

Tianjin Mandarin

Tones and Tunes



Cong Zhang

St. Anne's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2018

Abstract

Lexical tones and intonational tunes are both mainly realised through pitch modulation. What role does intonation play in a language which has a lexical tonal contrast? Can one separate ‘tone’ from ‘intonation’? If yes, how do lexical tones interact with intonational tunes? In order to answer these questions, this thesis investigates how tone and intonation interact during production and perception in Tianjin Mandarin, by means of examining the components of different intonational tunes under the Autosegmental-Metrical (AM) Framework (Pierrehumbert, 1980), and the cues native listeners use during the tune identification process.

Chapter 1 – 3 are the introductory chapters: Chapter 1 introduces the topic of research, and sketches the three research goals for this thesis – the theoretical goal, the documentation goal, and the methodological goal; Chapter 2 addresses the theoretical foundation of this thesis – the AM theory; and Chapter 3 outlines the linguistic background of Tianjin Mandarin.

Chapter 4 presents production studies of the tune of intonational Yes/No questions (IntQ) in Tianjin Mandarin. A total of six native Tianjin speakers were recorded for monosyllabic words in isolation ($\text{Mono}_{(\text{ISO})}$) and monosyllabic words as sentence prominence ($\text{Mono}_{(\text{SEN})}$), with statement tune and IntQ tune, respectively. The results show that when a monosyllabic word is produced in isolation, the IntQ tune has a raised register, and a floating $H\%$ boundary tone at the end of the intonational phrase. When a monosyllabic word is in sentence prominence position, the IntQ tune also has a raised register, a floating $H\%$ boundary tone, as well as a H^* pitch accent coming from the focus and a post-focus compression. The IntQ tune is:

[H^* pitch accent + (post-focus compression) + floating $H\%$ boundary tone] higher register.

To further investigate how the IntQ tune is represented, three perception experiments were conducted on monosyllabic words in isolation, monosyllabic words as sentence prominence, and sentences with monosyllabic words as prominence in Chapter 5. A total of 28 native Tianjin Mandarin speakers participated in the experiments. They were asked to identify the tunes (yes-no question or statement) of the audio stimuli. The accuracy of their responses and reaction time together show that they strongly prefer the H-Rising lexical tone for IntQs, and L-Falling lexical tone for statements, which indicate that they look for the low register information during the identification of statements, and a H boundary tone for the IntQ tune.

Another important tune, chanted call (CC) tune, was also studied to further investigate the possibilities of intonational tunes in a tonal language in Chapter 6. Six native speakers' production of monosyllabic words and disyllabic words were recorded. The results show that there is a L% boundary tone at the end of the intonational phrase, regardless of the lexical tones. Different from the IntQ data, the L% boundary tone is phonetically manifested and overrode the lexical tone contours. A H* pitch accent was found to be associated with the H of each lexical tone. Lengthening was also found in the CC tune. The CC tune in Tianjin Mandarin can be represented as follows:

[[H*]_{sustained}]_{higher register} + L%.

Keywords: tone; tune; intonation; Tianjin Mandarin; Autosegmental-Metrical Theory

Acknowledgements

Completing this thesis has been a wonderful journey, in particular as it has involved many wonderful people along the way who have directly or indirectly assisted me or offered me helpful advice. I do owe a great debt of gratitude to every one of them.

First and foremost, I would like to thank my supervisor, Professor Aditi Lahiri, for her patience, guidance, and support. Her words and actions have always been the sources of my inspiration.

My sincere gratitude goes to Dr. Lei Liang and Dr. Hui Feng in China for their extremely generous support during my data collection trips to Tianjin along with the graduate assistants who helped me with recruiting participants for my research. I would also like to thank all my participants for their time and effort – your voices are now immortalised.

I wish to thank my examiners, Dr. Mary Baltazani and Dr. Yiya Chen, for their valuable comments on my thesis, which provoked me to further reflect on prosody research. I would also like to thank the following linguists, whose insights and advice had significant influence on my final thesis: Professor Amalia Arvaniti, Professor Annie Rialland, Professor Henning Reetz, Professor Larry Hyman, Professor Pilar Prieto, and Professor Paula Fikkert. I would also like to give my special thanks to the organisers and speakers of Aix Summer School on Prosody for the immensely informative program.

This thesis would not have been possible to complete without the support I received from China Scholarship Council, Faculty of Linguistics, Philology and

Phonetics, St. Anne's College, and Santander Travel Awards for funding my studies as well as the data collection trips.

I am indebted to many of my colleagues at the Language and Brain Lab, for their help – academically (teaching me statistics and proofreading my work), emotionally (sharing joy and pain), materially (with food, drink, and even shelter). Professor Allison Wetterlin, Dr. Beinan Zhou, Dr. Colin Brooks, Dr. Holly Kennard, Dr. Sandra Kotzor, Dr. Hilary Wynne, Dr. Adam Roberts, Dr. Stephen Politzer-Ahles, Dr-to-be Yoolim Kim, Dr-to-be Sherry Yong Chen, and the numerous other people at the university whose details I might have left out.

My deepest gratitude goes to Mr. Qingwen Wang and Ms. Xuemei Li for their altruistic support. You are my true mentors, who guided me through many difficult times. I would also like to thank my best friend Ms. Weiyao Zhang, for always being there for me since I can ever remember. My sweetest thanks go to my parents – my dad, Yuelai Zhang, and my mom, Yumei Liu. Without their boundless love and support, I would never have reached this far. I would like to thank my husband, Dr-to-be Azad Maudaressi, a true linguist at heart, for asking me all those great questions that eventually turned into my thesis (e.g. “How do you ask a question in a tonal language?” and “Why does your mom call you with an /o/ attached at the end?”). I hope this thesis gives you the answers, in great length. Thank you for going through this journey with me. And to Sultan our cutest doggie – thanks for all the furry hugs whenever I needed them the most.

This thesis marks the end of my student years. I dedicate this thesis to all those who have taught, educated, and mentored me throughout my life and my education. A part of who I am, is you.

Pour avoir découvert le monde à travers le langage, je pris longtemps le langage
pour le monde.

Les Mots, Jean-Paul Sartre

致吾师

To my teachers and mentors

Table of Contents

Abstract	i
Acknowledgements	iii
List of Figures	xiii
List of Tables	xviii
List of Abbreviations	xx
Chapter 1 Introduction	1
1.1 Issues in Discussion.....	1
1.2 Research Goals	2
1.3 Structure of this Thesis.....	3
Chapter 2 Theoretical Background: Autosegmental-Metrical Theory	6
2.1 Introduction	6
2.2 Models of Intonation Research.....	6
2.3 Autosegmental-Metrical Theory	10
2.3.1 History and Development of Autosegmental-Metrical Theory	10
2.3.1.1 Origin.....	10
2.3.1.2 Original AM Model	12
2.3.1.3 Developments of AM Theory	16
2.3.2 Key Terms and Concepts in AM Theory.....	18
2.3.2.1 Tune vs. Intonation	18
2.3.2.2 Prominence vs. Stress	19
2.3.2.3 Pitch Accent vs. Lexical Tone	20
2.3.2.4 Tonal Association, Tonal Alignment, Tonal Scaling	21
2.3.3 Tenets of AM Theory.....	22
2.3.3.1 Sequential tonal structure	23
2.3.3.2 Distinction between pitch accent and stress	23
2.3.3.3 Analysis of pitch accents in terms of level tones.....	25
2.3.3.4 Local sources for global trends.....	26
2.4 Summary	28

Chapter 3	Tianjin Mandarin	29
3.1	Introduction	29
3.2	Origin and Typology	30
3.3	Segmental Features.....	34
3.3.1	Syllable Structure and Sound Inventory	34
3.3.2	<i>Chiyin</i> (Dentalisation).....	36
3.3.3	<i>Erhua</i> (Rhotacisation).....	37
3.4	Prosodic Features	39
3.4.1	Description of Lexical Tones.....	39
3.4.2	Tonal Alternation Rules.....	43
3.4.3	Prominence	46
3.4.3.1	Lexical Prominence	46
3.4.3.2	Sentence Prominence.....	48
3.4.4	Prosodic Grouping	48
3.5	Choice of Tianjin Mandarin	51
Chapter 4	Production Studies: Intonational Yes/No Question Tune in Tianjin Mandarin	55
4.1	Introduction	55
4.2	Literature Review: Yes/No Question Tunes	57
4.2.1	YNQ Tunes in World Languages	57
4.2.2	Tunes in Mandarin Chinese	59
4.2.2.1	“Big Wave” – Global F0 Trend.....	60
4.2.2.2	“Small Ripple” – Local F0 Events.....	64
4.3	General Methodology.....	68
4.3.1	Speakers	68
4.3.2	Procedures.....	68
4.3.3	Data Processing and Analyses	69
4.4	Production Study 1: Monosyllabic Words in Isolation (Mono _(ISO))... 73	
4.4.1	Methods.....	74
4.4.1.1	Materials	74
4.4.1.2	Procedures and Informants	74
4.4.1.3	Statistical Analyses	75

4.4.2	Results.....	75
4.4.2.1	Duration	79
4.4.2.2	Register	81
4.4.2.3	Pitch Accent.....	87
4.4.2.4	Edge/ Boundary Tone	89
4.4.3	Discussion.....	93
4.4.3.1	Lengthening in IntQ Tune	93
4.4.3.2	Register and Pitch Accent.....	94
4.4.3.3	Floating Boundary Tone	94
4.4.4	Summary of Mono _(ISO) Results	103
4.5	Production Study 2: Monosyllabic Words in Sentence Prominence (Mono _(SEN)).....	105
4.5.1	Methods.....	105
4.5.1.1	Materials	105
4.5.1.2	Procedures and Informants	109
4.5.1.3	Statistical Analyses	109
4.5.2	Results.....	109
4.5.2.1	Duration	114
4.5.2.2	Register	117
4.5.2.3	Pitch Accent.....	124
4.5.2.4	Edge/ Boundary Tone	126
4.5.3	Discussion.....	127
4.5.3.1	Boundary-Induced Lengthening	127
4.5.3.2	Register Lift	128
4.5.3.3	Pitch Accents	129
4.5.3.4	Floating Boundary Tone	132
4.5.4	Summary of Mono _(SEN) Results.....	133
4.6	Conclusions for Production Studies of IntQ Tune in Tianjin Mandarin	134

Chapter 5	Perception Experiments: Intonational Yes/No Question Tune in Tianjin Mandarin	136
5.1	Introduction	136
5.2	Literature Review: Identification of Tunes	137

5.3	Methodology	139
5.3.1	Participants.....	139
5.3.2	Stimuli.....	140
5.3.3	Procedures.....	143
5.3.4	Data analysis	145
5.4	Experiment 1: Monosyllabic Word in Isolation (Mono _(ISO)).....	147
5.4.1	Results.....	147
5.4.1.1	Accuracy	147
5.4.1.2	Reaction Time.....	148
5.4.2	Discussion.....	149
5.5	Experiment 2: Monosyllabic Word in Sentence Prominence (Mono _(SEN)).....	153
5.5.1	Results.....	153
5.5.1.1	Accuracy	153
5.5.1.2	Reaction Time.....	154
5.5.1.3	Dissimilation Effect.....	155
5.5.2	Discussion.....	158
5.6	Experiment 3: Sentences with Monosyllabic words as prominence (SEN)	161
5.6.1	Results.....	161
5.6.1.1	Accuracy	161
5.6.1.2	Reaction Time.....	162
5.6.1.3	Dissimilation Effect.....	164
5.6.2	Discussion.....	166
5.7	Discussion and Conclusions	167
Chapter 6	Production Studies: Chanted-Call Tune in Tianjin Mandarin.	171
6.1	Introduction	171
6.2	Literature Review: Chanted Call Tunes in World Languages	172
6.3	General Methodology.....	181
6.3.1	Speakers	181
6.3.2	Recording Procedures	181
6.3.3	Data Selection	182

6.4	Production Study 3: Monosyllabic Chanted Calls	183
6.4.1	Methods.....	183
6.4.1.1	Materials	183
6.4.1.2	Procedures and Informants	184
6.4.1.3	Statistical Analyses	184
6.4.2	Results.....	185
6.4.2.1	Visual Inspection	185
6.4.2.2	Duration	188
6.4.2.3	Register	190
6.4.2.4	Pitch Accent.....	192
6.4.2.5	Edge/Boundary Tone	195
6.4.3	Discussion.....	196
6.4.3.1	Lengthening	196
6.4.3.2	Pitch Accent.....	197
6.4.3.3	Boundary Tone	197
6.4.4	Summary for Monosyllabic Chanted Calls.....	198
6.5	Production Study 4: Disyllabic Chanted Calls	198
6.5.1	Methods.....	199
6.5.1.1	Materials	199
6.5.1.2	Procedures and Informants	202
6.5.1.3	Statistical Analyses	202
6.5.2	Results.....	202
6.5.2.1	Visual Inspection	204
6.5.2.2	Duration	214
6.5.2.3	Mean Pitch	218
6.5.2.4	Tonal Scaling	221
6.5.2.5	Tonal Alignment.....	229
6.5.3	Discussion.....	238
6.5.3.1	Lengthening	238
6.5.3.2	Register	240
6.5.3.3	Pitch Accent.....	241
6.5.3.4	Boundary Tone	242
6.5.4	Summary for Disyllabic Chanted Calls	243
6.6	Conclusion.....	245

Chapter 7	Final Discussion and Conclusions	246
7.1	Summary of Key Findings	246
7.2	Further General Discussion	247
7.2.1	Application of AM Theory in a Tonal Language	247
7.2.2	Tone-Tune Interaction	253
7.2.3	Consistency of Terms and Notation in AM Framework.....	253
7.3	Future Research	254
7.4	Concluding Remarks	255
References		258
Appendices		280
	Appendix 1: Figures from Production Study 2: Monosyllabic Words in Sentence Prominence (Mono _(SEN))	280
	Appendix 2: Figures of Chanted Call Tunes in World Languages	284

List of Figures

Figure 2.1	Pitch accents in Pierrehumbert (1980)	14
Figure 2.2	Tunes of English in Pierrehumbert (1980).....	16
Figure 2.3	Hierarchy of tunes according to Ladd (2008, p. 285)	17
Figure 2.4	Ladd’s modified intonation model (Ladd, 1996, p. 217)	18
Figure 2.5	Tonal space between topline and baseline (Bruce & Gårding, 1978)	26
Figure 2.6	Two ways of analysing intonation contour (adapted from Ladd, 1996, p. 75)	27
Figure 3.1	Map of Chinese dialect groups	32
Figure 3.2	Tianjin Mandarin and Neighbouring Accents (adapted from Shiyu Li & Han, 1991)	34
Figure 3.3	Vowels in Tianjin Mandarin (Q. Li et al., 2017, pp. 6&8).....	36
Figure 3.4	Vowel change in rhotacisation	39
Figure 3.5	Lexical tones of (a) Tianjin Mandarin (Shi, 1986) and (b) Standard Mandarin (Chao, 1930).....	41
Figure 3.6	Values of Tianjin Mandarin lexical tones in (a) Zhang & Liu (2011) and (b) Li et al. (2017)	42
Figure 3.7	Values of Tianjin Mandarin lexical tones: Data from the current thesis	43
Figure 4.1	Mandarin sentence types	55
Figure 4.2	Tunes in Standard Mandarin (X. S. Shen, 1990).....	63
Figure 4.3	Neutral and narrow focus in Tianjin Mandarin: Normalised F0 (Duan & Jia, 2014)	67
Figure 4.4	Typical artefacts at the end of a statement (left) and an IntQ (right).	71
Figure 4.5	Creakiness of Tone 3 (LH Tone).....	72

Figure 4.6	Statement (left) and IntQ (right): Mono _(ISO) – Tone 1 (L Tone)	77
Figure 4.7	Statement (left) and IntQ (right): Mono _(ISO) – Tone 2 (H Tone)	77
Figure 4.8	Statement (left) and IntQ (right): Mono _(ISO) – Tone 3 (LH Tone)	78
Figure 4.9	Statement (left) and IntQ (right): Mono _(ISO) – Tone 4 (HL Tone)	78
Figure 4.10	Duration: Mono _(ISO) onsets	80
Figure 4.11	Duration: Mono _(ISO) rhymes	80
Figure 4.12	Mean Pitch: Mono _(ISO) onsets	82
Figure 4.13	Mean Pitch: Mono _(ISO) rhymes	82
Figure 4.14	Stylised lines: Mono _(ISO) onsets and Mono _(ISO) rhymes	83
Figure 4.15	Measurements of F0 maximum and minimum in a rhyme	84
Figure 4.16	Average maximum pitch: Mono _(ISO) rhymes	86
Figure 4.17	Average minimum pitch: Mono _(ISO) rhymes	86
Figure 4.18	Further Stylised lines: Mono _(ISO) rhymes	89
Figure 4.19	F0 Range: Mono _(ISO) onsets	92
Figure 4.20	F0 Range: Mono _(ISO) rhymes	92
Figure 4.21	Bengali offering tune (Hayes & Lahiri, 1991)	96
Figure 4.22	Greek declarative tune (Arvaniti & Baltazani, 2005)	96
Figure 4.23	Cantonese questions and statements (J. K.-Y. Ma et al., 2006)	97
Figure 4.24	Boundary tones from Lin's (2004) study on Standard Mandarin	98
Figure 4.25	Thai tones and tunes (Luksaneeyanawin, 1998, p. 383)	102
Figure 4.26	Stylised tracings of F0 in statements (left) and questions (right) in Danish (Grønnum, 1998, p. 143)	103
Figure 4.27	Pre-nucleus, nucleus, and post-nucleus	110
Figure 4.28	Statement (left) and IntQ (right): Mono _(SEN) – Tone 1 (L Tone)	112
Figure 4.29	Statement (left) and IntQ (right): Mono _(SEN) – Tone 2 (H Tone)	112

Figure 4.30	Statement (left) and IntQ (right): Mono _(SEN) – Tone 3 (LH Tone)...	113
Figure 4.31	Statement (left) and IntQ (right): Mono _(SEN) – Tone 4 (HL Tone)...	113
Figure 4.32	Duration: Mono _(SEN) nuclei.....	116
Figure 4.33	Duration: Mono _(SEN) post-nuclei.....	116
Figure 4.34	Mean Pitch: Mono _(SEN) pre-nuclei	118
Figure 4.35	Mean Pitch: Mono _(SEN) nuclei	120
Figure 4.36	Mean Pitch: Mono _(SEN) post-nuclei	120
Figure 4.37	F0 Range: Mono _(SEN) post-nuclei	126
Figure 4.38	Schematised lines for register differences in Mono _(SEN)	128
Figure 4.39	Schematised lines for pitch accent differences in Mono _(SEN)	129
Figure 4.40	Statement (left) and IntQ (right) with neutral and narrow foci (Amended from Liu & Xu, 2005, p. 75).....	130
Figure 5.1	Examples of stimuli.....	141
Figure 5.2	Procedures of the Perception Experiments.....	145
Figure 5.3	Accuracy of responses by rhymes of stimuli.....	146
Figure 5.4	Mono _(ISO) : Accuracy.....	152
Figure 5.5	Mono _(ISO) : Reaction Time.....	152
Figure 5.6	Dissimilation Effect – Mono _(SEN) : Accuracy	157
Figure 5.7	Dissimilation Effect – Mono _(SEN) : Reaction Time.....	157
Figure 5.8	Mono _(SEN) : Accuracy	160
Figure 5.9	Mono _(SEN) : Reaction Time	160
Figure 5.10	SEN: Accuracy	163
Figure 5.11	SEN: Reaction Time.....	163
Figure 5.12	Dissimilation Effect – SEN: Accuracy.....	165
Figure 5.13	Dissimilation Effect – SEN: Reaction Time	165

Figure 6.1	Statement (left) and CC (right): Monosyllabic – Tone 1 (L Tone) ..	186
Figure 6.2	Statement (left) and CC (right): Monosyllabic – Tone 2 (H Tone)..	186
Figure 6.3	Statement (left) and CC (right): Monosyllabic – Tone 3 (LH Tone)	187
Figure 6.4	Statement (left) and CC (right): Monosyllabic – Tone 4 (HL Tone)	187
Figure 6.5	Duration: Monosyllabic onsets.....	189
Figure 6.6	Duration: Monosyllabic rhymes.....	189
Figure 6.7	Mean Pitch: Monosyllabic onsets.....	191
Figure 6.8	Mean Pitch: Monosyllabic rhymes.....	191
Figure 6.9	F0 Maxima Alignment in CC tune and Statement Tune	193
Figure 6.10	Alignment of Maximum F0 in CC Tune and Statement Tune	194
Figure 6.11	Alignment of Minimum F0 in CC Tune and Statement Tune.....	195
Figure 6.12	Statement (left) and CC (right): Disyllabic (Set1) – T1T1	207
Figure 6.13	Statement (left) and CC (right): Disyllabic (Set1) – T1T2	207
Figure 6.14	Statement (left) and CC (right): Disyllabic (Set1) – T1T3	207
Figure 6.15	Statement (left) and CC (right): Disyllabic (Set1) – T1T4	207
Figure 6.16	Statement (left) and CC (right): Disyllabic (Set1) – T2T1	208
Figure 6.17	Statement (left) and CC (right): Disyllabic (Set1) – T2T2	208
Figure 6.18	Statement (left) and CC (right): Disyllabic (Set1) – T2T3	208
Figure 6.19	Statement (left) and CC (right): Disyllabic (Set1) – T2T4	208
Figure 6.20	Statement (left) and CC (right): Disyllabic (Set1) – T3T1	209
Figure 6.21	Statement (left) and CC (right): Disyllabic (Set1) – T3T2	209
Figure 6.22	Statement (left) and CC (right): Disyllabic (Set1) – T3T3	209
Figure 6.23	Statement (left) and CC (right): Disyllabic (Set1) – T3T4	209
Figure 6.24	Statement (left) and CC (right): Disyllabic (Set1) – T4T1	210
Figure 6.25	Statement (left) and CC (right): Disyllabic (Set1) – T4T2	210

Figure 6.26	Statement (left) and CC (right): Disyllabic (Set1) – T4T3	210
Figure 6.27	Statement (left) and CC (right): Disyllabic (Set1) – T4T4	210
Figure 6.28	Statement (left) and CC (right): Disyllabic (Set2) – T1T0	212
Figure 6.29	Statement (left) and CC (right): Disyllabic (Set2) – T2T0	212
Figure 6.30	Statement (left) and CC (right): Disyllabic (Set2) – T3T0	213
Figure 6.31	Statement (left) and CC (right): Disyllabic (Set2) – T4T0	213
Figure 6.32	Duration: First syllable (Set 1)	216
Figure 6.33	Duration: Second syllable (Set 1).....	216
Figure 6.34	Duration: First syllable (Set 2)	217
Figure 6.35	Duration: Second syllable – T0 (Set 2)	217
Figure 6.36	Mean F0: Second syllable – T0 (Set 2)	220
Figure 6.37	Set 1: Alignment of Maximum F0 in the First Syllables	231
Figure 6.38	Set 1: Alignment of Maximum F0 in the Second Syllables	231
Figure 6.39	Set 1: Alignment of Maximum F0 (First Syllables of T1Tx).....	232
Figure 6.40	Set 2: Alignment of Maximum F0 in the First Syllables	234
Figure 6.41	Set 2: Alignment of Maximum F0 in the Second Syllables (T0)....	234
Figure 6.42	Set 1: Alignment of Minimum F0 in the Second Syllables.....	237
Figure 6.43	Set 2: Alignment of Minimum F0 in the Second Syllables (T0).....	237

List of Tables

Table 2.1	Models for intonation research (Arvaniti, 2011; Y. Xu, 2015).....	8
Table 2.2	Labels for pitch accents under AM framework.....	12
Table 3.1	Consonants in Tianjin Mandarin (Q. Li et al., 2017, p. 2)	35
Table 3.2	Dentalisation in older version of Tianjin Mandarin (based on X. Zhang, 1987)	37
Table 3.3	Values of Tianjin Mandarin lexical tones (Ma, 2012).....	41
Table 4.1	Markers of YNQ (Rialland, 2007, pp. 37–38)	58
Table 4.2	Mono _(ISO) words.....	74
Table 4.3	Order of presentation of results: Mono _(ISO)	76
Table 4.4	Alignment of F0 Maxima & Minima: Mono _(ISO)	88
Table 4.5	Order of presentation of results: Mono _(SEN)	111
Table 4.6	Mean pitch of Mono _(ISO) and Mono _(SEN) onset.....	121
Table 4.7	Average F0 extrema of Mono _(SEN) rhymes	124
Table 4.8	Alignment of F0 Maxima & Minima: Mono _(SEN)	124
Table 5.1	Stimuli: Mono _(ISO) and Mono _(SEN)	140
Table 5.2	Results of Mono _(ISO)	149
Table 5.3	Results of Mono _(SEN)	158
Table 5.4	Results of SEN	167
Table 6.1	Lexical Tones in Thai (adapted from Luksaneeyanawin, 1998)	176
Table 6.2	CC Tunes in World Languages	180
Table 6.3	Monosyllabic Calls.....	184
Table 6.4	F0 Scaling: Monosyllabic CC (Unit: ERB).....	192
Table 6.5	Disyllabic Calls (Set 1).....	201
Table 6.6	Disyllabic Calls (Set 2).....	201

Table 6.7	Tone Name vs Tone Height.....	203
Table 6.8	Mean Pitch of Disyllabic Words: Set1 (Unit: ERB)	219
Table 6.9	Mean Pitch of Disyllabic Words: Set2 (Unit: ERB)	221
Table 6.10	Average Maximum F0 Values (Set 1).....	226
Table 6.11	Average Minimum F0 Values (Set 1)	227
Table 6.12	Average Maximum F0 Values (Set 2).....	228
Table 6.13	Average Minimum F0 Values (Set 2)	228
Table 6.14	Summary of Key Results.....	244

List of Abbreviations

AM	Autosegmental-Metrical
CC	Chanted Call
C	Chanted call tune
H	High
IntQ	Intonational Yes/No question
L	Low
Ma-Q	Yes/No question with particle “ma”
Mono _(ISO)	Monosyllabic words in isolation
Mono _(SEN)	Monosyllabic words as sentence prominence
S	Statement tune
SEN	Full sentences with monosyllabic words as prominence
Syll.1	First syllable
Syll.2	Second syllable
T0	Neutral tone; Toneless tone
T1	Tone 1: L
T2	Tone 2: H
T3	Tone 3: LH
T4	Tone 4: HL
YNQ	Yes/No question

Chapter 1 Introduction

1.1 Issues in Discussion

Prosody is a universal linguistic feature that all languages share. Cross-linguistically, different languages employ different means of prosody to convey meanings, among which pitch modulation is the most common method. Tone, intonation, stress, and many other prosodic features are realised by way of the pitches, be it completely or partially. This thesis investigates the interaction between lexical tone and intonation: two prosodic features that both heavily rely on pitch in the tonal language Tianjin Mandarin. In order to capture the phonologically meaningful components of the tunes, this thesis works under the Autosegmental-Metrical framework. Two tunes are studied in this thesis – the tune of intonational Yes/No questions, and the tune of chanted calls, both of which are traditionally considered as being related with a H in many languages: the tune of intonational Yes/No questions are often generalised as having a H% boundary tone cross-linguistically, while the chanted calls in many languages have been observed to have a H* pitch accent. This thesis also investigates whether in Tianjin Mandarin the two tunes are perceived in the same manner as in the languages that have been studied and reported. The general research questions for this thesis are: What are tunes composed of in a tonal language? Are these components also the cues for tune perception? Specific research questions for each chapter will be presented in the relevant chapters.

1.2 Research Goals

This thesis is motivated by the following three research goals: the theoretical goal, the documentation goal, and the methodology goal. These three goals are further elaborated as follows:

First of all, from the perspective of prosodic phonology, this thesis aims to contribute to the Autosegmental-Metrical framework with analyses from a contour tone language. The findings of the current thesis will serve as the basis for the evaluation of whether Autosegmental-Metrical theory is applicable in a tonal language, with appropriate and sufficient descriptors for their intonation. The following questions are examined in order to reach this goal: since a tonal language like Tianjin Mandarin has already exploited the element of pitch on the lexical level, can there be any additional usage of pitch in the same way as in intonational languages? On the one hand, can the components of intonational tune which are proposed by the Autosegmental-Metrical Theory, be applied to Tianjin Mandarin? On the other hand, are those components sufficient in describing the intonation in a tonal language? To be specific, are there any pitch accents or boundary tones in Tianjin Mandarin tunes? Are there any other features that need to be included in the Autosegmental-Metrical theory, such as register changes or temporal features?

The second goal of this research is from the perspective of language documentation. Tianjin Mandarin is going through a rapidly changing stage in history, due to the influences from Standard Mandarin. Documentation work is therefore urgently needed. Additionally, most studies in Tianjin Mandarin prosody are on the lexical-level prosody, such as tonal alteration rules. This thesis makes the first attempt to describe Tianjin

Mandarin intonational tunes, which will in turn help with the description and documentation of this dialect. The following questions are to be answered – does Tianjin Mandarin have a rising intonation for the tune of intonational Yes/No questions? Does the intonational tune override the contours of the lexical tone in Tianjin Mandarin? How well can native speakers perceive these tunes?

Last but not least, from the perspective of research methodology, this thesis aims to integrate phonetic details with phonological analyses. It is often difficult to balance between the abstractness of phonology and the concreteness of phonetics. In this thesis, efforts are made to gain support from phonetic data for phonological analyses; meanwhile, all phonetic data are extracted based on the need of phonological analyses, in prevention of being trapped by excessive surface data which steer the analyses away from the true linguistics purpose.

1.3 Structure of this Thesis

Four production studies and three perception experiments are presented in this thesis in order to realise the goals presented in the previous section. This thesis is structured as follows:

Chapter 2 introduces the basics of the theoretical framework: Autosegmental-Metrical Theory. It starts with a summary of the existing models of intonation research and the reason for the usage of the Autosegmental-Metrical framework in the current thesis. It continues with introducing the theory, including its history and development, its key terms, and the four tenets of the Autosegmental-Metrical theory.

Chapter 3 presents the background information about Tianjin Mandarin, including the origin and typology, the segmental features, and the prosodic features of Tianjin Mandarin. It also outlines the reasons for the choice of Tianjin Mandarin in this thesis.

Chapter 4 investigates the intonational Yes/No question tune in Tianjin Mandarin, by means of two production studies. The first one studies the tune in monosyllabic utterances when they are produced in isolation. The second study investigates the same tune in longer sentences, which have the previous monosyllabic words in the sentence prominence position. Various phonetic properties are investigated, including the duration, register, F0 range, tonal scaling, tonal alignment, in order to support the phonological analyses of register change, pitch accents, and boundary tones. This chapter highlights the analyses of floating boundary tone, and the components of sentence prosody.

Chapter 5 examines the identification of the intonational Yes/No question tune, in order to investigate how native Tianjin Mandarin speakers perceive the tune. This chapter includes three experiments: one experiment uses monosyllabic words in isolation as stimuli, another uses sentences with these monosyllabic words in their prominence position as stimuli, and the third uses the monosyllabic words which are extracted from the sentences as stimuli. The results from this chapter corroborate with the analyses from the previous chapter: speakers make use of register and boundary information in both production and perception.

Chapter 6 explores the tune of chanted call in Tianjin Mandarin. Again, two studies were conducted: one that uses monosyllabic names and kinship terms as the speech

materials, and another one that uses disyllabic names. The possibility of pitch accents and boundary tones are examined through visual inspection and phonetic data. This chapter features the distinction between boundary-induced lengthening and sustainment in tune, the analysis of an actual boundary tone (c.f. floating boundary tone), as well as the association of the pitch accent.

Chapter 7 summarises the key findings from the previous chapters and presents the final general discussion. How the research goals of this thesis are met are sketched in the final chapter. Finally, the thesis ends with the suggestions for future research.

Chapter 2 Theoretical Background:

Autosegmental-Metrical Theory

2.1 Introduction

This chapter aims to present a brief introduction to the theoretical framework, Autosegmental-Metrical (hereinafter **AM**) Theory, used in this thesis. The chapter is organised as follows: In Section 2.2, different existing models are briefly listed in order to elaborate on how AM Theory fits into the bigger picture of intonation research. Section 2.3 introduces various aspects of AM Theory, including its history and development in §2.3.1, the key terms and concepts in §2.3.2, and the four basic tenets of AM Theory in §2.3.3. Finally, Section 2.4 presents a brief summary of all the sections in this chapter.

2.2 Models of Intonation Research

Prosody researchers have proposed various models of different natures for studying intonation. Ladd and Cutler (1983) divide these models into two broad categories – “concrete” and “abstract”. The “concrete” models pay more attention to phonetic details in the sense they regard the form-function relationship as a direct mapping between the physical forms (i.e. the acoustic parameters such as pitch, intensity, and duration) and functions of prosody; the “abstract” models, however, address more structural phonological issues and treat detailed acoustic details as “cues to abstract formal categories” (Ladd & Cutler, 1983, p. 2). Arvaniti (2011) and Xu (2015) reviewed these models and categorised them according to different criteria: Arvaniti’s (2011) categorisation was based on how the intonational tunes are formed, while Xu (2015) contrasted different properties of different models. A summary of their categorisations

is presented in Table 2.1, which provides the readers with a clearer picture of the field of intonation research, as well as where AM Theory, the framework this thesis adopts, stands in the field.

In Arvaniti (2011), the division was drawn firstly in terms of using dynamic tones such as rises and falls as primitives (i.e. “configurational”) and using level tones such as highs and lows as primitives (“pitch level as primitives”); then, the configurational models were further separated according to whether they make use of the contours as a whole (“melodies as gestalts”) or they different layers of tones (“superpositional”).

Similarly, Xu (2015) also compared the intonation models in terms of whether they are linear or superpositional. The linear models can use dynamic tones, level tones, or even gestalts as primitives, as long as they do not treat intonations as “decomposable” “layers”. Another division was between models of “acoustic” and “articulatory” natures, which differ from the mechanisms that the models were based on – as their names suggest, some models captures the acoustic features, while some others simulates the articulatory pitch targets. Xu (2015) also compared the “formal” models and “functional” models. While the “formal” models consider the forms, or F0 shapes, as their defining properties, the “functional” models treat functions and meanings as the defining properties of their prosodic units.

Table 2.1 Models for intonation research (Arvaniti, 2011; Y. Xu, 2015)

<p>Arvaniti (2011)</p>	<p><u>Configurational:</u></p> <ul style="list-style-type: none"> - Melodies as gestalts: Bolinger (1951); Jones (1972); Cooper and Sorensen (1981); INTSINT (International Transcription System for Intonation): Hirst and Di Cristo (1998); OXIGEN (Oxford Intonation Generator): Grabe et al. (2003); PENTA (Parallel Encoding and Pitch Target Approximation): Xu (2005). - Superpositional models: The IPO system; Fujisaki (1983,2004); Gårding (1983,1987); and Thorsen (1980, 1985, 1986). <p><u>Pitch level as primitives:</u></p> <ul style="list-style-type: none"> - Early level-based models: American structuralists (Pike 1945; Trager and Smith 1951; Hockett 1955; Trager 1961) - The AM model (Pierrehumbert, 1980)
<p>Xu (2015)</p>	<p><u>Linear vs. Superpositional:</u></p> <ul style="list-style-type: none"> - Linear: British nuclear tone tradition (Crystal, 1969; O’Connor & Arnold, 1961; Palmer, 1922; Wells, 2006); AM Theory (Pierrehumbert, 1980); IPO (‘t Hart et al., 1990). - Superpositional: Bailly & Holms, 2005; Fujisaki, 1983; Thorsen, 1980; van Santen et al., 2005 <p><u>Acoustic vs. Articulatory:</u></p> <ul style="list-style-type: none"> - Acoustic: Andruski, 2004; Gandour, Tumtavitikul & Satharnnuwong, 1999; Grabe, Kochanski & Coleman, 2007; Hirst, 2005; Liu, Surendran & Xu, 2006 - Articulatory: Fujisaki, 1983; the Stem-ML (Kochanski & Shih, 2003); PENTA (Xu, 2005). <p><u>Formal vs. Functional:</u></p> <ul style="list-style-type: none"> - Formal: AM Theory (Pierrehumbert, 1980), the IPO model (‘t Hart et al., 1990) and the tilt model (Taylor, 2000). - Functional: Superposition of Functional Contours (SFC) model (Bailly & Holm, 2005).

One of the most widely used phonetic models for studying Mandarin intonation is PENTA (Y. Xu, 2005). Xu (2015) defined his own model as a “quasi-superpositional, articulatory and functional model”. It works well for synthesising new intonational contours since this process fits the “articulatory” and “functional” goals; however, phonologically speaking, this model does not provide any abstract representations. In particular, the syllable-sized unit is over-detailed for a phonological analysis of intonational tunes, since lexical tones are lexically specified and should not be taken into account during intonational analyses (see a heated-debate with evidence from non-tonal languages: Arvaniti & Ladd, 2009, 2015; Y. Xu, Lee, Prom-on, & Liu, 2015). Chao’s (1930, 1968) analogy of “small ripples riding on big waves” is also often quoted in Mandarin intonational research. His approach is a superpositional approach in that he considers the global trend of the intonation as well as the local perturbations. Other popular models of tone and intonation include Stem-ML (Kochanski & Shih, 2003), and StaticTarget model (see Y. Chen & Gussenhoven, 2008, pp. 726–728 for a detailed review).

The Autosegmental-Metrical model (Pierrehumbert, 1980) is currently the most widely adopted phonological model for intonation research. It captures phonologically meaningful events such as pitch accents and edge tones. It makes describing intonation patterns possible without having to rely on excessive phonetic details (Gussenhoven, 2004, p. 123). Moreover, it offers a framework that all languages can use, regardless of whether they are tone languages, pitch accent languages, or stress languages, to describe their intonational phonologies, as well as being able to make cross-linguistic comparisons. The next section is dedicated to describing AM Theory in terms of its origin, components, and four tenets.

2.3 Autosegmental-Metrical Theory

2.3.1 History and Development of Autosegmental-Metrical Theory

2.3.1.1 Origin

As its name suggests, AM Theory was originated from Autosegmental Phonology and Metrical Phonology. Autosegmental Theory was proposed by Goldsmith (1976), based on the concept that tones belong to a separate tier from consonants and vowels by Williams (1976) and Leben (1973). It establishes that tones are autonomous from segments and are linked with their tone-bearing units by “association lines” (Goldsmith, 1976, p. 45), as shown in (1). Since tones and segments are represented on different tiers, tonal changes and segmental changes do not necessarily affect each other.

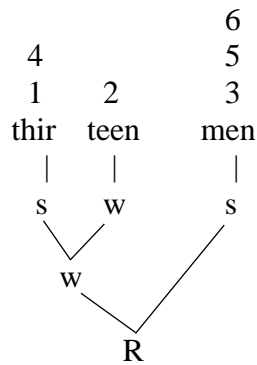
(1)

ma	<i>segmental tier</i>
	<i>association line</i>
HL	<i>tonal tier</i>

Following this tradition, AM Theory also regards the tones which make up intonational tunes as another tier from the texts, and addresses the issue of tune-text association.

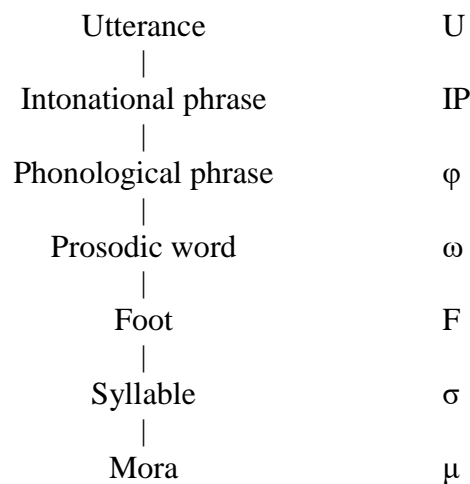
Metrical Phonology (Lieberman, 1975; Lieberman & Prince, 1977) provides a theory for organising segments into groups of prosodic prominence, as shown in (2). In AM Theory, prominence is where the tones associate with the text.

(2) Liberman & Prince (1977, p. 316)



Metrical Phonology is also the basis for Prosodic Hierarchy (Nespor & Vogel, 1986; Selkirk, 1972, 1980), as in (3). Prosodic hierarchy decides where the boundaries of different prosodic levels are. In AM Theory, the phrase accents and boundary tones are associated with these prosodic levels. There are also some language-specific prosodic levels such as the intermediate phrase in Japanese (Beckman & Pierrehumbert, 1986) and accentual phrase in Korean (Jun, 1993) and French (Jun & Fougeron, 2000).

(3) Prosodic hierarchy



AM Theory is widely adopted in that: (a) it makes describing intonation patterns possible without having to rely on phonetic details (Gussenhoven, 2004, p. 123); (b) it offers a framework that all languages can use to analyse their intonational phonologies,

regardless of whether they are tonal languages, pitch accent languages, or stress languages; these analyses are thus able to be compared cross-linguistically.

2.3.1.2 Original AM Model

Pierrehumbert (1980) was the first to integrate Autosegmental Theory and Metrical Theory and apply them in intonation research. She proposed three components for the intonational tune: pitch accent, phrase accents, and boundary tone. In this section, I briefly introduce the components in the original AM model.

All pitch accents in the original AM model consist of a H(igh) or a L(ow) tone, or one of the bitonal combinations of both H and L. A tone with an asterisk, or the starred tone (H* or L*), indicates that this tone is a central tone, which is associated with the sentence prominence. A preceding “leading” tone or a following “trailing” tone is written with a raised hyphen, i.e. H^ˉ or L^ˉ. Boundary tones, written as H% or L%, are single tones, which are associated with the end of intonational phrases. Table 2.2 summarises the symbols for single tones.

Table 2.2 Labels for pitch accents under AM framework

SYMBOL	MEANING	EXAMPLES
ASTERISK, T*	The central tone in pitch accents	H*, L*
RAISED HYPHEN, T^ˉ	Phrase accents	H ^ˉ , L ^ˉ
PERCENTAGE SIGN, T%	Boundary tones	H%, L%

Pitch accents

Pierrehumbert (1980) specifies seven types of pitch accents, namely H^* , L^* , H^*+L^- , H^-+L^* , L^*+H^- , L^-+H^* , H^*+H^- . A verbal description of these accents will be given below and, to further illustrate the features of these pitch accents, a graphic description will be subsequently presented in Figure 2.1:

- 1) H^* refers to a local peak aligned with an accented syllable;
- 2) L^* refers to a local valley aligned with an accented syllable;
- 3) H^*+L^- is a falling contour, which remains high for most of the accented syllable and then has a sharp fall into the following unaccented syllable;
- 4) L^*+H^- is similar to No. 3), referring to a rising contour, which remains low for most of the accented syllable and then has a sharp rise into the following unaccented syllable;
- 5) H^-+L^* shows a falling contour, in which the preceding unaccented syllable is higher than the tone on the accented syllable;
- 6) L^-+H^* reflects a rising contour, in which the preceding unaccented syllable is lower than the tone on the accented syllable.
- 7) H^*+H^- is a high-pitch contour, in which the accented syllable is high, followed by a relatively high-pitch transition, going into a following high accent. However, this pitch accent was eliminated in Beckman and Pierrehumbert (1986).

