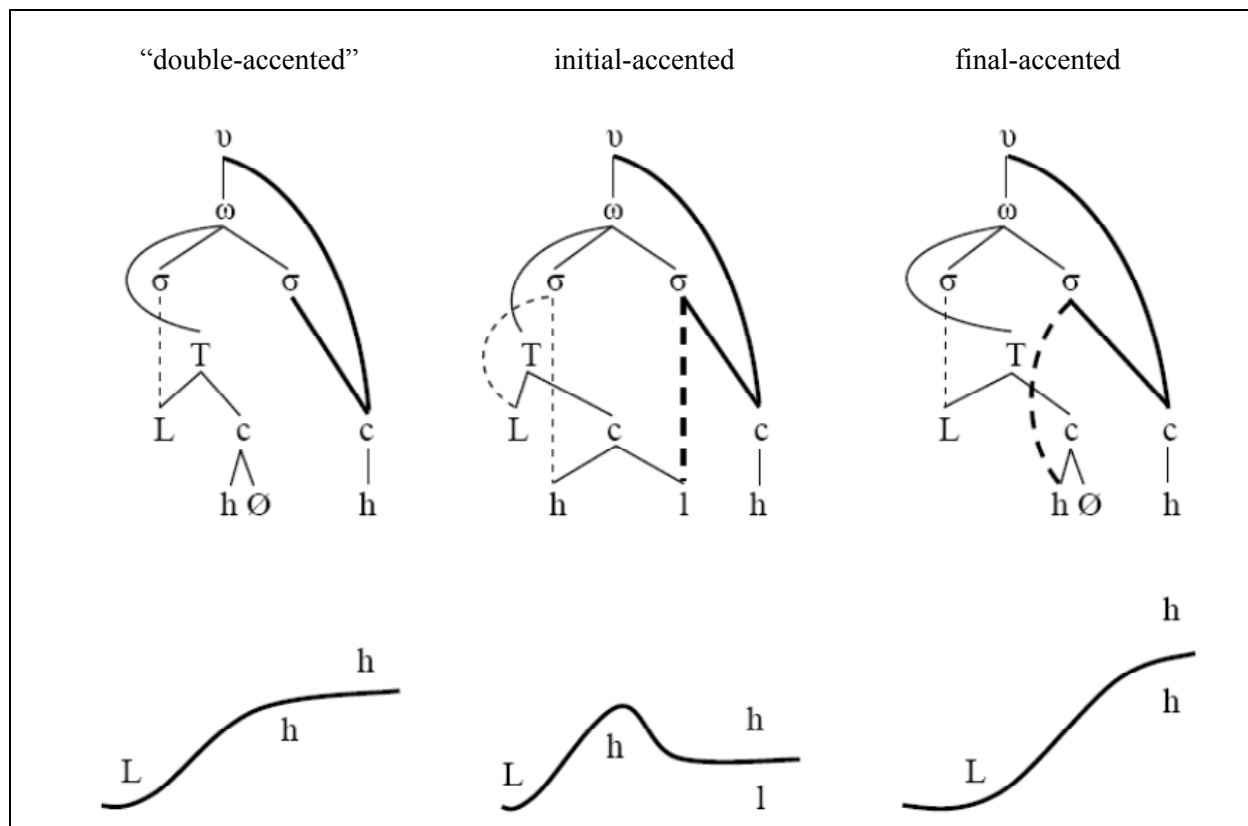


**Figure 6.19: Underlying representation for the melody of the NKK utterance *Eunhi-neun nam-i?* (‘Eunhi-TOP horse-NOM?’).**



**Figure 6.20: Surface representation for the melody of the NKK utterance *Eunhi-neun nam-i?* (‘Eunhi-TOP horse-NOM?’). The word *nam-i* is initial-accented.**

Following the convention established for Mandarin, the phonetics then implements the intonational **h** tone in parallel with the lexical **l** tone, resulting in a mid-range target. What makes the phonological interaction of the two melodic functions in NKK more complicated than that in Mandarin is that the right-hand constituent with which the intonational tune associates is not the TBU, but rather a sub-constituent of the TBU. For this reason, it is worth breaking down the various possible interactions by lexical tonal category; this breakdown is given in Figure 6.21. As shown in the figure, the theoretical choice has been made to assume that the **l** tone of the lexical tune is deleted from the “double-accented” and final-accented tunes in this environment, in order to account for the alternations discussed for NKK in 4.3.1 of Chapter 4. This can be thought of as another instance of sandhi, characterized structurally as the deletion of a sub-

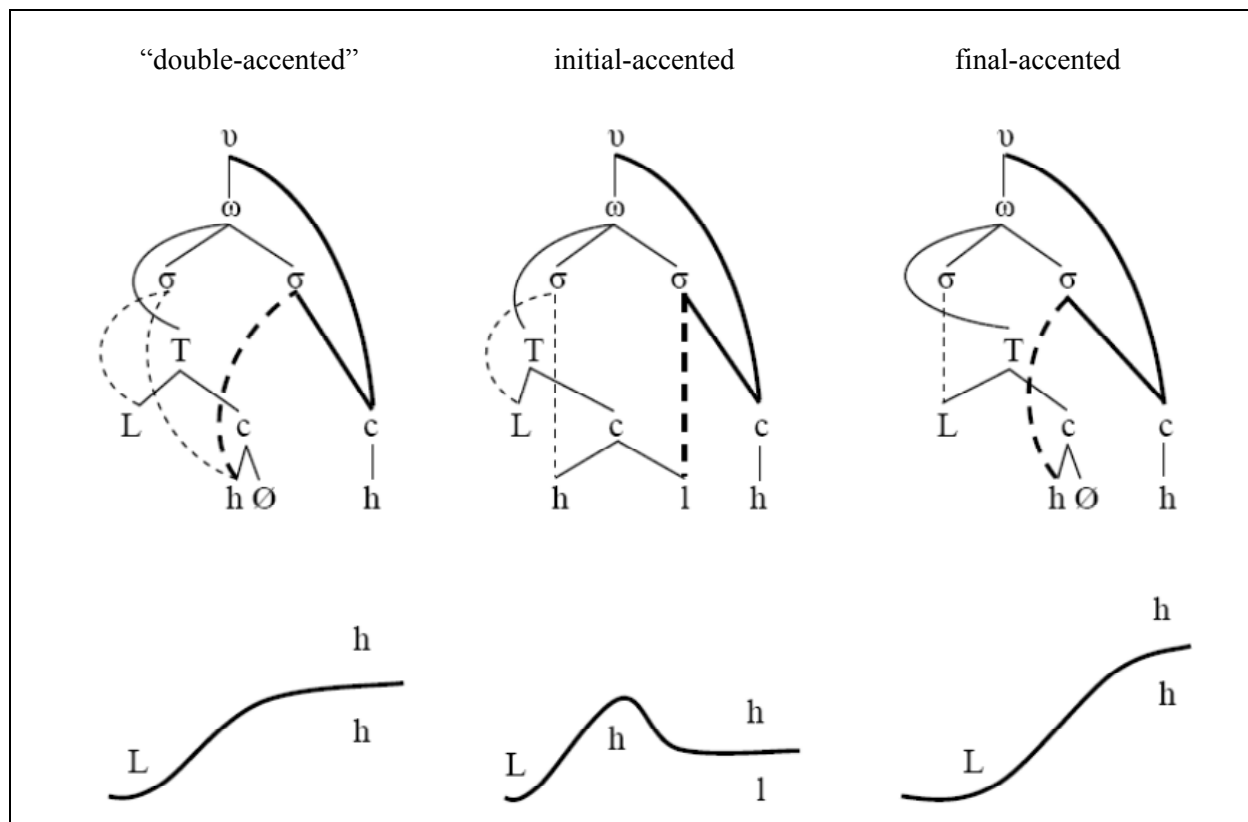


**Figure 6.21: Proposed surface representations for three lexical tunes in an echo question context in NKK, with accompanying schematic representations for phonetic implementation.**

constituent of the lexical tune<sup>65</sup>. One drawback of this analysis is that it is not clear how to characterize the *context* for **I** tone deletion in the double-accented tune<sup>66</sup>. Given this issue, an alternative (more traditional) representation that makes double-accented and final-accented disyllabic words more structurally similar at the right edge might be more desirable; one such representation is given in Figure 6.22. The issue that arises here is that, now that the right edges of disyllabic double-accented words and that of final-accented words look so similar structurally, we need to explain why the echo question contours end in different pitch heights. One solution

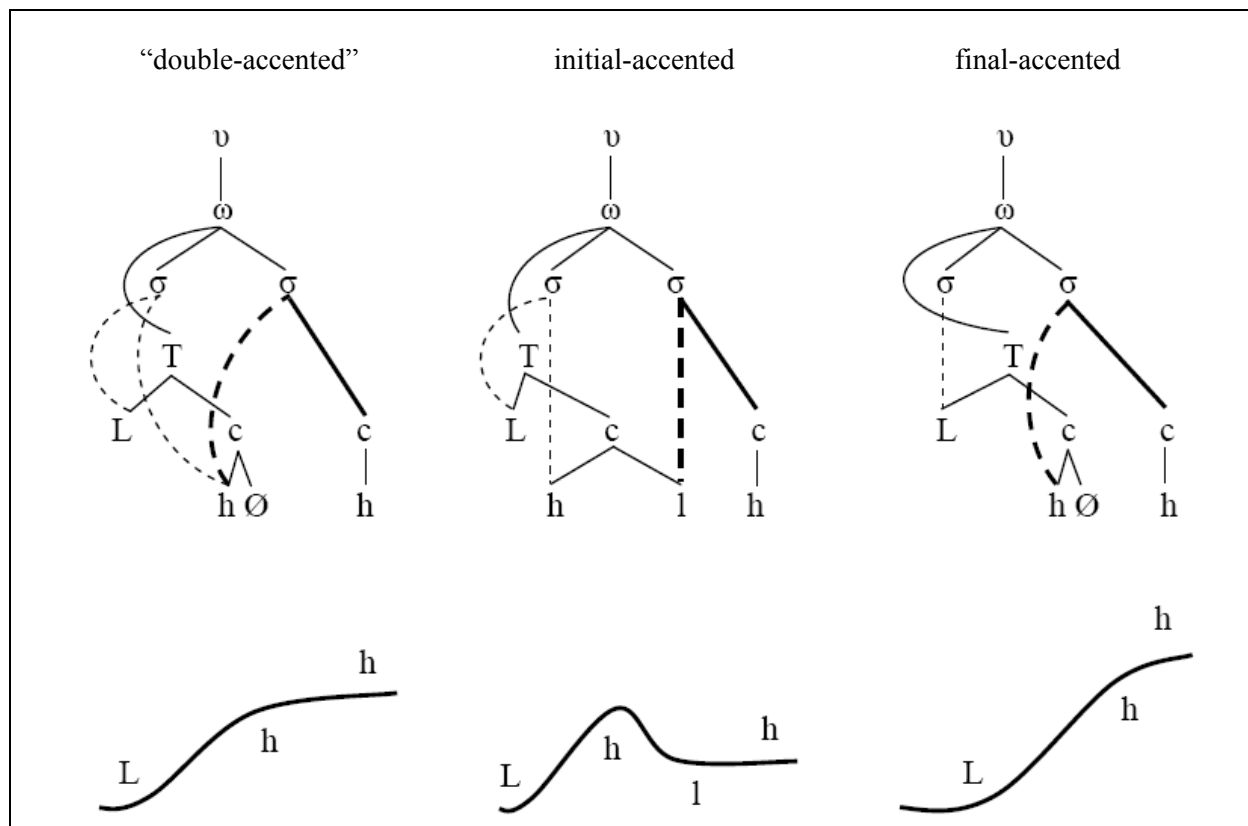
<sup>65</sup> An alternative analysis of the final-accented case is to say that NKK is like Tokyo Japanese in prohibiting the simultaneous association of the **h** and **I** lexical tones with a single terminal prosodic node regardless of the intonational environment; in that case the fall in pitch observed at the end of final-accented words in a declarative environment would be attributed to the presence of an intonational **I** tone

<sup>66</sup> One (perhaps radical) possibility is that the **h** tone in the double-accented tune remains unassociated postlexically but the **I** tone tries to associate with a third syllable if there is one and otherwise it gets deleted. In a personal communication, Hye-Sook Lee informed me that the contour at the right edge of a trisyllabic double-accented word is similar to disyllabic initial-accented words in an echo question context, which would suggest that the **I** tone is indeed associated with the third syllable in that context.