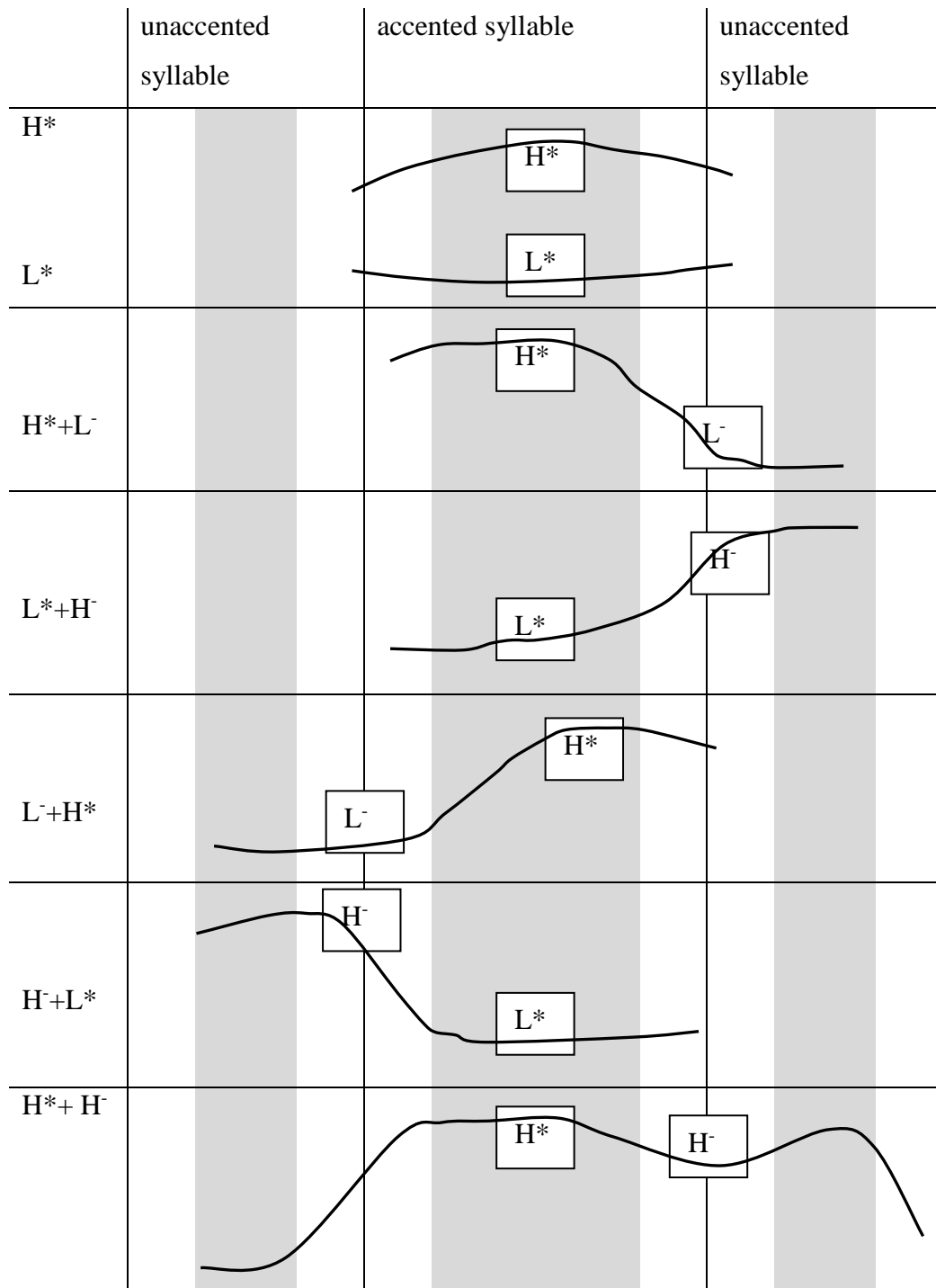


Figure 2.1 Pitch accents in Pierrehumbert (1980)

Edge tones

Edge tones in the original version of AM Theory (Pierrehumbert, 1980) consist of two types: boundary tones (T%) and phrase accents (T⁻). The definition of boundary tones is the same in the original (Pierrehumbert, 1980) and the revised version (Beckman and

Pierrehumbert, 1986) of the AM Theory – single tones which are associated with the edges (not the actual syllable) of intonational phrases, while phrase accent underwent some changes.

Originally, phrase accent was to describe the leading and trailing tones of the accented word. In Hayes and Lahiri's (1991) words, phrase accents are “extra tones found between the rightmost pitch accent and the final boundary tone”. In Beckman and Pierrehumbert's (1986) Tokyo Japanese intonation research, however, they discovered “accentual phrase” in Japanese, a prosodic unit based on a single pitch-accent, which cannot be found in English. They, therefore, proposed a revision of the original AM Theory, and organised accentual phrases into a higher-level unit, “intermediate phrase (ip)”. All intermediate phrases together form an intonational phrase (IP). Hayes and Lahiri (1991), however, suggested that intermediate phrases were actually phonological phrases in Prosodic Hierarchy (Nespor & Vogel, 1986; Selkirk, 1980).

Combining the pitch accents and the edge tones, the original AM model was illustrated as a finite grammar system as in Figure 2.2.

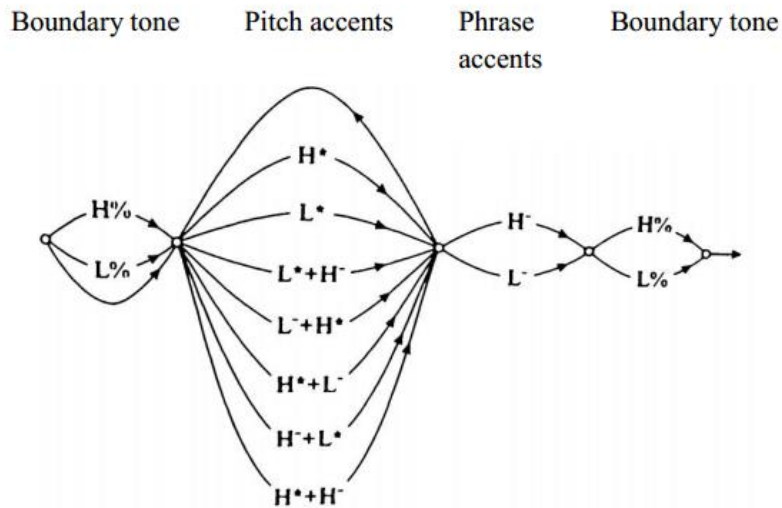


Figure 2.2 Tunes of English in Pierrehumbert (1980)

2.3.1.3 Developments of AM Theory

Many linguists have later modified or developed AM Theory, including Ladd (1996, 2008), Gussenhoven (2004), etc. In this section, I will briefly introduce Ladd's (1996, 2008) proposal. His division of “pre-nuclear”, “nuclear”, and “post-nuclear” accents is adopted in the current thesis.

Ladd (2008, pp. 282–283) raised the problem that the same tune could be represented with different tonal sequences depending on the lengths of the utterances. (4) below is an example of such problem (Ladd, 2008, p. 283): “H*...H*...L...L%” in (a) and “H*... H*...H*... L...L%” in (b) can both be declarative tunes. The difference is due to the extra syllable in the first utterance. To resolve the issue, he modified AM Theory with ideas from the traditional British school (see Palmer, 1922).

(4)

(a) H* H* LL%
I read it to Julia.

(b) H* H* H* LL%
I wanted to read it to Julia.

Ladd's new proposal mainly has the following three features: Firstly, Ladd integrated pre-nuclear accents into AM Theory. A pre-nuclear accent can be one or more accents which are identical to each other, but pre-nuclear accents are optional. Secondly, post-nuclear accents are obligatory, but the surface form can be either a boundary tone or a phrase accent. Thirdly, the structure of a tune is not linear but hierarchical, as illustrated in Figure 2.3.

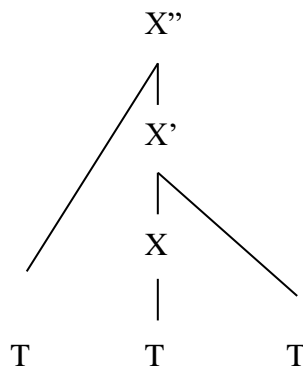


Figure 2.3 Hierarchy of tunes according to Ladd (2008, p. 285)

“X” represents the nuclear accent of an utterance – when an utterance only consists of one syllable, the nuclear accent is the only pitch accent. X’ is the post-nuclear accent which is almost always present – though it is possible to surface as edge tones in some language. X’’ is a pre-nuclear accent, which can be completely absent. Ladd’s version of the finite tune grammar is as shown in Figure 2.4. The pre-nuclear accent “H* in (4) is therefore iterable.

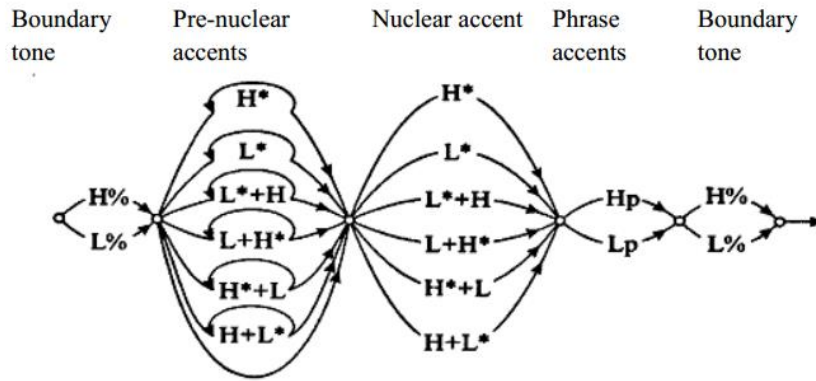


Figure 2.4 Ladd's modified intonation model (Ladd, 1996, p. 217)

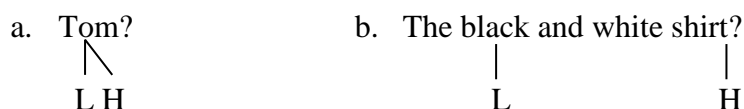
2.3.2 Key Terms and Concepts in AM Theory

2.3.2.1 Tune vs. Intonation

“Tune” and “intonation” both refer to sentence-level prosody. They are the two sides of the same “sentence-level prosody” coin – the tune is the abstract side, while intonation is the concrete side.

“**Intonation**” is a more general term for sentence level prosody. It views sentence-level prosody as a continuum that includes not only F0 properties, but also the duration, phonation, and all the other relevant sentence-level prosodic properties. However, a “**tune**” is the abstract representation of sentence-level prosody. As defined in Hayes and Lahiri (1991, p. 48), tunes are the “formal entities separate from the linguistic text” which “reside in an ‘intonational lexicon’ and convey (often highly elusive) intonational meanings”. For example, in the English examples below, the tune for a Yes/No question in (5) is LH, while the HLH tune in (6) represents a surprise tune:

(5)



(6)



In (5)a and (5)b, the Yes/No question tune remains the same regardless of the length of the utterances. In (6)a and (6)b, although the association methods for different words are different, the tune does not change.

In AM Theory, the term “tune” refers to an iterable sequence of tones, which is a phonologically significant event based on one or more pitch accents and edge tones. Tune is iterable in that

Pierrehumbert (1980) produced a finite-state grammar to generate tunes of English intonation, which is a linear H-L tonal sequence composed of pitch accents, phrase accents, and boundary tones. Section 2.3.3 will further elaborate on the principles that AM Theory follows in intonation studies.

2.3.2.2 Prominence vs. Stress

In the Metrical Theory (Lieberman & Prince, 1977), the term “stress” is used as a phonological concept that labels “the head of a metrical unit, either the head of a foot, or the head of the whole word” (Van der Hulst, 2012). It is completely abstract in that it only provides a potential docking point for the sentence prominence to be associated

with (Ladd, 2008, p. 51). While stress is lexical, prominence can be post-lexical. Note that the term “post-lexical” is only used in this thesis to separate relevant concepts from the perspective of phonological hierarchy. The adoption of the term “post-lexical” does not suggest any temporal order of pitch assignment in processing – which one of lexical prosody or sentence prosody is processed first is a separate and important issue for investigation. Relative prominence is sometimes modified by the structure of an utterance and can also be modified due to contextual reasons. In other words, on a prosodic level that is higher than the lexical level, not all lexical stresses remain the prominence of the higher prosodic level. More on the difference between stress and prominence can be found in Section 2.3.3.2.

2.3.2.3 Pitch Accent vs. Lexical Tone

In AM Theory, “pitch accent” refers to the intonational pitch accent, which is different from lexical pitch accent. Intonation pitch accent, lexical pitch accent, and lexical tone are similar prosodic properties with their own features. Essentially, lexical tones are tones that are specified lexically in languages such as Mandarin or Yoruba, while lexical pitch accents are lexically specified tones and stresses in languages such as Tokyo Japanese or Norwegian. An intonational pitch accent is similar to a lexical pitch accent in that they both contain the elements of tones and stresses; however, intonational pitch accents are post-lexically specified “phonological elements of pitch contours” that are usually associated with the prominence (Hyman, 2006; Ladd, 2008, pp. 48–49). In this thesis, unless otherwise specified, pitch accent refers to intonational pitch accent.

Both lexical tones and pitch accents are autosegmental. For lexical tones, the association is between tones and tone-bearing units (TBU), while for pitch accents, the association is between tones and sentence prominence. Neither the association between lexical tones and segments, nor the association between pitch accents and segments is one-to-one. Both lexical tones and pitch accents can be composed of one or more tones. For example, in many African tonal languages, the tones are mono-tonal level tones such as L or H, while contour tones such as LH, HL, or even more complicated tones, can be found in Chinese dialects. Pitch accents can also be bitonal such as H*+L or L+H* in intonational languages, e.g. in English (Pierrehumbert, 1980).

Pitch accents and lexical tones are not mutually exclusive. In a non-tonal language, due to the absence of lexically specified tones, pitch accents are directly manifested on sentence intonation, as it is observed from the contour of F0 lines. However, in tonal languages, pitches can be assigned lexically as well as post-lexically. In a tonal language, it is no less likely for pitches to be assigned post-lexically than a non-tonal language. The combination of lexical pitches and post-lexical pitches are then shown as an integrated pitch contour. To study a tonal language, it is therefore important to find ways of teasing these two elements apart.

2.3.2.4 Tonal Association, Tonal Alignment, Tonal Scaling

“Tonal association” is a concept that is derived from Goldsmith’s (1976) Autosegmental Theory as mentioned earlier in Section 2.3.1.1. It refers to the way tonal events relate to segments phonologically, which are usually metrically strong tonal targets. “Tonal alignment”, however, is the phonetic realisation of parts of an F0

contour in relation to the time course of a segmental string. “Tonal scaling”, in short, is the height or scale of F0 as its name suggests.

Tonal alignment is one of the most direct evidence for the tonal association. However, it is not guaranteed that the tones are strictly aligned with the associated segmental unit – the peak may appear earlier or later, or may even extend over the syllable boundary.

Tonal alignment and tonal scaling are the horizontal and vertical changes of an F0 turning point (e.g. maxima, minima, elbow, etc.). A tonal alignment change may suggest a pitch accent alternation; a tonal scaling change may change the H or L of the tones. In intonation research, it is essential to examine these two properties to investigate potential tonal associations.

2.3.3 Tenets of AM Theory

Ladd (2008, pp. 44–45) extracted four basic tenets for AM Theory, as quoted in (7):

(7) Four tenets of AM Theory (Ladd, 2008, p. 44)

- (a) Sequential tonal structure;
- (b) Distinction between pitch accent and stress;
- (c) Analysis of pitch accents in terms of level tones;
- (d) Local sources for global trends.

These four tenets are the major assumptions which make AM Theory different from past theories and models. In this section, these four tenets will be explained based on Ladd’s arguments. Chapter 7 will discuss how these four tenets are reflected in tonal languages.

2.3.3.1 Sequential tonal structure

Although AM Theory is based on two non-linear theories (i.e. Autosegmental Theory and Metrical Theory), one of the most important features of AM Theory is its sequential tonal structure. AM Theory only depicts the phonologically meaningful local *events* and deliberately ignores the underspecified *transitions* between the local events. In a non-tonal language, such as English and German, the term “intonational events” refers to pitch accents, phrase tones and boundary tones, as summarised in Section 2.3.1.2.

By adopting the linear structure, AM Theory has different assumptions and functions from the superpositional or overlay models in modelling F0. Ladd (2008, pp. 23–31) raised a few points on the shortcoming of superpositional models such as the lack of “quantitative characterisation”. An example given by Ladd is the slope and width of the “grid” proposed by Gårding (1984), used as a cue to distinguish different prosodic tunes. Since a grid contains both tonal/ accentual information and intonational contour, it is difficult to provide systematic quantitative criteria to define “grid” in such models. Without quantitative characterisation, the model fails to provide any more information than visual inspection.

Admittedly, AM Theory does not seem to be able to provide a better solution to these issues. Rather, AM Theory reduces the uncertainties and keeps the minimal abstract representation as a string of tonal events. This is in accordance with Occam’s razor law.

2.3.3.2 Distinction between pitch accent and stress

The distinction between pitch accents and stress is the second tenet of AM Theory. Although stress is an abstract phonological representation of the head of the metrical

structure, it is considered as phonetic details in sentence-level prosodic analyses since it represents the prominence on the lexical level. In contrast, pitch accent directly associates with the sentence prominence. Relevant definitions and features have already been elaborated in Section 2.3.2.2 and 2.3.2.3. In this section, the distinction will be further explained with an example in (8) and (9) from Ladd (2008, pp. 49–50). This example is to show how word stresses are not necessarily a part of the sentence prosody, and how the pitch accent is aligned with the sentence prominence.

(8)

a.

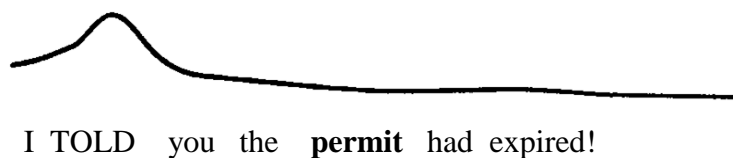


b.

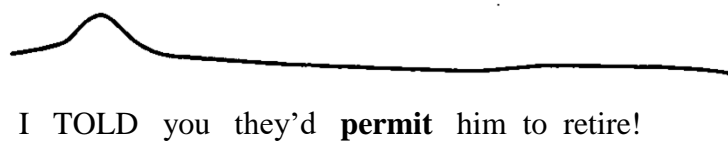


(9)

a.



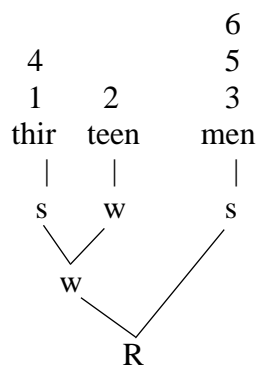
b.



In (8)a and (8)b, the alignments of the F0 peaks of the noun and verb are clearly different due to the different stress patterns. However, in (9), the difference in stress disappears in sentence condition with another word as the sentence prominence. The local phonetic performances are merely transitions and are not considered in AM Theory.

As illustrated in the beginning of this chapter, sentence-level prominence is derived from the metrical strength of each prosodic domain, as shown in (10).

- (10) Liberman & Prince (1977, p. 316) [identical with (2) in Section 2.3.1.1, p. 10]



2.3.3.3 Analysis of pitch accents in terms of level tones

AM Theory's answer to the level-vs-configuration debate is using level tones as primitives. Instead of using pitch contours to analyse intonations as in the traditional British school, or using the four-level "pitch phonemes" by early American structuralists such as Pike (1945), Wells (1945) and Trager and Smith (1951), AM theory only uses two level tones, H(igh) and L(ow), to describe intonational tunes. However, this raises the question: are two tones enough?

The reason for reducing the number of tones to two is that the two-tone system only predicts the necessary contrasts and avoids over analysis. These predictable contrasts in AM Theory using the two level tones are phonologically significant levels. The traditional four-tone system sometimes makes predictions that are not linguistically necessary. For example, a 231 tone and a 241 tone are phonologically the same and the over-analysis is neither enough for more detailed phonetic analysis nor significant for

phonological analysis. Moreover, some simple configurations can be recorded with bi-tonal accents. For example, a falling pitch accent can be noted as H+L. Depending on the association between the tones and prominence, a star can be added to the nuclear tone. Perceptually, empirical studies also support the discrete tones instead of pitch contours since the hearers tend to perceive the intonation categorically (Dilley & Brown, 2007).

2.3.3.4 Local sources for global trends

The scaling of the L and H tones is relative rather than absolute, since AM Theory offers a way of making phonological abstractions instead of acoustic analysis. The scaling of Hs and Ls are estimated relative to a speaker's "tonal space" (Ladd, 1992), which is the space between the "topline" and the "baseline" (Bruce & Gårding, 1978) of a pitch contour as shown in Figure 2.5. All the local pitch peaks are Hs and the valleys are Ls.

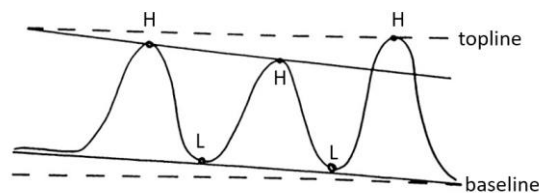


Figure 2.5 Tonal space between topline and baseline (Bruce & Gårding, 1978)

Apart from the speakers' tonal spaces, global trends need to be taken into consideration too. For an utterance that contains a global declination, the Point B in Figure 2.6a is substantially lower than the Point A. In a phonetic model that contains a quantitative description of declination (e.g. Fujisaki's model), the global contour is regarded as a whole as in Figure 2.6a and Point B is noted as a L. However, in AM Theory, the tonal scaling is dependent on the local sources. Point B is noted as a H tone since it is a local

pitch maximum. The analysis method is shown in Figure 2.6b, in which the local downsteps are considered independent from the global trend.

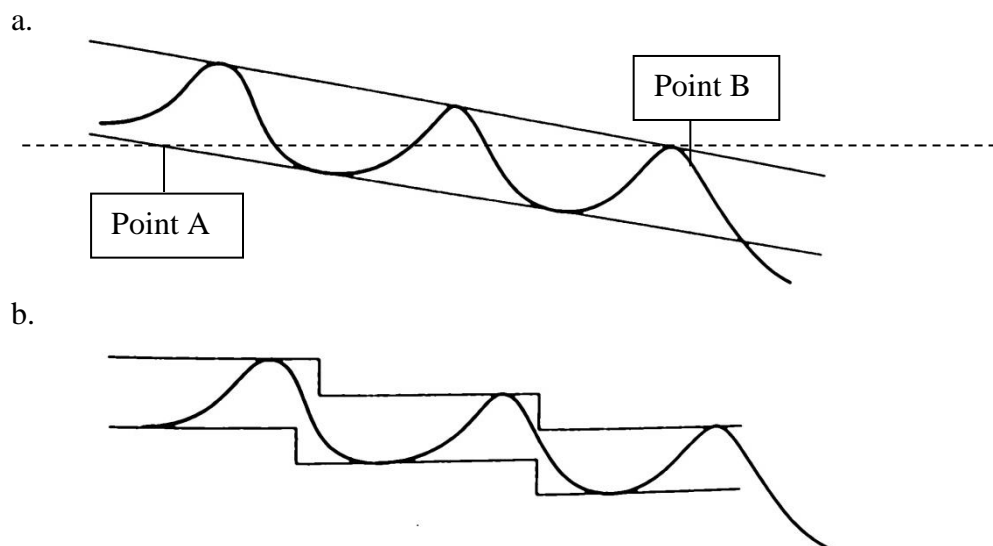
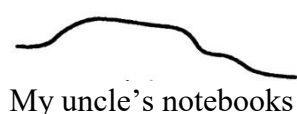


Figure 2.6 Two ways of analysing intonation contour
(adapted from Ladd, 1996, p. 75)

The utterance in (11) is an example given in Ladd (1996) to illustrate how to decide on the scaling of locally high tone in a globally declining contour:

(11)



This utterance, with a neutral focus on the whole phrase, is in the context of “my uncle’s notebooks, not any of the family furniture” (p. 75). The pitch of “notebooks” is obviously not as high as “uncle”, but it still counts as a H, since in its local source, it is high; otherwise, “uncle” will be the focus of the utterance. All in all, this tenet emphasises that phonetically low pitches are not necessarily phonologically low tones and vice versa.

2.4 Summary

The aim of this chapter was to provide an introduction and detailed review of the framework under which this thesis is working, i.e. Autosegmental-Metrical Theory. Section 2.2 listed a number of existing models for intonational study, according to the categorisation by Arvaniti (2011) and Xu (2015) respectively. This section placed AM Theory in the field of intonation research and explicitly pointed out that the AM model is a formal phonological model that uses pitch levels as primitives.

Section 2.3 covered various aspects of AM Theory in detail. The history and development of AM Theory were discussed in §2.3.1, including the theories from which the AM Theory developed, the original AM grammar and its components by Pierrehumbert (1980), and later modification by Ladd (1996, 2008). In §2.3.2, the key terms and concepts were elaborated: tune vs intonation, prominence vs stress, pitch accent vs lexical tone, tonal association, alignment, and scaling. Finally, in §2.3.3, the four basic tenets of AM Theory that Ladd summarised were discussed. The four tenets are: (a) sequential tonal structure; (b) distinction between pitch accent and stress; (c) analysis of pitch accents in terms of level tones; (d) local sources for global trends. These tenets will be revisited in Chapter 7 for a discussion about AM Theory's application in tonal languages.

Chapter 3 Tianjin Mandarin

3.1 Introduction

In this thesis, Tianjin Mandarin is studied, both as the object of study, and as a medium that connects AM theory with tonal languages. Therefore, this chapter outlines the basic background information on Tianjin Mandarin. In this chapter, comparisons between Tianjin Mandarin and Standard Mandarin are made whenever possible, since there is a much larger inventory of research of Standard Mandarin.

The goal of this chapter is threefold: firstly, this chapter contextualises the current research by providing background information on aspects of Tianjin Mandarin; secondly, in this chapter, some debatable issues such as lexical prominence in Section 3.4.3.1 are discussed, the stance taken in the discussion serves as the basis of discussion for the rest of the thesis; and finally, this chapter presents important dialectal features of Tianjin Mandarin, such as dentalisation in Section 3.3.2, in order to contribute to the conservation of the dialect.

This chapter is organised as follows: Section 3.2 introduces the origin and typology of Tianjin Mandarin, which includes the division of Chinese dialects; Section 3.3 presents a selection of segmental features distinctively distinguish Tianjin Mandarin from Standard Mandarin; Section 3.4 sketches the prosodic features of Tianjin Mandarin which serve as the foundation of intonation research, including its tonal values, tonal alteration rules, prominence, and prosodic grouping. Finally, in Section 3.5, the reasons behind the choice of Tianjin Mandarin for the current thesis are discussed.

3.2 Origin and Typology

Chinese can be divided into ten¹ major dialect groups (Wurm, Li, Baumann, & Lee, 1987), as shown in Figure 3.1: Mandarin (官话), Jin (晋), Wu (吴), Hui (徽), Gan (赣), Xiang (湘), Min (闽), Ke (客, Hakka), Yue (粤, Cantonese), and Ping (平). Mandarin Chinese is a dialect group of Chinese and can be further divided into eight sub-dialect groups, according to Wurm et al. (1987). traditional grouping. The division of the sub-groups of Mandarin is partially based on the typology as well as the diachronic changes of the lexical tones in different dialects.

Standard Mandarin, or *putonghua* ('common speech'), is a standardised lingua franca used in mainland China. Its vocabulary is based on northern Mandarin vocabulary and its grammar is based on written vernacular Chinese. Its pronunciation was designed to be based on Beijing dialect, while the first speech samples were collected between 1955-1956 in three towns in Luanping (滦平), Hebei Province (Wang et al. 2016). Luanping dialect is regarded as the dialect that sounds the most similar to Standard Mandarin, while differences still exist (L. Wu & Wang, 2014). Only a very small number of people can speak impeccable Standard Mandarin in the whole country. People who pass Level 1A in *Putonghua Proficiency Test* are qualified to be state-level or province-level news broadcasters; teachers are required to have 2B and above. Most university students' Standard Mandarin proficiency is above 2B. Informants recruited for Standard Mandarin production studies, although often not reported, are generally above Level 2A, while only above Level 2B for perception studies. Level-2 Standard Mandarin proficiency allows the speakers to carry dialectal features.

¹ Chinese dialects are divided into six to ten groups according to different views. The ten-group division has become comparatively popular among researchers.

In this thesis, “Mandarin” refers to Mandarin Chinese in general, while “Standard Mandarin”, “Tianjin Mandarin”, or “Beijing Mandarin” are used for referring to the specific Mandarin varieties.

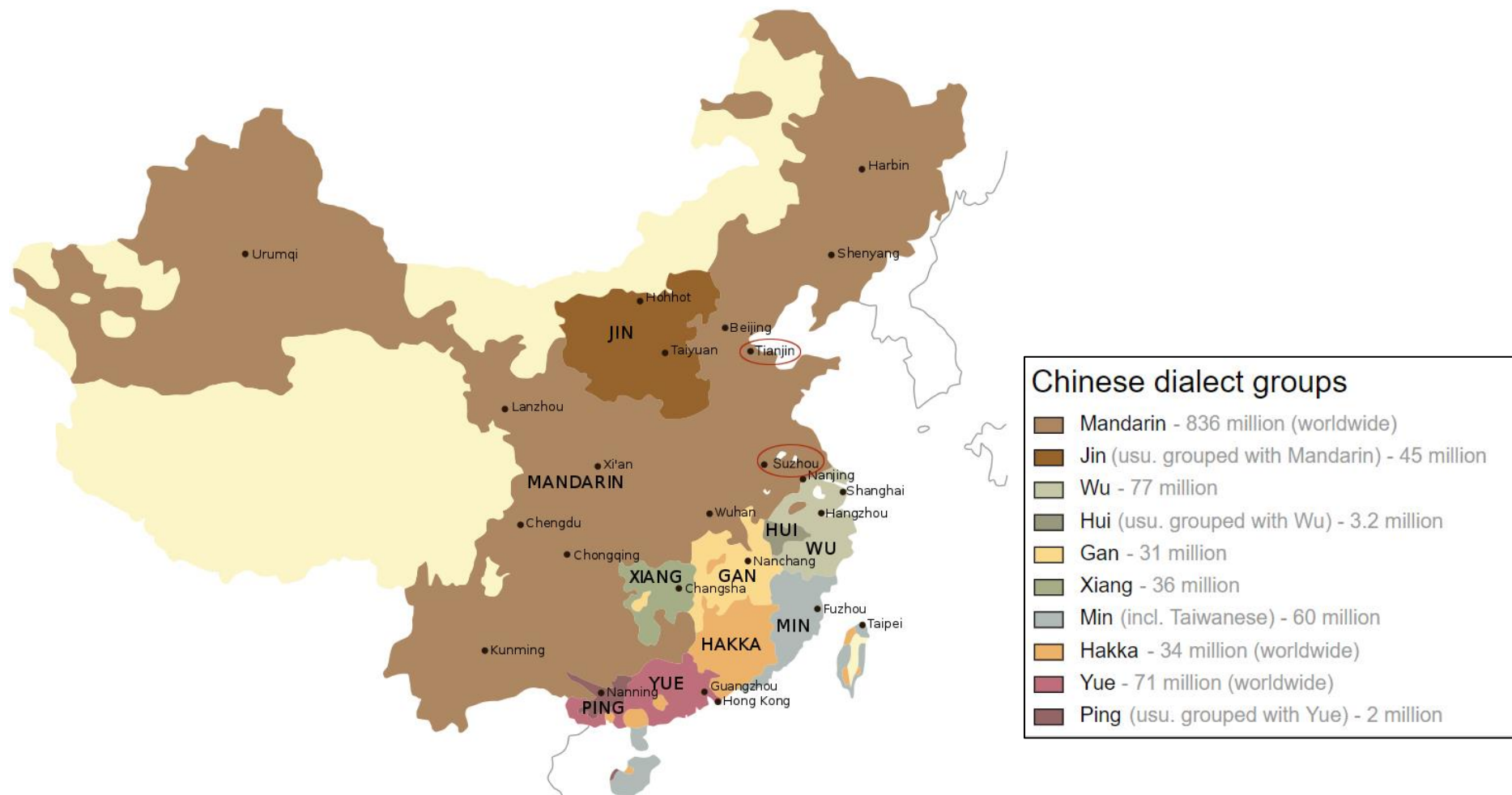


Figure 3.1 Map of Chinese dialect groups

(based on Wurm et al., 1987; image adapted from Wyunhe, 2011)

Tianjin Mandarin is spoken in the urban area of Tianjin, which is a municipality located in approximately 140km to the southeast of Beijing and is cradled by Hebei Province. However, Tianjin Mandarin belongs to Jilu Mandarin Group, while Beijing Mandarin belongs to the Beijing Mandarin Group. The suburban areas of Tianjin also speak dialects which belong to other dialect groups. In this sense, Tianjin Mandarin is a linguistic island. Surprisingly, Tianjin Mandarin resembles Suzhou (宿州) dialect in Anhui Province, a town 750km away from Tianjin in the south of China (Shiyu Li & Han, 1991). The resemblances between these two dialects are both in terms of pronunciation and dialectal vocabularies, including the lexical tone values, etc. Some linguists (e.g. Shiyu Li & Han, 1991; Xiaoyu Zeng, 2010) therefore posit that Tianjin Mandarin originated from northern Anhui dialects² based on evidence from similarities as such, as well as historical events. The linguistic island of Tianjin Mandarin was delineated by Li and Han (1991), as shown in Figure 3.2 as Area 1. In the rural areas to the far north of the linguistic island (Area 2 in Figure 3.2), locals speak Wuqing dialect, which is more similar to Beijing Mandarin since Wuqing County borders Beijing. Area 3a and 3b have two different North Suburban accents, while Area 5 has an East Suburban accent, which is similar to Tangshan accent of Hebei Province. Area 4a and 4b, which are areas neighbouring Jinghai County, have two different varieties of Jinghai accent.

The linguistic island of Tianjin Mandarin covers 38% of the total area of Tianjin and has an estimated³ population of 5.19 million, out of the 15-million total population of

² Some scholars such as Wang, Jiang, & Tang (2009) do not agree with such postulation, and insist that Tianjin Mandarin is a local accent influenced by dialects from nearby provinces.

³ Estimation method: Percentage of population of six urban districts in the 2010 report (Tianjin Municipal Bureau of Statistics, 2010) * total population of Tianjin in 2015, since the 2015 report did not give details on the population of each district.

Tianjin residents in 2015 (Tianjin Municipal Bureau of Statistics, 2015). Even within this area, not everyone speaks Tianjin Mandarin. This is especially true with regards to the younger generations born after 2000, not many of whom speak Tianjin Mandarin, even at home.

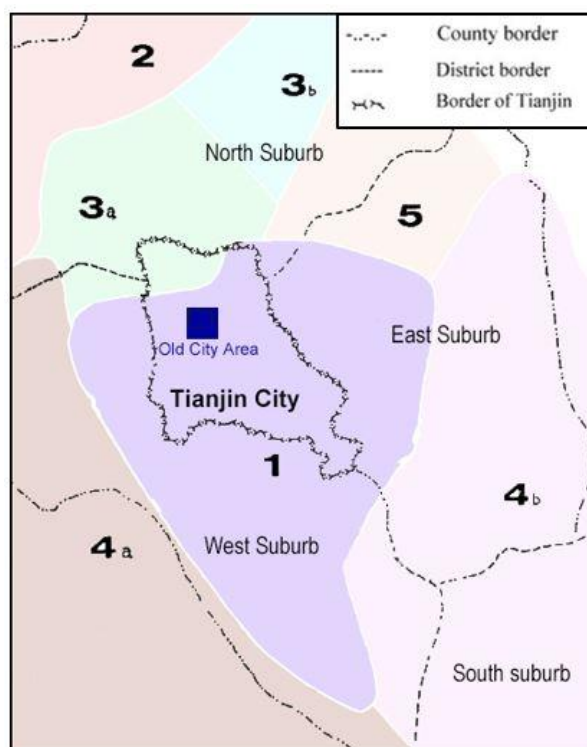


Figure 3.2 Tianjin Mandarin and Neighbouring Accents
(adapted from Shiyu Li & Han, 1991)

3.3 Segmental Features

3.3.1 Syllable Structure and Sound Inventory

Tianjin Mandarin syllable structure is identical to the syllable structure of Standard Mandarin, and many northern variations of Mandarin. A syllable can be as short as only one vowel (V), or up to four sounds—CGVX, in which C is a consonant, G a glide, X an offglide of a diphthong or a nasal (either [n] or [ŋ]). The onset C is optional in

Mandarin⁴. The G is sometimes analysed as a part of the vowel (see Duanmu, 2011 for a detailed review of different analyses). Duanmu (2000, pp. 79–81) argues that when C and G appear at the same time in Standard Mandarin, they are realized as C^G. All syllables with lexical tones are full syllables, while the toneless syllables are reduced syllables. For instance, a full CV syllable ma [ma:] (马, ‘horse’) is a heavy syllable, while a toneless syllable [ma] (吗, question particle) is light.

As for the sound inventory, Li, Chen, and Xiong (2017) presented a full description of the inventories of Tianjin Mandarin consonants (Table 3.1) and vowels (Figure 3.3, with the arrows showing the trajectories of the gliding in diphthongs. Note, there are also two syllabic consonants, [z] and [ʒ]. They only follow the dento-alveolar and post-alveolar affricates and fricatives. [z] only follows z[tʂ], c[tʂʰ] and s[ʂ], while [ʒ] is preceded by zh[tʂ], ch[tʂʰ], sh[ʂ], or [ʒ] itself when there is no vowel in the syllable.

Table 3.1 Consonants in Tianjin Mandarin (Q. Li et al., 2017, p. 2)

	Bilabial	Labio-dental	Dento-alveolar	Post-alveolar	Palatal	Alveolo-palatal	Velar
Plosive	p p ^h		t t ^h				k k ^h
Affricate			tʂ tʂ ^h	tʃ tʃ ^h		tɕ tɕ ^h	
Nasal	m		ɳ				ŋ
Fricative		f	s	ʃ		ɕ	x
Approximant	w			ɻ	j ɥ		
Lateral			l				

⁴ Traditionally, the zero-onset syllables with mid or low vowels in Standard Mandarin had a nasal [n] or [ŋ] as onsets in Tianjin Mandarin (Cui & Li, 1995), e.g. the word /ai/ ‘love’ is /nai/ in Tianjin Mandarin. It is very likely to be reflective of diachronic change of syllable structure in Mandarin Chinese (see Duanmu, 2000, p. 72 for related discussion in Standard Mandarin). However, this phenomenon is no longer as common in younger generation’s speech (Gu, 2008).

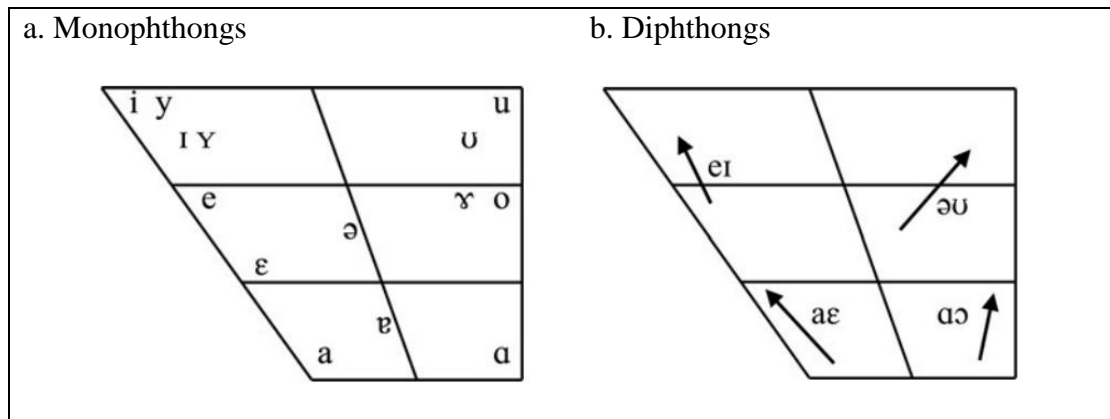


Figure 3.3 Vowels in Tianjin Mandarin (Q. Li et al., 2017, pp. 6&8)

3.3.2 *Chiyin* (Dentalisation)

Chiyin (齿音, ‘dental sound’) or dentalisation, describes the preference of the dento-alveolar fricatives and affricates over the post-alveolar ones in many Mandarin dialects. It is one of the most salient segmental features of Tianjin Mandarin that differs from Standard Mandarin. In Tianjin Mandarin, it is common to find a sound which would be a post-alveolar consonant ($zh[tʃ]$, $ch[tʃ^h]$, $sh[ʃ]$) in Standard Mandarin to be pronounced as a dento-alveolar consonant ($z[tʃ̺]$, $c[tʃ̺^h]$, $s[ʃ̺]$). This feature has been closely related to sociolinguistic factors such as generation, gender, social status. The dentalisation phenomenon is originated from Middle Chinese (c. 5th-10th century), and went through changes several times. Tianjin Mandarin does not simply replace all post-alveolar fricatives and affricates with dento-alveolar ones; instead, at different times in history, the speakers have used different consonants. In 1980s, Tianjin Mandarin was going through the shift from post-alveolar into dento-alveolar consonants (X. Zhang, 1987). Therefore, a division between older and newer version of Tianjin Mandarin occurred in late 1980s. In the syllables that should have a post-alveolar onset (e.g. $zh[tʃ]$), speakers of the older version of Tianjin Mandarin used different consonants for syllables with different rhymes, as shown in Table 3.2.

Table 3.2 Dentalisation in older version of Tianjin Mandarin

(based on X. Zhang, 1987)

Index	Onsets	Rhymes
(a)	Only $z[t_{\text{ɬ}}]$	$a[ɐ]$, $ai[aɛ]$, $uai[waɛ]$, $ua[wa]$
(b)	$z[t_{\text{ɬ}}]$, or $zh[t_{\text{ʃ}}]$	$uan[wan]$, $uang[wɑŋ]$
(c)	Mostly $zh[t_{\text{ʃ}}]$	$e[ɣ]$, $uei[weɪ]$, $u[u]$, $ao[ɑʊ]$, $ou[əʊ]$, $an[an]$, $uan[wan]$, $ang[ɑŋ]$, $en[ən]$, $uen[wən]$, $eng[əŋ]$
(d)	Only $zh[t_{\text{ʃ}}]$	$zhi[t_{\text{ʃ}}z]$

Speakers of the newer version of Tianjin Mandarin in the late 1980s only kept (d) and merge (a), (b) and (c) into the dento-alveolars. However, *shi*[t_ʃz] with Tone 4 is an exception and can be pronounced as *s*[sɿ]. Currently, owing to the influences from Standard Mandarin, dentalisation is fast disappearing in the younger generations' production of Tianjin Mandarin. The majority, if not all, of speakers of the younger generation without speech disorders, are all capable of using the post-alveolar consonants correctly. In the current thesis, the carrier sentences in Chapter 4 contains the word 是 *shi* [t_ʃz] ('be'-copula). The informants produced either pronunciations depending on their own preferences. The choice of the said segments was not controlled for the purpose of this thesis, since it would not lead to any considerable intonation change.

3.3.3 *Erhua* (Rhotacisation)

Erhua (儿化) refers to the process that a rhyme becomes r-coloured or rhotacised in Mandarin. In a number of Mandarin dialects, mostly northern, there is a morpheme “儿” (*er* [ɛ], which means ‘son’ when used in isolation). This is a feature that exists in Standard Mandarin but used for a wider range of words in Tianjin Mandarin. Moreover,

er suffix is a well-kept feature in Tianjin Mandarin, and many people find it to be unnatural not to rhotacise in some words. For instance, the carrier sentences in Chapter 4 end in the word “*zi*” (字, ‘character’). It is perfectly normal in Standard Mandarin for the word *zi* to be pronounced as it is; however, the Tianjin informants pronounced it in its rhotacised form without any orthographical cue in a pilot study. Therefore, in the studies in Chapter 4, I included the *er* suffix in the text to keep it consistent across different speakers.

As a suffix, *er* originates from three morphemes (Chao, 1968; Sijing Li, 1986) as shown in (12):

(12) Sources of *er* suffix

- a. the diminutive *er* (儿, ‘son’), e.g. 小手儿, *xiaoshou-er*, ‘little hand’;
- b. the noun *ri* (日, ‘day’), e.g. 今儿, *jin-er*, ‘today’;
- c. the locative *li* (里, ‘inside’), e.g. 这儿, *zhe-er*, ‘here’.

Phonologically, although the underlying form of *er* suffix is a full syllable [ə], it coalesces into one syllable with the preceding rhyme, and surfaces as a feature [+retroflex]. Detailed phonetic analysis for *er* suffix in Tianjin Mandarin can be found in Li et al. (2017).

The rhymes [ɑ, u, o, ʏ, ɛ] are back vowels which can be pronounced together with the apical retroflex feature, so rhymes with these vowels are directly rhotacised, e.g. [ma] → [ma̠]. Front vowels such as [i, y] as well as the two syllabic consonants [z] and [ʒ],

clash with the retroflex since the tip of the tongue is also used, so a central vowel [ə] is inserted after the vowel and then is rhotacised, e.g. [ɛi] → [ɛiə̃]. For the rest of the diphthong rhymes, the ones ending with high front vowels turns into an [ə] before rhotacisation, e.g. [wei] → [wə̃], [y] → [yə̃]; the ones ending with mid or vowels replaces the vowels with an [ə], e.g. [h^wa] → [h^wə̃]. If a rhyme that contains a nasal coda, a [n] coda is deleted before changes following the same rules as the above, e.g. [pan] → [pã] (the same surface form as rhotacised [pai]); a [ŋ] coda is also deleted, and the rhymes also go through the above processes, then the vowels turn into nasalised vowels, [fəŋ] → [fə̃̃]. The processes of vowel changes are shown in Figure 3.4, with the green line indicating the insertion of the [ə], and the blue line the replacement of the vowels with the target [ə] and [ə̃]. The red circles show back vowels that receive the retroflex feature without going through any other changes.