**Figure 6.22: Proposed surface representations for three lexical tunes in an echo question context in NKK (with a double-linked h representation for the double-accented tune), with accompanying schematic representations for phonetic implementation.**

is simply to say that the **h** tone originating from the final-accent tune is interpreted differently from the one originating from the double-accent tune because the former is associated lexically while the latter is associated postlexically (as Lee 2008 proposed to account for the exceptional alignment behavior of the double-accented contour in declarative contexts), and therefore the contour resulting from the parallel implementation of the two **hs** in one case is lower than in the other case. Another possibility is that the presumed parallel implementation of the **h** intonational tone is actually incorrect, and that the implementation of the melodic units on the last syllable is actually serial. This third analysis is given in Figure 6.23. If we need to stipulate that the phonetics is sensitive to whether associations are lexical or postlexical in order to account for the alignment difference between the double-accent and the final-accent in the declarative context anyway (see Conclusion II in (5.16) in Chapter 5), we can account for the difference in slope and

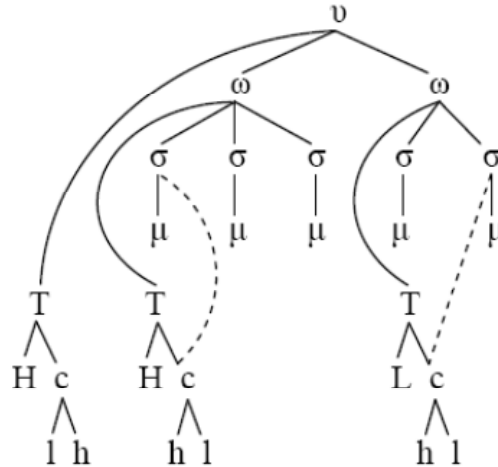


**Figure 6.23: Alternative surface representations for three lexical tunes in an echo question context in NKK (with serial implementation of melodic units on the last syllable), with accompanying schematic representations for phonetic implementation.**

pitch height at the right edge of those two lexical categories in the echo question context by saying that both pairs of **h** tones are realized in series but that the actual phonetic target for the lexical **h** tone is closer to that for the intonational **h** tone in the final-accented case, and therefore they reinforce each other more, resulting in a higher pitch. All of the above analyses include stipulative elements; absent more relevant data, the choice remains one of personal taste.

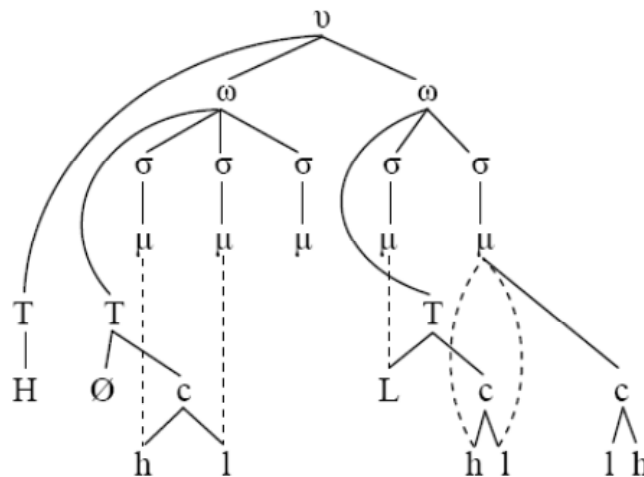
Finally, let us turn to SJ as a representative Kansai dialect. The proposed underlying representation for the melody of the utterance *Aya-ga ame?* ('Aya-NOM rain?')<sup>67</sup> is shown in Figure 6.24. As was done for NKK, the intermediate phrase node that is presumed to dominate the word nodes has been omitted for simplicity's sake. Recall that the **T** nodes are associated with the word nodes and the contour nodes are associated with syllable nodes (since the syllable

<sup>67</sup> The name *Aya* is initial-accented just like *Naoya*, which was the actual name used in the experiment with Shiga Subject C. *Aya* is used here simply because it is shorter!



**Figure 6.24: Underlying representation for the melody of the SJ utterance *Aya-ga ame?* ('Aya-NOM rain?').**

is the accent-bearing unit). Since *Aya-ga* is high-beginning, the register tone for that word is **H**, and since *ame* is low-beginning, the register tone for that word is **L**. In essence, then, these H and L register tones are the equivalents of pH and pL in the Kori and P&B analyses in Section 5.3.2. The surface representation for this melody is shown in Figure 6.25. Once again, the intonational tune on the left remains associated solely with the utterance node. Just as in



**Figure 6.25: Surface representation for the melody of the SJ utterance *Aya-ga ame?* ('Aya-NOM rain?').**

Cantonese, the registerless contour node is re-associated with the final mora of the utterance, but not with the utterance node. The phonetics then interprets the **h** and **l** tones from the lexical tune

and the intonational contour tune in series, resulting in a rise, a fall, and another rise on the last mora. The dialect-specific phonology dictates that both the **h** and **l** tones in the lexical tune on the last word get secondarily associated with the final mora and therefore realized; in Kyoto Japanese, of course, the **l** tone would fail to associate with the final mora in that environment, and therefore it would get deleted. The language-specific phonetics dictates that all of the melodic targets in the surface-phonological representation are fully realized in SJ, which for some speakers entails a lengthening of that final mora. This is in contrast to Cantonese, in which the serially realized melodic targets on the last syllable were short-changed (apparently in favor of maintaining a consistent syllable duration).

Finally, it should be noted that the left-edge phrase boundary tones discussed in Section 5.3.2 (H% and L% in the Kori analysis and L% in the P&B' analysis) have been left out of the current analysis. They could presumably be included here either as tunes consisting of a **L** register tone and a **l** contour terminal tone, with the phonetics dictating that this low target is realized lower before a **L** register tone and higher before a **H** register tone, or they could be expressed as registerless rising contour tunes, ensuring that there is always a rising contour leading into the word but the starting level of the contour being unspecified and predictable from the starting register of the word. Either way they would be inserted postlexically and associated with word-initial morae, and implemented in series with any other melodic units associated with those morae.

#### **6.4 Phonetic Awareness of the Tree Structure**

It was shown in Chapter 4 that there is sometimes a need for the phonetic implementation of utterance-level intonation to be tonal-category-dependent. While some of these category-dependent alternations (such as the context-sensitive deletion of low targets in NKK and Kyoto Japanese) are handled via structural interactions at the phonological level in the current model, others (such as the unique interpolation shape of utterance-final T1 in Cantonese echo questions)

are left for the language-specific phonetics to handle. One advantage of the structural framework outlined in Section 6.3 is that, in addition to giving us a way to parameterize which phonetic implementations happen in series and which happen in parallel, it also provides us with a way to constrain what aspects of the phonology the phonetics can access. Although lexical melodic units are not overtly distinguished in the surface representation we have proposed (i.e. we do not have different node labels for lexical tunes and intonational tunes), their origins can be encoded to a limited degree through the use of association lines. One testable stipulation we can make is that, if tunes are sisters (i.e. they share the same parent), the phonetic implementation of each of those tunes can be sensitive to the identity of the other one. We can even take this notion further and say that if a tune is *c-commanded* by another tune, the phonetic implementation of it can be sensitive to the category of that other tune. The robustness of this principle warrants further testing against a wider range of data from a variety of languages.

## 6.5 Summary and Conclusion

The structural account given in Sections 6.2 and 6.3 is by no means the only conceivable analysis given the facts; it is but one possible story we can tell that allows us to maintain a unified template for lexical tonal representation and tone-intonation interaction within an autosegmental framework. The basic autosegmental schema of hierarchical nodes and association lines has proven quite useful for explaining a wide range of melodic processes for decades, and the current account builds on the traditional schemata by introducing the mechanisms given in (6.4):

### (6.4) Summary of mechanisms in the scopal model of speech melody

1. Any node in the lexical tonal geometry may associate with any node dominated by the TBU node (defined as the node with which the **T** node is lexically associated). The phonology in a given language dictates which types of associations in which combinations are permitted, required, lexically specified, postlexically instantiated, etc.



2. Utterance-level intonational tunes are associated underlyingly with the utterance node in the prosodic hierarchy, and in the surface representation those tunes or daughter nodes of those tunes may re-associate with lower-level nodes, either in addition to or instead of that highest-level association.
3. The phonetics is sensitive to the various combinations of associations a given tune has (e.g. it will treat a tune that is associated with both the utterance node and a syllable node differently from both a tune that is solely associated with the utterance node and a tune that is solely associated with that syllable node). Specifically:
  - i) If a tune is solely associated with the utterance node, its effect will be observed over the entire utterance (i.e. it will be interpreted in parallel with all the tunes of the utterance).
  - ii) If a tune is associated both with the utterance node and with a lower-level node, its effect will be observed over the entire constituent dominated by the lower level node (i.e. it will be interpreted in parallel with any tunes dominated by that lower-level node).
  - iii) If a tune is solely associated with a lower-level prosodic node, its effect may only be observed over part of the duration of that prosodic constituent (i.e. it will be interpreted in series with its sister tunes and tones).
4. If Tune A c-commands Tune B in the autosegmental tree, the phonetic implementation of Tune B can be sensitive to the lexical category of Tune A.
5. If two nodes are sisters then the left-hand node cannot be associated with a node to the right of the right-hand node's parent (just as in traditional autosegmental frameworks). This illicit configuration is schematized in Figure 6.26.



## II. Tone-specific parameters

- a. In NKK, the realization of the post-accentual low target in echo-question-final polysyllabic initial-accented words but not in echo-question-final monosyllabic initial-accented words, double-accented disyllabic words, or final-accented words
- b. In Kyoto Japanese, the failure of the post-accentual low target of a final-accented word to be realized when the word is echo-question-final

This leaves a number of melodic phenomena to be handled by the phonetics. Where possible, these tasks for the phonetics were mentioned in passing in previous sections, but for the sake of clarity they are summarized in (6.6):

### (6.6) Melodic phenomena handled by the phonetics

#### I. Language/dialect/speaker-specific phenomena

- a. How the contrasts encoded in the autosegmental tune geometry get realized
- b. How serially implemented melodic units get realized (e.g., melodic units on the echo-question-final mora in SJ vs. those on the echo-question-final syllable in Cantonese)
- c. How parallel-implemented melodic units get realized

#### II. Tone-specific phenomena

- a. In Mandarin and Henanhua, the degree to which various parts of the contour are raised or steepened under the influence of a parallel-implemented **h** tone
- b. In Cantonese, the degree of pitch excursion executed on the final **l-h** tail in echo questions; also the interpolation shape of the serially implemented lexical-tune-intonational-tune sequence at the ends of echo questions
- c. In NKK, the realization of singly-associated as opposed to doubly-associated **h** tones (the relevant distinction could also be lexical vs. postlexical association, as Lee (2008) suggests)

- d. In Shiga Japanese, the strategy for interpolation between the register tone target and the echo question **l-h** tail in echo-question-final unaccented words

The formulation of an explicit phonetic component for a comprehensive model of speech melody is beyond the purview of this dissertation, but it is likely that one or more of the already-existing phonetic models could be adapted to handle all of the phenomena listed in (6.6). For example, many of the overlay models discussed in Chapter 1 and Chapter 4 have a way of controlling the degree of pitch excursion for a given tonal command—there is the tone command amplitude parameter in the command-response model, the strength parameter in Stem-ML, and the pitch range parameter in PENTA. Most phonetic implementation models also have some kind of phrase-curve parameter for controlling the overall pitch level of an utterance. The crucial argument being made here regarding the phonetic component is that, whatever the precise mechanisms invoked to produce these general effects, the settings of these parameters must be allowed to be sensitive to certain properties of the structure that is the output of the phonological component outlined in this chapter.

## CHAPTER 7: CONCLUSION

The interaction of tone and intonation is a relatively poorly understood component of language, and previous literature on the topic has tended either to be overly-general, glossing over language-specific quirks that complicate the overall picture, or too narrow, producing models that are tailored to a single language. This dissertation, in an attempt to bridge the gap between the two extremes, investigated the interaction of lexical tone and utterance-type intonation at the right edge of an utterance in several different languages. Qualitative and quantitative results from a series of production and perception experiments were presented, and the results painted quite a complex typological picture of speech melody. The results showed that, while some loose typological generalizations can be made (e.g., a rising contour that is the realization of one melodic function may be misinterpreted as the realization of another melodic function), but many attributes of melodic systems are “cross-cutting” (e.g., whether the lexical and intonational tunes at the right edge of an utterance are implemented in parallel or in series is not predictable based on whether the language in question is a word-tone or a syllable-tone language).