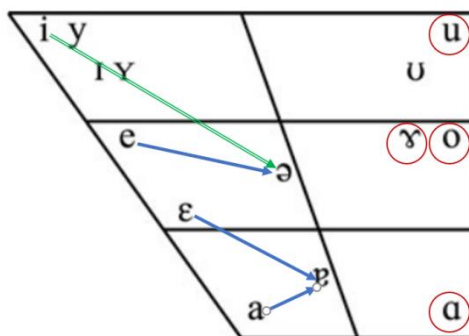


Figure 3.4 Vowel change in rhotacisation

3.4 Prosodic Features

3.4.1 Description of Lexical Tones

In Standard Mandarin, there are four lexical tones – high tone, rising tone, low tone, and falling tone. Chao (1930) designed a tone letter system, “tone number” or “Chao

letters”, that could represent the tones according to their relative pitch levels. In this system, “1” is the lowest level and “5” is the highest. The pitch levels of the four citation tones are 55, 35, 214 and 51 respectively. The four tones resemble music notes to some extent, but the pitches of music tones are absolute, while those of Mandarin lexical tones are relative. A fifth “tone”, if it can be counted as a tone, is the neutral tone, (a.k.a. toneless tone), which means a syllable does not carry any underlyingly specified tone, and the values of such a tone are dependent on the preceding and succeeding syllables.

The most noticeable difference between Tianjin Mandarin and Standard Mandarin lies in their lexical tones. Tianjin Mandarin has four lexical tones (and a neutral tone) as Standard Mandarin does; however, both the contours and the values are different in Tianjin Mandarin from Standard Mandarin, as shown in Figure 3.5. Tianjin Mandarin has a low tone (L, T1), a high tone (H, T2), a rising tone (LH, T3), and a falling tone (HL, T4), which correspond to the high tone (H, T1), the rising tone (LH, T2), the low⁵ tone (LH, T3), and the falling tone (HL, T4) respectively in Standard Mandarin.

⁵ Low tone in Standard Mandarin is also called “dipping tone” or “contour tone”. Its tone number is 214.

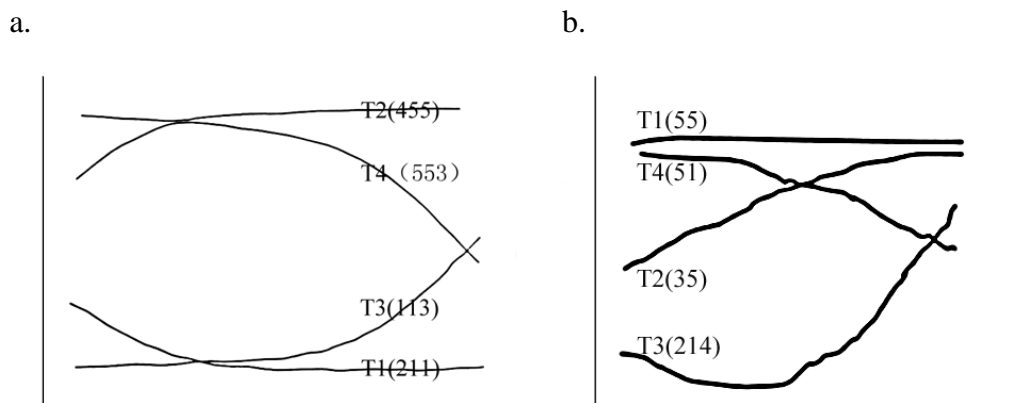


Figure 3.5 Lexical tones of (a) Tianjin Mandarin (Shi, 1986) and (b) Standard Mandarin (Chao, 1930)

Ma (2012) summarised Tianjin tone values with tone numbers from various experimental work before the 1990s, as shown in Table 3.3. Despite the differences between the tone numbers, the four tones are consistently L, H, LH, and HL.

Table 3.3 Values of Tianjin Mandarin lexical tones (Ma, 2012)

	Tone 1 L tone	Tone 2 H tone	Tone 3 LH tone	Tone 4 HL tone
S. Li, (1956)	11	55	(214)	(51)
Davison (1982)	21/11	45	24	53
X. Li and Liu, (1985)	21	34	213	53
Shi (1987)	21	45	13	53

However, in more recent studies (e.g. Q. Li et al., 2017; J. Zhang & Liu, 2011), Tone 1, as shown in Figure 3.6, has become a steeper fall than what it used to be described as in the literature. Zhang and Liu (2011) transcribed Tone 1 as 41, and Li et al. (2017), as 31. Judging from the above studies, the value of Tone 1 in Tianjin Mandarin seems to exhibit a diachronic change of having increasingly high pitch for the beginning.

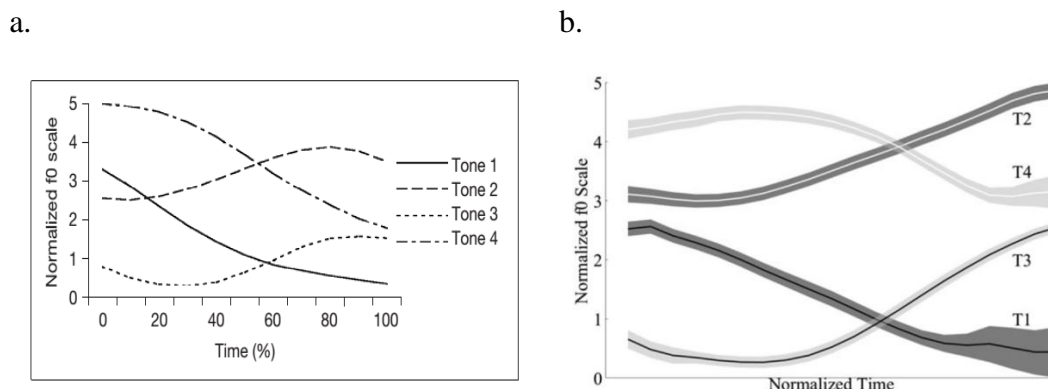
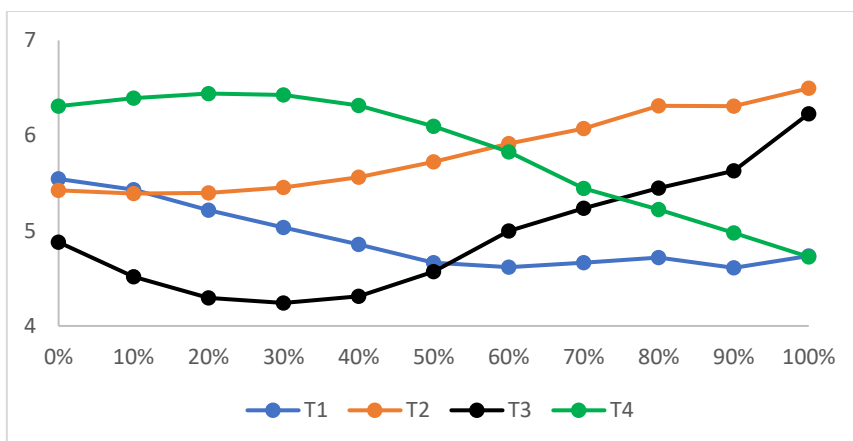


Figure 3.6 Values of Tianjin Mandarin lexical tones in (a) Zhang & Liu (2011) and (b) Li et al. (2017)

Data from the 6 six speakers of Study 1 in the current thesis (see Chapter 4 for detailed methods) are presented in Figure 3.7, with the values showing the mean pitches in ERB at each 10% of the rhymes of all the data. The contours of the lexical tones resemble those in Zhang and Liu (2011) and Li et al. (2017) in Figure 3.6. However, it is still not hard to see that Tone 1 has a slower slope than Tone 4, and Tone 2 has a slower slope than Tone 3.

The goal of the current thesis is to investigate the phonological components of the intonational tunes in Tianjin Mandarin, so whether a tone value is a 31 or a 21 does not make a significant difference. What is important is that Tone 4 is a high falling tone, and Tone 1 is a low falling tone; Tone 2 is a high rising tone, while Tone 3 is a low rising tone. Therefore, in this thesis, I will stick to the usage of L, H, LH, and HL to represent the four lexical tones.



(Y-axis unit: ERB)

**Figure 3.7 Values of Tianjin Mandarin lexical tones:
Data from the current thesis**

3.4.2 Tonal Alternation Rules

Tone alternation rules are commonly known as “tone sandhi”, especially in Standard Mandarin literature, since it has a widely-known tone sandhi rule – the third tone sandhi in (13):

$$(13) \quad LH + LH \rightarrow H + LH$$

However, since the word “sandhi” is from Sanskrit, meaning “joining”, it should denote such “joining” processes as assimilation; using the word “sandhi” to refer to a deviating process such as dissimilation would thus be inaccurate. In this thesis, I will point out the specific tonal alternation process, such as tonal assimilation or tonal dissimilation.

Tianjin Mandarin has four tonal dissimilation rules for disyllabic sequences (Davison, 1991b; Q. Ma, 2012), as shown in (14):

(14) Tonal Dissimilation Rules in Tianjin Mandarin

- a. $L \rightarrow LH(25) / _ L$ e.g. *bing(L) xiang(L)* \rightarrow *bing(LH) xiang(L)* ‘fridge’
- b. $LH \rightarrow H(55) / _ LH$ e.g. *wu(LH) dao(LH)* \rightarrow *wu(H) dao(LH)* ‘dance’
- c. $HL \rightarrow L(21) / _ HL$ e.g. *da(HL) di(HL)* \rightarrow *da(L) di(HL)* ‘earth’
- d. $HL \rightarrow H(55) / _ L$ e.g. *bi(HL) xu(L)* \rightarrow *bi(H) xu(L)* ‘must’

These four rules, all of which are results of the OCP (Obligatory Contour Principle), are widely accepted by linguists, except for Wee, Yan, and Lu (2005), who claimed there are two other tone sandhi rules in Tianjin Mandarin; however, their claim was rejected by Ma and Jia (2006) with evidence from an acoustic experiment and a perceptual experiment. In a number of studies (e.g. Q. Li & Chen, 2016), no significant change can be detected in the T4T4 sequence, since it is closely related to sociolinguistic factors such as age, occupation, etc. (H. Feng, 2015).

Not only the T4T4 sequence has been found to be changing, all the others are also merging with Standard Mandarin. As shown in (15), even though the rules have not changed yet, the values of the targets have changed into the values in Mandarin. For example, (15)b is a T3T3 sequence, which changes into a T2T3 as in Standard Mandarin. Traditionally, the first T3 syllable changes into a H (55), which is close to the value of T2 in Tianjin Mandarin; however, the new version changes the first syllable into a LH (35), which is the value of T2 in Standard Mandarin.

(15) New Tianjin Tonal Dissimilation Rules:

- a. $L \rightarrow H(55) / _L$
- b. $LH \rightarrow LH(35) / _LH$
- c. $HL \rightarrow HL(53) / _HL$
- d. $HL \rightarrow H(55) / _L$

Tianjin tone alternation is even more interesting when trisyllabic sequences are involved, since out of a total of 64 possible combinations of the four lexical tones in trisyllabic combinations, 27 of them can go through tonal dissimilation (Q. Ma, 2005, 2012).

X. Li and Liu (1985) first brought up the “paradox” of different tone sandhi in trisyllabic combinations in Tianjin Mandarin. Chen (1986, 1987) further brought this issue up to the attention of the world. In short, it is about the directionality of the application of the disyllabic tonal dissimilation rules in trisyllabic sequences. In some cases, the rules can be applied from left to right, as in (16):

(16) $T3 + T3 + T3 \rightarrow T2 + T2 + T3$ e.g. lao3 ling3 dao3, ‘old leader’

However, the left-to-right application cannot explain (17):

(17) $T4 + T4 + T4 \rightarrow T4 + T1 + T4$ e.g. da4 huai4 dan4, ‘big scoundrel’

The same problem happens to the right-to-left application, and cyclic application: each can explain a number of sequences, but never all. The directionality of the application of tonal dissimilation rules in trisyllabic sequences in Tianjin Mandarin remains a mystery. Chen (1986) claimed that “the phenomenon in Tianjin dialect severely drains

the descriptive and explanatory capability of any current linguistic theory”. Since then, a number of studies have been done from either the perspective of phonetic description⁶ (e.g. Q. Li & Chen, 2012, 2016) or the perspective of phonological rule ordering (e.g. Jansche, 1998; H. Lin, 2008; Q. Ma, 2005, 2012; Q. Ma & Jia, 2006; F. Shi, 2009; Wee, 2005). However, despite the fact that phonologists have tried explaining the issue with the traditional derivational approach, optimality theory, and other approaches, they nevertheless failed to “explain [it] without proposing ad hoc directionalities of rule application and/or arbitrary rule orderings among the elementary rules” (H. Lin, 2008).

Since the trisyllabic tonal sequence has not been fully understood, this thesis will try to avoid such trisyllabic sequences in order to achieve better control of the studies.

3.4.3 Prominence

3.4.3.1 Lexical Prominence

Lexical prominence is commonly referred to as “stress” in intonational languages. However, stress is only one type of lexical prominence. Hyman (2009) addresses the structural nature of stress, stating “... stress is identified with abstract metrical structure ... Stress is a structural property in which syllables are metrically hierarchized as relatively strong vs. weak (however this contrast is realized phonetically).” Whether Mandarin Chinese makes use of lexical stress is a heavily debated issue: many believe that stress does exist in Chinese (Chao, 1968; Duanmu, 1990, 2000; S. Feng, 1995, 1997), while some argue that Mandarin does not have stress at all (Bao, 2003; Hyman, 2009; X. S. Shen, 1990; Yip, 1980).

⁶ Li & Chen (2016) examined data from six native speakers who were born in the 80s. The results showed substantial difference from the literature. As also mentioned in their conclusions, the gap is very likely to be a result of language evolution: quite a number of tonal alteration rules do not, or seldomly, manifest in younger generation’s speech, while they can still be found in older generation’s speech.

In Mandarin, it is more likely that the lexical prominence is not structural as in a language like English. However, there is lexical level prominence. Wang and Feng (2006) studied the prominence distribution in a large corpus of Beijing Mandarin disyllabic full-toned words based on native-speaker judgement, and concluded that the right-heavy syllables are the default ones, while the left-heavy ones should be listed in dictionaries, if lexical prominence ever becomes a part of the dictionary components. They also observed a type of words that can be heavy in either syllable, which can be merged into the right-heavy words. From their result, it can be inferred that the underlying form of lexical prominence in Beijing Mandarin (which can possibly be applied to Standard Mandarin, and some other Mandarin varieties including Tianjin Mandarin) is right-prominent, and the left-prominent ones are specified lexically.

There are a few ways of testing whether lexical prominence exists. One way is based on the fact that lexical prominence can differentiate meanings in some cases, although the number of such cases is small. For example, *'shuiping* (left-prominent, 水平, level/capability) and *shui'ping* (right-prominence, 水平, level/ horizontal) have different meanings. Another way is specifically for Tianjin Mandarin. Tianjin Mandarin has four tonal dissimilation rules as introduced in Section 3.4.2. They only work when the prominence is on the right. For example, the word Tianjin (L Tone + L Tone) should become H Tone + L Tone, according to the tonal dissimilation rules. However, since Tianjin has a lexically specified left-prominence, the tonal dissimilation rule does not apply to this word.

Although in Mandarin, no convincing rules can be found in lexical prominence yet, a few types of words are observed to have consistent lexical prominence. The first type is when a neutral-tone syllable is involved, it is automatically left-prominent, since all full-toned syllables are heavy, and all neutral-tone syllables are light. For example, in *jiaozi* (T3T0), the T3 syllable *jiao* carries the lexical prominence. Disyllabic full names with full tones are usually right-prominent. For example, in the full name *Wang Wei* (surname + given name), the second syllable carries the prominence. In Chapter 6, disyllabic full names are used as speech materials for the examination of chanted call tune: the full-toned names are all right-prominent; and the ones with neutral-tone syllables are all left-prominent.

3.4.3.2 Sentence Prominence

The second type of prominence is the sentence-level prominence. In Mandarin, it is usually created by sentence focus. Focus can come from three sources: one is based on syntactic structure, e.g. *ta xie X zi* ('He writes the character X', a carrier sentence in Chapter 4); another one is based on the contrastiveness – illustrating with an example from English: 'it is MINE, not HIS'); the third source is the same type of right-end default prominence as the lexical prominence – for instance, in a neutral-focus sentence, the last syllable receives an automatic sentence prominence. Later in Section 4.2.2.2, a more detailed review of focus will be presented.

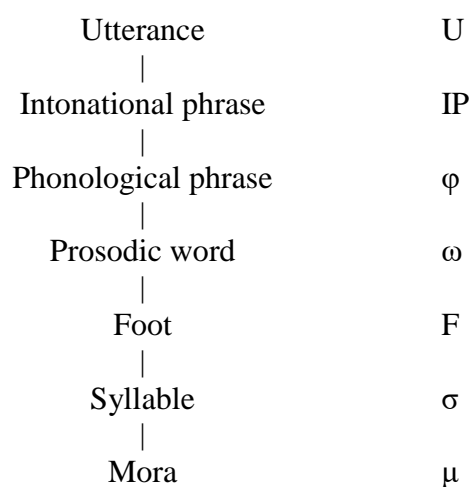
3.4.4 Prosodic Grouping

Lexical prominence is the prominence of a lexical word. But the question that arises from the discussion here is, what is a “word” in Mandarin? This section introduces the

prosodic grouping in Mandarin Chinese in order to facilitate the understanding of the internal structure of a sentence from the perspective of prosody.

In the previous chapter (Section 2.3.1.1), I introduced the structure for prosodic hierarchy, reproduced as (18):

(18) Prosodic hierarchy (identical with (3))



This structure is mostly based on European languages. Is it applicable to Tianjin Mandarin? Since no study has investigated the prosodic grouping of Tianjin Mandarin, proposals based on Standard Mandarin are reviewed in this section instead. Due to the similarities between Tianjin Mandarin and Standard Mandarin, the overall tendencies are expected to be the same.

As mentioned above, Selkirk (1980, 1986) proposed a five-level prosodic hierarchy on and above word level, shown in (18). These five levels under utterance include intonational phrase (I or IP), phonological phrase (φ), prosodic word (ω), foot (F) and syllable (σ). Nespor & Vogel (1986) has a more comprehensive presentation of the idea, with their own proposal of adding clitic group into the original hierarchy. Beckman and

Pierrehumbert (1986), as mentioned in Section 2.3.1.1, also proposed the intermediate phrase. Moreover, according to the Selkirk's (1980) "strict layering hypothesis", the higher unit consist exclusively of the next lower level, which means there is no multiple dominations, no heterogeneous sisters, no skipping of levels, no unlabelled nodes, and no recursion, according to the summary in Ladd (1996, p. 239). However, Ladd (1996) argued against the strict layering hypothesis by proposing the concept of "compound prosodic domain" (e.g. "telephone call", see Ladd, 1996, p. 244)

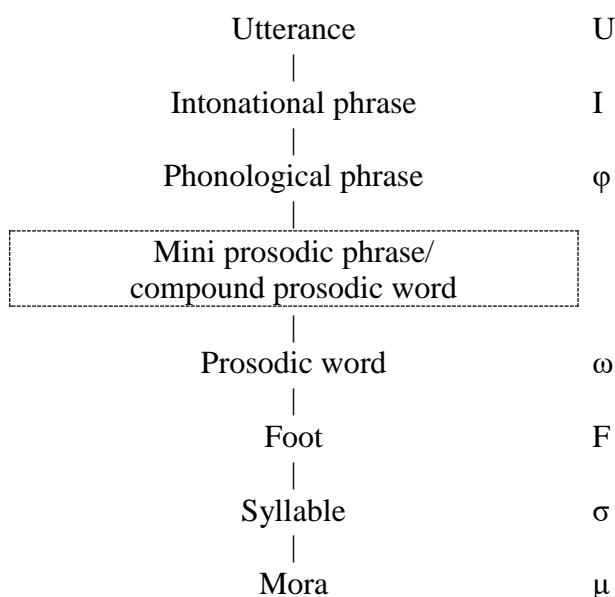
Many linguists (e.g. Cao, 2001; S. Feng, 1996, 1998; M. Lin & Yan, 2000; H. Wang, 2000) have studied the prosodic structure of Mandarin. Most of them believe that there are at least three levels in Mandarin: prosodic word, prosodic/ phonological phrase, and intonational phrase. Feng (1995) believes a prosodic word in Mandarin contains at least two syllables, i.e. two characters. He argues that monosyllabic foot ("Degenerate Foot") and trisyllabic foot ("Super Foot") also exist in Mandarin (also see Chen, 1979; Guo, 1938; Shih, 1986b), with most of the monosyllabic degenerated prosodic words being functional words or monosyllabic verbs (Cao, 2001). The prominence of prosodic phrases is usually decided by the syntactic structure of the phrase; for example, in a V-O phrase, the Object is more likely to be stressed (Zheng, Wang, Yang, Lu, & Cao, 2000). The concept of the intonational phrase is the same with that in English, which usually refers to a clause of either a complex utterance or a simple utterance.

A special case is the prosodic unit which contains four syllables. In Chinese, there are many quadrisyllabic idioms or set expressions. Morphologically speaking, some of those expressions are words, while some others are phrases. Phonologists disagree with each other in terms of their prosodic unit. H. Wang (2000) selected four types of

quadrisyllabic expressions into prosodic words, while some other linguists categorised these quadrisyllabic words and phrases into a new prosodic unit, which falls between prosodic word and prosodic phrase. Names for this level vary, for instance, “mini prosody phrase”(Cao, 2001) and “compound prosodic word”(S. Feng, 1997).

In summary, Mandarin prosodic structure can be shown in (2.6):

(19) Mandarin prosodic hierarchy



3.5 Choice of Tianjin Mandarin

As introduced in the previous sections, Tianjin Mandarin shares many similarities with Standard Mandarin, which may bring about concerns with regards to the significance of the current study – why is Tianjin Mandarin chosen instead of Standard Mandarin, when the two are so similar to each other? The reasons for choosing Tianjin Mandarin as the instrument for the current study are presented as follows:

First of all, Tianjin Mandarin was selected over Standard Mandarin due to the fact that it is a natural language so the influences from other languages are easier to control, while Standard Mandarin is a lingua franca, with speakers of different dialects and levels of proficiency (see Section 3.2). In the previously mentioned *Putonghua Proficiency Test*, speakers' accuracy of producing the lexical tones are explicitly evaluated, while there is no clear criterion for intonation – it only needs to be “natural”. Other aspects of prosody such as rhythm are not mentioned at all. However, prosody is one of the subtlest properties of language, which is acquired first in the first language, and is also one of the most difficult features to become native-like in a second/ foreign language acquisition. In perception, as young as new-borns can perceive lexical stress to a certain extent (e.g. Sansavini, Bertocini, & Giovanelli, 1997). In production, Mandarin-speaking children acquire lexical tones much earlier than segments (C. N. Li & Thompson, 1977). Zhu and Dodd (2000) reported that only the youngest group of their participants, who were aged 1;6-2;0, made two tone errors. This indicates that by 1;6, lexical tones are already mastered by Mandarin-speaking children. Meanwhile, even advanced second language learners find aspects of prosody difficult. Jun and Oh (2000) investigated the production of intonation produced by Korean-learning English speakers. Even the most advanced learners only produced 62.5% of wh- phrasing. Lexical tones are even more notorious for being difficult to acquire by Mandarin learners. Miracle (1989) reported an overall error rate of 42.9% in the performance of some second-year American learners of Mandarin. Therefore, Standard Mandarin, as a lingua franca, contains uncertain dialectal influences. By studying a language with more controllable variables, I hope to investigate the tones better.

Secondly, the tonal system and rules in Tianjin Mandarin are more interesting. The four lexical tones in Tianjin Mandarin form two groups of mirror images – one H tone and one L tone; one HL tone and one LH tone. Standard Mandarin, however, has three lexical tones that end in H. It is therefore more likely for studies in Standard Mandarin to be biased towards H tones. In Tianjin Mandarin, there are also intricate tonal dissimilation rules, which has already attracted linguists' attention. This thesis examines the effect of these dissimilation rules on intonation, and it also lays the foundation for future work on more complicated tonal changes, e.g. trisyllabic tonal changes, or the intonational tunes which contain the interactions between such lexical tonal changes and post-lexical prosodic features.

Thirdly, because of the contact with Standard Mandarin, all Chinese dialects are fast changing and even disappearing. Dialects always carry historical information, both from the perspective of both linguistics studies and historical studies. Many northern Mandarin dialects are particularly prone to change because of the similarities with Standard Mandarin, as well as the lower social status of dialects in general. Many endangered languages in the world disappear at an astonishing rate. For example, the number of speakers of Breton, a Celtic language spoken in northern France, decreased by around 8.5% within one decade's time (Kennard, 2013, p. 1). Although there still are a comparatively large number of speakers of Tianjin Mandarin right now, due to the strong influence from Standard Mandarin and better connectivity within the country and the world, it is not hard to imagine that one day Tianjin Mandarin will merge with Standard Mandarin, and potentially even English. This study therefore is also an effort to contribute to the preservation of the dialect.

Lastly, because of the similarity with standard Mandarin, I hope the results in this thesis can serve as the basis of solid research hypotheses, and provide applicable research methods under the framework of AM Theory for future studies of Standard Mandarin intonation.

Chapter 4 Production Studies:

Intonational Yes/No Question Tune in Tianjin Mandarin

4.1 Introduction

The previous chapters presented the theoretical framework and research background of this thesis. In this chapter, we move on to an analysis of question intonation, and particularly of syntactically unmarked Yes/No questions, in comparison to regular declarative statements.

In Standard Mandarin, as well as many other varieties of Mandarin, there are at least five ways of asking a question. Figure 4.1 is a brief outline of Standard Mandarin interrogative sentence types, based on Li & Thompson (1981) and Li & Cheng (2008). The five categories of interrogative sentences in Standard Mandarin are: (i) YNQs (a.k.a. “polar questions”), (ii) question-word questions (wh- questions), (iii) disjunctive questions (including V-not-V questions and A-or-B questions), (iv) tag questions (... , ok? / ..., right?), and (v) rhetoric questions.

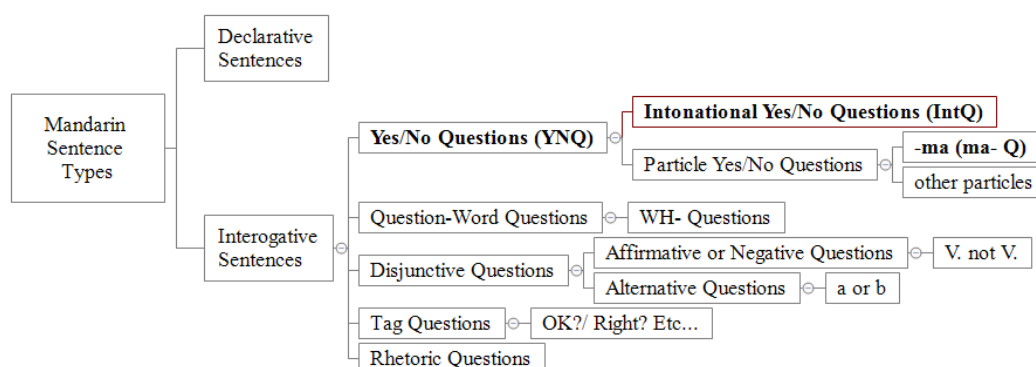


Figure 4.1 Mandarin sentence types

YNQs are further divided into two types – particle questions (hereinafter **ma-Q**) and intonational Yes/No questions (hereinafter **IntQ**). For instance, (21) is a ma-Q for the statement in (20). 吗 *ma* in (21) is a question particle. Other question particles include 吧 *ba*, 啊 *a*, and so forth, with variations in the degrees of incredulity. Both ma-Q and IntQ can be used for seeking information and echoing the previous sentence (Lee, 2005, p. 6).

(20)

这	是	咖啡。
<i>zhe</i>	<i>shi</i>	<i>kafei</i>
this	be	coffee

“This is coffee.”

(21)

这	是	咖啡	吗?
<i>zhe</i>	<i>shi</i>	<i>kafei</i>	<i>ma?</i>
this	be	coffee	Q-par?

“Is this coffee?”

An IntQ, as shown in (22), is syntactically identical to its declarative counterpart.

(22)

这	是	咖啡?
<i>zhe</i>	<i>shi</i>	<i>kafei?</i>
this	be	coffee?

“This is coffee?”

Since IntQs only differ from the statements by their intonations, how do the lexical tones behave? Does intonation override the lexical tone information in IntQs? What is the IntQ tune composed of?

This chapter is structured as follows: Section 4.2 reviews the previous studies on YNQ tunes in different languages; Section 4.3 presents the general methodology used in the subsequent two production studies of IntQs in this chapter; Section 4.4 reports a production study of monosyllabic words in isolation (**Mono_(ISO)**); Section 4.5 reports another production study of sentences which use the same monosyllabic words as their sentence prominences (**Mono_(SEN)**); and finally, Section 4.6 discusses the results of both studies and presents a summary of the conclusions.

4.2 Literature Review: Yes/No Question Tunes

4.2.1 YNQ Tunes in World Languages

YNQ tune is frequently associated with a High or rising ending, by linguists and by laymen. Linguists proposed this association as a linguistic universal based on different reasons and from different perspectives. For instance, Ohala (1983, 1984) proposed a universal “frequency code” from an ethological view, suggesting that high pitch was associated with “smallness” and thus was used for question or politeness prosody. Gussenhoven (2002) extended on this concept and argued that the universality lay in the paralinguistic aspect of form-meaning relationship rather than a direct linguistic mapping.

However, whether such a “universal” exists at all and to what extent it is true remains a question. Bolinger (1978) reported a few languages that did not use high pitch in

questions, although he made one of the earliest universalist claims. Rialland (2007) further surveyed YNQ tunes in 78 African languages and found a substantial number of languages that did not associate high pitch with question prosody. Moreover, she listed 11 ways of forming questions, as shown in Table 4.1, a few of which were not even related to pitch.

Table 4.1 Markers of YNQ (Rialland, 2007, pp. 37–38)

High-pitched Yes/No question markers	(a). cancellation/ reduction of downdrift, register expansion
	(b). raising of last H(s) (not necessarily sentence-final)
	(c). cancellations/ reduction of final lowering
	(d). final High tone or rising intonation (final H%)
	(e). final HL melody (listed in this category as it involves a high component)
Non-high- pitched Yes/No question markers	(f). final Low tone or falling intonation (final L%)
	(g). final polar tone or M tone
	(h). lengthening: V (a vocalic mora) or V... (considerable vocalic lengthening)
	(i). breathy termination
	(j). cancellation of penultimate lengthening
	(k). [open] vowel

Among the 11 markers in Table 4.1, (d), (e), (f), and (g) all involve the alternation of the boundary tones. Final HL boundary tone in (e) – also described as rise-fall – is observed in Bengali [L*+HL%] (Hayes & Lahiri, 1991), Greek (Arvaniti, 2009), and so on. In some other languages such as Puerto Rican Spanish [!H*+L% / H+L*+L%] (Armstrong, 2012), and Bari Italian [L+H*/ H+L* + L-L%] (Grice & Savino, 2003), a falling or low boundary tone is used to mark an IntQ.

Apart from boundary phenomena, other types of pitch variations (a, b, and c in Table 4.1) are also used in a number of languages. For example, in Wolof (non-tonal, Rialland & Robert, 2001), Hausa (tonal, Lindau, 1986), and Danish (non-tonal, Grønnum, 1998), reduction of downdrift as in (a) is used for making an IntQ.

The rest (h, i, j, and k in Table 4.1) are not pitch-related. These have only been found in African languages so far. For example, Cahill (2013) examines polar questions in six Ghanaian languages and suggests that Adele and Safaliba make use of final vowel lengthening as in (h), and Chumburng employs breathy termination as in (i).

Among contour-tone languages such as Chinese, Thai, and Vietnamese, register change and the extent of rise or fall of the lexical tones are commonly seen. In Vietnamese, “greater high tone movement” is found in IntQ tune (Dung, Huong, & Boulakia, 1998, p. 405); in Thai, “narrower pitch range” is observed in IntQ tune (Luksaneeyanawin, 1998, p. 385). Mandarin has similar results as these contour-tone languages. In the following section, the tunes in Mandarin Chinese will be reviewed in greater detail.

4.2.2 Tunes in Mandarin Chinese

Intonation studies in Mandarin date back to Chao (1933) where he described Mandarin intonation as “small ripples riding on big waves”. Many subsequent studies viewed Mandarin intonation from these two different perspectives – a global perspective (“big wave”) and a local perspective (“small ripple”). The early studies were mainly impressionistic, while the more recent ones mostly concentrated on the phonetic properties of the intonation (e.g. degree of declination, topline vs. bottom line), which made use of superpositional models of intonation. In this section, I will review the

previous studies about Mandarin intonation from these two perspectives, in order to identify the major disagreements among linguists on Mandarin intonation. Note that some of these studies are on IntQ tune specifically, while some studied YNQ tune in general without differentiating IntQs from ma-Qs.

4.2.2.1 “Big Wave” – Global F0 Trend

Past studies failed to reach an agreement on the global trends of Mandarin tunes, partly due to the vastly different research approaches and goals. There are mainly two disputes over the declarative intonation and IntQ intonation – (a) whether there is declination in declarative intonation; (b) whether the IntQ intonation rises.

Declarative Intonation: Decline or Not

Declination refers to a gradual downward tendency of the global F0 during the course of an utterance, which has been found in many languages (Cohen & Hart, 1965; Ladd, 1984; Pike, 1945; Shih, 1997). In Mandarin, whether utterances have a declination effect attracts much debate.

One view is that declination is incidental. Tseng (1981) studied the spontaneous and reading speech of Taiwan Mandarin. The result of this study showed that only 20% of the utterances in spontaneous speech had declination effect. She therefore argued that declination effect was not common in Mandarin. Xu (1999) supported this view by analysing the lab recorded declarative utterances with neutral focus of eight native standard Mandarin speakers. The downtrend of the F0 of an utterance with all high tones was only 0.9 Hz. The more L tones an utterance had, the more the sentence

declined. Xu therefore argued that it was not an overall declination effect, but a downstep of low tones.

However, a number of other acoustic studies show that declination does exist in Mandarin. Shih (1997, 2000) tested four Mandarin speakers, two speaking mainland Mandarin and two speaking Taiwan Mandarin, using two utterances with different focuses. The lexical tones of the first utterances were Low (214) + Rising (35) + High (55) + High (55) + High (55), while the second utterance had a final neutral-tone particle to make the target HHH tone non-final. The results showed a global declination effect, with a greater declination at the beginning. Although this study has been taken as counterevidence against Xu's (1999) study, it did not actually differ from Xu's conclusion, since the beginning of the utterance had a low tone which may have induced a greater declination effect as in Xu's study.

M. Wang and Lin (2003) studied declination for different intonation types in a Standard Mandarin telephone conversation corpus. They found that more than 80% of the declarative sentences, wh- questions, and ma-Qs showed declination trend. Only IntQs had 26% of rising global F0. Judging from the topline⁷, both wh- questions and ma-Qs had steeper declination than statements. They used spontaneous speech as Tseng (1981) did with Taiwan Mandarin, but obtained extremely different results from Tseng (1981).

A. Wang, Chen, and Lv (2004) investigated a less spontaneous speech corpus. They examined 78 standard Mandarin declarative utterances from a broadcast speech

⁷ Topline: A line that connects all local F0 peaks.

database, recorded by one female speaker. They measured the baseline⁸ of the utterances and found a declination (more than 50Hz difference) effect on most utterances. In their data, the baseline was reset at intonational phrase boundaries. They further agreed with J. Shen (1992, 1994) that the topline and the baseline measured different properties – the topline was responsible for stress and the baseline was for rhythm. B. Wang (2002) supported this division with evidence from acoustic and perception experiments.

Yuan and Liberman (2010) did their research on a bigger Mandarin broadcast speech corpus and compared it with an English broadcast corpus. They selected 8383 Mandarin utterances and 5652 English utterances. Four main conclusions were drawn from the study. First, there was a strong correlation between declination slope and utterance length – the shorter the utterance, the steeper the declination. Shih (1997, 2000) also detected a correlation between length and F0 pitch – the shorter the sentence, the lower the initial tone and the higher the last tone. Shih (2000) further investigated the relationship between utterance length and declination slope, but did not find any correlation. The second conclusion in Yuan and Liberman (2010) was that the correlation holds even when the initial and final 500ms were excluded in order to eliminate the effects of initial rising and final lowering. They claimed this correlation indicated there was pre-planning of declination in speech production. The third conclusion was that both the topline and the baseline showed declination in Mandarin, with the baseline close to a straight line. This conclusion contradicted the study by A. Wang, Chen, and Lu (2004). Both studies nevertheless supported that toplines and baselines were independent of each other.

⁸ Baseline: A line that connects all local F0 valleys.

YNQ intonation: Rise or not

Most studies in Standard Mandarin agree that IntQ has a rising slope, whereas ma-Q does not necessarily rise. However, they differ in terms of how and to what extent the IntQ intonation rises.

X. S. Shen (1990) found that both types of YNQs had a higher pitch at the beginning than their statement counterparts. As shown in Figure 4.2, for IntQs (Tune 2), both the beginning and the end were raised; while for ma-Qs (Tune 3), the end was almost only as high as the beginning of the corresponding statement tune. Ni & Kawai (2004) experimented with the same set of materials and got the same conclusion as X. S. Shen (1990).

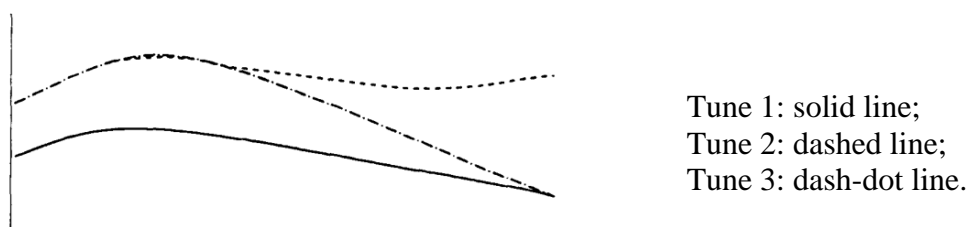


Figure 4.2 Tunes in Standard Mandarin (X. S. Shen, 1990)

Gårding (1984, 1987) analysed intonation using “grid” and “pivot”. She generalised the YNQ tune in Standard Mandarin as a rising ‘grid’. Many argued against this ‘grid’ model – for instance, Shih (1988) contended that one uniformed grid failed to explain all the variations brought by the different lexical tones in Mandarin.

Liu and Xu (2005) elicited YNQs with sequences of syllables of the same lexical tones in Standard Mandarin and concluded that there was exponential or double-exponential

raising for the questions and argued that accelerated final rise was merely a part of the global pitch raise.

While the idea of rising intonation for IntQ prevails, some studies did not fully support this claim. In Shi's (1980) early impressionistic description of Mandarin statement and IntQ tunes, IntQ was not described as rising; instead, IntQ was depicted as having both a higher beginning and a higher ending than statements. Shen (1985) observed that, in Beijing Mandarin, the IntQ's bottom line got raised, while the topline did not – this was different from IntQs in such non-tonal languages as English or Russian. Zeng, Martin, and Boulakia (2004) showed that IntQs had a rising slope; however, it was not significant for the two rising tones, Tone 2 (LH) and Tone 3 (LLH), which indicated that the IntQ tune did not rise much at all. This reflects that Standard Mandarin, with only one falling tone, sometimes is more representative of the lexical tones which ends in H, and the falling tones and low tones are hard to detect. This is one reason why Tianjin Mandarin was chosen as an example of contour tonal language (for further discussion, see Chapter 3, pp. 51-55).

4.2.2.2 “Small Ripple” – Local F0 Events

Previous studies of Mandarin intonation mainly discussed the local F0 differences between YNQs and statements in terms of boundary tones and pitch accents. Boundary tones were discussed in the existing studies of Mandarin from the perspectives of whether the boundary tones of YNQs were realised as raised beginnings and/or raised endings, as well as what the associated unit for the boundary tones in Mandarin is. Pitch accents are mainly elicited by the sentence focus in Mandarin, due to the association

with the prominence of the utterance (see Chapter 2 for detailed definitions on *pitch accent* and *prominence*).

Boundary tones

Lin (2004, 2006) examined IntQs in a telephone conversation corpus and found that the registers of the starting and ending points of the last non-neutral lexical tones were higher. He thus believes that the boundary tone in Standard Mandarin is the most crucial component for IntQ tune. Lee (2005) elicited conversation data and concluded that the boundary tone at the last NP of IntQs got raised in Beijing Mandarin. X. S. Shen (1990) not only observed a final boundary tone for IntQs, but also proposed that the first few syllables of YNQs – both IntQ and ma-Q – had higher boundary tones than their corresponding statement intonations. However, although some believe that the boundary tone in IntQ tune gets raised, it does not mean that the boundary tone is a H%, since in Mandarin there are lexical tones. Especially for a falling tone, the boundary tone usually remains low, despite the fact that it is raised. This point will be further discussed in Section 4.4.3.3.

Focus

Liu and Xu (2005) investigated the intonation of Mandarin statements and five types of questions (including IntQs⁹ and ma-Qs¹⁰) with neutral and narrow foci. Their results showed that (i) the pitch range of the focused words was expanded; (ii) the post-focused words were compressed and lowered; (iii) the pre-focus words were largely unaffected. They also examined whether the location of the focus made any difference. In a

⁹ IntQs are referred to as “yes/no question” in Liu and Xu (2005).

¹⁰ Ma-Qs are referred to as “particle question” in Liu and Xu (2005).

question with initial focus¹¹, the whole pitch level was lifted, compared with the corresponding statements. For sentences with medial focus, the questions had a moderate lift in F0 range starting from the focused words, compared with the F0 range of statements. The intonation of the sentences with final focus resembled that of neutral focus the most – sentence type had a larger influence on the different F0 contours of questions and statements. This result is in accordance with Xu and Xu's (2005) study on English declarative intonation.

Chen and Gussenhoven (2008) studied the effect of different levels of emphasis on focus duration and pitch range change. They found the durational change between the NoEmphasis condition and the Emphasis condition was comparable to that between the Emphasis condition to MoreEmphasis condition. The F0 expansion was bigger between the NoEmphasis and the Emphasis conditions than that between the Emphasis condition and the MoreEmphasis condition. Interestingly, they considered the post-focus intonation to be a “lack of sharp or precise F0 contours that are characteristic to the lexical tones” rather than what Liu and Xu's (2005) argument of post-focus compression. The two proposals described the same phonological phenomenon of the post-focus pitch contour being flatter.

Duan and Jia (2014) studied narrow focus realisation of declarative sentences in five northern Mandarin dialects, including Dalian, Harbin, Jinan, Xi'an, and Tianjin Mandarin. For Dalian, Harbin, Jinan, and Xi'an Mandarin, F0 range was raised on the focused words, with narrow focus being raised higher than neutral focus. The F0 of

¹¹ Initial focus: The location of the focus word is at the beginning of the sentence. Medial and final focuses are located in the middle and final parts of the sentence respectively.

narrow-focus sentences then fell rapidly in the post-focus region until the end of the sentence, while neutral-focus sentences remained in their original shape. The pre-focus region of narrow focus remained largely the same within neutral-focus sentences. The results for these varieties of Mandarin were in conformity with Liu and Xu's (2005) study on Standard Mandarin.

However, although the conclusions from the other dialects largely held true, Tianjin Mandarin had different results from the other four dialects when comparing the narrow-focus sentences with the neutral-focus sentences. The pre-focus regions were different in narrow-focus sentences and neutral-focus sentences in Tianjin Mandarin, as shown in Figure 4.3, while the other dialects had almost the same contours for the two types. In the post-focus region, Tianjin Mandarin had the first syllable of the narrow-focus sentence higher than that of the neutral focus, while in other dialects, the narrow-focus F0 was lowered from the first syllable onwards, compared with the neutral-focus F0. Moreover, the second and third syllables (the syllables in the dotted box in Figure 4.3) were largely different between the two types in Tianjin Mandarin, while the major difference lay in the F0 height in other dialects.

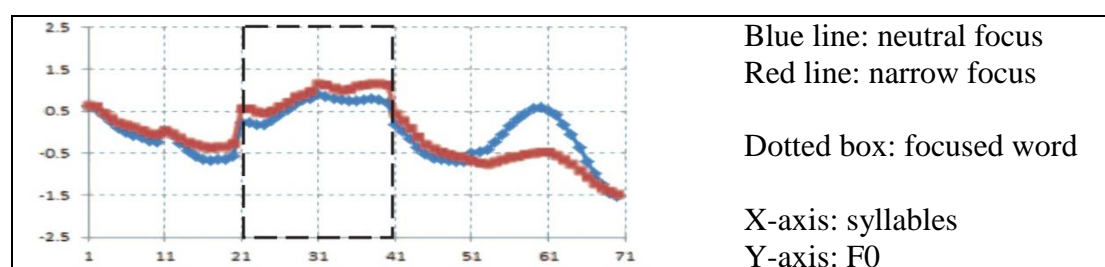


Figure 4.3 Neutral and narrow focus in Tianjin Mandarin: Normalised F0

(Duan & Jia, 2014)

The previous studies on Standard Mandarin and its other varieties provided us with a general impression of how intonation works in a tonal language, from both the perspective of global F0 trend and the perspective of local F0 events. In the remaining sections of this chapter, two production studies were therefore designed to examine the IntQ tune in Tianjin Mandarin in order to scrutinise the similarities and differences between the IntQ tunes in Tianjin Mandarin and Standard Mandarin, and to fit the tunes in Tianjin Mandarin, or even Mandarin in general, in AM framework.