Admittedly, the scope of investigation in these experiments was quite narrow, shining a light on only a small corner of the vast realm of speech melody in each language. However, it proved to be quite an effective litmus test for assessing the viability of existing models of speech melody. It was shown that phonetics-only overlay models are not powerful enough to handle the tone-dependent intonational phenomena that abound in tone languages; meanwhile, it was suggested that sequential models couched in the traditional AM framework lack the machinery needed for comparing different types of melodic systems in a useful way.

A scopal model that exploits the strengths of the traditional AM framework and augments it with structural concepts borrowed from several orthogonally related theories (e.g., feature geometry, tonal anchoring, scopal domains) was proposed. While the model does not constitute the only theoretical possibility for reconciling all of the results presented, it does address the

shortcomings of other models. The model also makes certain predictions in terms of the types of melodic systems that should be observed in the world's languages and, at the same time, as more cross-linguistic data become available, there is room built into it for adapting and evolving accordingly.

It goes without saying that the possibilities for relevant future investigation are innumerable. Obviously, data from more tone languages are needed to flesh out the melodic typology along the lines of the current study, and data from more speakers of all of the languages included here—especially the less-well-studied ones—need to be collected in order to determine which phenomena are language-specific and which are dialect- and speaker-specific. Eventually, other melodic functions, including those that encode contrastive focus, should also be systematically controlled for and investigated across a variety of languages.

In order to place certain phenomena squarely in either the phonological or phonetic component of the model, it would be helpful to find examples of phonetic and phonological doublets<sup>69</sup> in this realm of speech melody. One promising direction for this was seen in the preliminary speech rate results obtained for NKK, which were discussed in Section 4.3.1 of Chapter 4—it appears that NKK displays both a process of phonological target deletion and a tendency for targets to be less than fully realized. The first process appears to be categorical and triggered by “tone crowding” with respect to syllable structure, while the second process appears to be gradient and affected by factors like speech rate and degree of emphasis.

Last but not least, if we assume that many of the instances of tone-dependent intonation that have been observed in the current study are indeed governed by the phonetics, it is worth asking what the driving force behind these phenomena are. In some cases, the answer may have to do with articulation—it is known that different sets of muscles may be used for raising and lowering pitch, respectively (Ohala 1972), and so in some cases the realizations of seemingly parallel combinations of pitch targets may be asymmetrical due to the disparate sets of motor

---

<sup>69</sup> A term coined by Cohn (1998) to describe cases where “similar but distinct effects of both a categorical and gradient nature are observed in the same language” (Cohn 2006, p. 29); see footnote 50.

commands involved. In other cases, there may be a paradigmatic perceptual basis for certain irregularities—speakers may try to maintain or maximize lexical tonal contrasts when the melodic signal is simultaneously transmitting intonational information. This idea has been proposed by Chen and Gussenhoven (2008) and Smiljanić (2004) to account for tone-dependent effects observed in the realization of focus in Mandarin and Serbo-Croatian, respectively.

## APPENDIX

**Table A: Complete word list for the Kansai Japanese production experiment, with expected tonal categorizations based on entries for Kyoto Japanese in Martin (1987).**

word	gloss	register	accented mora	trad. tone pattern
<i>ame</i> <i>momo</i> <i>mai</i> <i>mi</i>	‘candy’ ‘peach’ ‘dance’ ‘body’	H	unaccented	HH
<i>ne</i> <i>umi</i> <i>ni</i> <i>nuno</i> <i>yane</i>	‘root’ ‘sea’ ‘two’ ‘cloth’ ‘roof’	L	unaccented	LH
<i>yama</i> <i>me</i> <i>nami</i> <i>nawa</i> <i>nen</i> <i>nou</i> <i>mimi</i>	‘mountain’ ‘eye’ ‘wave’ ‘rope’ ‘year’ ‘noh (theatre)’ ‘ear’	H	1	HL
<i>ame</i> <i>mame</i> <i>mae</i>	‘rain’ ‘bean’ ‘front’	L	2 (final)	LHL̄
<i>nayami</i> <i>mimono</i> <i>mirin</i>	‘trouble’ ‘(a) sight’ ‘mirin’ (type of sweet rice wine)	H	unaccented	HHH
<i>nunome</i> <i>minwa</i> <i>meiro</i> <i>youi</i>	‘texture’; ‘weave (of cloth)’ ‘folklore’ ‘facial expression’ ‘easy’	L	unaccented	LLH
<i>omoi</i> <i>nyoui</i> <i>nioi</i> <i>ningyo</i> <i>minami</i> <i>yumemi</i>	‘thought’ ‘urge to urinate’ ‘smell’; ‘scent’ ‘mermaid’ ‘south’ ‘having a dream’	H	1	HLL
<i>onna</i>	‘woman’	H	2	HHL
<i>namami</i> <i>yamai</i> <i>youi</i>	‘living flesh’ ‘illness’ ‘preparation’	L	2	LHL



**Table A (cont'd): Complete word list for the Kansai Japanese production experiment, with expected tonal categorizations based on entries for Kyoto Japanese in (Martin 1987).**

word	gloss	register	accented mora	trad. tone pattern
<i>deppa</i> <i>noppo</i>	'protruding tooth' 'tall lanky person'	L	3 (final)	LLH̄L
<i>naiyou</i> <i>noumen</i> <i>yamaimo</i>	'subject'; 'contents' 'noh mask' 'Japanese yam' (lit. 'mountain yam')	H	unaccented	HHHH
<i>nagaimo</i> <i>namamono</i> <i>ningyou</i>	'Chinese yam' (lit. 'long yam') 'raw food' 'doll'	L	unaccented	LLLH
<i>nairon</i> <i>nanaman</i> <i>yonman</i>	'nylon' 'seventy thousand' 'forty thousand'	H	1	HLLL
<i>yamayama</i> <i>nomimono</i>	'mountains' 'beverage'	H	2	HHLL
<i>omonaga</i> <i>nougyou</i>	'having a long (oval-shaped) face' 'agriculture'	L	2	LHLL
<i>kagaribi</i> <i>tomarigi</i>	'fire in an iron basket'; 'bonfire' 'perch'	H	3	HHHL
<i>irogami</i>	'colored paper'	L	3	LLHL
<i>nanamagari</i>	'winding path'	H	unaccented	HHHHH
<i>nairugawa</i>	'the Nile (River)'	H	3	HHHLL
<i>youmuin</i>	'orderly'; 'janitor'	L	3	LLHLL
<i>yamanoie</i>	'mountain house'	H	4	HHHHL
<i>miyoumimane</i>	'learning by watching'	H	unaccented	HHHHHH

## REFERENCES

- Arvaniti, A., D. R. Ladd, et al. (1998). "Stability of tonal alignment: the case of Greek prenuclear accents." *Journal of Phonetics* 26(1): 3-26.
- Arvanti, A., D. R. Ladd, et al. (1998). "Stability of tonal alignment: the case of Greek prenuclear accents." *Journal of Phonetics* 26(1): 3-26.
- Bao, Z. (1990). *On the nature of tone*. Doctoral dissertation. Massachusetts Institute of Technology, Cambridge, MA.
- Bao, Z. (1996). "The syllable structure in Chinese." *Journal of Chinese Linguistics* 24: 312-354.
- Bard, E. G., A. H. Anderson, et al. (2000). "Controlling the Intelligibility of Referring Expressions in Dialogue." *Journal of Memory and Language* 42: 1-22.
- Belotel-Grenié, A. and M. Grenié (2004). The Creaky Voice Phonation and the Organization of Chinese Discourse. *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*. Beijing.
- Bolinger, D. (1964). "Intonation: around the edge of language." *Harvard Educational Review* 34: 282-296.
- Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund, Gleerup.
- Bruce, G. and E. Gårding (1978). A prosodic typology for Swedish dialects. *Nordic Prosody*. E. Gårding, G. Bruce and R. Bannert. Lund: 219-228.
- Chang, S.-E. (2005). F0 Timing in North Kyungsang Korean. *2005 Annual Meeting of the Linguistic Society of America*. San Francisco, CA.
- Chao, Y. R. (1930). "A System of Tone Letters." *Le Maître Phonétique* 30: 24-27.
- Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley, CA, University of California Press.
- Chen, A., C. Gussenhoven, et al. (2004). "Language-Specificity in the Perception of Paralinguistic Intonational Meaning." *Language and Speech* 47(4): 311-349.

- Chen, M. Y. (2000). *Tone Sandhi: patterns across Chinese dialects*. Cambridge, Cambridge University Press.
- Chen, Y. and C. Gussenhoven (2008). "Emphasis and tonal implementation in Standard Chinese." *Journal of Phonetics* 36: 724-746.
- Chen, Y. and Y. Xu (2006). "Production of Weak Elements in Speech - Evidence from F<sub>0</sub> Patterns of Neutral Tone in Standard Chinese." *Phonetica* 63: 47-75.
- Cheung, K.-h. (1986). *The Phonology of Present Day Cantonese*. Doctoral dissertation. University College, London.
- Cho, N.-h. (2002). *Hyöndaekugö sayong pindo chosa: Han'gugö haksübyong öhwi sönjöng ül wihan kich'ö chosa*. Söul T'ökyölsi, Kungnip Kugö Yön'guwön.
- Chomsky, N. and M. Halle (1968). *The Sound Pattern of English*. New York, Harper and Row.
- Chuang, C. K., S. Hiki, et al. (1972). The acoustical features and perceptual cues of the four tones of standard colloquial Chinese. *The 7th International Congress of Acoustics*. Budapest, Akademiai Kiado. 3: 297.
- Clements, G. N. (1979). "The Description of Terraced-Level Tone Languages." *Language* 55(3): 536-558.
- Clements, G. N. (1985). "The geometry of phonological features." *Phonology Yearbook* 2: 225-252.
- Clements, G. N. (1990). The status of register in intonation theory: comments on papers by Ladd and by Inkelas and Leben. *Papers in Laboratory Phonology I*. J. Kingston and M. E. Beckman. Cambridge, Cambridge University Press: 58-71.
- Cohn, A. C. (1998). "The phonetics-phonology interface revisited: Where's phonetics?" *Texas Linguistic Forum* 41: 25-40.
- Cohn, A. C. (2006). Is there gradient phonology? *Gradience in Grammar: Generative Perspectives*. G. Faneslow, C. Féry, M. Schlesewsky and R. Vogel. Oxford, Oxford University Press: 25-44.

- Connell, B. A. and D. R. Ladd (1990). "Aspects of pitch realisation in Yoruba." *Phonology* 7: 1-30.
- Davidson, D. S. (1991). "Stress and Tonal Targets in Tianjin Mandarin." *Working Papers in Phonetics (UCLA)* 78: 50-57.
- Dilley, L. C., D. R. Ladd, et al. (2005). "Alignment of L and H in bitonal pitch accents: testing two hypotheses." *Journal of Phonetics* 33(1): 115-119.
- Duanmu, S. (1990). *A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages*. Doctoral dissertation. MIT, Cambridge, MA.
- Edmondson, T. and J. T. Bendor-Samuel (1966). "Tone patterns of Etung." *Journal of African Languages* 5: 1-6.
- Fiengo, R. (2007). *Asking Questions: Using Meaningful Structures to Imply Ignorance*. Oxford, Oxford University Press.
- Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. *The Production of Speech*. P. F. MacNeilage. Heidelberg, Springer-Verlag: 39-55.
- Fujisaki, H. and K. Hirose (1982). Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. *Preprints of Papers, Working Group on Intonation, Thirteenth International Congress of Linguistics*. Tokyo: 57-70.
- Fujisaki, H. and K. Hirose (1984). "Analysis of voice fundamental frequency contours for declarative sentences of Japanese." *Journal of the Acoustical Society of Japan* 5(4): 233-242.
- Fujisaki, H. and S. Nagashima (1969). "A model for synthesis of pitch contours of connected speech." *Annual Report of the Engineering Research Institute, University of Tokyo* 28: 53-60.
- Gårding, E. (1979). "Sentence intonation in Swedish." *Phonetica* 36: 207-215.