4.3 General Methodology

4.3.1 Speakers

Six native speakers of Tianjin Mandarin (3 male and 3 female) were recorded. All speakers were born and raised in the city area of Tianjin (for map, see Figure 3.2, p.34), and spoke Tianjin Mandarin on a daily basis. The age range of the informants was between 22 and 50. All were capable of speaking Standard Mandarin in formal situations, though with different degrees of Tianjin accents. None of them reported any speech disorders. All informants were given financial compensation for their time.

4.3.2 Procedures

Recordings were made in a soundproof booth in the Language and Brain Laboratory at the University of Oxford, United Kingdom and in a quiet room at Nankai University in Tianjin, China. The reading was recorded using a Rode NT-USB Microphone with Audacity onto a PC, at a sampling rate of 44.1 kHz.

The informants were given time to briefly familiarise themselves with the materials before the experiment started. The test materials were presented as Chinese characters

without context in a Microsoft PowerPoint presentation. The informants were asked to read from the screen as naturally as they could, by producing a statement when seeing a Chinese full stop “。” at the end of the utterances, and an IntQ when seeing a Chinese question mark “?” . For the IntQs, they were told to imagine that the experimenter had made the corresponding statement and they were checking whether it was what the experimenter had said. For example, when they saw “麻” ([ma], H Tone, *hemp*) on the screen, the informants were asked to respond with a non-syntactically marked question with the intention of asking the experimenter to confirm whether this was the word the experimenter had uttered. The intonations of information-seeking IntQs and confirmation IntQs do not have perceivable differences in Tianjin Mandarin.

Each token was repeated 3 times and all the materials were randomised so that the three repetitions did not follow each other. To avoid any potential list effect, each utterance was placed on a separate slide. The experimenter controlled the pace of the experiment, leaving a short pause after each utterance. The words and sentences were placed in the middle of the screen to reduce head movement. Each item was repeated three times. The Mono_(ISO) task lasted approximately 10 to 20 minutes, and the Mono_(SEN), for 20 to 25 minutes.

4.3.3 Data Processing and Analyses

The majority of existing studies under the AM framework study non-tonal languages, with a few on level-tone languages. Contour-tone languages differ from non-tonal languages and level-tone languages in that the contours of lexical tones are important besides the H and L levels of their intonational events. However, since both the lexical

contours and intonational events are both encoded in pitches, it is extremely hard to tease them apart.

Few studies, if any, described the methods of processing production data of intonation in contour-tone languages in detail, which makes the reproducibility of the studies extremely varied, as well as difficult. In this section, I will present a step-by-step description of how the production data were processed.

Step 1: Segmentation

Two out of three repetitions were randomly selected. The acoustic data were manually segmented and annotated in Praat (Boersma & Weenink, 2014) according to the waveforms and spectrograms. Monosyllabic data were analysed first, and multiple tiers were included in the annotation file. Since the tone-bearing unit in Mandarin is commonly recognised as the syllable or the rhyme (Yip, 1989), a syllable tier and an onset-rhyme tier were specified. For sentence data, the pre-nuclear, nuclear, and post-nuclear parts were divided. The onsets and rhymes of the nucleus were also segmented in a different tier.

Step 2: Visual Inspection

In order to establish what the key points of analyses were, visual inspection was first conducted in Praat (Boersma & Weenink, 2014). Some small artefacts were observed at the end of both IntQs and statements (as the examples from Speaker B's data in Figure 4.4), they were considered as errors caused by occasional creakiness or breathiness at the end of utterances, instead of distinct boundary tones that distinguish IntQs from statements. Apart from the artefacts, the only visual difference between

IntQs and statements in Tianjin Mandarin is register difference. No extra meaningful interpolated tones could be observed in both studies in the current chapter.

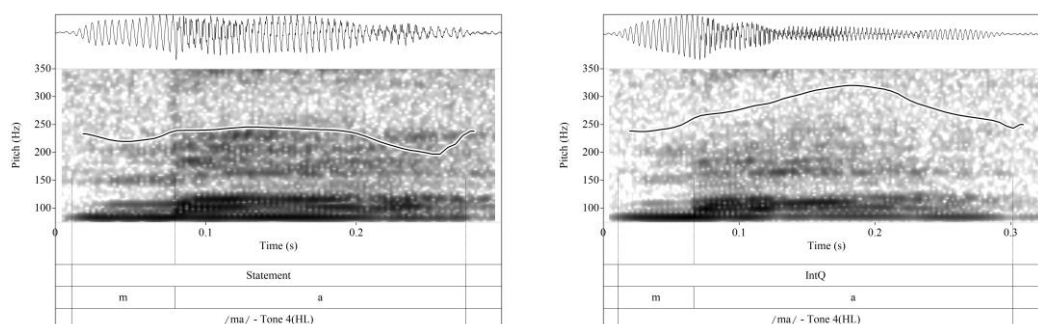


Figure 4.4 Typical artefacts at the end of a statement (left) and an IntQ (right)

Step 3: Measurement

Data were extracted with the help of various Praat scripts and manually checked to exclude artefacts. In Praat, the pitch range was set as 75Hz-500Hz for all sound files. Although male and female voices have different ranges, it is important to keep the lower limit consistent so that it has consistent analysis windows (Intro 4.2, Boersma & Weenink, 2014).

Obstruents and the creakiness of the LH tone (see Figure 4.5) often cause errors and undetectable pitch values, which is frequently an issue in both Tianjin Mandarin and Standard Mandarin prosody research (Chao, 1968; Davison, 1991a). Together with the artefacts detected during visual inspection, all such data were taken as missing data points during the analyses.

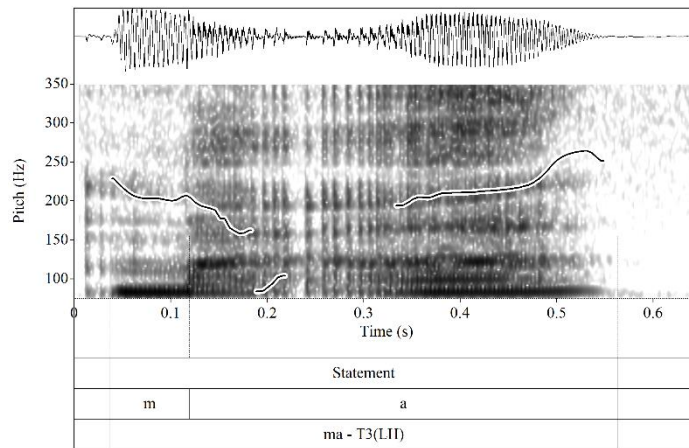


Figure 4.5 Creakiness of Tone 3 (LH Tone)

The data was analysed from four aspects: duration, register, pitch accent, and edge/boundary tone, since these are important elements of intonation changes across different types of tunes. For **duration** analysis, the lengths of the monosyllabic syllables, as well as the onsets and rhymes respectively, were measured. With regard to **register** change, I studied F0 scaling by calculating the mean F0 values to examine the overall register change and the F0 maxima and minima, to understand whether the pitch register was raised or expanded. The existence of **pitch accent** was investigated by means of identifying the alignments of the pitch maxima and minima. Pitch accents are difficult to quantify – even in intonational languages which do not have word-level pitch manifestations, pitch accents are often derived directly from visual inspection. This thesis used the tonal alignment, i.e. the absolute distance from the F0 extrema to the rhyme onsets. Studying where the F0 extremas locate in relation to a whole tone-bearing rhyme provided us with the information of whether the peaks and valleys of the F0 contours had significant movements across intonation types, which consequently enabled us to understand whether the lexical tones were overridden by intonational tunes. Finally, in order to ascertain whether there was any **edge or boundary tone**, the

conclusions from visual inspection were verified by analysing the F0 range change (a.k.a. pitch excursion), since it supplied us with the information of directionality of the pitch change across tones and tune types.

All F0 data were converted to ERB (Equivalent Rectangular Bandwidth) units for statistical analysis. ERB is a scale that was derived from human auditory selectivity and is more representative of human perception than Hertz (Nolan, 2003). It also makes the data from male speakers and female speakers more comparable (Daly & Warren, 2001). The formula used here is as follows (Hermans & van Gestel, 1991):

$$\text{ERB} = 16.7 \log_{10} \left(1 + \frac{f}{165.4} \right)$$

Step 4: Statistical Analysis

Currently, the best statistical model for analysing data from production studies is the mixed-effect model since production studies include many factors. In this study, the fixed factors were lexical tones, and intonational tune types, as well as rhyme variations. Informant variation and repetition of the items were taken as random factors. More detailed statistical analysis methods will be later elaborated in the method section of each study.

4.4 Production Study 1: Monosyllabic Words in Isolation (Mono_(ISO))

The goal of this study is to (i) present the citation tone contours (i.e. statement tune) of all four lexical tones in Tianjin Mandarin; (ii) to investigate the IntQ tune for monosyllabic words in isolation (hereinafter abbreviated as **Mono_(ISO)**).

The organisation of the current section (Section 4.4) is as follows: Section 4.4.1 presents the specific methods for this production study of Mono_(ISO). Section 4.4.2

reports the results for Mono_(ISO): duration (§4.4.2.1), register (§4.4.2.2), pitch accent (§4.4.2.3), and edge and boundary tone (§4.4.2.4). Section 4.4.3 analyses and discusses the results. Finally, Section 4.4.4 concludes the whole Section 4.4 with a summary of all results and analyses for Mono_(ISO).

4.4.1 Methods

4.4.1.1 Materials

12 monosyllabic words in isolation (Table 4.2), which consisted of three syllables ([ma], [mi], [mao]) with four different tones (L, H, LH, HL), were recorded as both statements and IntQs. The statement tune for isolated monosyllables is their citation lexical tones.

Table 4.2 Mono_(ISO) words

SYLLABLES	L TONE (TONE 1)	H TONE (TONE 2)	LH TONE (TONE 3)	HL TONE (TONE 4)
[ma]	妈 'mother'	麻 'hemp'	马 'horse'	骂 'scold'
[mao]	猫 'cat'	毛 'fur'	铆 'rivet'	帽 'hat'
[mi]	咪 'meow'	迷 'riddle'	米 'rice'	蜜 'honey'

4.4.1.2 Procedures and Informants

Six native Tianjin speakers participated in this study. Detailed information about the informants and procedures were provided in Section 4.3.1 and 4.3.2.

4.4.1.3 Statistical Analyses

Statistical analyses in this section were conducted using R (R Core Team, 2015), and a mixed-effect model was used to analyse the relationship between various independent variables (duration, mean pitch, F0 range, etc.) and Tune TYPE and TONE. TYPE (Statement, Question), TONE (L-Tone, H-Tone, LH-Tone, HL-Tone), TYPE:TONE interaction, RHYME ([a], [aʊ], [i]), and GENDER (Male, Female) were taken as fixed-effect factors. SUBJECT (6 different informants) and ITEM (24 tokens) were held as random-effect factors, with intercepts for both SUBJECT and ITEM, as well as by-SUBJECT random slopes for the effect of TYPE.¹² The factors that are important to this study are TYPE, TONE, and their interactions – TYPE*TONE. The focus of this section therefore is mainly on reporting these results. The error bars in all figures under Section 4.4.2 are ± 1 standard error.

4.4.2 Results

The four lexical tones in Tianjin Mandarin are L (Tone 1), H (Tone 2), LH (Tone 3), and HL (Tone 4), as described in Chapter 2. Figure 4.6 to Figure 4.9 are instances from Speaker A (female) for all four tones in both declarative and IntQ tunes. Comparing the figures on the left (statements) and the ones on the right (IntQ), together with figures of other data collected, we cannot detect any visually clear and systematic difference between the IntQ tune and the statement tune.

According to AM theory, tunes are made up of three major elements – pitch accents, phrase accents, and boundary tones. Figures 5.2 to 5.5 show that IntQs keep the same contours as the statements. Unlike many non-tonal languages, in which the IntQ

¹² Models did not converge with by-ITEM random slopes.

contours are dramatically different from the contours of the statements, in Tianjin Mandarin, no pitch accents could be discerned from the F0 contours. Edge tones, which are below intonational phrase (IP) level, and final IP boundary tones cannot be detected simply from the figures either.

The only visual difference lies in the register of the tones. IntQs are observed to have higher overall register than the corresponding statements. In that case, is register the sole factor that distinguishes IntQs from statements? How otherwise do native Tianjin speakers tell whether this is a question or a statement? In the following sections, precise temporal and pitch data are analysed to confirm the observation of the figures.

The following subsections of Section 4.4.2 report the results from different perspectives. They are organised in the order shown in Table 4.3:

Table 4.3 Order of presentation of results: Mono_(ISO)

Section Number	Property of Intonation	Measurements
§4.4.2.1	Duration	<ul style="list-style-type: none"> • Length of onsets and rhymes
§4.4.2.2	Register	<ul style="list-style-type: none"> • Mean Pitch • Tonal Scaling: Values of F0 Maxima and Minima
§4.4.2.3	Pitch Accent	<ul style="list-style-type: none"> • Tonal Alignment: Distance of F0 Maxima and Minima to Rhyme Onsets
§4.4.2.4	Edge/ Boundary Tone	<ul style="list-style-type: none"> • F0 range

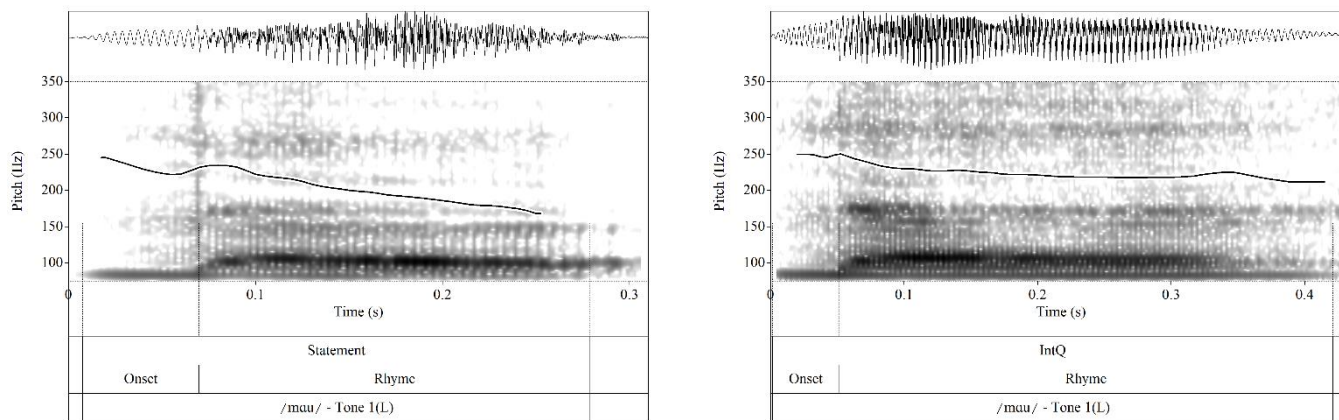


Figure 4.6 Statement (left) and IntQ (right): Mono_(ISO) – Tone 1 (L Tone)

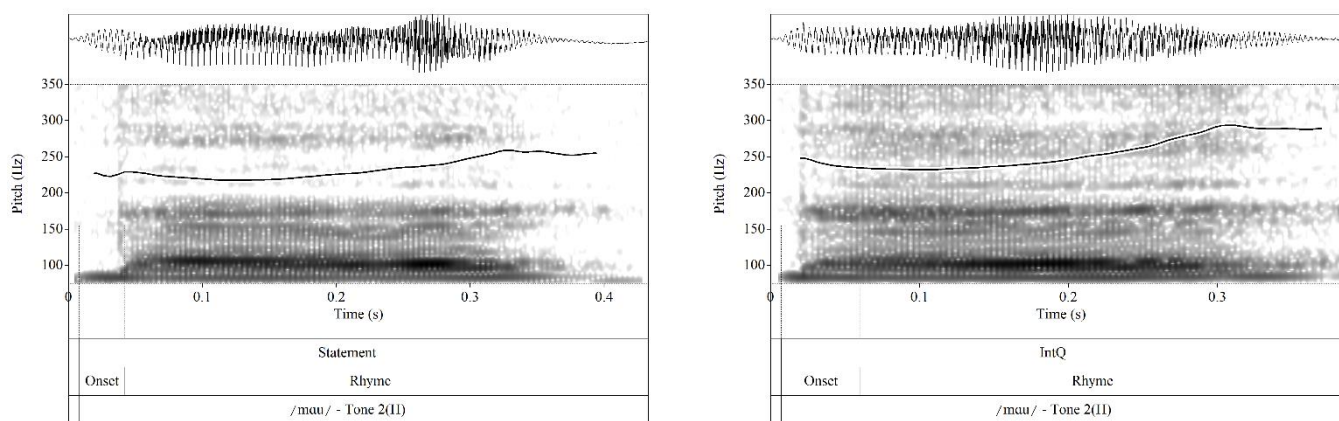


Figure 4.7 Statement (left) and IntQ (right): Mono_(ISO) – Tone 2 (H Tone)

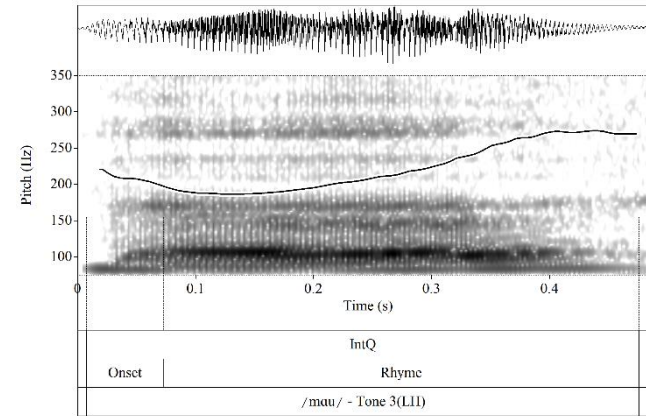
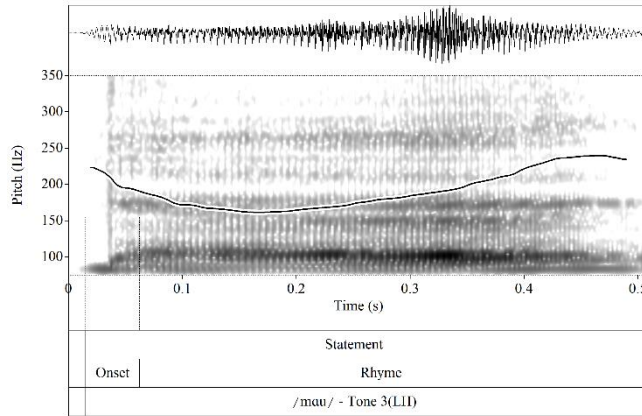


Figure 4.8 Statement (left) and IntQ (right): Mono_{ISO} – Tone 3 (LH Tone)

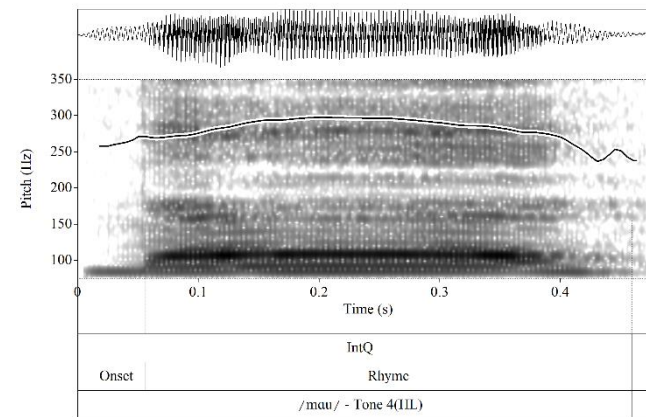
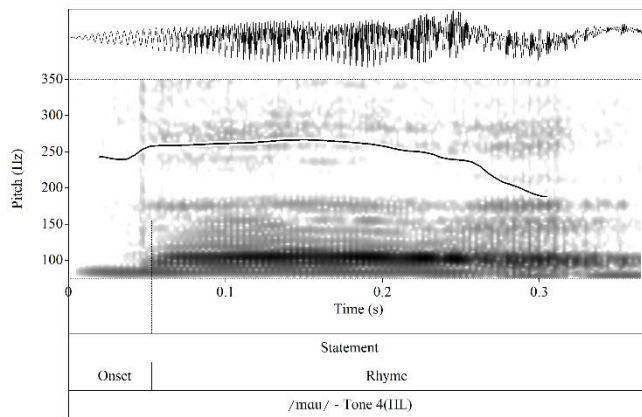


Figure 4.9 Statement (left) and IntQ (right): Mono_{ISO} – Tone 4 (HL Tone)

4.4.2.1 Duration

The average duration¹³ of the **whole Mono_(ISO) words** was longer in IntQs than in statements, except for the Tone 2 (H Tone) which had a longer statement than IntQ by an insignificant margin. The durational difference between Mono_(ISO) statements and IntQs was affected by TYPE ($\chi^2(1) = 4.$, $p < 0.29^{**}$), TONE ($\chi^2(3) = 100.49$, $p < 0.001^{***}$), and TYPE*TONE ($\chi^2(3) = 34.84$, $p < 0.001^{***}$). The average duration of the Mono_(ISO) statements was 363.6ms ($\overline{SD} = 66.8$), while the average duration of IntQ Mono_(ISO) words was 388.9ms ($\overline{SD} = 72.9$).

However, for **Mono_(ISO) onsets**, the duration of statements was marginally longer than that of IntQs, as shown in Figure 4.10. The duration for onsets was influenced by TYPE ($\chi^2(1) = 4.03$, $p = 0.045^{**}$), TONE ($\chi^2(3) = 15.90$, $p < 0.00012^{**}$), without the interaction between TYPE and TONE. Nevertheless, **Mono_(ISO) rhymes** (Figure 4.11) behaved the same as the whole syllable, with IntQs' duration longer than statements'. They were also influenced by TYPE ($\chi^2(1) = 5.58$, $p < 0.018^*$), TONE ($\chi^2(3) = 106.96$, $p < 0.001^{***}$), TYPE*TONE ($\chi^2(3) = 34.93$, $p < 0.001^{***}$).

The average duration of **Mono_(ISO) onset** was 67.5ms and 61.5ms in statements and IntQs respectively, while that of **Mono_(ISO) rhymes** was 295.0ms and 327.4ms in statements and IntQs respectively. The rhymes were approximately four to five times as long as the onsets.

¹³ RHYME is also a significant factor that influences duration change: on syllable level, $\chi^2(3) = 25.27$, $p < 0.001^{***}$; on onset level, $\chi^2(2) = 13.06$, $p = 0.0015^{**}$; on rhyme level, $\chi^2(2) = 17.37$, $p < 0.001^{***}$. This is obvious since different rhymes have different durations.

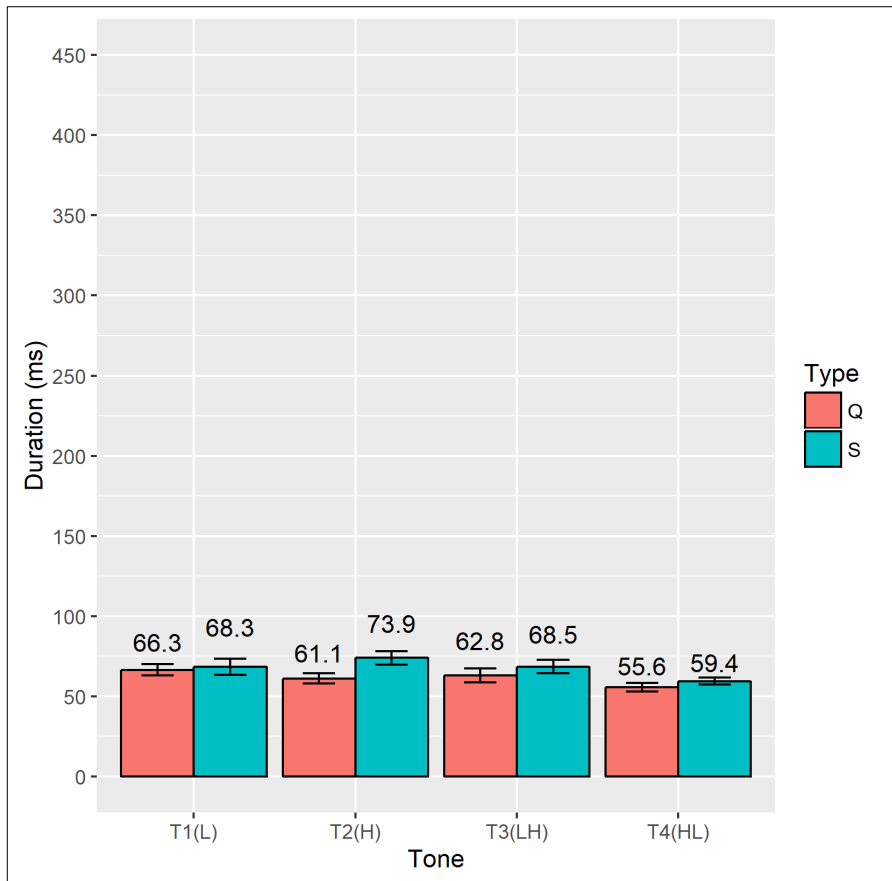


Figure 4.10 Duration: Mono(ISO) onsets

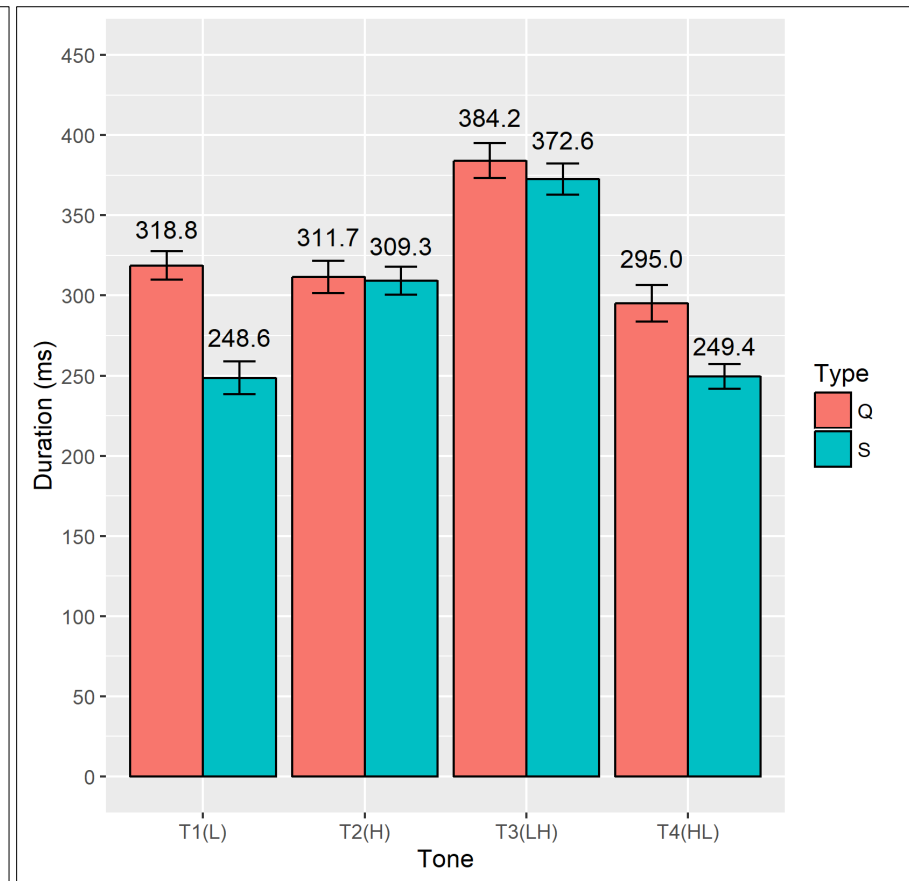


Figure 4.11 Duration: Mono(ISO) rhymes

4.4.2.2 Register

Mean Pitch

The mean pitch¹⁴ of the IntQs of **Mono(ISO) words** was significantly higher than that of the statements, TYPE ($\chi^2(1) = 18.47$, $p < 0.001^{***}$). TONE was also a factor that influenced mean pitch significantly, TONE ($\chi^2(3) = 105.1$, $p < 0.001^{***}$). On average, the mean pitch of IntQ was higher than statements by 0.65ERB.

The mean pitch of the **Mono(ISO) rhymes** had a very similar pattern with that of the whole word, as shown in Figure 4.13, since rhymes took up the greatest proportion of the whole syllables. The mean pitch of the rhymes was also affected by TYPE and TONE separately, TYPE ($\chi^2(1) = 18.04$, $p < 0.001^{***}$), TONE ($\chi^2(3) = 105.17$, $p < 0.001^{***}$).

They also had the same effect on **Mono(ISO) onset** (Figure 4.12), TYPE ($\chi^2(1) = 11.44$, $p < 0.001^{***}$), TONE ($\chi^2(3) = 150.04$, $p < 0.001^{***}$). Besides, onset has a significant interaction effect between TYPE and TONE – TYPE*TONE ($\chi^2(3) = 24.90$, $p < 0.001^{***}$). This significant interaction was owing to the fact that the mean pitches of Tone 2 (H Tone) and Tone 3 (LH Tone), which both started with lower onsets than the offsets, hardly changed across types; while the mean pitch of Tone 1 (L Tone) and Tone 4 (HL Tone), the two tones with higher onsets than offsets, raised more for IntQs than for statements. This demonstrates that the tones which started with Hs did not change much at the onset when turned into an IntQ, while the tones starting with Ls changed more.

¹⁴ Mean pitch is also significantly affected by GENDER and RHYME. On syllable level, GENDER ($\chi^2(1) = 11.70$, $p < 0.001^{***}$), RHYME ($\chi^2(3) = 9.68$, $p < 0.02^*$); on onset level, GENDER ($\chi^2(1) = 7.14$, $p = 0.008^{**}$), RHYME ($\chi^2(2) = 6.00$, $p = 0.0508$); on rhyme level, GENDER ($\chi^2(1) = 12.81$, $p < 0.001^{***}$), RHYME ($\chi^2(2) = 8.13$, $p = 0.017^{**}$). Different rhymes ([a], [aʊ], [i]) naturally have different pitch heights. Gender (Male, Female) also influences the pitch height which is consistent with previous findings and common knowledge that women's pitches are higher than men's.

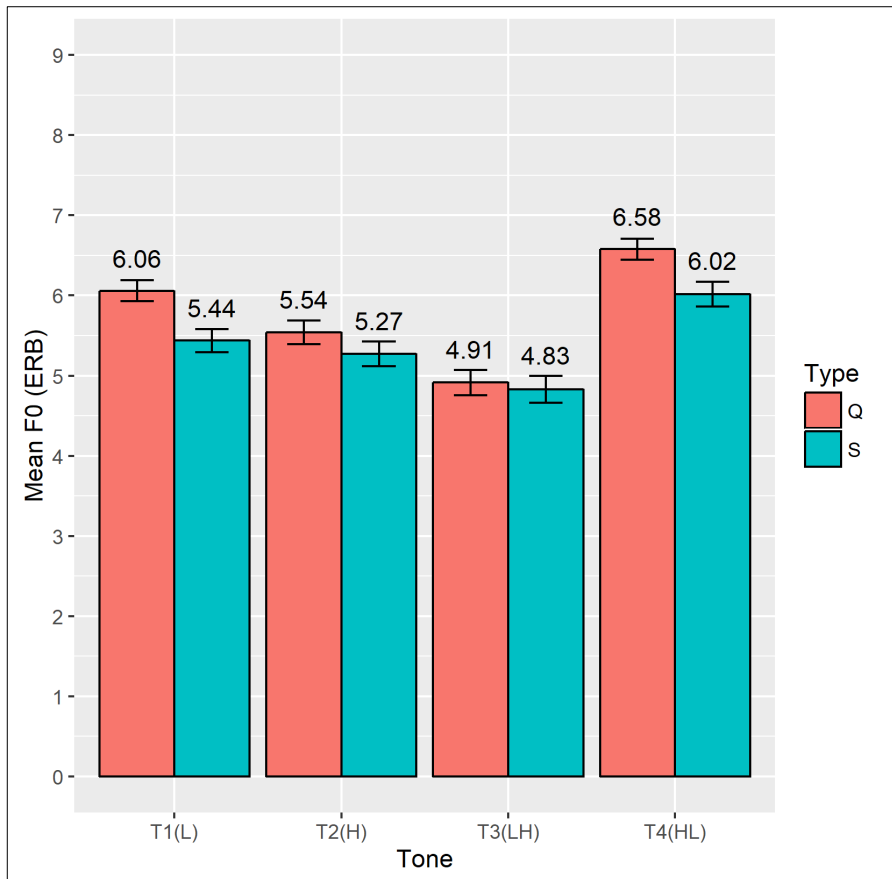


Figure 4.12 Mean Pitch: Mono_(ISO) onsets

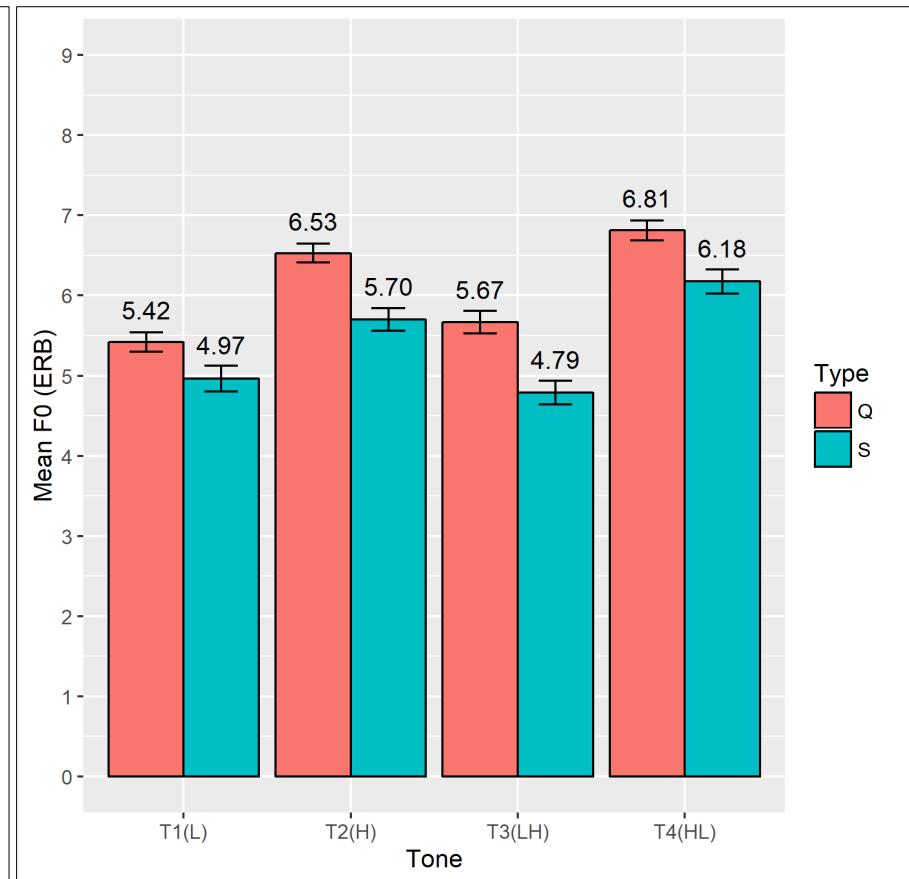


Figure 4.13 Mean Pitch: Mono_(ISO) rhymes

Mean pitches of all four lexical tones showed a higher mean pitch on the questions over the statements, which supported the fact that the **register** of questions was higher than statements. After examining onsets and rhymes separately, the significant results also indicated that the difference was not only in one segment but in the whole word. Different tones had different degrees of changes too. As shown in Figure 4.13, Tone 1 and Tone 3 were the lowest since L tone was the dominant tone in those tones, while Tone 2 and Tone 4 were higher since H tone was the dominant tone. Figure 4.14 illustrates the register differences with stylised lines. The vertical lines separate onsets and rhymes.

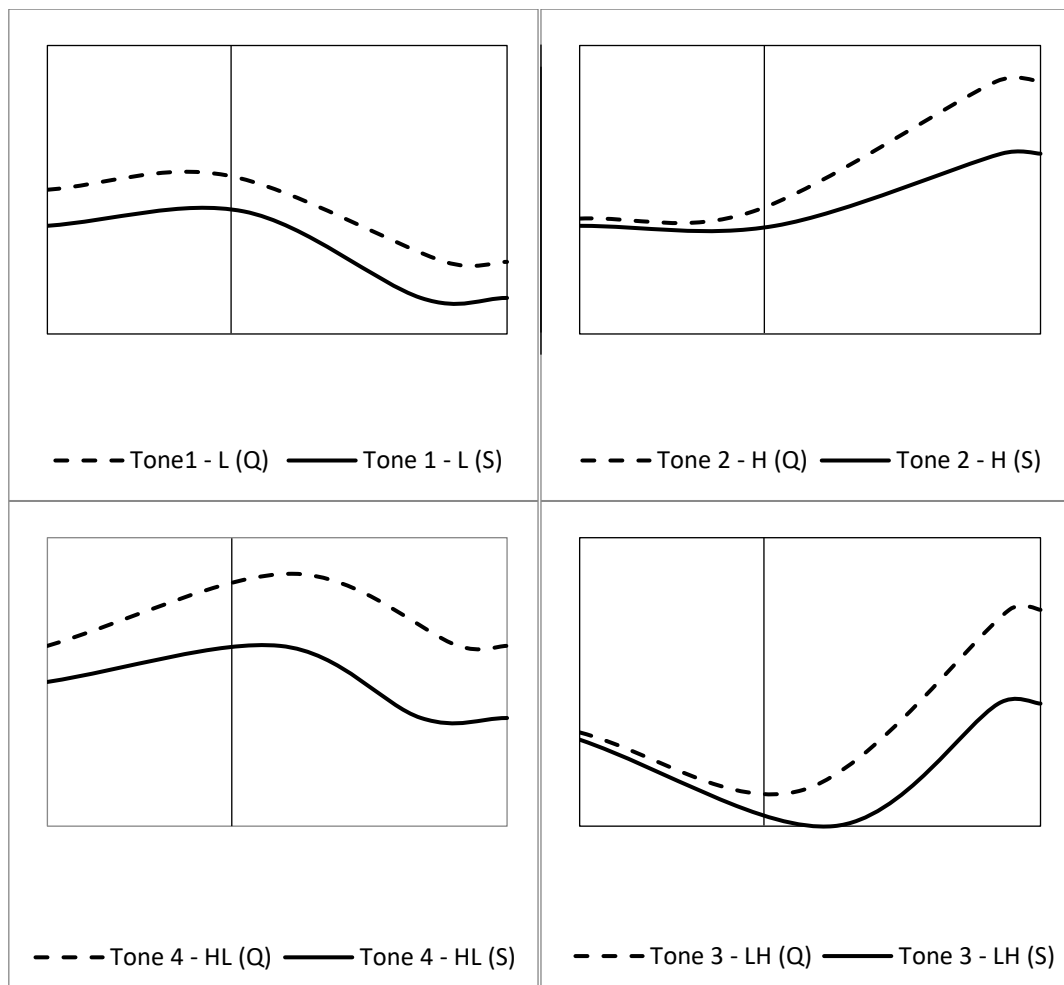


Figure 4.14 Stylised lines: Mono_(ISO) onsets and Mono_(ISO) rhymes

Tonal Scaling: Values of F0 Maxima and Minima

Since the mean pitch data indicated a register difference, the values of maximum F0 and minimum F0 of all Mono_(ISO) rhyme were measured to investigate whether the register was raised, which means that both the maxima and the minima were raised; or the register was expanded, which means that only the maxima were raised. The maximum pitch and the minimum pitch were extracted using Praat (Boersma & Weenink, 2014) and manually checked. Figure 4.15 is an illustration of where the pitch maximum (red) and minimum (blue) within a rhyme are.

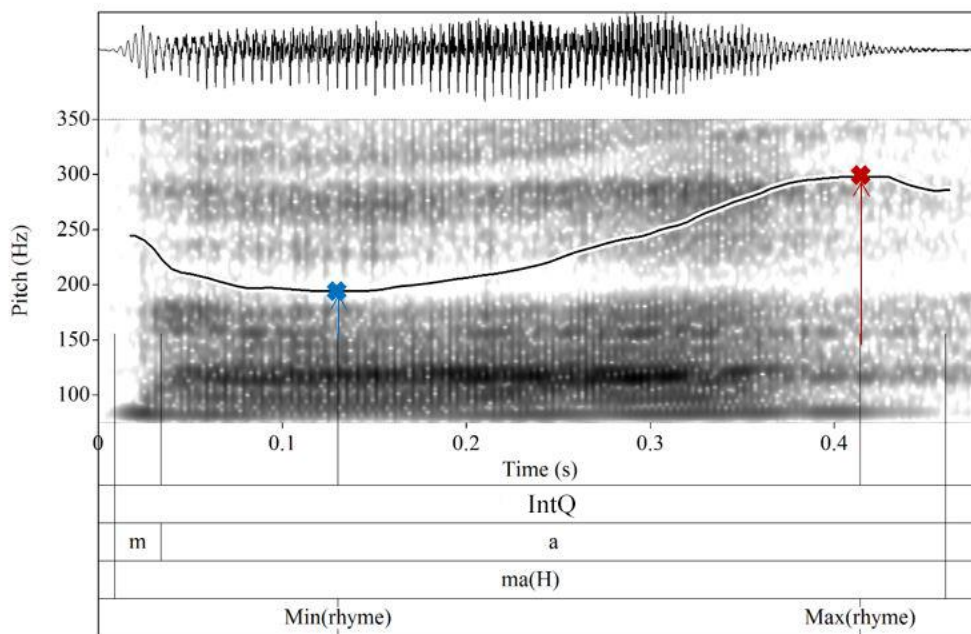


Figure 4.15 Measurements of F0 maximum and minimum in a rhyme

The results show that the register raises, rather than expanding (see Figure 4.16 for maximum pitches, and Figure 4.17 for minimum pitches). Regardless of the tones, the IntQs were always higher than the corresponding statements in terms of both minimum and maximum pitches of the rhymes (maximum pitch: TONE ($\chi^2(3) = 127.54$, $p < 0.001^{***}$), TYPE ($\chi^2(1) = 13.49$, $p < 0.001^{***}$), TYPE*TONE ($\chi^2(3) = 28.37$, $p <$

0.001***); minimum pitch: TONE ($\chi^2(3) = 124.33$, $p < 0.001$ ***), TYPE ($\chi^2(1) = 15.11$, $p < 0.001$ ***), TYPE*TONE ($\chi^2(3) = 24.03$, $p < 0.001$ ***)). The range of maximum pitch for the statements was 4.0ERB to 8.3ERB, while that for IntQs was 4.63ERB to 9.69ERB; minimum pitches ranged from 2.67ERB to 6.45ERB for the statements, while that for the IntQs lay between 2.73ERB and 7.82ERB. The maximum pitch of IntQs was higher than the statements by 1.03ERB on average; while the minimum was higher for 0.71ERB on average.