- Gårding, E. (1983). A Generative Model of Intonation. *Prosody, Models and Measurements*. A. Cutler and D. R. Ladd. Heidelberg, Springer-Verlag.
- Gårding, E. (1984). "Comparing Intonation." *Working Papers (Department of Linguistics, Lund University)* 21: 57-99.
- Gårding, E. and G. Bruce (1981). A Presentation of the Lund model for Swedish intonation. *Nordic Prosody II*. T. Fretheim. Trondheim, Tapir: 33-40.
- Gårding, E., J. Zhang, et al. (1983). "A generative model for tone and intonation in Standard Chinese based on data from one speaker." *Working Papers (Department of Linguistics, Lund University)* 25: 53-65.
- Godjevac, S. (2005). Transcribing Serbo-Croatian intonation. *Prosodic Typology: The Phonology of Intonation and Phrasing*. S.-A. Jun. New York, Oxford University Press: 146-171.
- Goldsmith, J. A. (1975a). *An Autosegmental Typology of Tone: and how Japanese fits in*. The Fifth Meeting of the North East Linguistics Society, Harvard University.
- Goldsmith, J. A. (1975b). *Tone Melodies and the Autosegment*. The Sixth Conference on African Linguistics, OSU WPL.
- Goldsmith, J. A. (1976). *Autosegmental Phonology*. Doctoral dissertation. Massachusetts Institute of Technology, Cambridge.
- Gordon, M. K. (2005). Intonational Phonology of Chicasaw. *Prosodic Typology: The Phonology of Intonation and Phrasing*. S.-A. Jun. New York, Oxford University Press: 301-330.
- Gu, W., K. Hirose, et al. (2006). "Modeling the Effects of Emphasis and Question on Fundamental Frequency Contours of Cantonese Utterances." *IEEE Transactions on Audio, Speech, and Language Processing* 14(4): 1155-1170.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*, Cambridge University Press.
- Haraguchi, S. (1977). *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo, Kaitakusha.

- Haraguchi, S. (1999). Accent. *The Handbook of Japanese Linguistics*. N. Tsujimura. Malden, Blackwell Publishers: 1-30.
- Hyman, L. (2001). Tone systems. *Language typology and language universals: An international Handbook*. M. Haspelmath, E. König, W. Oesterreicher and W. Raible. Berlin New York, Walter de Gruyter. 2.
- Jun, J., J. Kim, et al. (2006). "The Prosodic Structure and Pitch Accent of Northern Kyungsang Korean." *Journal of East Asian Linguistics* 15: 289-317.
- Jun, S.-A., Ed. (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. New York, Oxford University Press.
- Kim, N.-J. (1997). *Tone, Segments, and Their Interaction in North Kyungsang Korean: a correspondence theoretic account*. Doctoral dissertation. The Ohio State University, Columbus, OH.
- Kochanski, G. and C.-L. Shih (2000). Stem-ML: language-independent prosody description. *Proceedings of the International Conference on Spoken Language Processing 2000*. Beijing.
- Kochanski, G. and C.-L. Shih (2003). "Prosody modeling with soft templates." *Speech Communication* 39(3-4): 311-352.
- Kochanski, G., C.-L. Shih, et al. (2003). "Quantitative measurement of prosodic strength in Mandarin." *Speech Communication* 41(4): 625-645.
- Köhnlein, B. (2011). *Rule Reversal Revisited: synchrony and diachrony of tone and prosodic structure in the Franconian dialect of Arzbach*. Doctoral dissertation. Leiden University, Utrecht.
- Kori, S. (1987). "The tonal behavior of Osaka Japanese: An interim report." *Ohio State Working Papers in Linguistics: Papers from the Linguistics Laboratory* 36: 31-61.
- Kratochvil, P. (1968). *The Chinese Language today: Features of an Emerging Standard*. London, Hutchinson University Library.

- Ladd, D. R. (1996). *Intonational Phonology (First Edition)*. Cambridge, Cambridge University Press.
- Ladd, D. R. (2008). *Intonational Phonology (Second Edition)*. Cambridge, Cambridge University Press.
- Ladd, D. R., I. Mennen, et al. (2000). "Phonological conditioning of peak alignment in rising pitch accents in Dutch." *Journal of the Acoustical Society of America* 107: 2685-2696.
- Laniran, Y. O. (1992). *Intonation in tone languages: the phonetic implementation of tones in Yoruba*. Doctoral dissertation. Cornell University, Ithaca.
- Law, S.-P. (1990). *The Syntax and Phonology of Cantonese Sentence-Final Particles*. Doctoral dissertation. Boston University.
- Leben, W. (1973). *Suprasegmental phonology*. Doctoral dissertation. Massachusetts Institute of Technology, Cambridge.
- Lee, H.-S. (2008). *Pitch accent and its interaction with intonation: experimental studies of North Kyeongsang Korean*. Doctoral dissertation. Cornell University, Ithaca.
- Lee, O. J. (2005). *The Prosody of Questions in Beijing Mandarin*. Doctoral dissertation. The Ohio State University, Columbus, OH.
- Liberman, M. (1975). *The intonational system of English*. Doctoral dissertation. MIT, Cambridge.
- Liu, F. and Y. Xu (2005). "Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation." *Phonetica* 62: 70-87.
- Ma, J. K.-Y., V. Ciocca, et al. (2006). "Effect of intonation on Cantonese lexical tones." *Journal of the Acoustical Society of America* 120: 3978–3987.
- Maddieson, I. (1971). "Tone in generative phonology." *Research Notes (University of Ibadan)* 3.
- Martin, S. E. (1987). *The Japanese Language Through Time*. New Haven and London, Yale University Press.
- Matthews, S. and V. Yip (1994). *Cantonese: A Comprehensive Grammar*. London and New York, Routledge.

- McCawley, J. D. (1964). What is a tone language? *Paper presented at the Summer Meeting of the Linguistic Society of America.*
- McCawley, J. D. (1970). Some tonal systems that come close to being pitch-accent systems but don't quite make it. *Papers from the Sixth Regional Meeting, Chicago Linguistic Society:* 526-532.
- Myers, J. and J. Tsay (2003). "Investigating the Phonetics of Mandarin Tone Sandhi." *Taiwan Journal of Linguistics* 1(1): 29-68.
- Ohala, J. (1972). "How is pitch lowered?" *Journal of the Acoustical Society of America* 52(1A): 124-124.
- Ohala, J. J. (1983). "Cross-language use of pitch: An ethological view." *Phonetica* 40: 1-18.
- Ohala, J. J. (1984). "An ethological perspective on common cross-language utilization of F0 of voice." *Phonetica* 41: 1-16.
- Okuda, K. (1971). *Accentual Systems in Japanese Dialects*. Doctoral dissertation. UCLA, Los Angeles.
- Patin, C. (2008). Tone and Intonation's Waltz in Shingazidja Polar Questions. *3rd TIE Conference on Tone and Intonation (TIE3)*. Lisbonne.
- Peng, S.-h., M. K. M. Chan, et al. (2005). Towards a Pan-Mandarin System for Prosodic Transcription. *Prosodic Typology: The Phonology of Intonation and Phrasing*. S.-A. Jun. New York, Oxford University Press: 230-270a.
- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Doctoral dissertation. Massachusetts Institute of Technology, Cambridge, MA.
- Pierrehumbert, J. B. and M. E. Beckman (1988). *Japanese Tone Structure*. Cambridge, MA, MIT Press.
- Pike, K. L. (1948). *Tone languages: a technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion*. Ann Arbor, University of Michigan Press.