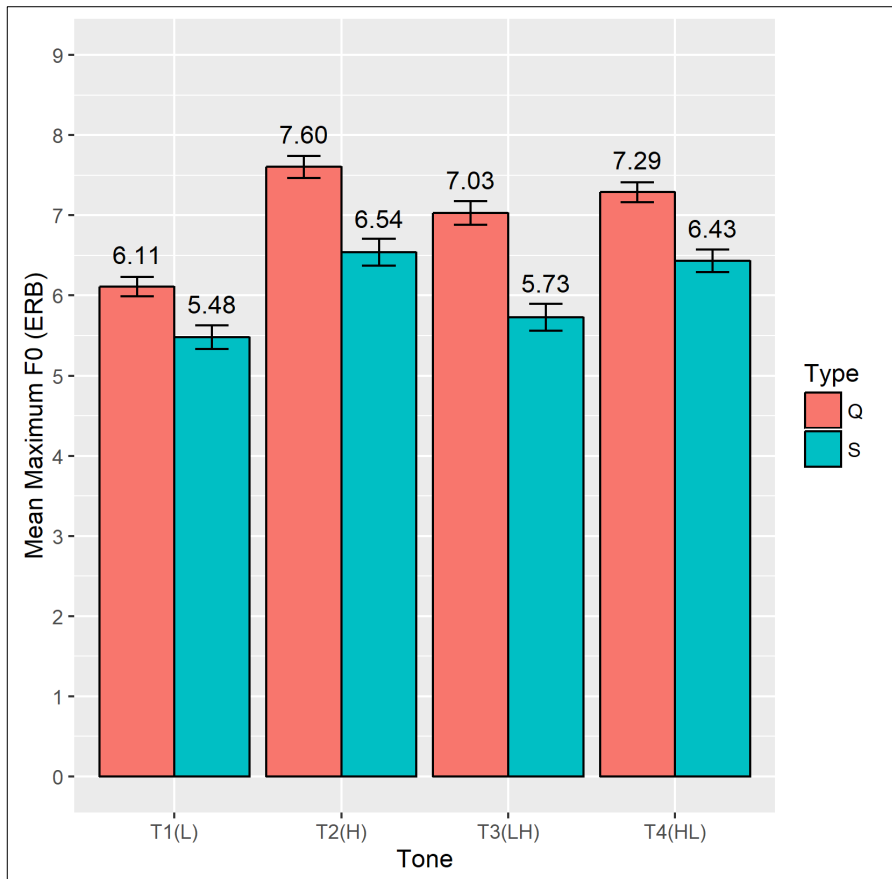


Figure 4.16 Average maximum pitch: Mono(ISO) rhymes

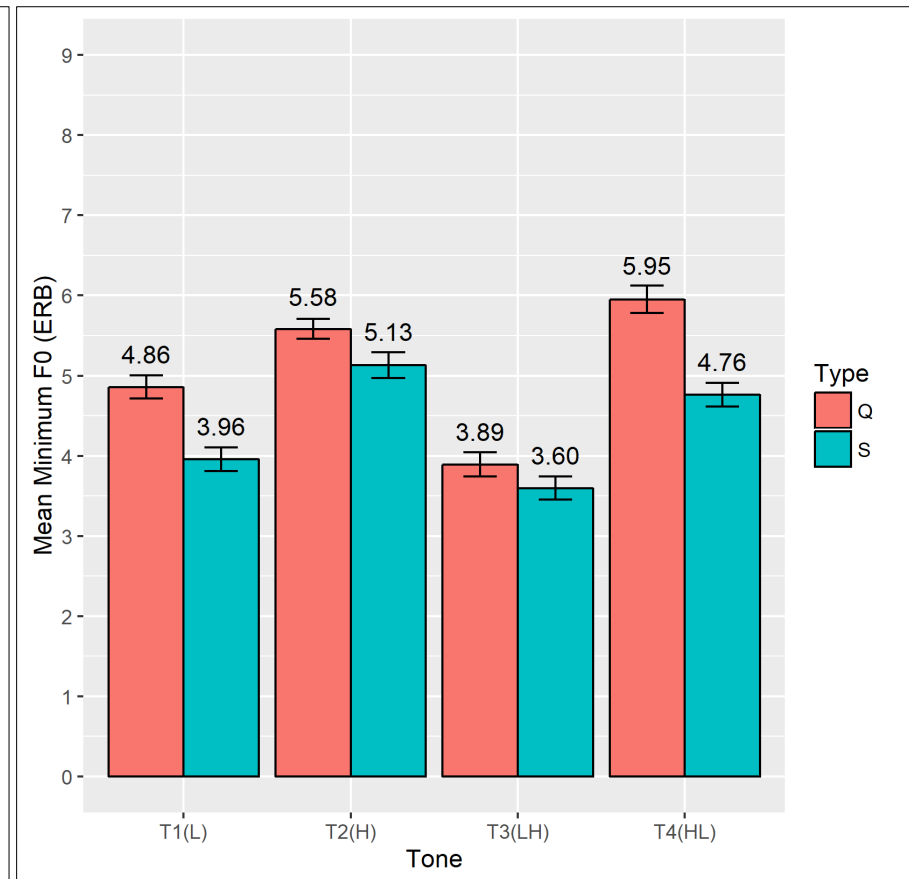


Figure 4.17 Average minimum pitch: Mono(ISO) rhymes

4.4.2.3 Pitch Accent

Tonal Alignment: Distance of F0 Maxima and Minima to Rhyme Onset

The mean pitch data, pitch maxima, and pitch minima sketch how the register behaves in statements and in IntQs. It is yet unknown whether there is any pitch accent overriding the lexical tones. From the observation of the pitch contours, we cannot find any clear pitch accents in the same way that we can easily detect them in many non-tonal languages. Therefore, the tonal alignment changes across tune types were analysed to investigate whether the tonal peaks and valleys were significantly moved, by mean of measuring the distance of pitch maxima and minima from the temporal onset point of the Mono_(ISO) rhymes, using the following formula:

Distance of Maximum Pitch (or Minimum Pitch) to Rhyme Onset =

$$\frac{\text{Time of Maximum Pitch (or Minimum Pitch)}}{\text{Duration of rhyme}} * 100\%$$

Table 4.4 suggests that the tonal alignment does not differ significantly between statements and IntQs. For Tone 1 (L Tone), the maximum pitch is at the beginning of the rhyme, while the minimum is towards the end. Tone 4's (HL Tone) maximum pitch is later than Tone 1, but the minimum resembles Tone 1. Tone 2 (H Tone) is the opposite of Tone 1 – the maximum pitch is at the end of the rhyme, while the minimum is at the beginning. The minimum pitch of Tone 3 (LH Tone) is later than Tone 2 but the maximum is similar to Tone 2.

Table 4.4 Alignment of F0 Maxima & Minima: Mono_(ISO)

	TONE	TYPE	MAX F0 DIST.	MIN F0 DIST.
<i>FALLING</i>	TONE 1 (L)	Q	1.69%	83.98%
		S	9.43%	85.02%
	TONE 4 (HL)	Q	35.06%	83.55%
		S	30.93%	76.34%
<i>RISING</i>	TONE 2 (H)	Q	90.15%	14.26%
		S	90.53%	23.03%
	TONE 3 (LH)	Q	93.52%	21.93%
		S	84.83%	28.39%

Whether the F0 maxima/ minima distances to the rhyme onsets differ significantly between different TYPES was tested with a mixed effect model (see Section 4.4.1.3, p. 75 for model details). For F0 maxima alignment, TYPE ($\chi^2(1) = 0.006$, $p = 0.94$, n.s.); for F0 minima alignment, TYPE ($\chi^2(1) = 0.028$, $p = 0.87$, n.s.). These results indicate that the pitch alignment does not change significantly across intonation types. The locations of the pitch minima and maxima of IntQ are very similar to those of statements. Lexical tones keep their contours regardless of what intonation tunes are. Combining the results from the previous section on the values of pitch maxima and minima with the finding in the current section, it can be concluded that there is no pitch accent in Mono_(ISO).

In Figure 4.18, I further stylised the contours, combining the mean pitch data, locations and values of pitch extremes. The red spots show where the pitch maxima are in relation to the duration of the rhymes, and the blue ones, pitch minima. The IntQs exhibit the same contours as the statements – the Hs remain high, and the Ls are kept low.

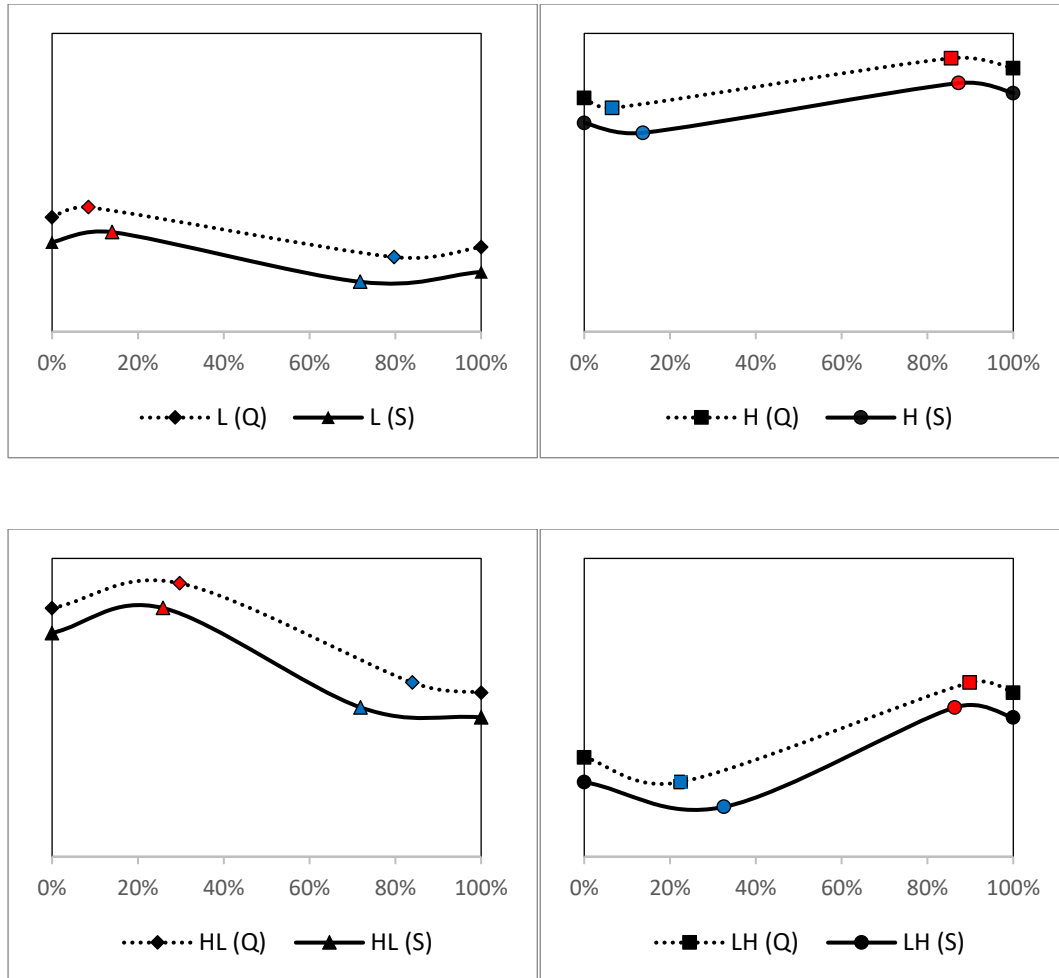


Figure 4.18 Further Stylised lines: Mono_(ISO) rhymes

4.4.2.4 Edge/ Boundary Tone

Since no visible edge and boundary tones can be found from observing the contours, I examined the pitch excursion by investigating the F0 range. The F0 range¹⁵ of the **whole Mono_(ISO) IntQs** differs from that of the statements significantly. The factors that concern the differences are TYPE, TONE, as well as TYPE*TONE interaction. TYPE ($\chi^2(1)$

¹⁵ RHYME is also a significant factor that influences the F0 range of the rhymes ($\chi^2(2) = 14.42, p < 0.001$ ***) and the whole monosyllabic utterances ($\chi^2(3) = 7.88, p < .05$ *). However, regardless of the change in rhymes, onset pitch range is not affected significantly. Besides, GENDER is not an influential factor either, which suggests that male and female voices have the same pitch range across types and tones.

= 6.27, $p = 0.012$ **), TONE ($\chi^2(3) = 40.98$, $p < 0.001$ ***), TYPE*TONE ($\chi^2(3) = 53.61$, $p < 0.001$ ***).

While **Mono_(ISO) onset** only shows a marginally significant difference on TONE ($\chi^2(3) = 11.04$, $p = 0.012$ *), the **Mono_(ISO) rhyme** generates the same pattern (Figure 5.16) as the whole Mono_(ISO) utterance (TYPE ($\chi^2(1) = 5.93$, $p = 0.015$ *), TONE ($\chi^2(3) = 52.29$, $p < 0.001$ ***), TYPE*TONE ($\chi^2(3) = 51.15$, $p < 0.001$ ***)). The smallest F0 range of the Mono_(ISO) rhymes in monosyllabic statements is 0.47ERB and the highest is 5.77ERB; while the mean pitch for IntQs range from 0.67ERB to 7.72ERB.

For Tone 1 (L Tone) and Tone 4 (HL Tone), the F0 range of the IntQs is smaller than that of the statements; while for Tone 2 (H Tone) and Tone 3 (LH Tone), the F0 range of the IntQs is bigger than that of the statements. These results suggest that the tones that end with a L tone (L, HL), which are in fact both falling, do not fall as much as they do in their declarative form; instead, the falls are gentler. On the contrary, the tones that end with a H tone (H, LH), which has rising components at the end of both tones, rise more than before. This cannot be considered as a H boundary tone since it is a part of the lexical tones themselves.

The Mono_(ISO) onsets (Figure 4.19) are not significantly affected by TYPE or TONE*TYPE. This means that the onset only changes across different tones slightly. Tone 1 (L tone) and Tone 2 (H tone), which are nearly level, only had very slight differences between their statements and IntQs. Tone 3 (LH tone) and Tone 4 (HL tone) changed more from statements to IntQs. These results confirm that the slopes of Tone 1 and Tone 2 are gentler than the other two tones. However, even the pitch range of

Tone 1 rhyme and Tone 2 rhyme changed significantly, as shown in Figure 4.20. Moreover, the F0 range difference between Tone 1 and Tone 2 is much bigger in IntQ than in statements.

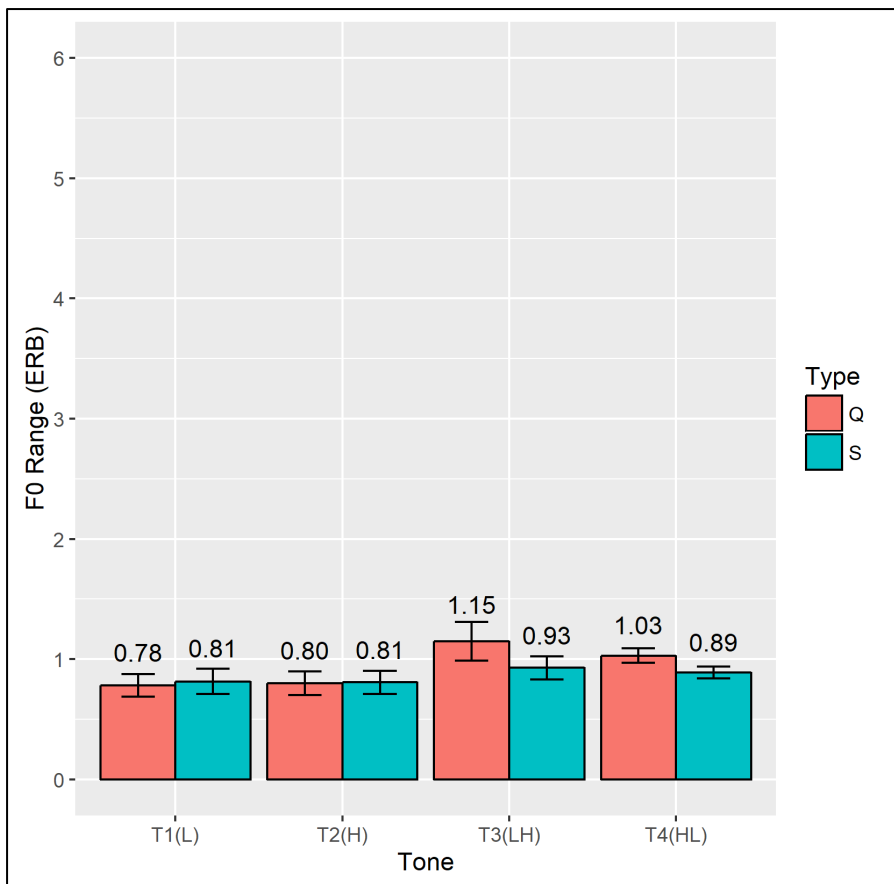


Figure 4.19 F0 Range: Mono(ISO) onsets

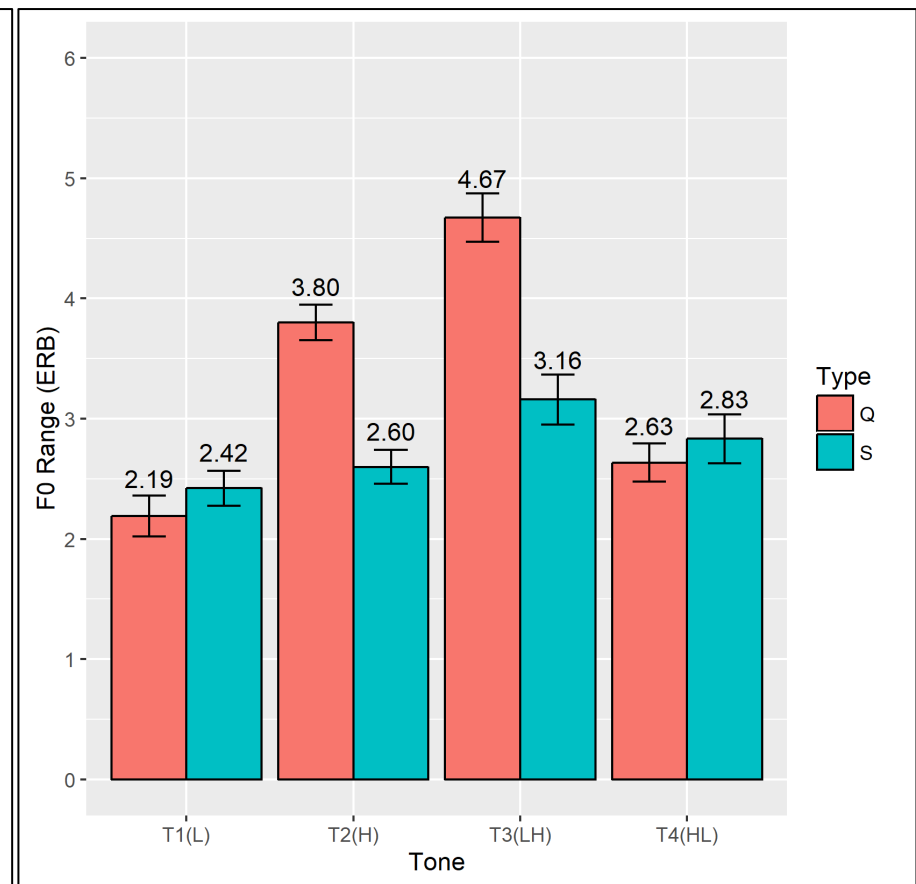


Figure 4.20 F0 Range: Mono(ISO) rhymes

4.4.3 Discussion

4.4.3.1 Lengthening in IntQ Tune

The duration differences between IntQs and statements are statistically significant – not only across different types, but also systematically different across different lexical tones. Yuan (2012) studied the duration of sentence-final syllable in IntQ and statements in Standard Mandarin. He found that in sentence-final position, Tone 4 (falling tone, HL) was longer in IntQ than in statements; while Tone 2 (rising tone, LH) did not change significantly. A study on Cantonese by Ma, Ciocca, & Whitehill (2006) also had similar results: questions had longer durations than statements at sentence-final position. In some non-tonal languages, questions were also observed to have longer durations than statements, especially with the last part of the utterances; for example the last segment in Italian (Cangemi & D’Imperio, 2013) and last vowel in Galician-Asturian (Muñiz Cachón, Rodríguez, Gómez, & Pedrero, 2006). Our $\text{Mono}_{(\text{ISO})}$ duration results exhibited the same pattern as these studies, especially Yuan’s on Standard Mandarin: IntQs are longer than statements; the difference between the durations of IntQs and statements manifest more significantly in falling lexical tones than rising lexical tones. In both Cantonese (J. K.-Y. Ma, Ciocca, & Whitehill, 2011) and Italian (Cangemi & D’Imperio, 2013), follow-up perception experiments were conducted to investigate whether duration was a cue for perceiving the difference between IntQ and statements: in both languages, duration was not a useful cue. These experiments left whether durational difference should be included in the phonological description of tunes in doubt. Moreover, the great variability of duration within each tune (212ms to 598ms for statements, and 191ms to 629ms for IntQ), together with the fact that the average difference is merely 50ms, also made it difficult to conclude that duration is a phonologically meaningful factor in distinguishing IntQs from statements.

4.4.3.2 Register and Pitch Accent

The overall register, indicated by mean pitch data, is higher. Both the minimum F0 and the maximum F0 of the IntQs are higher than those of the statements'. The mean pitch data and F0 extrema indicated that the whole register was lifted. Register raising is not a rare phenomenon in questions in tonal language questions. For example, in Xhosa, a tonal Bantu language, a raised register at the beginning of sentences indicates a question tune (Jones, Louw, & Roux, 1998). Thai and Vietnamese, both being contour-tone languages, also have a similar register lift in their question tunes (Dung et al., 1998; Luksaneeyanawin, 1998). Many other studies on Standard Mandarin also show that global register raising is very important in Mandarin questions (J. Shen, 1992; Shih, 2000).

In this study, it is hard to tease apart register and pitch accent, since there is only one syllable. It is possible that there is also a H* pitch accent. To further test the hypothesis, longer utterances with more controlled variations are needed. In Section 4.5, a study of longer sentences with the same monosyllabic words in prominence positions will be reported, to further examine pitch accents in Tianjin Mandarin.

4.4.3.3 Floating Boundary Tone

The pitch range data showed that the falling lexical tones (L and HL) had a smaller range for the IntQ than statements; while the other two rising tones (H and LH) had a bigger range. This indicates that the fall in the IntQs was not allowed to fall as much as in the statements, and the rising tones were encouraged to rise higher. In this thesis, I propose that the reason behind the said phenomenon is that a H floating boundary tone

is at the end of the IntQ anchoring the pitches – deterring the falls from falling and facilitating the rises with rising.

This boundary phenomenon is a H phonological floating boundary tone rather than a simple phonetic rise/ fall since a phonetic rise/ fall would result in a same phonetic realisation for all four lexical tones – all IntQs of the four tones would end with either a H or a L. However, in the case of the Tianjin Mandarin IntQ, although the IntQs all end higher than statements, they keep the mixture of ending Hs and Ls from the original lexical tone contours.

This section will provide further discussions of the boundary tone, floating tone, and the necessity for the floating boundary tone analysis respectively.

Boundary Tone

The concept of ‘boundary tone’ is mostly used in intonation literature. The definition of ‘boundary tone’ was made explicit in Ladd (1996, p. 103): “The H% boundary tone always indicates a final rise, which typically takes place at the very end of the phrase or utterance.” This suggests that if a H% boundary tone exists, the end of the prosodic unit is phonetically high; if the boundary tone is a L%, then the boundary physically falls low. Most existing studies follow this criterion: Figure 4.21 shows the offering tune in Bengali (Hayes & Lahiri, 1991, p. 66), in which the H_i (same with H%) lands high. Figure 4.22 shows a declarative utterance in Greek – the boundary tone is L% since the actual pitch is low (Arvaniti & Baltazani, 2005, p. 90), regardless of its slightly rising trend.

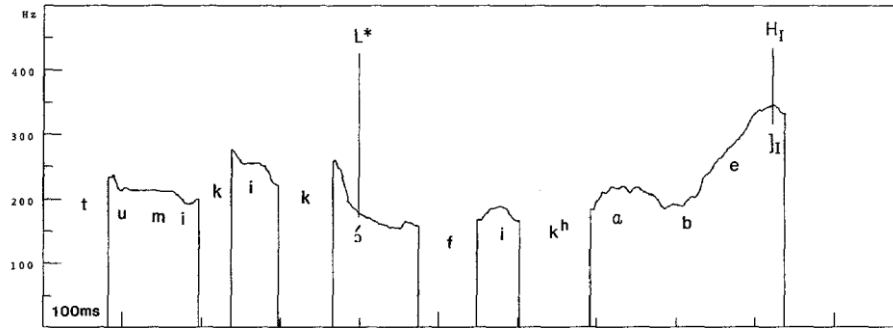


Figure 4.21 Bengali offering tune (Hayes & Lahiri, 1991)

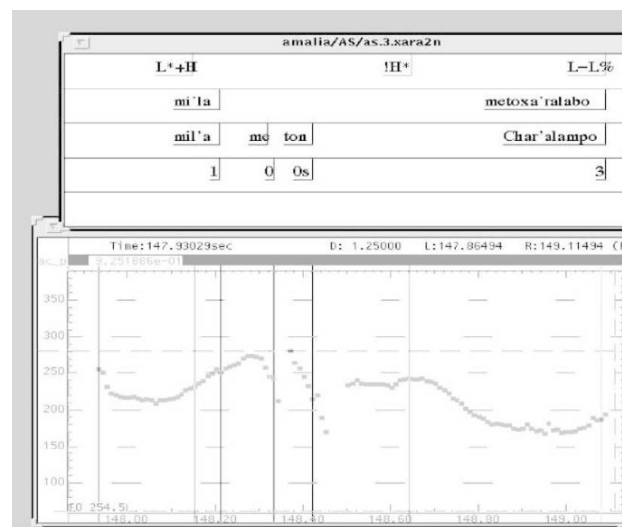


Figure 4.22 Greek declarative tune (Arvaniti & Baltazani, 2005)

Such definition of a boundary tone can be applied to tonal languages without modification. Figure 4.23 shows in Cantonese that the boundary tone H% in question intonation overrides its lexical tone and is manifested as a clear and separate high tone, especially in (e) where the lexical tones are low-falling.

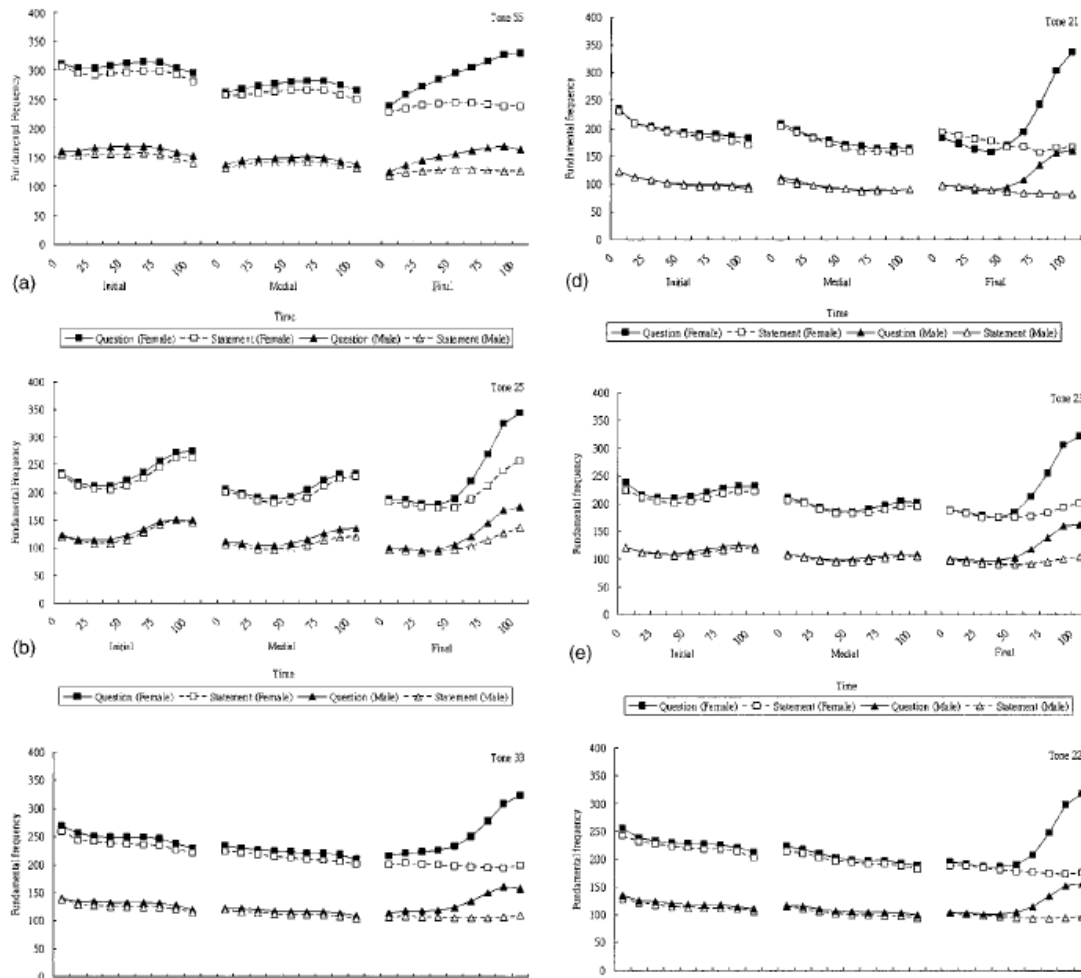


Figure 4.23 Cantonese questions and statements (J. K.-Y. Ma et al., 2006)

The literature of boundary tones in Standard Mandarin has divided opinions. One view posits that boundary tone is not necessary to be included in the phonological representation in Mandarin (P. Wang & Shi, 2010; Yuan, 2011; Yuan & Shih, 2004); the other side argues that boundary tone is essential (Flemming & Nie, 2016; Ho, 1977; M. Lin, 2004, 2006). However, this is not actually a division in term of research results, but reflects different choices of the definitions of ‘boundary tone’. For example, Lin (2004) argued that there was a H boundary tone in the interrogative tune (the upper line in each graph), as shown in Figure 4.24. However, for the falling lexical tone in the last row, the boundary remained a L tone regardless of the raise from the statements.

Therefore, the boundary tones claimed by these studies are not actual H% boundary tones, at least not under AM framework.

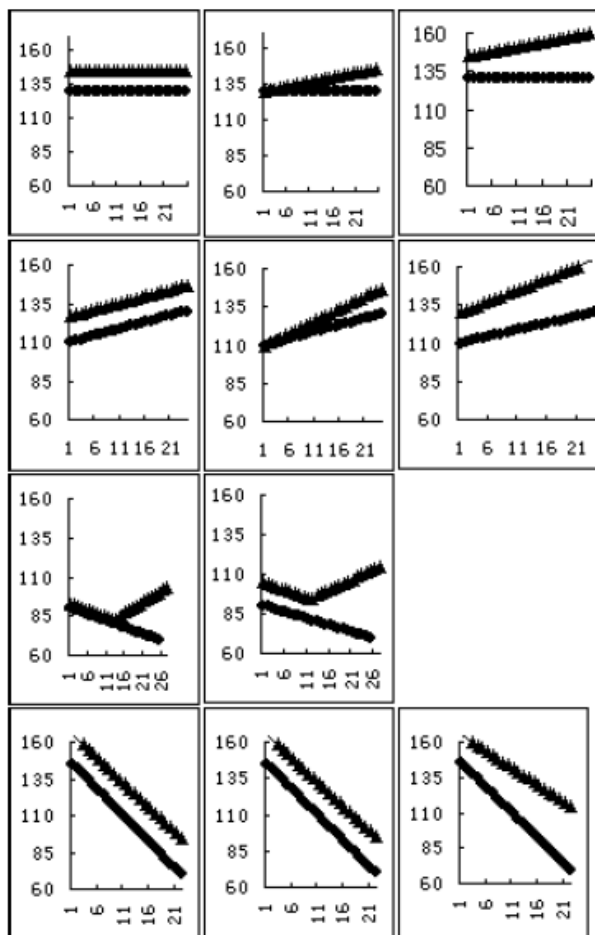


Figure 4.24 Boundary tones from Lin’s (2004) study on Standard Mandarin

Similar to Standard Mandarin, Tianjin Mandarin does not possess any clear boundary tones. Instead, it has an invisible anchor point that modifies the lexical tone, which is very similar to the concept of a floating tone in tonal literature.

Floating Tone

A floating tone is symbolised with a small circle under the tone, as “T̚”, “H̚” or “L̚”.

Hyman and Tadadjeu (1976, pp. 60–61) divided floating tones into two categories –

grammatical floating tones and lexical floating tones. Grammatical floating tones act as ‘tonal morphemes’ (Welmers, 1959). In such cases, “... a tone can exist without a TBU, and vice versa. When a morpheme consists solely of a tone, it is referred to as a tonal morpheme” (Hyman, 2011). Lexical floating tones, however, do not serve any grammatical function. The floating tones are often manifested as upstep/downstep of the neighbouring tones.

The floating H% boundary tone of IntQ tune in Tianjin Mandarin is a blend of both grammatical and lexical floating tones. Note that although one of the most well-known features of a floating tone is that it is not associated with any tone-bearing unit, I do not address its association here by bringing in the term “floating”, since “boundary” tone already denotes this point. The floating H boundary tone in this study has the function of a grammatical floating tone since it constitutes an essential part of the prosodic grammar and changes the sentence types; however, it does not serve any morphological or syntactical function as in tonal literature. It has the phonological effect of a lexical floating tone on the neighbouring tones – upstepping the final part of the final lexical tone. A very similar case in which floating tone changes the last lexical tone can be found in Tiv, a tonal Benue-Congo language, presented by Arnott (1984, cited in Leben, 2011) as in (23). A floating L tone at the end of *mba* in (23)a causes downstepping for the following word *kasev*; however, when *mba* is in phrase-final position, the floating L falls on *mba* and changes its high tone into a falling tone, as shown in (23)b.

(23)

a) mbá[↓] kásév ‘there are women’
 [— — —]

b) kásév mbâ ‘there are women’
 [— — \]

While floating tones are very common in register tone languages, they are rare in contour-tone languages (Hyman, 2010, p. 284). It is not hard to understand since most contour-tone languages, such as Tianjin Mandarin, have tones that are composed of more than one tone associated with one tonal node (“branching tone”) as in (24) (Yip, 1989), while register-tone languages often have sequences of tones (“tonal clusters”), which are more likely to be separated and thus allow floating tones. On the post-lexical level, however, the tonal nodes do not affect the floating tones anymore since the floating boundary tone associates with the IP, which is much bigger than one single tone-bearing unit.

(24)

a. branching tone



b. tonal clusters



Necessity of the Term “Floating Boundary Tone”

The term “floating boundary tone” is not only important in Tianjin Mandarin, it is also useful in many other dialects and languages. It can surely apply to other Mandarin dialects, including Standard Mandarin (Flemming & Nie, 2016), and Sichuan Mandarin

(Chang, 1958, cited from Ladd 2008: 159). For example, in Sichuan Mandarin, the lexical tones go through different changes in questions, as shown in (25):

- (25) Changes of the lexical tones of final syllables in questions:
- high-rising: remains high and often ends higher than usual
 - low-falling: becomes low level
 - high-falling: becomes high level
 - low-falling-rising: becomes low rising

The changes to the first three tones are identical with the changes in Tianjin Mandarin: the rising tone rises higher in questions than in statements, and the falling tones do not fall in question as much as in statements. The IntQ tune in Sichuan Mandarin clearly also presents a floating H% at the IP boundary.

Floating tones can also explain phenomena in other languages. In contour-tone languages such as Thai, question intonation seems to behave in a similar way as in Tianjin Mandarin. Figure 4.25 shows the schematised lines of declarative tune (Tune 1) and interrogative tune (Tune 2) in Thai (Luksaneeyanawin, 1998, p. 383). Mid and Low tones' changes from statement tune to question tune are particularly similar to what the floating tone does in Tianjin Mandarin.











Tune contrasts		Tune 1	Tune 2
Phonemic tone contrasts			
I High	1.	high narrow continuous rise and fall	extra high level
			
	2.	mid narrow continuous fall	high level
Mid			
	3.	low narrow continuous fall	mid level
Low			
	4.	low wide delayed rise	mid narrow delayed rise
II Rising			
	5.	high wide continuous or delayed fall	high or extra high narrow delayed fall
Falling			

Figure 4.25 Thai tones and tunes (Luksaneeyanawin, 1998, p. 383)

Languages other than contour-tone languages can certainly benefit from this term too. For example, the reduction of downdrift in African question (Rialland, 2007) may be as a result of floating tones. In a non-tonal language, Danish, the intonation contours of statement and questions are largely the same with a potential floating tone at the end to raise the pitch, based on results presented in Grønnum (1998, p. 143) as shown in Figure 4.26.

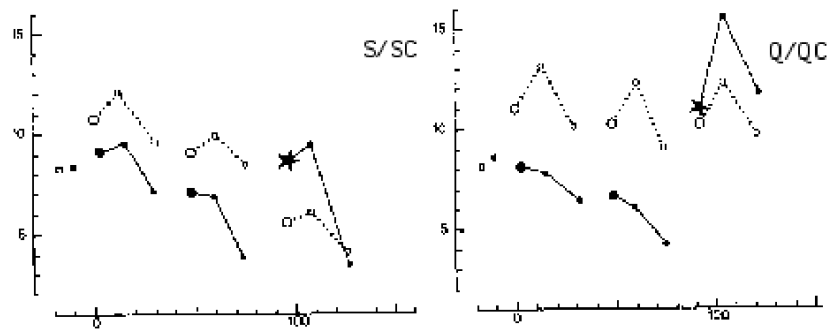


Figure 4.26 Stylised tracings of F0 in statements (left) and questions (right) in Danish (Grønnum, 1998, p. 143)

All in all, the floating boundary tone analysis is based on (i) its differences from a real boundary tone in tonal and non-tonal literature; (ii) its similarity with grammatical and lexical floating tone in tonal literature; and (iii) the necessity in having such a term to explain prosodic phenomena that would otherwise not be able to be categorised and need to be lengthily described.

4.4.4 Summary of Mono_(ISO) Results

In Section 4.4, the results of Mono_(ISO) utterances were reported and discussed in terms of duration, register, pitch accent, and boundary tone by examining acoustic parameters of length, mean pitch, tonal scaling, tonal alignment, and F0 range. The results are summarised below:

Duration: On Mono_(ISO) utterance level (also the whole syllable), IntQs' duration was longer than that of statements. The duration of Mono_(ISO) rhymes was also longer in IntQs than in statements. The Mono_(ISO) onsets were slightly longer in statements than in IntQs. However, further evidence from psycholinguistic experiments is needed to determine whether the duration is a phonologically meaningful cue.

Register: Both Mono_(ISO) onsets and Mono_(ISO) rhymes were higher with IntQ tune than with statement tune. This suggests that the register of IntQs is higher than statements. Both maxima and minima were higher in IntQs than in statements. This indicates that the register was lifted, instead of merely being expanded.

Pitch accent: The alignments of the pitch maxima and minima of both IntQ and statement are consistent, and did not show any overriding pitch accent in this study. However, the monosyllabic data do not reject the possibility of having a H* pitch accent aligned with the prominence. Further examinations are needed to confirm the pitch accents.

Edge/ Boundary tone: There was a floating H% boundary tone at the right edge of the Mono_(ISO) words. The F0 range of Mono_(ISO) onset did not differ between IntQ and statements. F0 range of Mono_(ISO) rhymes displayed an interesting pattern: the IntQ F0 range of falling tones (Tone 1 and Tone 4) showed a decrease from statements, while that of rising tones (Tone 2 and Tone 3) showed an increase. This suggests that the fall was deterred, while the rise was facilitated by a floating H% boundary tone.

In summary, the results from the production of Mono_(ISO) utterances in Tianjin Mandarin conclude that:

- a) The statement tune is the citation lexical tone of the monosyllabic word;
- b) The IntQ tune is lexical tone + higher register + floating H% boundary tone.

4.5 Production Study 2: Monosyllabic Words in Sentence Prominence

(**Mono**_(SEN))

Having investigated the statement and IntQ tones of monosyllabic utterances, we now move on to examine longer sentences with the same monosyllabic words in sentence prominence position (hereinafter abbreviated as **Mono**_(SEN)) in this section. The overall goal of the current study is to examine the post-lexical prosody change caused by the variation of lexical tones of the monosyllabic sentence prominence. By using carrier sentences with different lexical tones, we can also investigate the possible interaction between lexical tonal dissimilation rules and sentence tones.

The structure of the current section (Section 4.5) follows that used earlier in Section 4.4: Section 4.5.1 presents the specific methods for this production study of **Mono**_(SEN). Section 4.5.2 reports the results for **Mono**_(SEN): duration (§ 4.5.2.1), register (§ 4.5.2.2), pitch accent (§ 4.5.2.3), and edge/boundary tone (§ 4.5.2.4). Section 4.5.3 analyses and discusses the results, and finally, Section 4.5.4 concludes with a summary of all results and analyses for **Mono**_(SEN).

4.5.1 Methods

4.5.1.1 Materials

In this study, the monosyllabic words used in the previous production study of **Mono**_(ISO) (i.e. the **Mono**_(ISO) words in Table 4.2, p. 74) were placed in two carrier sentences. The carrier sentences were designed to serve the following three purposes:

Firstly, the carrier sentences are used to detach the changing lexical tones from the intonational phrase boundaries. The production study of **Mono**_(ISO) could not show any

pitch accent due to the fact that the lexical tones, boundary tones, and any potential pitch accents were all associated with one single syllable. One purpose of the current study is to discover potential pitch accents so that it is important to create a condition that separates the two post-lexical prosodic elements – boundary tones and pitch accents. Many existing studies (e.g. Grice, Roettger, Ridouane, & Fougeron, 2011) also used carrier sentences for easier identification of tonal association of pitch accents.

Secondly, the carrier sentences create a fixed sentence prominence by way of a syntactic structure that automatically sets the target words as the narrow foci of the sentences. The variation of the lexical tones of the word in prominence position and the potential post-lexical pitch accents which are associated with the prominence will then have a direct interaction. Potential pitch accents are thus easier to detect through the manifestation of the interaction. Moreover, Mandarin speakers tend to emphasise a part of the sentence in an unpredictable manner when reading a broad-focus sentence without any given context (Y. Xu, 1999), and mostly towards the right IP boundary which would defeat the purpose of the carrier sentences.

The third purpose of the carrier sentences is to vary the lexical tones of the syllable preceding the nucleus (hereinafter **PreTone**¹⁶) in order to investigate the possible prosodic change caused by tonal dissimilation rules in Tianjin Mandarin. To control the effect of tonal dissimilation rules, the carrier sentences used in this study did not use sequences of syllables with the same lexical tone. Tianjin Mandarin has complicated tonal dissimilation rules, which have not been well understood yet, especially in tri-tone sequences (M. Y. Chen, 2000; Q. Ma, 2005; Q. Ma & Jia, 2006). Therefore, the

¹⁶ PreTone: The tone of the word preceding the sentence prominence.

current study uses two different tones for the syllables preceding the target words – Tone 4 (HL) in (26) and Tone 3 (LH) in (27). In Tianjin Mandarin, a HL tone is expected to undergo tone dissimilation before either a HL tone or a L tone; a LH tone is expected to undergo tone dissimilation before a LH tone. By using both LH and HL tones in the carrier sentences, we are able to examine whether tone dissimilation rules have a post-lexical effect on the tones in Tianjin Mandarin.

The two carrier sentences are as follows, shown in pinyin with their lexical tones in brackets:

(26)

a. Statement 1:

ta(L) *shi*(HL) ***ma/mao/mi*** *zi*(HL)-*er*(0).

[t^ha ʂʐ ma / maʊ / mi tsz̥]

It be ma/mao/mi character.

它是“**ma**”/“**mao**”/“**mi**”字儿。

“It is the character “**ma**”/“**mao**”/“**mi**”.

b. Intonational polar question 1 (IntQ 1)

ta(L) *shi*(HL) ***ma/mao/mi*** *zi*(HL)-*er*(0)?

[t^ha ʂʐ ma / maʊ / mi tsz̥]

It be ma/mao/mi character?

它是“**ma**”/“**mao**”/“**mi**”字儿?

“It is the character “**ma**”/“**mao**”/“**mi**”?”

(27)

c. Statement 2:

ta(L) *xie*(LH) ***ma/mao/mi*** *zi*(HL)-*er*(0).

[t^ha ɕjɛ ma / maʊ / mi tsz̥]

S/He write ma/mao/mi character.

他写“**ma**”/“**mao**”/“**mi**”字儿。

“S/He writes the character **ma/mao/mi**.”

d. Intonational polar question 2 (IntQ 2):

ta(L) *xie*(LH) ***ma/mao/mi*** *zi*(HL)-*er*(0)?

[t^ha ɕjɛ ma / maʊ / mi tsz̥]

S/He write ma/mao/mi character?

他写“**ma**”/“**mao**”/“**mi**”字儿?

“S/He writes the character **ma/mao/mi**?”

4.5.1.2 Procedures and Informants

Six native Tianjin speakers participated in this study, and detailed information about the informants and procedures has already been provided in Section 4.3. They are the same informants who participated in the previous Mono_(ISO) study.

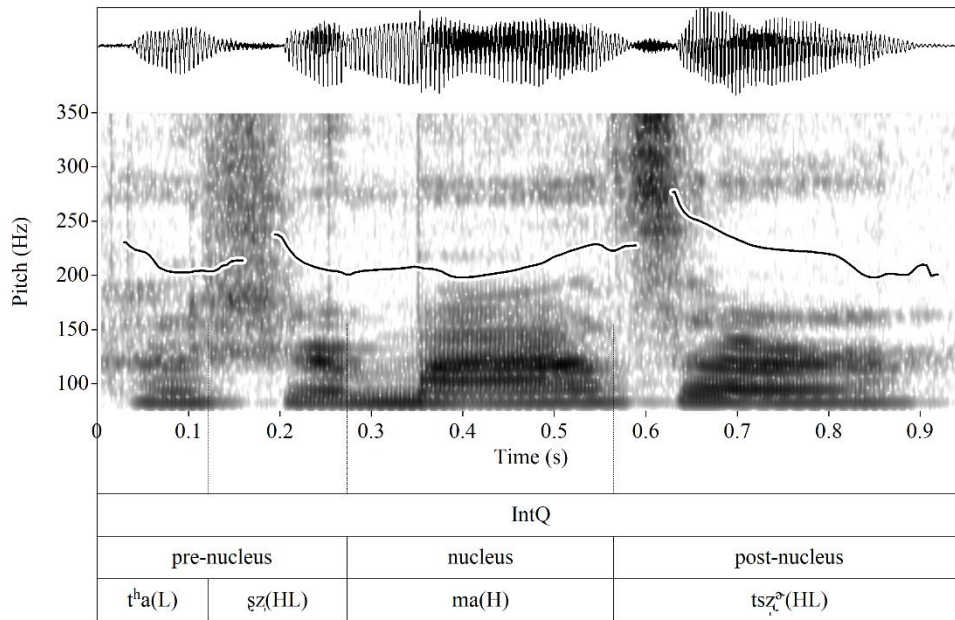
4.5.1.3 Statistical Analyses

Statistical analyses for Mono_(SEN) were also similar to those for Mono_(ISO). A mixed-effect model was used. For Mono_(SEN), apart from TYPE (Statement, Question), TONE (L-Tone, H-Tone, LH-Tone, HL-Tone), their interactions – TYPE*TONE, RHYME ([a], [au], [i]) and GENDER (Male, Female) as in Mono_(ISO) utterances, another fixed factor, PRETONE (LH-Tone, HL-Tone) was also taken into account. PRETONE is the lexical tone of the character before the nucleus Mono_(SEN) in each sentence, which triggers tone dissimilation under certain conditions. SUBJECT (6 different informants) and ITEM (48 items) were held as random-effect factors, with intercepts for both SUBJECT and ITEM, as well as by- SUBJECT random slopes for the effect of TYPE. RHYME and GENDER were usually not important factors in the current study since there are intrinsic differences between pitches of different genders, lengths of different rhymes, and so on. Therefore, these two effects were not reported in the text. PRETONE was important in some aspects of the results so was only reported when it was theoretically motivated.. The error bars in this section are also ± 1 standard error as in the previous study.

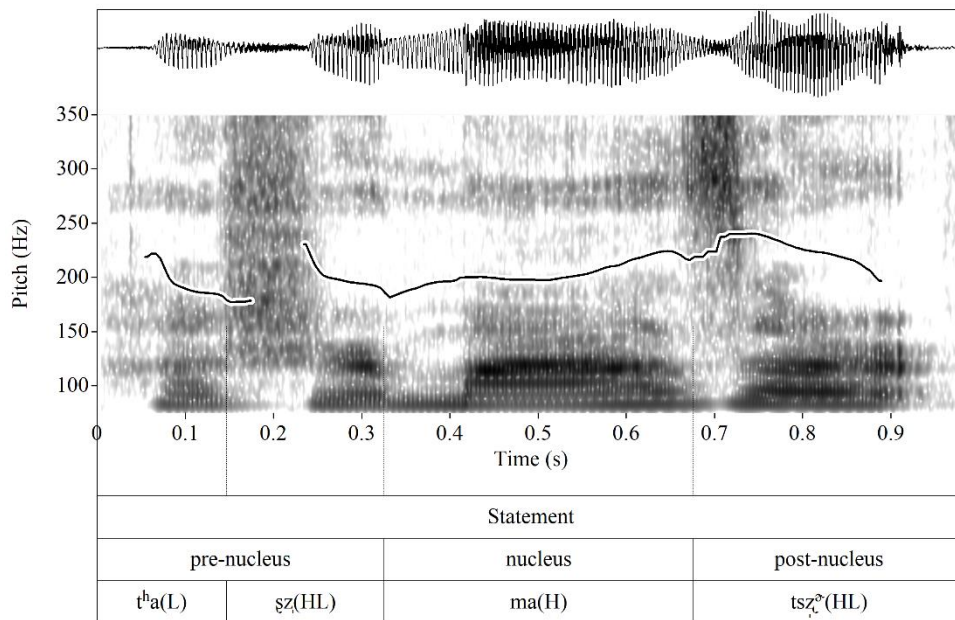
4.5.2 Results

Since the Mono_(ISO) words are the nuclei in this study, the components before and after the nuclei become pre-nucleus and post-nucleus. As the examples showing Mono_(SEN) uttered by Speaker A (female) in Figure 4.27, pre-nucleus refers to the “*ta shi*” or “*ta*

xie” part, while post-nucleus is always *zi-er*. The results are presented for all three parts of the sentences in each of the following section, with more emphasis on the nucleus.



a.



b.

Figure 4.27 Pre-nucleus, nucleus, and post-nucleus

Through visual inspection, the contours of statements and IntQs in Mono_(SEN) again show great resemblance, as found in Mono_(ISO). Figure 4.28 to Figure 4.31 show example statements and IntQs uttered by Speaker A (see Appendix 1 for more figures). No clear pitch accent or edge/ boundary tone in Mono_(SEN) can be clearly detected.

The following subsections of Section 4.5.2 report the results from different perspectives.

They are organised in the order shown in Table 4.5:

Table 4.5 Order of presentation of results: Mono_(SEN)

Section Number	Property of Intonation	Measurements
§4.5.2.1	Duration	<ul style="list-style-type: none"> Length of pre-nucleus, nucleus, and post-nucleus
§4.5.2.2	Register	<ul style="list-style-type: none"> Mean Pitch Tonal Scaling: Values of F0 Maxima and Minima
§4.5.2.3	Pitch Accent	<ul style="list-style-type: none"> Tonal Alignment: Distance of F0 Maxima and Minima Post-Nuclear Compression
§4.5.2.4	Edge/ Boundary Tone	<ul style="list-style-type: none"> F0 range for post-nucleus

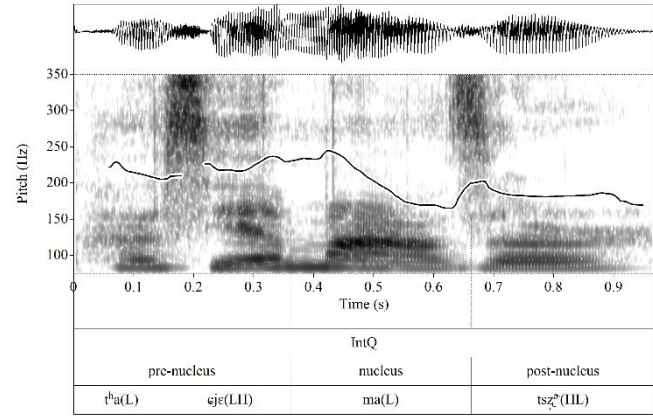
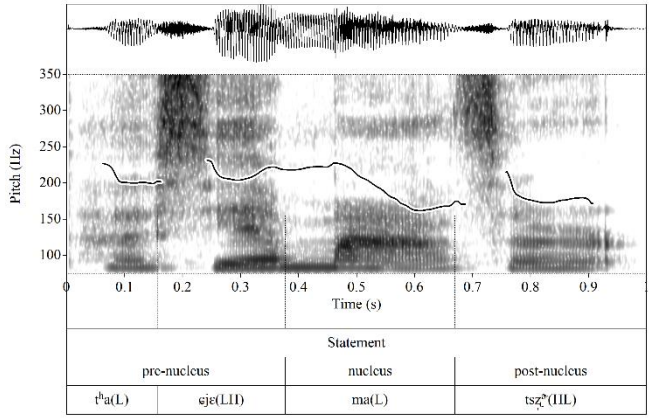


Figure 4.28 Statement (left) and IntQ (right): Mono_(SEN) – Tone 1 (L Tone)

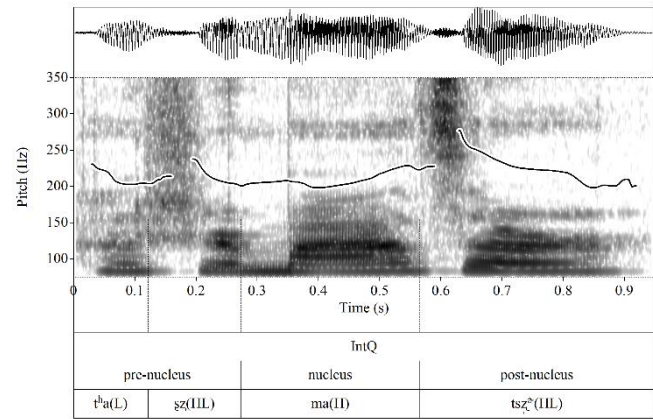
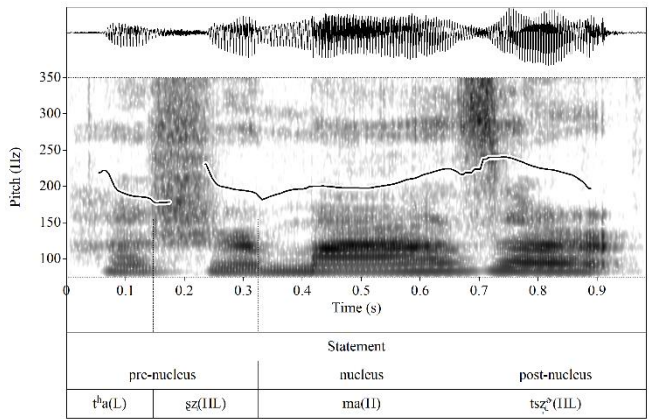


Figure 4.29 Statement (left) and IntQ (right): Mono_(SEN) – Tone 2 (H Tone)

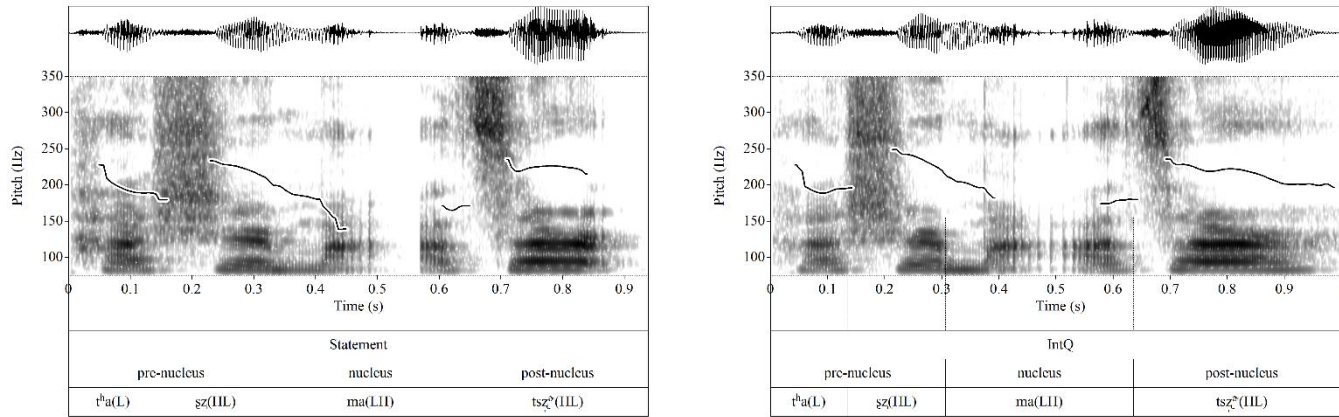


Figure 4.30 Statement (left) and IntQ (right): Mono_(SEN) – Tone 3 (LH Tone)

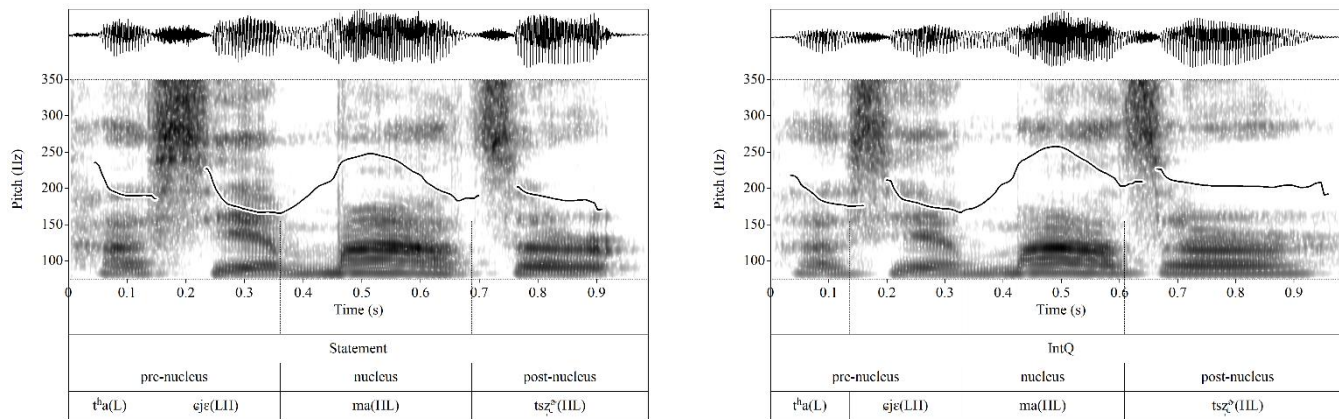


Figure 4.31 Statement (left) and IntQ (right): Mono_(SEN) – Tone 4 (HL Tone)

4.5.2.1 Duration

For the **pre-nucleus** part, statements had longer durations than IntQs across all four tones (TYPE: $\chi^2(1) = 6.92$, $p = 0.009^{**}$). The average duration of statement pre-nuclei was 315.0ms ($\overline{SD} = 67.4$), and that of IntQ pre-nuclei was 277.9ms ($\overline{SD} = 58.6$). The **nucleus** part (as shown in Figure 4.32) was also longer in statements than in IntQs (TYPE: $\chi^2(1) = 16.10$, $p < 0.001^{***}$). The average duration of statement nuclei and IntQ nuclei were 296.0ms and 274.7ms respectively (\overline{SD} s = 45.9 for statements and 42.7 for IntQs). However, for the **post-nucleus** part, as shown in Figure 4.33¹⁷, IntQs are significantly longer than statements (TYPE: $\chi^2(1) = 9.16$, $p = 0.002^{**}$), with the average duration being 320.5ms and 281.2ms for IntQ and statement post-nuclei respectively (IntQ: $\overline{SD} = 57.0$; statement: $\overline{SD} = 56.0$). One possible reason why the IntQs were shorter than statements for the pre-nuclear and nuclear parts of the sentences, while longer in the post-nuclear part, may be that the floating tone at the right boundary added length to the IntQs, while compressing the pre-nuclear and nuclear parts to highlight the prosodic prominence.

The **comparison between Mono_(SEN) and Mono_(ISO) data** also supported this explanation. The average duration for Mono_(ISO) statements was 363.6ms (see Section 4.4.2.1, p. 79), and that of the Mono_(SEN) statements was 296.0ms. In statements, the duration of Mono_(SEN) was shorter than Mono_(ISO). This is because the speech rate is faster in longer utterances than in a monosyllabic word. The duration of Mono_(ISO) IntQs increased to 388.9ms on average, while Mono_(SEN) IntQs decreased to 274.7ms on average. The increase in Mono_(ISO) IntQs also supports the idea that the floating H%

¹⁷ The x axis “Tones of Prominence” in Figure 4.33, and hereinafter, indicates the lexical tones of the lexical tones of their corresponding Mono_(SEN) nuclei.

boundary tone added length to the utterances since it was associated with the $\text{Mono}_{(\text{ISO})}$ itself.

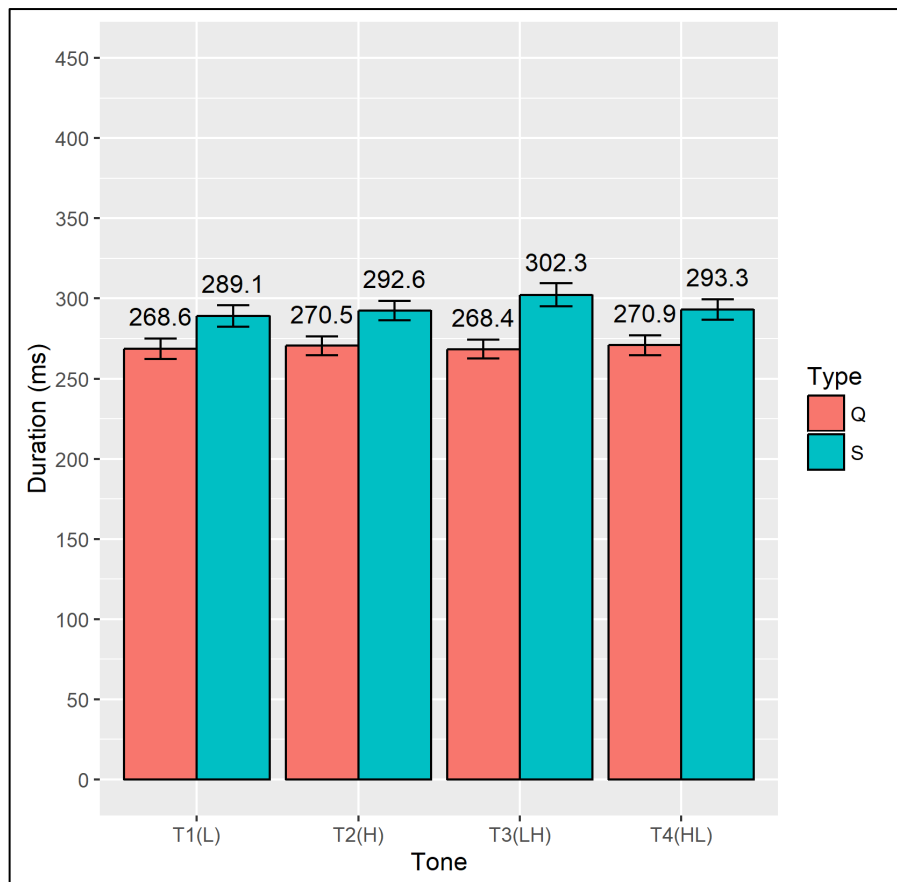


Figure 4.32 Duration: Mono_(SEN) nuclei

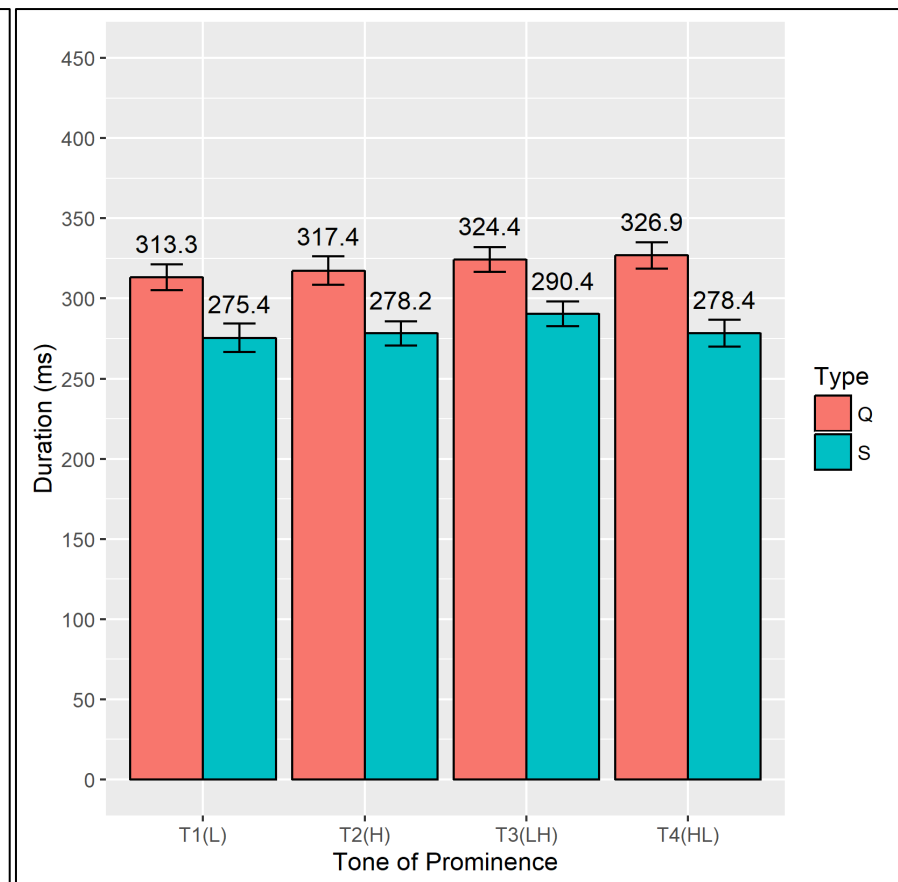


Figure 4.33 Duration: Mono_(SEN) post-nuclei

4.5.2.2 Register

Mean Pitch

The mean pitch of IntQs was significantly higher than that of statements in terms of the **full sentences**, which indicated a raise in the register. Across TYPE ($\chi^2(1) = 9.80$, $p = 0.0017^{**}$) and TONE ($\chi^2(3) = 15.01$, $p = 0.0018^{**}$), the mean pitch of IntQ was always significantly higher than statements. The mean pitch of whole statements was between 3.36ERB to 6.49ERB, while it was 3.95ERB to 7.02ERB for IntQs.

The mean pitch for the **pre-nuclear** part was influenced by the tune TYPE ($\chi^2(1) = 6.92$, $p = 0.009^{**}$) and TONE ($\chi^2(3) = 37.83$, $p < 0.001^{***}$). As shown in Figure 4.34, the IntQs across all four tones were higher than statements. The **nuclei** of IntQs were higher than statements too, as shown in Figure 4.35. The significant factors influencing the mean pitch of the nuclei were TYPE ($\chi^2(1) = 7.91$, $p = 0.005^{**}$) and TONE ($\chi^2(3) = 73.55$, $p < 0.001^{***}$), with a significant interaction between TYPE*TONE ($\chi^2(3) = 9.30$, $p = 0.03^*$). The mean pitch of the **post-nuclei** part was the same as that of nuclei's (TYPE ($\chi^2(1) = 11.22$, $p = 0.001^{***}$) and TONE ($\chi^2(3) = 142.59$, $p < 0.001^{***}$), TYPE*TONE ($\chi^2(3) = 20.82$, $p = 0.0001^{***}$), and the IntQs were also higher than statements for the post-nuclei portion as shown in Figure 4.36. Although IntQs are higher in terms of mean pitch throughout the sentences than the statements, different parts of the sentences have different changes. The average difference between IntQs and statements in pre-nuclear position was merely 0.14ERB, while the difference for nuclear and post-nuclear parts was 0.39ERB and 0.74ERB respectively.

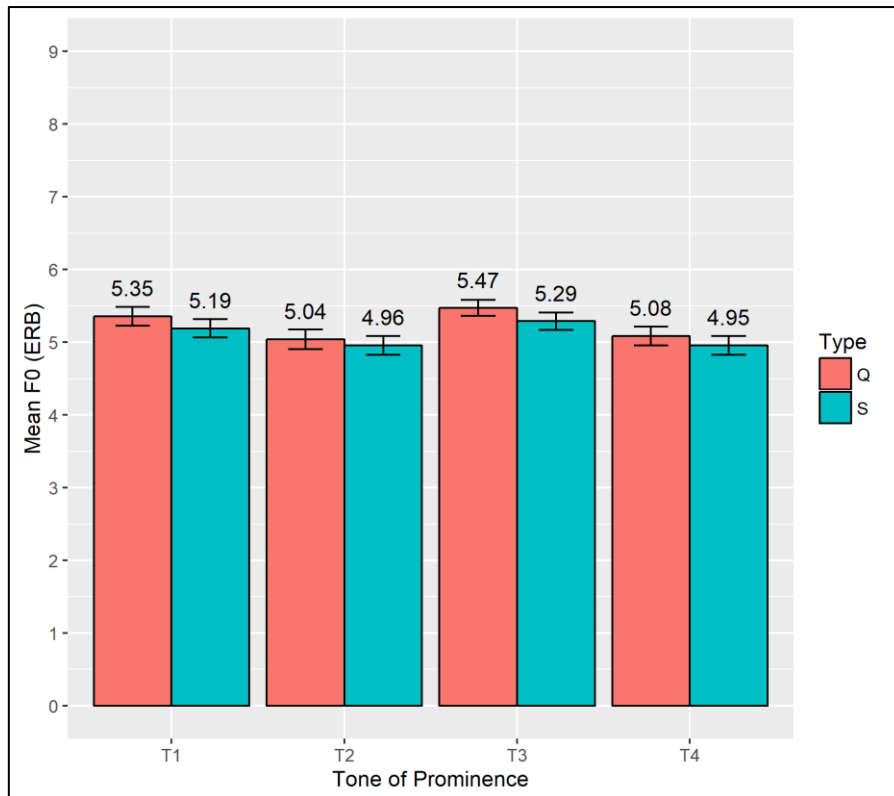


Figure 4.34 Mean Pitch: Mono_(SEN) pre-nuclei

Figure 4.35 and Figure 4.36 compare the mean pitch of **Mono_(SEN) nuclei** and **post-nuclei**. The word at the post-nuclear position was the final word of the carrier sentences. All tokens had the same final word, i.e. 字 (/tsz̥²¹⁴/ Tone 4, “character”), which had a HL lexical tone. The mean pitch for Tone 4 at nucleus position for the statement tune and IntQ tune were 5.31ERB and 5.82ERB respectively, as the rightmost group shown in Figure 4.35. The two carrier sentences varied at the syllable preceding the prominent syllable. The PreTones were Tone 3 (LH) and Tone 4 (HL). Therefore, when comparing the mean pitch of the post-nuclei with the nuclei, I only compared the post-nuclei that followed Tone-3 and Tone-4 nuclei to keep the comparison consistent by only keeping Tone3 + Tone 4 and Tone 4 + Tone 4 combinations. Therefore, in Figure 4.36, the rightmost two groups (T3 and T4) were the meaningful data. The average mean pitches of the statement tune and IntQ tune at post-nuclear position were 4.74ERB ((5.15ERB + 4.33ERB) / 2) and 5.55ERB((5.80ERB + 5.30ERB) / 2) respectively. The mean pitch

data of statements and IntQs were both lower at post-nuclear position than nuclear position. These results suggest that the mean pitch of the post-nuclei was lower than that of the nuclei.

The mean pitch data of post-nuclei displayed two distinct patterns depending on the lexical tone patterns of the preceding Mono_(SEN), especially in statements. When the preceding Mono_(SEN) nuclei ended with L tones (L, HL), the post-nuclei had lower mean pitch; while the mean pitch was higher when the preceding nuclei ended with H tones (H, LH). This was caused by phonetic assimilation, since the pitches of post-nuclei were directly affected by the pitch height of the preceding tone.

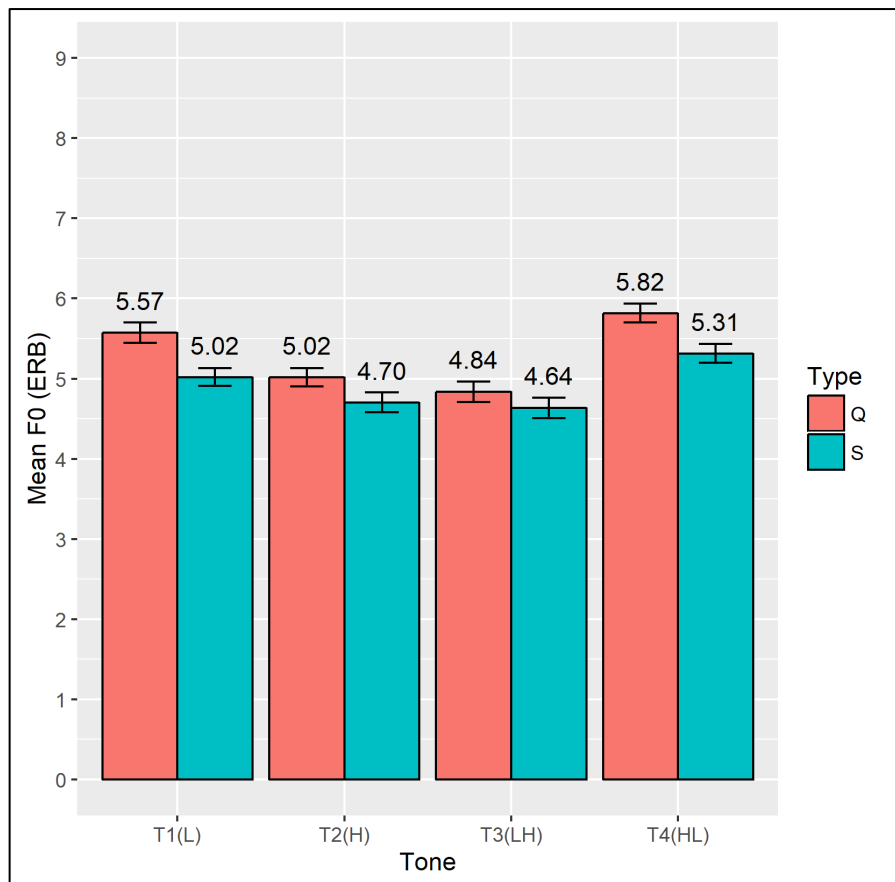


Figure 4.35 Mean Pitch: Mono(SEN) nuclei

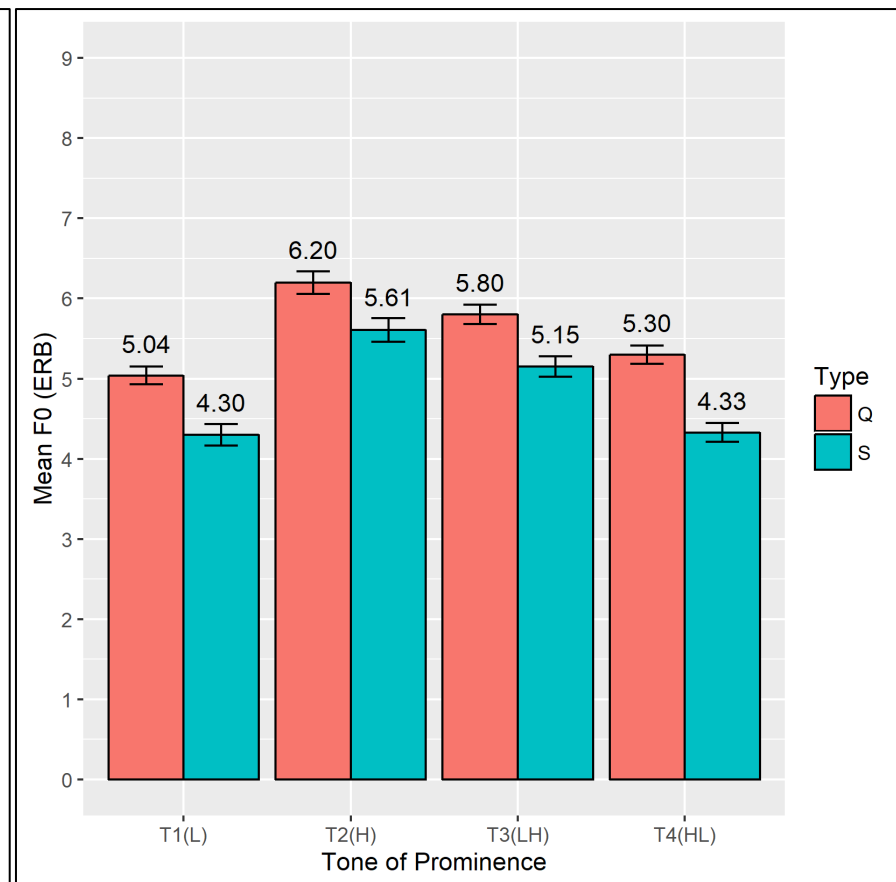


Figure 4.36 Mean Pitch: Mono(SEN) post-nuclei

The comparison between the **onset data of Mono_(SEN) and Mono_(ISO)**, as shown in Table 4.6, reveals the phonetic effect of the tonal dissimilation rules on the height of the mean pitch at the onset of the subsequent syllable, regardless of the types of tune. PRETONE significantly influences the mean pitch of Mono_(SEN) (PRETONE ($\chi^2(1) = 16.98$, $p < 0.001^{***}$). PreTones are the lexical tones in the pre-nuclear part of the sentence that was either a Tone-3 (LH Tone) or a Tone-4 (HL Tone), as shown in the carrier sentences in Section 4.5.1.1, p. 105.

Table 4.6 Mean pitch of Mono_(ISO) and Mono_(SEN) onset

(Unit: ERB)

		Mono _(ISO)	Mono _(SEN)	
			PreTone-3 (LH)	PreTone-4 (HL)
Tone 1 (L Tone)	S	5.44	5.38	5.61
	Q	6.06	5.87	5.94
Tone 2 (H Tone)	S	5.27	4.26	4.66
	Q	5.54	4.44	4.91
Tone 3 (LH Tone)	S	4.83	5.16	4.75
	Q	4.91	5.65	5.08
Tone 4 (HL Tone)	S	6.02	4.70	5.19
	Q	6.58	4.90	5.47

To better elaborate how the tonal dissimilation rules affect the pitch of the subsequent syllable's onset, I will explain the differences between the onset data of Mono_(SEN) and Mono_(ISO) shown in Table 4.6, one by one, according to the order of the relevant Tianjin Mandarin tonal dissimilation rules in (28) below. Only statement data (shaded in the table) will be explained since the same rules apply to both statements and IntQs.

- (28) The Tianjin tonal dissimilation rules relevant to this study:
- a. T3 (LH) + T3 (LH) → T2 (H) + T3 (LH)
 - b. T4 (HL) + T1 (L) → T2 (H) + T1 (L)
 - c. T4 (HL) + T4 (HL) → T1 (L) + T4 (HL)

In Table 4.6, the highest values within each row are shown in bold. It is thus easy to see that Tone 3 is the only tone that has a much higher mean pitch in Mono_(SEN) (5.16ERB) than in Mono_(ISO) (4.83ERB). This is due to the tonal dissimilation rule in (28)a, which changes the PreTone-3 into a H tone (Tone 2). The value of a Mono_(ISO) Tone-2 (5.27ERB) indeed was similar to the Tone-3 Mono_(SEN) with PreTone-3 (5.16ERB).

The mean pitch of Mono_(ISO) onset data for Tone-1 statements (5.44ERB) was approximately the same as the PreTone-3 (5.38ERB) in Mono_(SEN) data, while it was much higher in PreTone-4 (5.61ERB) than Mono_(ISO) and was similar to the mean pitch of Tone-2 Mono_(ISO) of 5.27ERB. Tonal dissimilation rule (28)b was thus reflected in the data – the HL Tone (Tone 4) preceding the L Tone (Tone 1) became a H Tone (Tone 2) and raise the pitch of the onset of the L Tone meanwhile.

(28)c was reflected in that the Tone-4 mean pitch data of Mono_(SEN) with a PreTone-4 (HL), 5.19ERB, was much lower than the corresponding Mono_(ISO) mean pitch of 6.02ERB, but closer to the Tone-1 Mono_(ISO). The PreTone-4 followed by another Tone-4 was realised as a Tone-1 (L) due to (28)c.

Apart from the effect of the three tonal dissimilation rules reported here, another reason for the lower mean pitch in Mono_(SEN) is due to the influence of the L tone in the preceding Tone 3 (LH) and Tone 4 (HL): often only the first half of the pre-nuclear

Tone 3 is produced during fast speech (J. Zhang & Lai, 2010, p. 162), so the PreTone 3 ends low; PreTone 4 allows the pitch to fall to a low level so the PreTone 4 also ends low. Therefore, the Mono_(SEN) data were not as high as the Mono_(ISO) counterparts that received no influence from any other tones since they were produced in isolation.

Tonal Scaling: Values of F0 Maxima and Minima

The pitch maxima and minima in the Mono_(SEN) rhymes exhibited the same pattern as the results of Mono_(ISO) pitch extrema. Both the minimum pitch and the maximum pitch of Mono_(SEN) rhymes were raised when changed from a statement to an IntQ (maximum pitch: TONE ($\chi^2(3) = 119.07$, $p < 0.001^{***}$), TYPE ($\chi^2(1) = 7.84$, $p = 0.0051^{**}$); minimum pitch: TONE ($\chi^2(3) = 124.33$, $p < 0.001^{***}$), TYPE ($\chi^2(1) = 12.10$, $p = 0.0005^{***}$), TYPE*TONE ($\chi^2(3) = 72.95$, $p < 0.001^{***}$)), as shown in Table 4.7.

The values in Mono_(SEN) changed considerably from Mono_(ISO), especially Tone 2 and Tone 3. As previously shown before in Figure 4.16 in Section 4.4.2.2 (p. 84), the Tone-2 Mono_(ISO) rhymes had a mean maximum pitch of 7.6ERB for IntQ and 6.54ERB for statements respectively. However, in the current study of Mono_(SEN) rhymes, the mean maximum pitches of IntQ and statement were only 5.83ERB and 5.32ERB. The IntQs' maximum pitch in Mono_(SEN) was lower than Mono_(ISO) by 1.77ERB, and there was a difference of 1.22ERB between the statements in Mono_(SEN) and Mono_(ISO), with Mono_(SEN) also being lower.

The reason for the lower pitch maxima in Mono_(SEN) was the same as the difference between Mono_(SEN) and Mono_(ISO) mean pitch data: the L tone in pre-nuclear lexical tones affects the pitch height of Tone 2.

Table 4.7 Average F0 extrema of Mono_(SEN) rhymes

	Tone	Type	Max F0 (ERB)	Min F0 (ERB)
<i>FALLING</i>	TONE 1 (L)	Q	6.18	4.74
		S	5.66	3.96
	TONE 4 (HL)	Q	6.49	5.38
		S	5.99	4.51
<i>RISING</i>	TONE 2 (H)	Q	5.83	4.70
		S	5.32	4.44
	TONE 3 (LH)	Q	5.07	3.80
		S	4.66	3.62

4.5.2.3 Pitch Accent

Tonal Alignment: Distance of F0 Maxima and Minima to Rhyme Onsets

The distance of the pitch maxima and minima to the rhyme onset of Mono_(SEN) nucleus rhyme did not differ much between IntQ and statements: maxima alignment – TYPE ($\chi^2(1) = 0.016$, $p = 0.9$, n.s.; minima alignment – TYPE ($\chi^2(1) = 0.68$, $p = 0.41$, n.s. The contours of the nuclei rhyme were the same as in Mono_(ISO) (cf. Table 4.4 in Section 4.4.2.3, p. 87). This indicates that no pitch accent has alternated the pitch contour of the lexical tone; however, it does not reject the possibility of having a H* pitch accent, since the mean pitch of the nuclear part of the sentences were raised.

Table 4.8 Alignment of F0 Maxima & Minima: Mono_(SEN)

	Tone	Type	Max F0 Dist.	Min F0 Dist.
<i>FALLING</i>	TONE 1 (L)	Q	7.40%	90.68%
		S	3.30%	86.27%
	TONE 4 (HL)	Q	37.92%	85.64%
		S	30.35%	88.93%
<i>RISING</i>	TONE 2 (H)	Q	94.43%	8.58%
		S	92.11%	14.09%
	TONE 3 (LH)	Q	25.42%	47.51%
		S	42.02%	53.06%

Post-Nuclear Compression

Unlike the sentence nuclei, the lexical tone contour changed considerably in the post-nuclear position. However, the change was not brought about by the type of tunes, but by the focus. The post-nuclear word was the same word in all sentences, *zi*, Tone 4 (HL). In Mono_(ISO) data, the range of a Tone-4 statement (also the citation tone) was 2.91ERB. The F0 range for Tone-4 nuclei of Mono_(SEN) statement tune was 2.55ERB. However, much lower than the F0 ranges of both the citation tone (i.e. Mono_(ISO)) and Mono_(SEN) nuclei, at the post-nuclear position of Mono_(SEN) sentences, the average F0 ranges were only 1.21ERB for IntQs and 1.07ERB for statements, which were only around half the citation tone's range. As shown in Figure 4.37, the smallest range was 0.77ERB and the largest was 1.56ERB on average. These data showed post-focus compression for both the statement tune and IntQ tune.

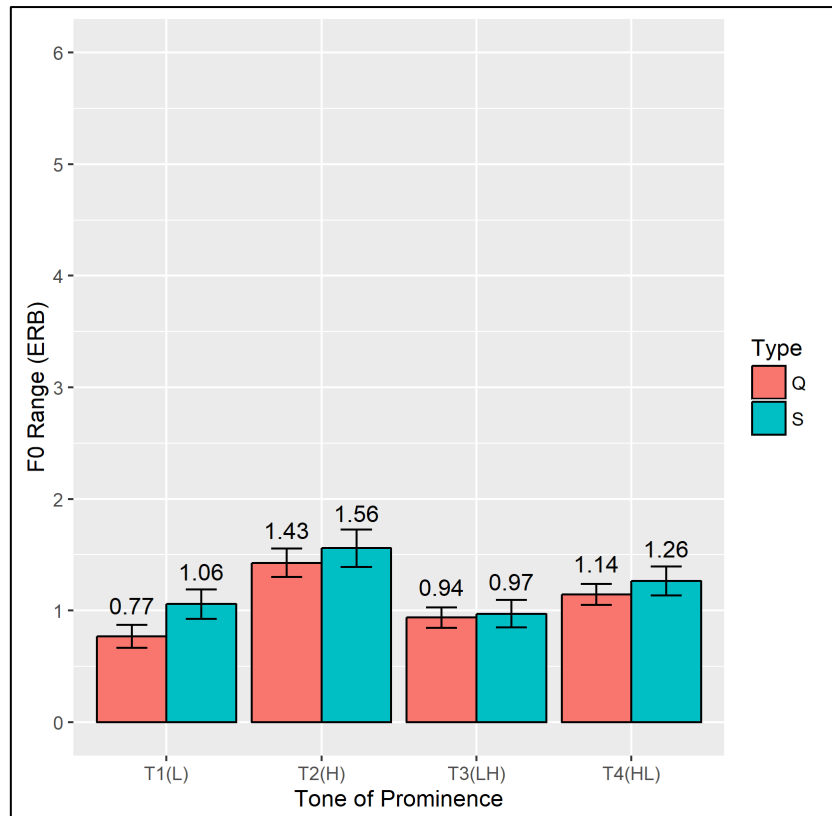


Figure 4.37 F0 Range: Mono_(SEN) post-nuclei

4.5.2.4 Edge/ Boundary Tone

Figure 4.37 above also shows that the IntQs had smaller F0 range than statements. Since the post-nuclear tone was Tone 4 for all sentences, the fact that IntQs had a narrower F0 range indicates that the falling tone did not fall as much the statements, which was consistent with the F0 range data of Tone 4 in Mono_(ISO) data. Moreover, the mean pitch of IntQs is higher than that of the statements in post-nuclei position by 0.74ERB, while that of the nuclear part was only higher by 0.39ERB, which also suggests that there was a H associated with the last part. Therefore, Mono_(SEN) had a floating H% boundary tone for the IntQ tune, as the Mono_(ISO) IntQ tune did.

4.5.3 Discussion

4.5.3.1 Boundary-Induced Lengthening

The duration of the Mono_(SEN) nuclei displayed the opposite trend from Mono_(ISO), although the monosyllabic words used were identical: the Mono_(SEN) nucleus was longer in statements than in IntQs, while Mono_(ISO) words were shorter in statements than in IntQs. However, Mono_(SEN) post-nucleus are longer in IntQs than in statements. What Mono_(SEN) post-nucleus and Mono_(ISO) had in common was that they both contain the IP boundary. As discussed in Section 4.4.3.1 (p. 93), many languages lengthen question finals for the IntQ tune. In the Cantonese production study by Ma et al. (2006), the results also show that at sentence initial or medial position, target words were longer in the statements than in the IntQs; but at sentence final position, the statements were shorter than in the IntQs. The same went for the Italian study (Cangemi & D’Imperio, 2013): questions had a longer final segment but a shorter initial segment than in statements; the middle segments did not significantly differ.

Another type of lengthening that has been observed in the question prosody literature is final vowel lengthening in Ncam, a three-tone Gur language (Rialland, 2009). This type of lengthening is one of the question markers specified by Rialland (2007) (i.e. (h) in Table 4.1, p. 58). Such lengthening is different from the type of lengthening observed in the current study since Ncam uses lengthening as a sole marker for questions; in the current study, however, lengthening in the IntQ tune is accompanied by register lift and interpolation of a floating boundary tone. The previously mentioned perception studies suggest that although questions are lengthened in production studies, duration is not necessarily a perceptual cue. We can therefore infer that the durational difference is brought about by the floating boundary tone at the right edge of IP.

4.5.3.2 Register Lift

The results of the register change in $\text{Mono}_{(\text{SEN})}$ data are in accordance with $\text{Mono}_{(\text{ISO})}$ data, as presented in Section 4.5.2.2. The IntQ tune has both higher mean pitch and higher F0 extrema than the statement tune. The new information we gained from this study was that the pre-nuclear accent, nuclear accent, and post-nuclear accent raised by different margins. The difference between IntQs and statements, in terms of the pre-nuclear register, was very small while the difference was largest for the post-nuclear register. The register change for the nuclear accent is larger than the pre-nuclear accent, but smaller than the post-nuclear accent. The schematised figure below (Figure 4.38) illustrates the register difference between different components of an IntQ tune and a statement tune. The statement tune, together with all the lexical tones are postulated to be level, in order to highlight the difference in terms of post-lexical differences in the tunes.