- Pittayaporn, P. (2005). Prosody of final particles in Thai: the interactions between lexical tones and boundary tones.
- Poser, W. (1984). *The Phonetics and Phonology of Tone and Intonation in Japanese*. Doctoral dissertation. Massachusetts Institute of Technology, Cambridge, MA.
- Prieto, P. and F. Torreira (2007). "The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish." *Journal of Phonetics* 35: 473-500.
- Rooth, M. (1992). "A Theory of Focus Interpretation." *Natural Language Semantics* 1: 75-116.
- Rowlands, E. C. (1959). *A Grammar of Gambian Mandinka*. London, School of Oriental and African Studies.
- Schrachter, P. and V. Fromkin (1968). "A phonology of Akan: Akuapem, Asante, and Fante." *Working Papers in Phonetics (UCLA)* 9.
- Shen, X. S. and M. Lin (1991). "A perceptual study of Mandarin tones 2 and 3." *Language and Speech* 34: 145-156.
- Shih, C.-L. (1987). *The Phonetics of the Chinese Tonal System*, AT&T Bell Labs.
- Shih, C.-L. and G. Kochanski (2000). Chinese tone modeling with Stem-ML. *International Conference on Spoken Language Processing*. Beijing.
- Silverman, K., M. E. Beckman, et al. (1992). *ToBI: a standard for labeling English prosody*. Proceedings, Second International Conference on Spoken Language Processing, Banff, Canada.
- Smiljanić, R. (2004). *Lexical, Pragmatic, and Positional Effects on Prosody in Two Dialects of Croatian and Serbian: An Acoustic Study*. New York, Routledge.
- Venditti, J. J., K. Maekawa, et al. (2008). Prominence marking in the Japanese intonation system. *The Oxford Handbook of Japanese Linguistics*. S. Miyagawa and M. Saito. New York, NY, Oxford University Press, Inc.: 456-512.
- Wang, W. S.-Y. (1967). "The phonological features of tone." *International Journal of American Linguistics* 33: 93-105.

- Welmers, W. E. (1962). "The phonology of Kpelle." *Journal of African Languages* 1: 69-93.
- Wightman, C. W. (2002). *ToBI or Not ToBI?* Proceedings of Speech Prosody, Aix-en-Provence, France.
- Williams, E. S. (1976). "Underlying Tone in Margi and Igbo." *Linguistic Inquiry* 7(3): 463-484.
- Wong, M. K.-S. (1982). *Tone Change in Cantonese*. Doctoral dissertation. University of Illinois at Urbana-Champaign, Urbana-Champaign.
- Wong, W. Y. P., M. K. M. Chan, et al. (2005). An Autosegmental-Metrical Analysis and Prosodic Annotation Conventions for Cantonese. *Prosodic Typology: The Phonology of Intonation and Phrasing*. S.-A. Jun. New York, Oxford University Press: 271-300.
- Woo, N. H. (1969). *Prosody and phonology*. Bloomington, Indiana Linguistics Club.
- Wu, K.-y. (1990). *A Linguistic Study of Interrogation in Cantonese: comparisons with English*. M.Phil. thesis. The University of Hong Kong, Hong Kong.
- Xie, G. (1974). "Yi yiqi yanjiu Guangdongren xuexi guoyu shi shengdiao fangmian suo zaoyu zhi kunnan. ('An instrumental study of the difficulties encountered by Cantonese people in learning Mandarin.')." *Jiaoyu yu Wenhua Yuekan* 422: 62-66.
- Xu, B. R. and P. Mok (2011). *Final rising and global raising in Cantonese intonation*. Proceedings of the International Congress of Phonetic Sciences, Hong Kong.
- Xu, Y. (2005). "Speech melody as articulatorily implemented communicative functions." *Speech Communication* 46: 220-251.
- Xu, Y. and X. Sun (2002). "Maximum speed of pitch change and how it may relate to speech." *Journal of the Acoustical Society of America* 111: 1399-1413.
- Xu, Y. and Q. E. Wang (2001). "Pitch targets and their realization: Evidence from Mandarin Chinese." *Speech Communication* 33: 319-337.
- Xu, Y. and C. X. Xu (2005). "Phonetic realization of focus in English declarative intonation." *Journal of Phonetics* 33: 159-197.
- Yang, R.-X. (2011). *The Phonation Factor in the Categorical Perception of Mandarin Tones*. ICPhS XVII, Hong Kong.

- Yip, M. (1980). *The Tonal Phonology of Chinese*. Doctoral dissertation. MIT, Cambridge, MA.
- Yip, M. (1989). "Contour Tones." *Phonology Yearbook* 6: 149-174.
- Yip, M. (1995). Tone in East Asian Languages. *The Handbook of Phonological Theory*. J. A. Goldsmith. Oxford, Basil Blackwell: 476-494.
- Yip, M. (2002). *Tone*. Cambridge, Cambridge University Press.
- Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling*. Doctoral dissertation. Cornell University, Ithaca, NY.
- Yuan, J., C.-L. Shih, et al. (2002). *Comparison of Declarative and Interrogative Intonation in Chinese*. *Speech Prosody* 2002.
- Zhang, Q., T. Chen, et al. (1993). *Henan fangyan yanjiu*. Kaifeng, Henan Daxue.
- Zsiga, E. and D. Zec (to appear). "Contextual Evidence for the Representation of Pitch Accents in Standard Serbian." *Language and Speech* 55.
- Zue, V. W. (1976). Some perceptual experiments on the Mandarin tones. *The 92nd Meeting of the Acoustical Society of America*. San Diego.