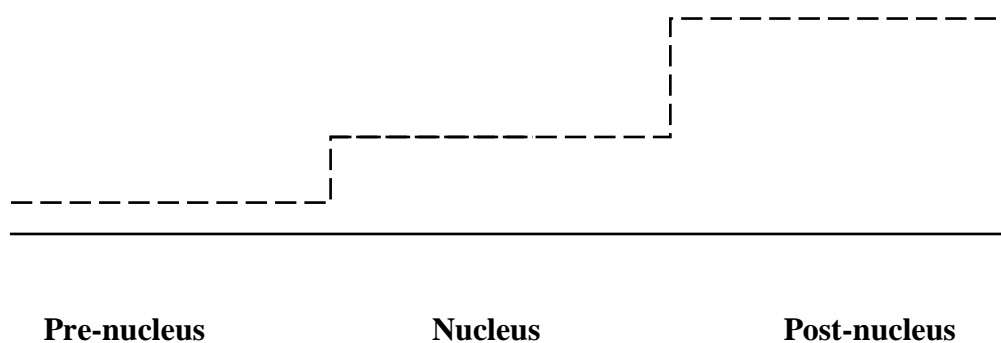


Figure 4.38 Schematised lines for register differences in $\text{Mono}_{(\text{SEN})}$

(Dashed line: IntQ tune; Solid Line: Statement tune)

4.5.3.3 Pitch Accents

Pitch Accent of Focus

Many previous studies of Standard Mandarin intonation have examined the acoustic properties of sentence focus, including the effect of focus types and focus locations. Liu and Xu (2005) observed from Mandarin IntQ data that the pitch range of the focused words was expanded. Chen and Gussenhoven (2008) also studied Standard Mandarin and found different degrees of pitch expansion for different levels of emphasis. Duan and Jia (2014) confirmed the expansion of pitch range for focused words in five northern Mandarin dialects, including Tianjin Mandarin. The comparison between the mean pitches of nuclei and post-nuclei in Section 4.5.2.2 suggested that the post-nuclei were lower than nuclei on average, regardless of the types of tunes. Therefore, it is safe to conclude that there is a H* pitch accent associated with the sentence prominence in both statement tune and IntQ tune. We can further schematise the tunes as Figure 4.39, combining the results: (i) the register differences between IntQ tune and statement tune at pre-nuclear, nuclear, and post-nuclear positions respectively, as discussed in Section 4.5.3.2; (ii) the H* pitch accent which was induced by the sentence focus.

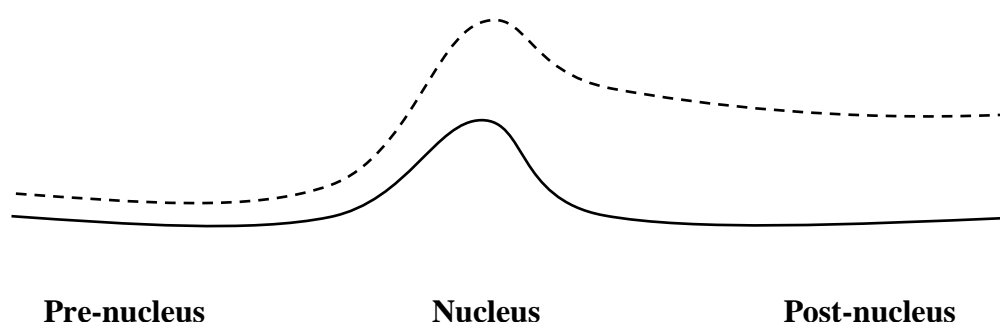


Figure 4.39 Schematised lines for pitch accent differences in Mono_(SEN)

(Dashed line: IntQ tune; Solid Line: Statement tune)

Post-Focus Compression

Apart from the expansion of pitch range on the focused words, a number of studies of intonation in Standard Mandarin and other Mandarin varieties have also detected post-focus lowering and compression (Duan & Jia, 2014; Liu & Xu, 2005; Y. Xu, Chen, & Wang, 2012). For example, Liu and Xu (2005) illustrated the pitch contours of the statement intonation and IntQ intonation in Standard Mandarin as in Figure 4.40. The tones used in this figure were all Tone 4 (HL). The figure on the right depicts the IntQ tune while the one on the left, the statement tune. The IntQ tune shows a clear register lift from the statement but no other pitch accent can be detected. The dotted lines in both statement tune and IntQ tune lines are substantially higher at the sentence-medial narrow focus position, while the post-focus was compressed at the post-focus position.

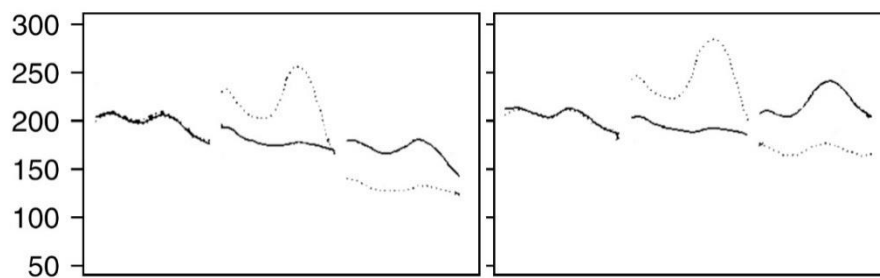


Figure 4.40 Statement (left) and IntQ (right) with neutral and narrow foci

(Amended from Liu & Xu, 2005, p. 75)

(Dotted line: Sentence-medial narrow focus; Solid Line: Neutral tune)

However, post-focus compression does not necessarily occur in every language. Even under different conditions within one language, the post-focus region shows different patterns. For example, in Hong Kong Cantonese (W. L. Wu & Xu, 2010), Taiwanese (Y. Xu et al., 2012), and Taiwan Mandarin (Y. Xu et al., 2012), post-focus compression does not occur. Chen (2010) found that depending on the preceding lexical tones, the

compression effect differed significantly in Standard Mandarin. Post-focus compression does not only modify lexical tone range; it has also been found in a number of non-tonal languages that the pitch range of the intonation contours was compressed (e.g. Bruce, 1982; Gussenhoven, 1983; Ladd, 2008; Pierrehumbert, 1980).

The comparisons among the F0 ranges of the citation form of Tone 4 (i.e. $\text{Mono}_{(\text{ISO})}$ statement, 2.83ERB), Tone-4 $\text{Mono}_{(\text{SEN})}$ nuclei (2.55ERB) with the statement tune, and $\text{Mono}_{(\text{SEN})}$ post-nuclei with the statement and with IntQ tune in Figure 4.37 (between 1.77ERB and 1.56ERB) show that the post-nuclear accent has substantially smaller F0 ranges than the same tone at sentence nucleus position, or in isolation as a citation tone. These results corroborate the post-focus compression phenomenon in the Standard Mandarin literature.

Future research can further attest the results in sentences with final syllables of other tones, with a rising tone in particular, to examine how other tones behave at the post-nuclear position. Instead of a more phonetic analysis of “compression”, there is a possibility that there is a floating pitch accent that phonologically neutralises the rising tones and the falling tones and makes them level. Moreover, whether this phenomenon is specifically focus-related or sentence final related in Tianjin Mandarin remains a question to be further investigated.

Tonal Dissimilation Rules

In the design of the $\text{Mono}_{(\text{SEN})}$ experiment, two carrier sentences were used to investigate the effect of tonal dissimilation on sentence tunes. Data and statistical analyses attested the existence of the tonal dissimilation rules in Tianjin Mandarin since

apparent phonetic effect was found on the F0 values of the neighbouring segments. However, tonal dissimilation rules did not create any systematic change on sentence tunes – no pitch accent, edge tone, or register change on the nuclear or post-nuclear accents are related to the tonal dissimilation rules.

This also sheds light on what kind of prosodic phenomena can be counted as intonation in Tianjin Mandarin, or even tonal languages in general. The lexical tones are not a part of the intonation since they are specified lexically and do not change according to different functions of the sentences. Neither do tonal dissimilation rules like the ones in Tianjin Mandarin count as a part of the intonation contour. Such tonal dissimilation rules do not change across sentence types, nor do they vary other parts of the tunes in a functional way. The only change is the phonetic F0 height change, i.e. tonal scaling, which can be within and across prosodic boundaries and occur anywhere in a sentence independent of the properties of the sentences.

4.5.3.4 Floating Boundary Tone

As discussed in Section 4.5.3.1, the floating H% boundary tone analysis was supported by the duration results. Apart from the evidence from the lengths, the F0 range results in Figure 4.37 also lent support to the floating H% analysis. The pitch range data showed that the IntQs did not fall as much as the statements. As shown in the schematised lines in Figure 4.39, there is a possibility that the post-focus compression of F0 range was partially, if not completely, caused by the boundary phenomenon (i.e. the floating H% boundary tone). The floating H% observed in the Mono_(ISO) data was a further support for the existence of the floating H% boundary tone at the right edge of the Mono_(SEN) IP boundary.

4.5.4 Summary of Mono_(SEN) Results

In Section 4.5, the results of Mono_(SEN) utterances were reported and discussed in terms of duration, register, pitch accent, and boundary tone by examining acoustic parameters of length, mean pitch, tonal scaling, tonal alignment, and F0 range. The results are summarised below:

Duration: At Mono_(SEN) utterance level, the duration of the IntQs is longer than that of the statements. The nuclear accent in Mono_(SEN) IntQs are shorter than the statements, while the post-nuclear accent is longer in the IntQ tune than in the statement tune. This is boundary-induced final lengthening, which is possibly not a perceptual cue.

Register: The register of IntQs is higher than statements as indicated by the mean pitch data. Moreover, both maxima and minima are higher in IntQs than in statements. This indicates that the register was lifted, instead of merely being expanded. The post-nuclear accent has the biggest difference between the mean pitch of the IntQ tune and the statement tune, while the pre-nuclear accent only marginally differed in statements and IntQs. The nuclear accent in the IntQ tune is higher than its statement counterpart, but the difference is not as big as in the post-nuclear accents.

Pitch accent: The alignments of the pitch maxima and minima of both the IntQ tune and the statement tune are consistent, and did not show any overriding pitch accent in this study, which is in conformity with Mono_(ISO) results. There is a potential H* associated with the prominence; the post-nuclear accent goes through post-focus compression. However, these differences are results of the focus. The difference in

tunes does not generate any pitch accent difference. Additionally, although tonal dissimilation rules bring a phonetic carry-over effect, it does not create any phonologically meaningful pitch accent or boundary tone.

Edge/ Boundary tone: There was a floating $H\%$ boundary tone at the right edge of the $\text{Mono}_{(\text{SEN})}$ utterances. The F0 range of $\text{Mono}_{(\text{SEN})}$ is much smaller than its $\text{Mono}_{(\text{ISO})}$ counterpart, which is likely to be brought by the floating $H\%$ boundary tone, together with post-focus compression.

In summary, the results from the production of $\text{Mono}_{(\text{SEN})}$ utterances in Tianjin Mandarin conclude that:

- a) The statement tune includes the citation lexical tones of all the words in the utterance, and H^* associated to its prominence, and a compressed post-nucleus;
- b) The IntQ tune is a combination of the statement tune, and higher register, and a floating $H\%$ boundary tone.

4.6 Conclusions for Production Studies of IntQ Tune in Tianjin Mandarin

This chapter contains two production studies, one on monosyllabic words in isolation (i.e. $\text{Mono}_{(\text{ISO})}$), and the other one on sentences with monosyllabic words as sentence prominence. Both temporal and F0 data were measured, following visual inspection. The results showed that the IntQ tune brings a lifted pitch register and a floating $H\%$ boundary tone on the basis of the statement tune.

This chapter highlights the proposal of the term “floating boundary tone” in intonation studies. The following are the principles for delineating a floating boundary tone:

(i) a floating boundary tone should be different from a real boundary tone – a real boundary tone is an interpolation of an extra tone, while a floating boundary tone does not realise as an extra tone.

(ii) a floating boundary tone should comply with grammatical and lexical floating tone in the tonal literature– a floating boundary tone should make a functional difference as a grammatical floating tone does (in Tianjin Mandarin, the floating H% boundary tone distinguishes tune types); and a floating boundary tone should modify the neighbouring tone(s) without having its own phonetic realisation as a lexical floating tone does.

Another important issue that has been brought into discussion in this chapter is the components of an intonational tune in tonal languages. Only the prosodic features on the post-lexical level that alter the function of tunes can be counted as components of tunes. I therefore conclude that the statement and IntQ tunes in Tianjin Mandarin are as in (29), where the part in brackets is optional.

(29)

Statement:

H* + (post-focus compression);

IntQ:

[H* pitch accent + (post-focus compression) + floating H_o% boundary tone] higher register

Chapter 5 Perception Experiments:

Intonational Yes/No Question Tune in Tianjin Mandarin

5.1 Introduction

The IntQ tune is regarded as one of the most basic and salient types of intonation in the majority of the languages that allow such YNQ, in that this type of questions is syntactically identical to its declarative counterpart, and intonation is the only cue to suggest that it is a question. However, it is not always the case that such questions can be perceived without failure, especially when the pitch of a language is also used for other properties. In tonal languages such as Tianjin Mandarin, as shown in the previous chapter, the differences between IntQs and statements are subtle. Communication failures often occur when an IntQ tune is used, even between native speakers, and even when the speaker and hearer are familiar with each other's speech.

In this chapter, I will investigate how lexical tones interact with intonational question tune. To be specific, the following questions will be answered:

- How well are IntQs perceived by native Tianjin speakers?
- Does lexical tone play a role in intonation identification? If yes, is there a tonal bias?
- Do tonal alternation rules have an effect on tune identification?

Three tune identification experiments were designed to answer these questions. Natural stimuli (i.e. not artificially manipulated and synthesised) were used in the experiments to investigate native speakers' ability to identify real utterances.

This chapter is organised as follows: Section 5.2 briefly reviews some relevant literature on the perception of Yes/No questions and perception of intonation in general; Section 5.3 presents detailed methodologies that are used in the three tune identification experiments, including the participant information (§5.3.1), the stimuli used in the three experiments respectively (§5.3.2), the procedures of the experiments (§5.3.3), and data analyses methods (§5.3.4). The following three sections, Section 5.4, Section 5.5, and Section 5.6 present the three experiments respectively. Under each section, the results of the accuracy of the responses and the reaction time are reported separately; a discussion also follows the report of results in each section. Finally, this chapter concludes with a general discussion and summary of the findings.

5.2 Literature Review: Identification of Tunes

In non-tonal languages, the perception of intonational tunes is rarely studied with the aim of finding out whether the native speakers can perceive a certain intonation, since it is fairly straightforward and there is no reason for native speakers not to recognise the tunes. Tune identification is therefore more commonly used with resynthesised materials to test certain hypotheses. For instance, in order to study whether the durational cue is used in the identification of tunes, Cangemi and D'imperio (2013) created a nine-step continuum using two base factors, i.e. F0 and duration. This type of studies are not rare – the manipulation is always made to either pitch scaling (vertical manipulation of F0 values on a time-pitch scale as Figure 3.7, p. 43) or pitch alignment (horizontal manipulation of F0 values in relation to time) (e.g. Makarova, 2001; Roettger, 2017).

Ni and Kawai (2004) generated a continuum to study the relationship between the perception of questions and the F0 height of the final syllable in Standard Mandarin. They discovered that for Tone 2 (LH Tone), the higher the pitch peaks were, the easier it was for the native speakers to consider them as questions; for Tone 4 (HL Tone), the utterance was only regarded as a question when the whole tone was placed at a higher register.

However, contrary results were found in a series of studies on Standard Mandarin intonation identification (Yuan, 2011; Yuan & Shih, 2004). In Yuan (2011), a corpus that varied in terms of the tones of the last syllables, the tunes (question tune or statement tune), as well as the focus location (initial, medial, final), were used as the stimuli. The corpus had 1040 sentences in total, which were recorded by 8 speakers. The results from these studies demonstrated that when the last syllable of an utterance carried a Tone 4, it would be easier for the utterances to be correctly identified as a question. On the contrary, those with Tone-2 final syllables were the most difficult ones to identify.

Xu and Mok (2012) studied the cross-linguistic perception of four types of sentences in Cantonese and in Standard Mandarin by Cantonese speakers and Mandarin speakers respectively. The four types of sentences were complete statements, cut-off statements (with the final syllable being cut off), complete questions, and cut-off questions (with the final syllable being cut off). Cantonese speakers had the most problems with identifying questions with the Cantonese Tone 1 (H Tone) in complete sentences, although the accuracy was still above 90%. It was much more difficult to identify the question tunes in the cut-off questions by native Cantonese speakers, the tones with

higher register gained higher accuracy. The Mandarin tune identification task by native Mandarin speakers had the same results as Yuan's (2011): T4 (HL Tone) had higher accuracy than T2 (LH Tone), in both complete and cut-off sentences. The cut-off sentences achieved the same level of accuracy as the complete sentences in Mandarin. They therefore concluded that in Mandarin it was the global contour that mattered the most in question identification.

Ma et al. (2011) also investigated the perception of statements and IntQs in Cantonese. They varied the last syllable of a sentence. The results from this study showed that Tone 33 (mid-level Tone) had the highest accuracy, while Tone 55 and 25 had the lowest accuracy (note: the other tones in Cantonese are 21, 23, 22). The results partly supported Yuan (2011) in that the tones ending with H were the most difficult to distinguish in Cantonese as well.

The above studies had diverging results. In some cases, a high tone facilitated the identification of the question, while in some other cases, T4 which is a falling tone, was easier to be identified as a question. In the remaining sections of this chapter, three experiments are conducted in Tianjin Mandarin in order to find out how Tianjin Mandarin IntQs and statements are identified.

5.3 Methodology

5.3.1 Participants

A total of 28 native Tianjin speakers, including 13 female and 15 male speakers, participated in this perception experiment. The participants were students at Tianjin University, Nankai University, and Tianjin College of Commerce. Their age ranged

from 20 to 28 years old (mean age = 20.25). None of them had any hearing loss or speech disorders. The participants were not the same participants as the informants in the production studies in Chapter 4. All 28 participants took part in all three experiments and received payments for their participation.

5.3.2 Stimuli

All three experiments used the same tune identification task, but differed in their stimuli. The three sets of stimuli used in the experiments were all from the production data in Chapter 4. The data were from two randomly chosen speakers, one female speaker (Speaker C) and one male speaker (Speaker D). The stimuli were all finely cut in windows of 0.01s ~ 0.05s in Praat (Boersma & Weenink, 2014).

The first set of stimuli were the monosyllabic words in isolation (hereinafter **Mono_(ISO)**), e.g. /ma/, were produced in Section 4.4. An example figure can be found in Figure 5.1a – the pink shaded area was one of the finely cut Mono_(ISO) stimuli. Table 5.1 is a summary of all tokens in Mono_(ISO).

Table 5.1 Stimuli: Mono_(ISO) and Mono_(SEN)

SYLLABLES	L TONE (TONE 1)	H TONE (TONE 2)	LH TONE (TONE 3)	HL TONE (TONE 4)
[ma]	妈 'mother'	麻 'hemp'	马 'horse'	骂 'scold'
[mau]	猫 'cat'	毛 'fur'	铆 'rivet'	帽 'hat'
[mi]	咪 'meow'	迷 'riddle'	米 'rice'	蜜 'honey'

The second set of stimuli were the monosyllabic words in the sentence prominence position (hereinafter **Mono_(SEN)**), e.g. the /mi/ syllable cut from the sentence “*ta xie mi zi-er*” in (31) below. They were extracted from the sentence data from Section 4.5 by means of extracting the pink-shaded area of Figure 5.1b.

The third set of stimuli were the same sentences produced in Section 4.5 but in full (hereinafter abbreviated as **SEN**), e.g. “*ta xie mi zi-er*” (as illustrated in Figure 5.1b as the whole sentence). The sentences used in SEN, which were also the source of Mono_(SEN) are presented in (30) and (31).

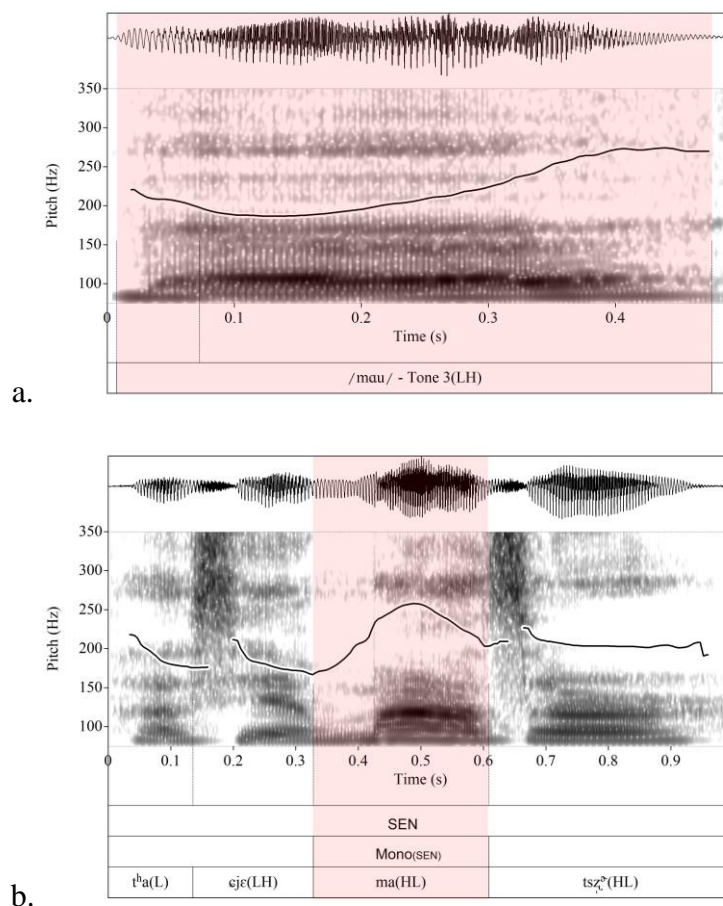


Figure 5.1 Examples of stimuli
(a. Mono_(ISO); b. Mono_(SEN) & SEN)

(30)

a. Statement 1:

ta(L) *shi*(HL) **ma**/*mao*/*mi* *zi*(HL)-*er*(0).

[^hta sʒ **ma** / maʊ / mi tsʒ[◌]]

It be **ma**/ *mao*/ *mi* character.

它是“**ma**”/“**mao**”/“**mi**”字儿。

“It is the character “**ma**”/“**mao**”/“**mi**”.

b. Intonational polar question 1 (IntQ 1)

ta(L) *shi*(HL) **ma**/*mao*/*mi* *zi*(HL)-*er*(0)?

[^hta sʒ **ma** / maʊ / mi tsʒ[◌]]

It be **ma**/ *mao*/ *mi* character?

它是“**ma**”/“**mao**”/“**mi**”字儿?

“It is the character “**ma**”/“**mao**”/“**mi**”?”

(31)

a. Statement 2:

ta(L) *xie*(LH) **ma**/*mao*/*mi* *zi*(HL)-*er*(0).

[^hta ɛjɛ **ma** / maʊ / mi tsʒ[◌]]

S/He write **ma**/ *mao*/ *mi* character.

他写“**ma**”/“**mao**”/“**mi**”字儿。

“S/He writes the character **ma**/ *mao*/ *mi*.”

b. Intonational polar question 2 (IntQ 2):

ta(L) *xie*(LH) **ma**/*mao*/*mi* *zi*(HL)-*er*(0)?

[^hta ɛjɛ **ma** / maʊ / mi tsʒ[◌]]

S/He write **ma**/ *mao*/ *mi* character?

他写“**ma**”/“**mao**”/“**mi**”字儿?

“S/He writes the character **ma**/ *mao*/ *mi*?”

Experiment 1 had Mono_(ISO) as shown in Table 5.1 **Error! Reference source not found.** as stimuli. The aim of this experiment was to test how hearers dealt with monosyllabic utterances, which contained a mixture of lexical tones and intonational tunes within a single syllable. The total number of stimuli is 96: 3 monosyllables ([ma], [mau], [mi]) * 4 lexical tones * 2 tunes (IntQ, Statement) * 2 speakers (male voice, female voice) * 2 repetitions = 96 stimuli.

Experiment 2 had Mono_(SEN) as stimuli, with the aim of investigating whether the hearer was able to identify the tunes from only the prominence words, which had no boundary information. A total of 192 stimuli were presented: 3 monosyllabic nuclei ([ma], [mau], [mi]) * 4 lexical tones * 2 preceding tones (Tone 3, Tone 4) * 2 tunes (IntQ, Statement) * 2 speakers (male voice, female voice) * 2 repetitions = 192 stimuli.

Experiment 3 used SEN as the stimuli. The aim of this experiment was to test whether hearers would identify longer utterances with varied prominence tones in the same way as they did in the Mono_(ISO) utterances. A total of 192 stimuli were presented: 3 monosyllabic nuclei ([ma], [mau], [mi]) * 4 lexical tones * 2 carrier sentences * 2 tunes (IntQ, Statement) * 2 speakers (male voice, female voice) * 2 repetitions = 192 stimuli.

5.3.3 Procedures

The experiments were conducted in a quiet room at Tianjin University. The participants were given breaks between each experiment. To avoid the task effect, the orders of the three experiments were randomised between groups of participants. The participants were tested in groups of a maximum of five. The auditory stimuli were played through individual closed-ear headphones (Sennheiser PX200 stereo headphones). The

participants were asked to decide whether each stimulus was a question or a statement. They made their choices on custom-made individual two-button handsets which were labelled as “陈述。” (‘statement’ in Chinese, with a Chinese full-stop) above the left button, and “疑问?” (‘question’ in Chinese, with a Chinese question mark) above the right button. The stickers were switched for left-handed participants to keep the “question” choice on their dominant-hand side. The participants were explicitly instructed to press the buttons with their thumbs and make their responses as accurate and as fast as possible.

The stimuli were programmed in experimental software SPLICE (Reetz & Kleinmann, 2008). The participants heard three bleeps at the beginning of the experiment as a signal for the start of the experiment. For each stimulus, the participants heard two 200ms bleeps separated by a 100ms pause. Then a 300ms pause was inserted before the onset of the audio stimulus. The time given to participants for deciding whether a stimulus was a question or a statement and pressing the button was 1500ms from the offset of the audio stimulus. The reaction time was recorded from the onset of the audio stimulus. In later statistical calculation for reaction time, the duration of the audio stimulus was deducted from the reaction time recorded by SPLICE. The procedures are illustrated in Figure 5.2.

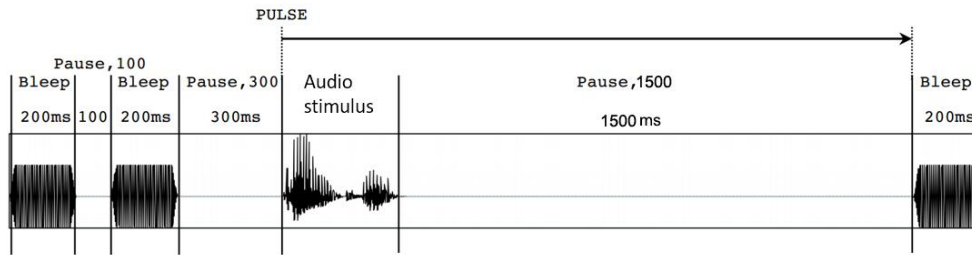


Figure 5.2 Procedures of the Perception Experiments

5.3.4 Data analysis

Responses that were made to stimuli that contained the syllable /ma/ were excluded in all three experiments. In Mandarin, multiple characters share the same pronunciation. One frequent token of /ma/ with Tone 4 is 骂 (“to scold”); however, in Tianjin Mandarin, the dialectal wh- word 嘛 (“what”) also has the same pronunciation. It thus created an opposite trend against the other rhymes (Figure 5.3 for an illustration of Tone-4 monosyllabic words with different rhymes – *a*, *ao*, and *i*). This trend shows that the semantic meaning of a word is much more influential in the identification of a question than the intonational cues. To prevent the influence from the semantics of *ma*, data of *ma* syllable, including all four tones across two tunes (as shown as shaded in Table 5.1 **Error! Reference source not found.**, (30), and (31)), were all excluded. 32 tokens of stimuli were excluded from Experiment 1, and 64 tokens were removed from Experiment 2 and 3 respectively.



Figure 5.3 Accuracy of responses by rhymes of stimuli

Responses that were made before voice onset were excluded from the analysis since the responses were made before the participants heard the audio stimuli.

All the accuracy data were analysed with generalised linear mixed models, using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2015). To analyse the accuracy of the responses, binomial regression models were constructed with the ACCURACY (Correct, Incorrect) as the dependent measure. Models including the main effects or the interaction were compared to the same models without main effects or interaction through Likelihood Ratio Tests, where the p-value is derived. When further examination of the interaction between the two main effects was needed, post hoc Tukey HSD tests were conducted. Detailed fixed and random factors are presented in each section.

The reaction time data were analysed with linear mixed effect models, with the same main factors and random factors as in the accuracy models. The α level is 0.05; when p-value is smaller than 0.05, this thesis describes the factor in examination as being “significant”.

To simplify the presentation of results, the tones are ranked by their descriptive results: the order of accuracy is descending, while that of the reaction time is ascending. Both orders indicate that on average the participants made more accurate and faster choices for the ones on the left than the ones on the right. The almost equal sign “ \approx ” is used for symbolising that the two elements connected do not differ significantly from each other. The greater-than and lesser-than signs are used to indicate significant differences.

5.4 Experiment 1: Monosyllabic Word in Isolation (Mono_(ISO))

5.4.1 Results

5.4.1.1 Accuracy

TONE (L, H, LH, HL), TYPE (IntQ, Statement), and their interaction TONE * TYPE were taken as fixed factors. SUBJECT and ITEM were taken as random factors, with intercepts for both SUBJECT and ITEM¹⁸. TYPE (of tunes) and TONE on their own were not significant factors that influenced the accuracy of the tune identification of Mono_(ISO): TYPE ($\chi^2(1) = 0.0043$, $p = 0.95$, n.s.), TONE ($\chi^2(3) = 4.41$, $p = 0.22$, n.s.); however, the interaction between TONE and TYPE was significant, TYPE*TONE ($\chi^2(7) = 61.26$, $p < 0.001$ ***).

¹⁸ Model did not converge with By- SUBJECT random slopes for the effect of TYPE .

The description data, as shown in Figure 5.4, exhibit that the accuracy rates of the identification of IntQ tune and statement tune almost followed the opposite trend: in IntQ, the most accurately identified lexical tone is T2 (H), while the lowest is T1 (L). The difference between T1 and T2 was very large, with T2 being identified correctly for most cases while the identification of T1 was only near chance level. A post hoc Tukey HSD test showed that the difference between T1 and T4 in IntQ was not significant; neither was the difference between T2 and T3. The relationship between the accuracy rates of different lexical tone in IntQ tune can be represented as in (32):

$$(32) \quad \text{IntQ: } T2(H) \approx T3(LH) > T4(HL) \approx T1(L)$$

As for the accuracy of the identification of statements (Figure 5.4), Tone 2 has significantly lower accuracy than all the other lexical tones. The accuracy of T2 in the statement tune was merely 41.7%, which was even lower than the lowest in IntQ tune, T1 (59.4%), and lower than chance level. With the results from the post hoc Tukey Test, the accuracy of the identification of statement tune is demonstrated in (33):

$$(33) \quad \text{S: } T3(LH) \approx T1(L) \approx T4(HL) > T2(H)$$

5.4.1.2 Reaction Time

TONE (L, H, LH, HL), TYPE (IntQ, Statement), and their interaction TONE * TYPE were taken as fixed factors. SUBJECT and ITEM were taken as random factors, with intercepts for both SUBJECT and ITEM, as well as by- SUBJECT random slopes for the effect of TYPE. TYPE, and TONE * TYPE interaction were the significant factors that affected the

reaction time of the responses: TYPE ($\chi^2(1) = 10.52, p = 0.001, **$), TONE ($\chi^2(3) = 6.03, p = 0.11, n.s.$), TYPE*TONE ($\chi^2(3) = 46.16, p < 0.001***$).

Figure 5.5 shows the reaction time results for the correct trials. The reaction time data showed the mirror pattern of the accuracy data. In IntQ tune, T1 had the longest reaction time of 611.16ms, while T2 was much shorter (457.99ms). Post hoc Tukey Test showed that T1 and T2, T4 and T2 differed significantly. It can be expressed with (34):

$$(34) \quad \text{IntQ: } T2(H) \approx T3(LH) < T4(HL) \approx T1(L)$$

The reaction time data for the statements also mirrors the accuracy results. T2 had a significantly longer reaction time than any other tones, as shown in (35)

$$(35) \quad \text{S: } T3(LH) \approx T1(L) \approx T4(HL) < T2(H)$$

5.4.2 Discussion

Table 5.2 Results of Mono_(ISO)

	Accuracy	Reaction Time
IntQ	$T2(H) \approx T3(LH) > T4(HL) \approx T1(L)$	$T2(H) \approx T3(LH) < T4(HL) \approx T1(L)$
S	$T3(LH) \approx T1(L) \approx T4(HL) > T2(H)$	$T3(LH) \approx T1(L) \approx T4(HL) < T2(H)$

Table 5.2 is a summary of the results of Experiment 1. The TYPE:TONE interaction exhibited an interesting general pattern that in IntQs, the rising tones (T2, T3) were

identified both more accurately and faster than the falling tones (T4, T1); while in statements, the low tones (T3, T1) were better identified than the high tones (T4, T2).

In IntQs, T2 (H), with both high initial and high ending tone, was correctly identified as questions most of the time. T3 (LH) ends with a H tone too; yet, the initial L tone of T3 became a distractor in IntQ identification. Therefore, although the accuracy of T3 was significantly above chance, it was not as high as T2. From these results, it is evident that the listeners were trying to look for a H ending tone for the IntQs. These results are consistent with the production data in Chapter 4, in which the IntQ tune was found to have a floating H% boundary tone associated with the IP boundary, despite the fact that the floating tone did not have a clear phonetic realisation. During the identification process, the listeners searched for this information from the audio stimuli. A H ending from the lexical tones are confusing during the process of IntQ tune identification.

On the contrary, for the statements, the crucial cue was not the right boundary but the left boundary. The initial L tone facilitated the identification of a statement. T1 and T3 both have an initial L, so they both achieved the highest accuracy. When the initial tone did not help with identification, such as T4 and T2, the ending tone interfered with the identification. T2 ends with a H tone that interferes with the identification, therefore the accuracy is extremely low. T4, although it does not have a facilitating L initial tone, it does not have an interfering H tone at the end either. What this L initial tone actually represents is the register. In the production study in Chapter 4, IntQ tune was observed to have a higher register than the corresponding statement tune. The results here show that the register information is also used during the perception process.

In summary, the results of Experiment 1 suggest that during the identification of IntQ tune, the floating H% is what the listeners search for; and the initial L tone (register) interferes with the identification. While the identification of statements regards the initial L (register) as the most crucial cue; the ending H interferes with the identification.

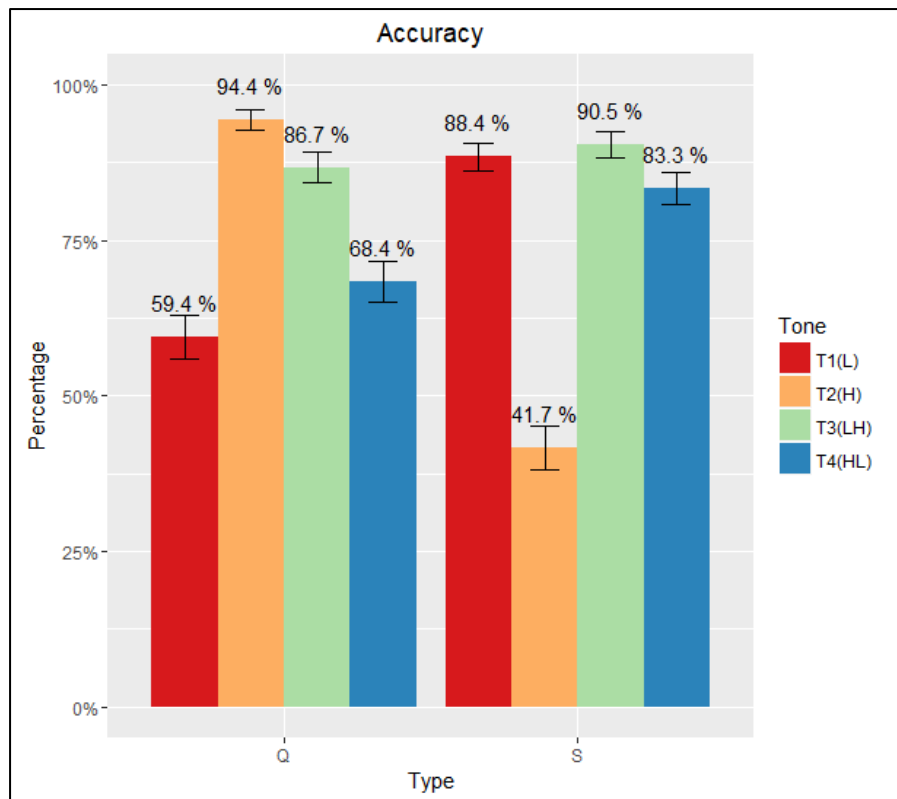


Figure 5.4 Mono_(ISO): Accuracy

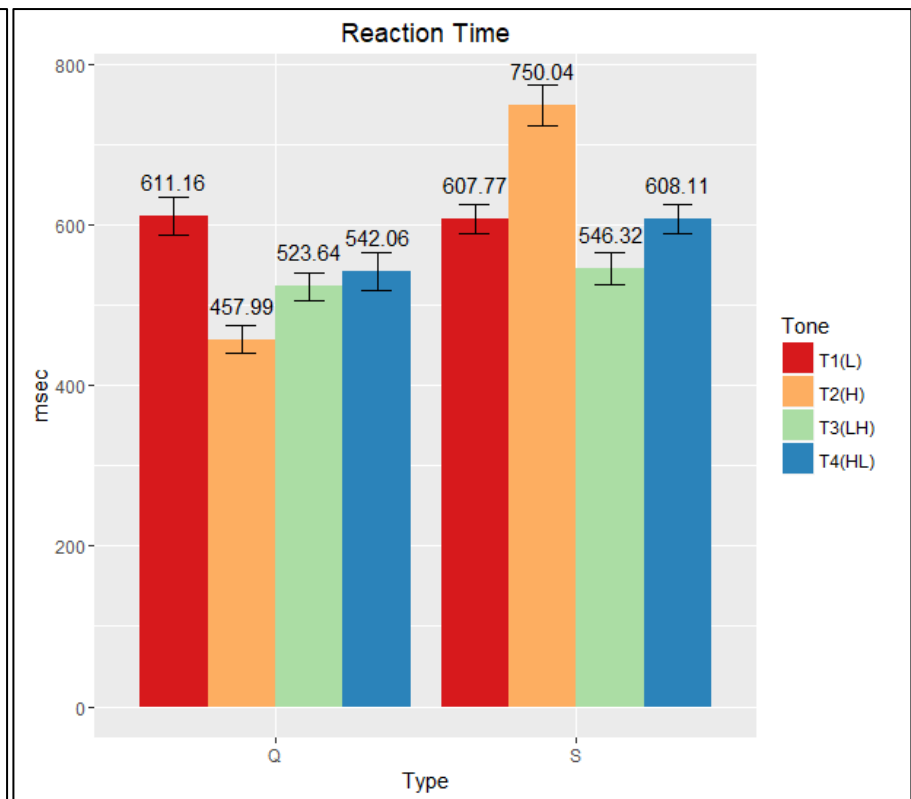


Figure 5.5 Mono_(ISO): Reaction Time

5.5 Experiment 2: Monosyllabic Word in Sentence Prominence (Mono_(SEN))

5.5.1 Results

5.5.1.1 Accuracy

The fixed factors were TONE (L, H, LH, HL), TYPE (IntQ, Statement), and their interaction TONE * TYPE. SUBJECT, and ITEM were the random factors, with intercepts for each of the two random factors. In this experiment, TYPE, TONE, and their interaction were all significant factors that influenced the accuracy of the tune identification of Mono_(SEN): TYPE ($\chi^2(1) = 46.72, p < 0.001, ***$), TONE ($\chi^2(3) = 12.55, p = 0.02, *$), TYPE*TONE ($\chi^2(3) = 85.22, p < 0.001***$).

The accuracy for Mono_(SEN) shown in Figure 5.8 is very different from the Mono_(ISO) results in Experiment 1. The average accuracy for IntQs was much lower than that of the statements. In the IntQ results, Tone 3's accuracy was the lowest, only 20.1%; Tone 4 achieved the highest accuracy, 60.8%. Based on the descriptive data and post hoc Tukey Test, the relationship between the accuracy rates of different lexical tone in IntQ tune can be represented as in (36):

$$(36) \quad \text{IntQ: } T4(\text{HL}) \approx T2(\text{H}) > T1(\text{L}) > T3(\text{LH})$$

As for the accuracy of the identification of statements (Figure 5.8), Tone 2 has significantly lower accuracy than all the other lexical tones. The accuracy of T2 in the statement tune was at chance level, 50.4%, while the other tones all had much higher accuracy. With the results from the post hoc Tukey Test, the accuracy of the identification of statement tune by lexical tones is demonstrated in (37):

$$(37) \quad S: T1(L) \approx T3(LH) > T4(HL) > T2(H)$$

5.5.1.2 Reaction Time

The fixed factors were TONE (L, H, LH, HL), TYPE (IntQ, Statement), and their interaction TONE * TYPE. PRETONE(LH, HL), SUBJECT, and ITEM are the random factors, with intercepts for all three random factors, as well as by- SUBJECT random slopes for the effect of TYPE. In contrast with the accuracy data, the reaction time data did not differ across TYPE or TONE. The interaction TYPE*TONE was the only significant factor that affected the reaction time of the responses: TYPE ($\chi^2(1) = 1.79, p = 0.18, n.s.$), TONE ($\chi^2(3) = 2.11, p = 0.55, n.s.$), TYPE*TONE ($\chi^2(3) = 12.01, p = 0.007^{**}$).

Figure 5.9 shows the reaction time results for the correct trials. The average reaction time for all the tones of both types of tunes was approximately the same, ranging from 593.54ms to 654.92ms. T1 and T2 in the statement tune was the only pair that differed from each other significantly. Nevertheless, T3 and T1 were longer than T2 and T4 in IntQ tune, which again showed mirror pattern with the accuracy. In the statements, T2 was significantly longer than other tones, followed by T4, T3, and T1. (38) and (39) illustrate the trend:

$$(38) \quad \text{IntQ: } T4(HL) \approx T2(H) \approx T1(L) \approx T3(LH)$$

$$(39) \quad S: T1(L) \approx T3(LH) \approx T4(HL) \approx T2(H)$$

5.5.1.3 Dissimilation Effect

The stimuli in this experiment were extracted from sentences with different tones (LH, HL) preceding the Mono_(SEN) syllables. This section is to examine the effect of the tonal dissimilation rules on the tone identification process. Note all the tonal dissimilation rules in Tianjin Mandarin work on the first syllable in disyllabic combinations. The effect was nevertheless examined to investigate possible carry-over effects. Since Tone 2 does not go through any tonal dissimilation rules, relevant data were excluded to keep the data comparable. Since Tone 4 + Tone 4 rule was reported to be not effective anymore by a number of studies (reviewed in Section 3.4.2, Chapter 3), this combination was not counted as a dissimilation rule.

Figure 5.6 and Figure 5.7 display the accuracy and reaction time results for the identifications with and without tonal dissimilation rules. Accuracy and reaction time were statistically tested with a mixed-effect model – TONE, TYPE, RULE of tonal dissimilation, and the interaction between TYPE*RULE were the fixed factors. SUBJECT and ITEM were taken as random factors, with intercepts for both SUBJECT and ITEM, as well as by- SUBJECT random slopes for the effect of TYPE. The interaction of TYPE*RULE was the most important factor in the current analysis.

The effect of the interaction of TYPE*RULE ($\chi^2(1) = 9.89, p = 0.0016^{**}$) was significant on the accuracy. However, a Tukey post hoc test found that within each TYPE, RULE did not make a significance difference. There was not any significant TYPE*RULE interaction on the reaction time data either.

In the descriptive data, the condition without tonal dissimilation rule in IntQs was higher than the condition with tonal dissimilation rules. This, to a certain degree, indicated that tone dissimilation added cognitive load on the identification of the non-default tune. A more plausible explanation was that the tonal dissimilation rules changed the first tone of a disyllabic combination (i.e. the PreTone in this experiment), and the carry-over effect of the first changed tone led to distortion of the following tone (i.e. Mono_(SEN) syllables in the current experiment).

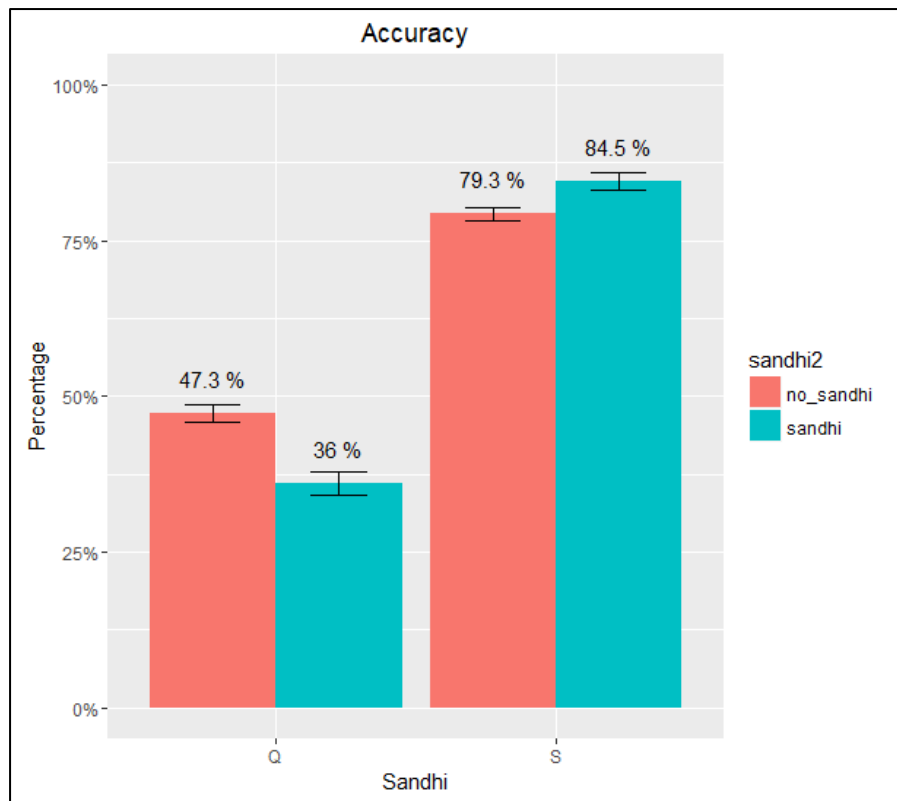


Figure 5.6 Dissimilation Effect – Mono(SEN): Accuracy

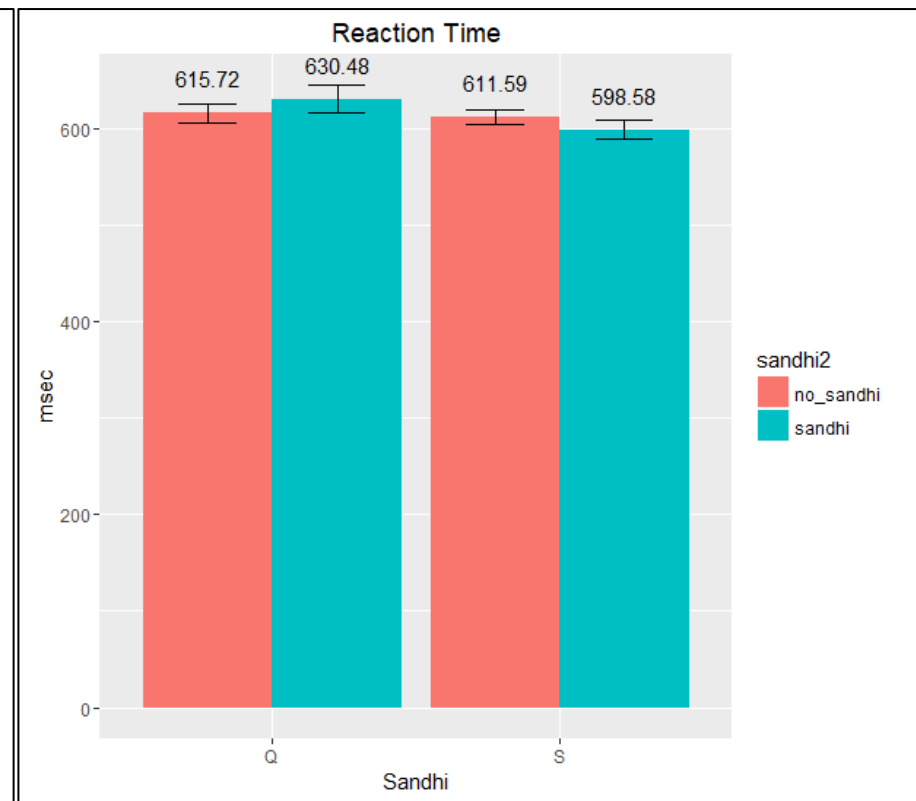


Figure 5.7 Dissimilation Effect – Mono(SEN): Reaction Time

5.5.2 Discussion

The aim of Experiment 2, as mentioned in Section 5.3.2, was to examine how native listeners identify the tunes from only the prominent words without any boundary information. The results showed that the identification rate of the high tones (T4 and T2) was significantly higher than that of the low tones (T1, and especially T3) in the IntQs, while the low tones were more accurately identified in statements. The reaction time results are coherent with the accuracy results: the more accurate the responses, the faster they are.

Table 5.3 Results of Mono_(SEN)

	Accuracy	Reaction Time
IntQ	$T4(HL) \approx T2(H) > T1(L) > T3(LH)$	$T4(HL) \approx T2(H) \approx T1(L) \approx T3(LH)$
S	$T1(L) \approx T3(LH) > T4(HL) > T2(H)$	$T1(L) \approx T3(LH) \approx T4(HL) \approx T2(H)$

The results of Experiment 2, as shown in Table 5.3 above, lent strong support for the conclusions that were drawn in Experiment 1. Native listeners only use the register information for prominent words which are away from the sentence boundaries. Even in IntQ, the speakers prefer the ones that start high, instead of searching for a H at the end.

The fact that the IntQ had such a low identification rate also suggests that the H% floating boundary tone is very important in IntQ tune. Without the boundary information, the identification of IntQ tune is extremely difficult.

Among all the results, T3 had the lowest accuracy in IntQ identification, even lower than T1 (L Tone) despite having a H ending tone. Three factors may have contributed to the low accuracy: (1) Tone 3 is often creaky – sometimes it loses its original contours in continuous speech due to the pitch lowering effect of creakiness; (2) the stimuli in this experiment were all cut from continuous speech, the contours of lexical tones were seriously affected by the preceding and the succeeding lexical tones due to coarticulation, as well as the sentence-level prosody; (3) T3 has the longest duration among all tones and goes through tonal changes – either by losing the first half or the latter half of the tone during fast speech (J. Zhang & Lai, 2010, p. 162). The combination of these three factors made T3's lexical tone identification hard and unpredictable. The difficulty in identifying the lexical tone itself thus leads to more confusion over the sentence level prosody. Therefore, whether the discrepancy between T1 and T3 provides insight for tone identification or tune identification is unclear.

The reaction times in this experiment did not differ from each other by a large margin. It is also because the difficulty of the task was so high that the reaction time demonstrated a ceiling effect.

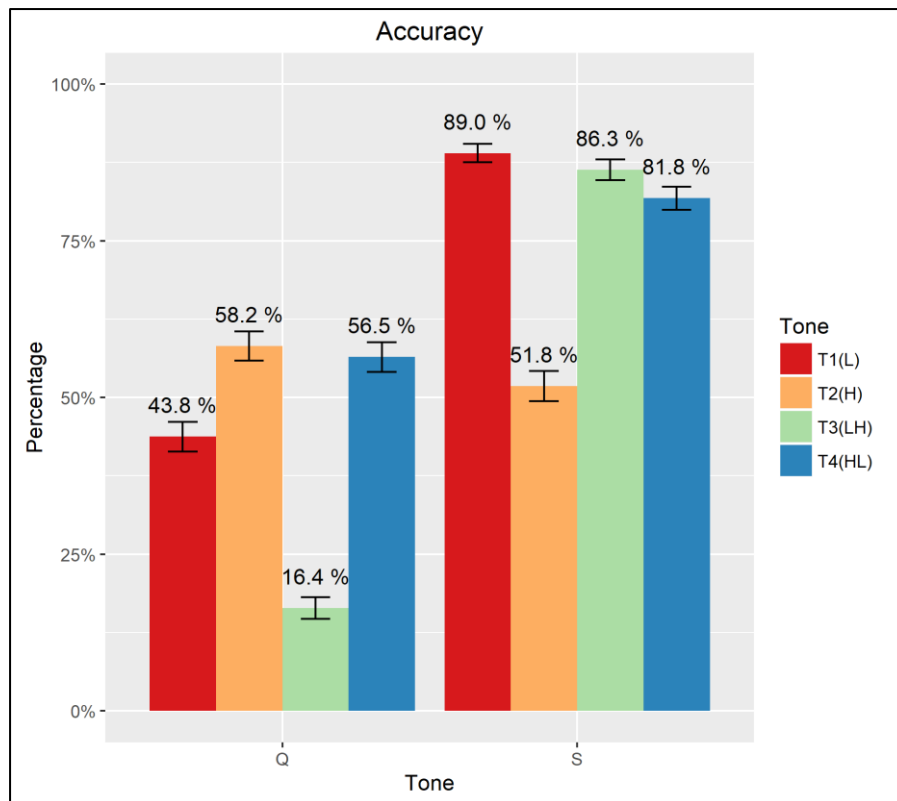


Figure 5.8 Mono_(SEN): Accuracy

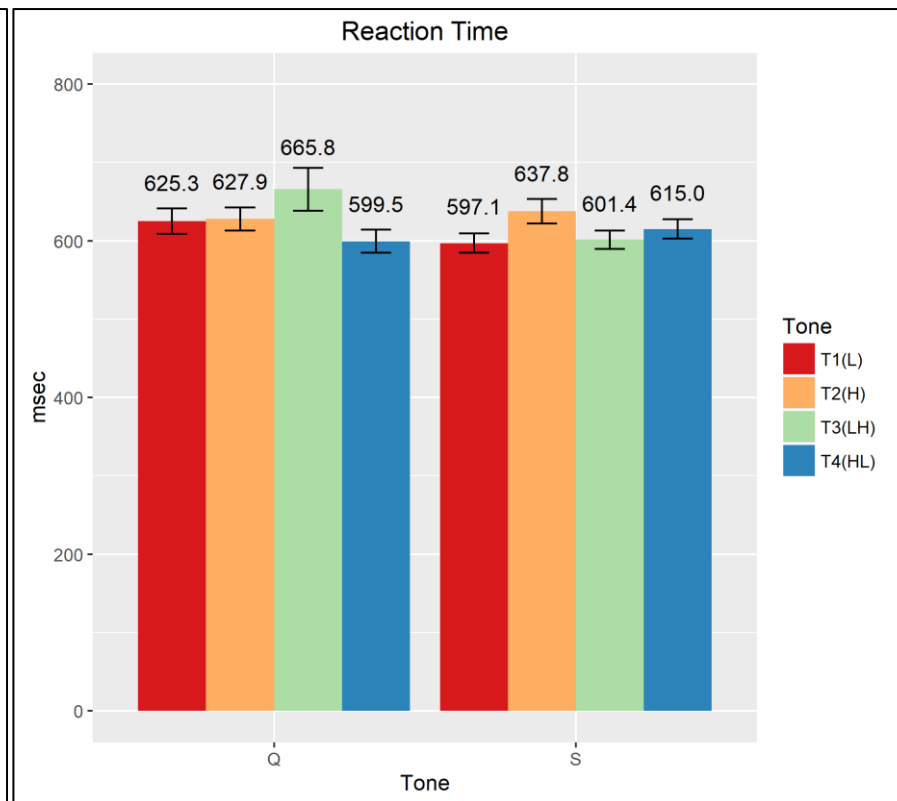


Figure 5.9 Mono_(SEN): Reaction Time

5.6 Experiment 3: Sentences with Monosyllabic words as prominence (SEN)

5.6.1 Results

5.6.1.1 Accuracy

The model used in this section is identical to the model for the accuracy of identifying $\text{Mono}_{(\text{SEN})}$ tunes: the fixed factors were TONE (L, H, LH, HL), TYPE (IntQ, Statement), and their interaction TONE * TYPE. SUBJECT, and ITEM were the random factors, with intercepts for each of the two random factors. Similar to the results of the accuracy rate in Experiment 1, TYPE and TONE on their own were not significant factors that influenced the accuracy of the tune identification of $\text{Mono}_{(\text{ISO})}$: TYPE ($\chi^2(1) = 0.20$, $p = 0.66$, n.s.), TONE ($\chi^2(3) = 5.74$, $p = 0.13$, n.s.); however, the interaction between TONE and TYPE was significant, TYPE*TONE ($\chi^2(3) = 36.21$, $p < 0.001^{***}$).

The accuracy in this experiment, as shown in Figure 5.10, is the highest among all three experiments. Similar to the results in Experiment 2, the accuracy of IntQ saw the highest in T2 and T4, both reaching over 90%; T3 and T1 were lower, but still reached more than 70%. For statements, the accuracy of T4, T3, and T1 were all more than 90%, but T2 once again fell to chance level, 53.9%.

Combining the descriptive data and the post hoc Tukey HSD test, the results can be demonstrated as in (40) and (41):

$$(40) \quad \text{IntQ: } T2(\text{H}) \approx T4(\text{HL}) > T3(\text{LH}) \approx T1(\text{L})$$

$$(41) \quad \text{S: } T4(\text{HL}) \approx T3(\text{LH}) \approx T1(\text{L}) > T2(\text{H})$$

5.6.1.2 Reaction Time

The model used in this section was the same model as for the reaction time of the identification of Mono_(SEN) tunes in Section 5.5.1.2) The fixed factors were TONE (L, H, LH, HL), TYPE (IntQ, Statement), and their interaction TONE * TYPE. PRETONE(LH, HL), SUBJECT, and ITEM were the random factors, with intercepts for all three random factors, as well as by- SUBJECT random slopes for the effect of TYPE. TYPE, and TONE*TYPE interaction were the significant factors that affected the reaction time of the responses: TYPE ($\chi^2(1) = 0.25$, $p = 0.62$, n.s.), TONE ($\chi^2(3) = 4.76$, $p = 0.19$, *), TYPE*TONE ($\chi^2(3) = 16.62$, $p = 0.0008$ ***).

Figure 5.11 shows the reaction time results for the correct trials. As with the highest accuracy rate, this experiment had the shortest reaction time too. The trend was consistent with that of the accuracy: T4, T2, T3 were all shorter than T1 in IntQ; T2 is longer than all other tones in statements.

It can be transcribed as (42) and (43):

$$(42) \quad \text{IntQ: } T4(\text{HL}) \approx T2(\text{H}) \approx T3(\text{LH}) < T1(\text{L})$$

$$(43) \quad \text{S: } T3(\text{LH}) \approx T4(\text{HL}) \approx T1(\text{L}) < T2(\text{H})$$

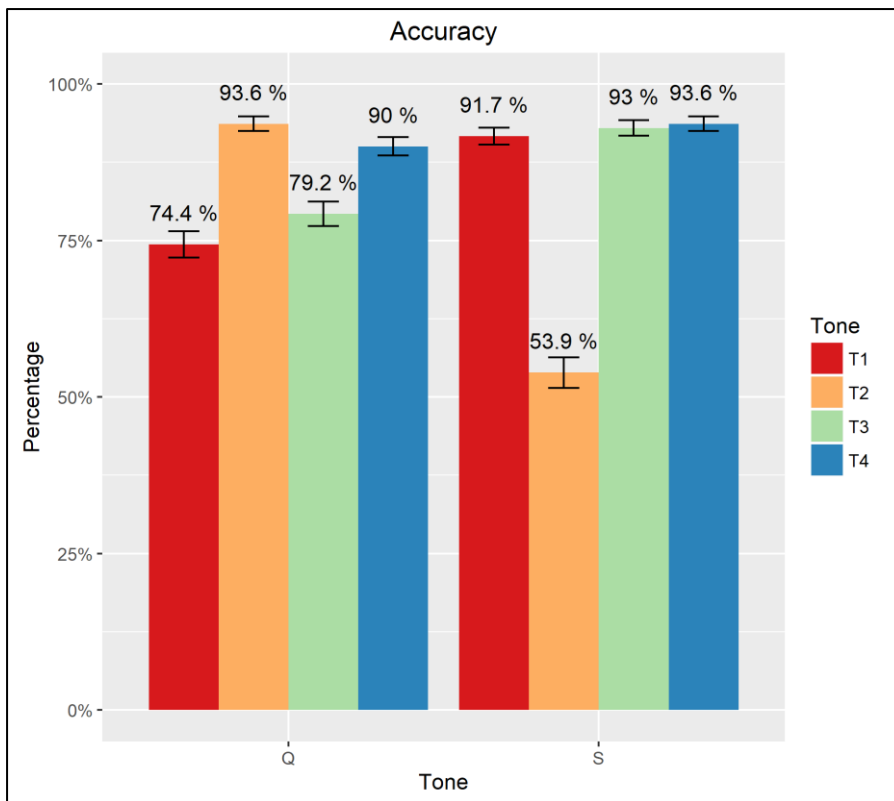


Figure 5.10 SEN: Accuracy

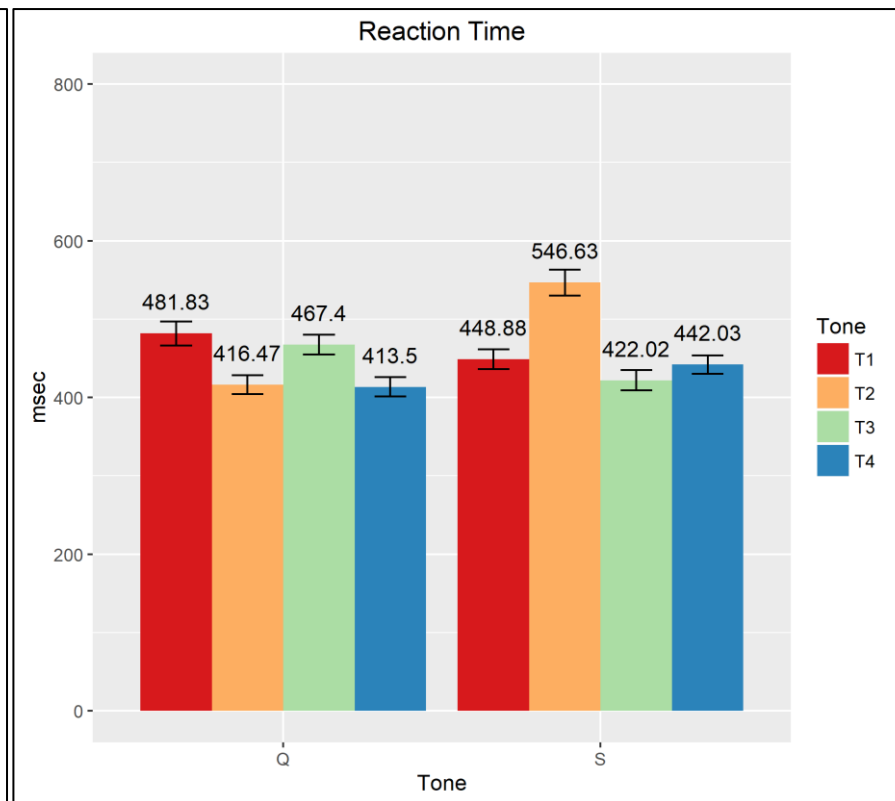


Figure 5.11 SEN: Reaction Time

5.6.1.3 Dissimilation Effect

Two carrier sentences with different tones (LH, HL) preceding the Mono_(SEN) syllables. The data selection criteria were the same as in Section 5.5.1.3. The accuracy and reaction time results are illustrated in Figure 5.12 and Figure 5.13. Accuracy and reaction time were statistically tested with a mixed-effect model – TONE, TYPE, RULE of tonal dissimilation, and the interaction between TYPE*RULE were the fixed factors. SUBJECT and ITEM were taken as random factors, with intercepts for both SUBJECT and ITEM, as well as by- SUBJECT random slopes for the effect of TYPE. The interaction of TYPE*RULE was the most important factor in the current analysis. The accuracy results showed a significant difference (TYPE*RULE ($\chi^2(3) = 30.97$, $p < 0.001^{**}$) between the tones that went through dissimilation and the tones that did not. In the statement tune, whether there were dissimilation rules or not did not create a huge difference, although the ones that went through tonal dissimilation had slightly lower accuracy by 3.7%. In the IntQ, however, this discrepancy was enlarged to 17.1%. The reaction time data did not differ between the two conditions (TYPE*RULE ($\chi^2(1) = 2.64$, $p = 0.104$, n.s.). The description data nevertheless showed a little increase for the tonal combinations that have to go through tonal dissimilation, although the difference was only 10.77ms.

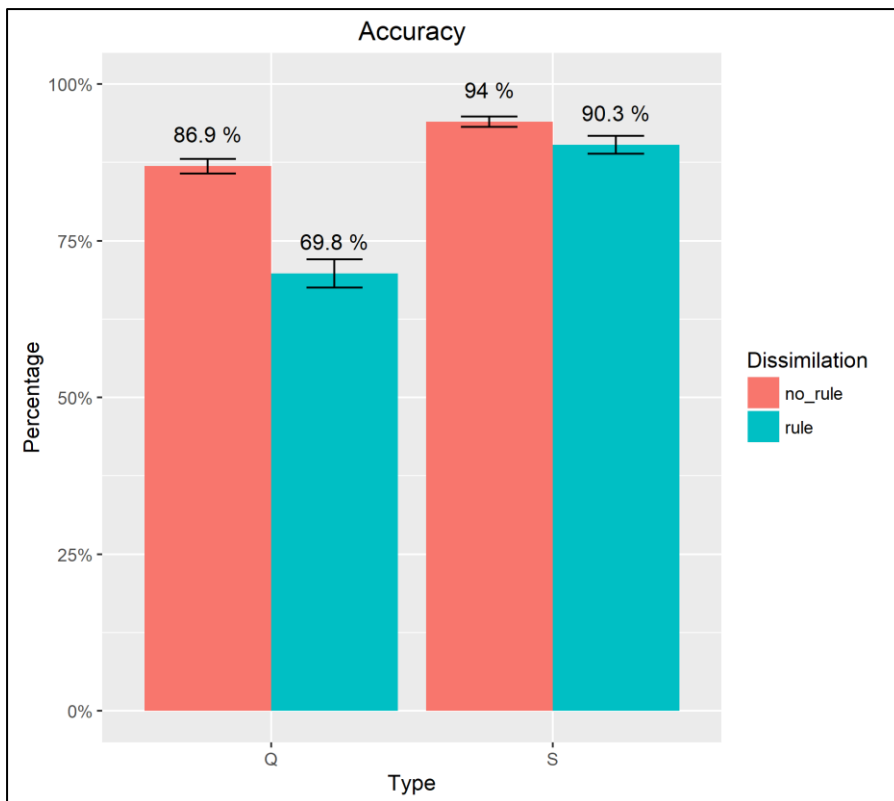


Figure 5.12 Dissimilation Effect – SEN: Accuracy

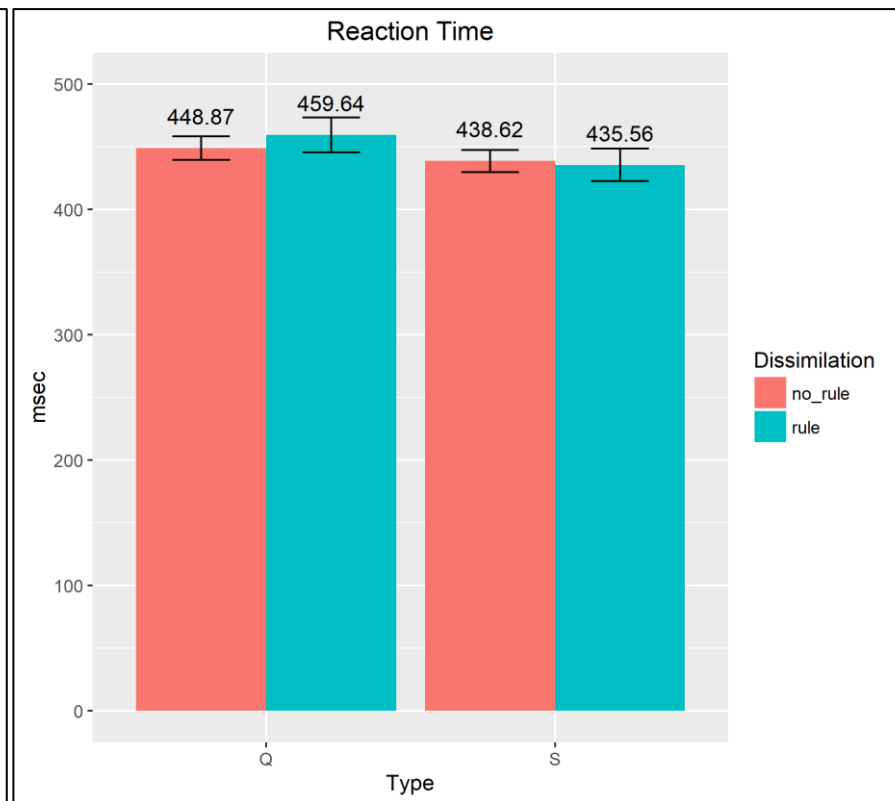


Figure 5.13 Dissimilation Effect – SEN: Reaction Time

5.6.2 Discussion

The results of Experiment 3 show that longer sentences are easier to identify, since they carry more information, especially pitch register information, since the register is raised as early as the pre-nuclear part of the sentences, according to the production study of IntQ in sentences in Chapter 4. The sentences also contain the floating H% boundary tone in the end, which further facilitates the identification of the tunes.

The fact that the reaction times in this experiment were shorter than the previous two experiments lends support for incremental processing of the intonation, i.e. the online processing of intonation which starts early on instead of waiting for the whole utterance to be completed. The short reaction time in this experiment indicates that processing already happened before the listeners were provided with the boundary information. The accuracy results also support this point. The major results are repeated in Table 5.4. The accuracy of the IntQ shows that the high tones were more accurate than the low tones. In Experiment 1, the listeners primarily searched for the H boundary tone in the process of identifying IntQs. However, the results of this study did not resemble the pattern in Experiment 1, but were more similar to Experiment 2, in which the boundary information was not provided so the listeners relied more on the register. The reason behind this is incremental processing. Combining the accuracy and the reaction time results, it is clear that as soon as the listeners heard an indicator for the register, either H or L, they start to make the decision. In many other studies, online methods were used to examine the time course. For example, Heeren et al. (2015) employed eye-tracking to examine the online processing of intonation and how it affected comprehension. The participants made their own decision early on hearing the first half

of the stimuli, and a delayed pattern showed when there was a mismatch between the participants' early decision and the final complete intonation contour.

Table 5.4 Results of SEN

	Accuracy	Reaction Time
IntQ	$T2(H) \approx T4(HL) > T3(LH) \approx T1(L)$	$T4(HL) \approx T2(H) \approx T3(LH) < T1(L)$
S	$T4(HL) \approx T3(LH) \approx T1(L) > T2(H)$	$T3(LH) \approx T4(HL) \approx T1(L) < T2(H)$

The last issue to discuss in this experiment is that tonal alternation rules have a significant effect on the identification of intonational tunes, despite the fact that they belong to lexical-level prosody. The effect is exaggerated when the cognitive load is increased in the IntQ tune, since the statement tune is the default tune and processing IntQ tune itself is already an effort. This is consistent with the results in Liu et al. (2016) that it was more difficult for the participants to perceive a question tune than a statement tune.

5.7 Discussion and Conclusions

This chapter contains three perception experiments to investigate how the IntQ tune and the statement tune are identified by native Tianjin speakers. The key findings are summarised as follows:

Experiment 1 examines the identification of the tunes with monosyllabic words in isolation, which contain both the register information and the boundary information that are important to the IntQ tune. The results suggest that when native listeners identify the IntQ, the floating H% is what the listeners search for; an initial L tone, which is an

indicator of the pitch register, interferes with the identification; when the native listeners identify the statement tune, the register indicator, the initial L, is the most crucial cue, while an ending H would interfere with the identification.

Experiment 2 examined the extracted monosyllabic sentence prominence words, which did not contain any boundary information. The listeners consequently only used the register information for the identification.

Experiment 3 examined longer sentences which contained both the register information and the boundary information. However, since the length of the sentences was long, it allowed incremental processing to happen before the boundary information was presented. The register cue therefore was the cue that was used in the identification process, while the boundary information only facilitated the identification in part of the cases. This experiment also revealed that tonal dissimilation rules, despite being a part of the lexical prosody, places more cognitive burden on the identification, especially in the IntQ tune, which, by itself, is already more difficult than the statement tune.

These results are not consistent with some of the findings in Standard Mandarin. Two Mandarin studies (B. R. Xu & Mok, 2012; Yuan, 2011) both found the IntQ tune of Tone 4 (HL Tone) to be the easiest to identify while that of the Tone 2 (LH Tone) was the most difficult. The results of the current study are completely opposite in that the high rising tone was the easiest to identify as the IntQ tune, regardless of what position of an utterance it is in. However, the results from the current study do support the findings in Ni and Kawai (2004) that a higher pitch peak for Tone 2 would indicate a question tune and a higher register facilitates with the identification of question tunes

with the falling Tone 4. Although the Mandarin results in Xu and Mok (2012) are different from the results of the experiments from this chapter, their Cantonese results saw some similarities with the findings in this chapter that the tones with higher register provided more information for the identification of questions with cut-off endings. Due to the very small number of comparable studies, it is difficult to draw any meaningful cross-linguistic conclusions based on the implications of some similarities and differences between these experiments. The perception, especially identification, of intonation tunes definitely needs to be studied more in-depth.

I end this chapter by providing some brief answers to the questions asked at the beginning of this chapter (Section 5.1).

Q1: How well are IntQs perceived by native Tianjin speakers?

A1: It depends on the length of the utterances and the information contained in the utterances. If the utterances contain both the register information and the boundary information, the identification rate is well above chance level; if boundary information is removed, then the identification rate is much lower. The length of the sentences matter too – longer sentences can have the identification rate as high as over 90%, while shorter utterances cannot reach so high.

Q2: Does lexical tone play a role in intonation identification? If yes, is there a tonal bias?

A2: Yes, the lexical tones create competition with the intonational tunes. The high rising tone is identified as having an IntQ tune correctly more often; similarly, the low falling tone is identified as having statement tune correctly more often.

Q3: Do tonal alternation rules have an effect on tune identification?

A3: Yes. Although tonal alteration happens on the lexical level in production, it has an effect on the identification of sentence tunes, since it serves as extra cognitive load.

Chapter 6 Production Studies:

Chanted-Call Tune in Tianjin Mandarin

6.1 Introduction

The previous two chapters investigated the production and perception of Intonational Yes-No questions in Tianjin Mandarin. The results demonstrated that intonational contours did not override lexical tone contours for the intonational Yes/No questions – a floating boundary tone without phonetic realisation itself was associated with the IP boundary of IntQ, which facilitates the rising of rising lexical tones and deters the falling lexical tones from falling substantially. Are all boundary tones in Tianjin Mandarin floating? Do all the boundary tones have no phonetic realisation? In this chapter, I study another tune that is usually exaggerated – chanted call – with the aims of (i) investigating further possibilities about the intonational tunes in Tianjin Mandarin, a tonal language; (ii) contributing to the intonational typology of the chanted call tune by supplementing phonetic data and phonological analyses in this tonal language.

Chanted call tune was observed as early as in Pike (1945), in which he called it “spoken chant” (pp. 71-72). This contour also has a number of other names such as “calling contour” (Gibbon, 1976), “vocative chant” (Lieberman, 1975), “stylised fall” (Ladd, 1978), and “chanted call” (Hayes & Lahiri, 1992). I adopt the term “chanted call” (hereinafter abbreviated as “CC”) since it captures both the prosodic feature (“chanted”) and the functional feature (“call”) of the tune in question.

This chapter is organised as follows: Section 6.2 reviews the previous studies on the CC tunes in different languages; Section 6.3 presents the general methodology used in

the subsequent two production studies of the CC tune in this chapter; Section 6.4 reports a production study of monosyllabic chanted calls; Section 6.5 details another production study of disyllabic chanted calls; and finally, Section 6.6 discusses the results of both studies and presents a summary of the conclusions.

6.2 Literature Review: Chanted Call Tunes in World Languages

Although the CC tune is one of the most studied tunes, the majority of the existing studies are on Indo-European languages (Arvaniti & Baltazani, 2005; Arvaniti, Žygis, & Jaskuła, 2016; Borràs-Comes, Sichel-Bazin, & Prieto, 2015; Fagyal, 1997; Frota, 2014; Frota et al., 2015; Grice, Baumann, & Benz Müller, 2005b; Gussenhoven, 1993; Hayes & Lahiri, 1992; Ladd, 1978; Sadat-Tehrani, 2007). Very few studies have investigated the CC tune outside Indo-European languages. For instance, Varga (2008) studied Hungarian's CC tune and compared it with the English counterpart. Rialland and Robert (2001) briefly reported the CC tune in Wolof, a non-tonal Niger-Congo language; Abe (1998) also reported an observation on Japanese CC tune; Luksaneeyanawin (1998) investigated it in Thai, and it is the only study that is on a tonal language; however, only the observations were outlined and without presenting any acoustic data.

In the abovementioned languages, the CC tunes share many features. The first shared feature is that there is a H* pitch accent involved. In some languages, the pre-prominence L is obligatory under different circumstances and in different languages, while in some others it is optional. The second shared feature is that the boundary tone is lower than the H* pitch accent. In some languages, it is labelled as a downstepped H, i.e. !H, while in some others, the boundary tone is simply a L%. The third feature is that

the boundary tones in the CC tune of these languages are mostly lengthened, or in another word, sustained. In general, the CC tunes in most of these languages have a scooped shape as in (44), i.e. LHL or LH!H. However, the specific associations and alignments are language-specific, and vary in the same language according to different pragmatic functions.

(44) typical CC tune contour



In English, the CC tune has been studied by different researchers, and consequently there have been a few different descriptions of the same contour. For example, Pierrehumbert (1980) proposed $H^{*+} L- H- L\%$ as the notation for the CC tune. Hayes and Lahiri (1992) used the analyses in Liberman (1975) and noted the CC tune in English as HM. The M tone was described as aligning with the strongest stress after H in their study, which suggested that the M was not considered as a boundary tone. Ladd (1996, 2008, pp. 117–118) analysed the CC tune as $(L)H^{*!}H$, as shown in (45):

(45)	Johnny	Jonathan	Rebecca	Jo-ohn			
	$H^{*} !H$	$H^{*} !H$	$L H^{*} !H$	$H^{*} !H$			

Hayes and Lahiri (1992) also studied the CC tune in Bengali, which shares the same underlying form as the English CC tune, i.e. HM in their analyses. However, their analyses showed that the Bengali CC tune differed from the English one in that Bengali followed syllable-based rules to map the tunes to the texts, while in English the H was always associated with the strongest beats. Moreover, in Bengali, there was a limit on

the length of the names – names longer than three syllables had uncertain tone assignments.

Another language that has very similar CC tune with English is Dutch, which is able to allow more than one downstepped !H as a sequence. One of the examples which has as many as three downstepped !Hs by Gussenhoven (1993, p. 42) is shown in (46):

(46) Dutch chanted call (adapted from Gussenhoven, 1993, p. 42)

nép-àlmanàk-je
| | | |
H !H !H !H

“fake-almanac + DIM”

Gussenhoven analysed this contour as HL, in which the H-Spreading, H-Splitting, further downstepping among all the steps, to become a sequence of !Hs. As for the boundary tone, Gussenhoven et al. (2012) proposed, in the latest version of ToDI (Transcription of Dutch Intonation), to use the boundary sign % alone to represent the sustained boundary tone, which kept unchanged from the previous !H.

The default CC tunes in different languages are summarised in Table 6.2¹⁹, with the tune for default calls being labelled in bold. With regard to the pitch accents, most languages have at least a H* associated with the prominence, except for Sulista Brazilian Portuguese (Frota et al., 2015), which has a L* pitch accent. Greek (Arvaniti & Baltazani, 2005) is also different in terms of the alignment of the L tone association:

¹⁹ Some figures taken from the original articles can be found in Appendix 2.

while the other languages mostly have an optional or obligatory L associated with the pre-stress syllables, the Greek pitch accent has the L aligned and associated with the prominence which is followed by a H.




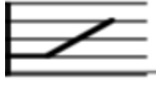
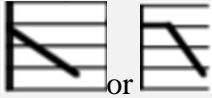
As for the boundary tones in the default CC tune, different notations have been used to describe the sustainment of the boundary tone. For example, Varga (2008) proposed a 0% to express the concept of lengthening, following Grabe (as cited in Varga, 1998); Gussenhoven et al. (2003) and Grice et al. (2005a) used the boundary tone sign “0%” alone to indicate a sustained boundary tone; Rialland noted the sustainment with text, as in “H(sust)%”. Although in Pierrehumbert’s (1980) original proposal, she used a H-phrase tone and a L% together to describe the lengthened boundary tone in the CC tune, it remains a debate whether phrase tones actually exist.

The contours of the boundary tones in the languages listed in Table 6.2 can be divided into three major groups: most languages have a boundary tone that contains a downstepped !H% (e.g. Greek, Polish, Portuguese, Wolof), some have a L% but appear to be very similar to the previous category (e.g. French), since both suggest a lowered plateau. Catalan (Borràs-Comes et al., 2015) has a bitonal HL% as a boundary tone for the default CC tune, despite the fact that in some non-default calls the boundary tones are also !H% or H%. In many languages presented in the table, a phrase accent is depicted too, with some being indicative of the lengthening of the boundary tone.

Luksaneeyanawin (1998) studied the calling contour in Thai, a contour tonal language with five lexical tones as shown in Table 6.1. In the CC tune, the static lexical tones (high tone, mid tone, and low tone) were levelled as higher mid, mid, and lower mid.

The pitch range of the dynamic tones (rising tone and falling tone) became smaller. Regrettably, no figures were provided in the study to help with our understanding of how the pitch range became smaller. Despite not knowing the exact contours, we are still able to conclude that in Thai CC tune, the lexical tones are compressed and overridden by the intonational contour.

Table 6.1 Lexical Tones in Thai (adapted from Luksaneeyanawin, 1998)

STONE	SCHEMATISED CONTOUR	DESCRIPTION
HIGH		High narrow continuous rise and fall
MID		Mid narrow continuous fall
LOW		Low narrow continuous fall
RISING		Low wide delayed rise
FALLING		High wide continuous or delayed fall

Apart from describing the tunes of the CCs, studies have also been done to investigate the effect of different pragmatic factors on the CC tunes. The politeness factor has been studied in Catalan (Borràs-Comes et al., 2015) and Greek (Arvaniti & Baltazani, 2005). Both languages showed a H% boundary tone, with the Greek boundary tone rising slightly at the end (!H-H%). The study of Catalan CC tune examined the politeness by including power (to superior or to subordinate) and social distance (at home or at work)

as sociopragmatic factors. Negative politeness strategy²⁰ was used abreast with the highest boundary tone (H%) among all three legitimate tunes, which is suitable for treating superior (power factor) and at work condition (social distance factor). A slightly lower boundary tone (!H%) was used with positive politeness strategy. This tune is more suitable for subordinate at home, such as children. HL% is an indicator of “bold on-record” politeness strategy, which was best to be used at home.

The factor of seriousness has been studied in Polish (Arvaniti et al., 2016), Persian (Sadat-Tehrani, 2007), and Wolof (Rialland & Robert, 2001). Unlike the politeness factor which seems to have some sort of correlation with H pitch based on the existing studies, seriousness creates two extremes. In Polish urgent calls, a L-L% was used after a H*, which creates a big fall in pitch. Persian has a similar pattern for its anger calls: L+H* h L%, a low boundary tone follows a h tone for the accentual phrase, which also created a fall to express the seriousness. However, in Wolof, an even higher H than in a default call was used at the end of the utterance to indicate a threatening tone with the intention to stop an action.

In a few Romance languages and dialects, the tune of the insistent calls was also examined (Catalan: Borràs-Comes et al., 2015; Portuguese: Frota, 2014; Frota et al., 2015). Insistent calls are not the calls that are used in greeting or the first time, but the second time or more, which indicates a subtle amount of insistence or even impatience. The insistent calls had the same tunes as the greeting/ default calls in most Portuguese dialects except for two European varieties, Lisbon Portuguese (Frota, 2014) and Porto

²⁰ Politeness strategies: “*bald on-record* (e.g., “Shut the window”), *positive* (e.g., “Hey, how about shutting the window, man?”), *negative* (e.g., “Would you mind shutting the window, please?”), or *off-record* (e.g., “I think I’m cold ...”).” (Borràs-Comes et al., 2015)

Portuguese (Frota et al., 2015). In both varieties, the L% boundary tone was further lowered in the insistent calls than the !H% in the greeting calls, and created an even bigger drop together with the H* in the pitch accents. In Catalan (Borràs-Comes et al., 2015), an insistent call either had a H% or a !H%, compared with the HL% in the first calls. Since the first call was a fall originally, the insistent call dissimilated from the original contour and raised to a high tone.

Physical distance between the speaker and listener is another factor that is of interest. Ladd (1978, pp. 518–519) presented different beliefs about physical distance role in the CC tune: Abe (1962) and other scholars alike believed that distance was a vital factor in the CC tune and considerable distance was needed to serve as the reason for the sustainment of the boundary tone. However, other scholars such as Pike (cited in Ladd, 1978, pp. 518–519) argued that, instead of distance, whether the speaker and the listener could see each other was the key to creating such a contour. The Catalan study also examined the effect of physical distance on the CC tunes. They discovered that, for different physical distances, different tunes were preferred. For shorter distances, the HL% and !H% boundary tones were used, with the latter indicating insistence; while for distant calls, a H% boundary tone was used. The result of this study suggests that distance may not be a factor that is decisive in creating a CC tune, but it is certainly a factor in what the specific components of the tune are. A study in Thai (Luksaneeyanawin, 1983) also touched upon the topic of distance, but did not discuss the effect of distance on the CC tune in Thai explicitly; however, they did categorise normal calling (such as calling someone for lunch in a room), and professional calling (such as street vendors and bus conductors), which entails longer distance than a room

setting. In my thesis, the distance factor influenced the production of the CC tune, which will be reported in detail in the method section (Section 6.3.2).

Table 6.2 CC Tunes in World Languages

Language	Author(s)	Types of Chanted Call	Pitch Accent	Phrase & Boundary Tones
English	Lieberman, 1975		HM	
	Pierrehumbert, 1980		H*+ L-	H-L%
	Ladd, 2008		(L)H*!H	
Dutch	Gussenhoven, 1993		HL	
	Gussenhoven et al., 2012		H* !H	%
Bengali	Hayes & Lahiri, 1992		HM	
German	Grice et al., 2005		(L+)H*	!H-%
French	Jun & Fougeron, 2000	Simple & Complex	H*	H-L% ²¹
Greek	Arvaniti & Baltazani, 2005	Stylised Call	L*H	!H-!H%
		Polite Stylised Call	L*H	!H-H%
Polish	Arvaniti et al., 2015	Routine Call	LH*	!H-H%
		Urgent Call	H*	L-L%
Catalan	Borràs-Comes et al. 2015	Negative Politeness; Distant Call; Insistent Call	L*	H%
		Bald On-Record Politeness; Shorter Distance; First Call;	L+H*	HL%
		Positive Politeness; Insistent Call in Short Distances	L+H*	!H%
Portuguese (Lisbon & Porto, European)	Frota, 2014, 2015	Greeting Call	(L+)H*	!H%
		Insistent Call	(L+)H*	L%
Portuguese (Castro Verde, European; Baiano, Brazilian)	Frota, 2015	Greeting & Insistent Call	(L+)H*	!H%
Portuguese (Sulista, Brazilian)	Frota, 2015	Greeting & Insistent Call	L*	HL%
Hungarian	Varga, 2008		H*	!H-0%
Persian	Sadat-Tehrani, 2007	Default & Anger Call	L+H*	h ²² L%
		Surprise Call	L+H*	l L%
Wolof	Rialland & Robert, 2001	Default Call	LHL	H(Sust)%
		Threatening Call with intention to stop an action	LHL	H(Sust)% [Enlarged Pitch Range]
Japanese	Abe, 1998		H*LH	
Thai	Luksaneeyanawin, 1998	Chanting, Professional Call, Recitation	All lexical tones' contours are levelled.	

²¹ Realised as a mid-high and an upstepped L% after H-

²² The h and l in lower case are accentual phrase tones.

6.3 General Methodology

6.3.1 Speakers

Six native speakers of Tianjin Mandarin (3 males and 3 females) were recorded. They were not the same speakers as those in Chapter 4. All speakers were born and raised in the city area of Tianjin, and spoke Tianjin Mandarin daily. The age range of the informants was between 22 and 28. All were capable of speaking standard Mandarin in formal situations, though with different degrees of Tianjin accent. None of them reported any speech disorders. All informants were given financial compensation for their time.

6.3.2 Recording Procedures

Recordings were made in a soundproof booth in Tianjin University, China. The reading was recorded monaurally using a Rode NT-USB Microphone with Audacity directly onto a PC, at a sampling rate of 44.1 kHz.

The informants were given time to briefly familiarise themselves with the materials before the experiment started. They were asked to read the materials, presented as Chinese characters, without context in a Microsoft PowerPoint presentation, as naturally as they could. The informants were asked to produce a statement when seeing a Chinese full stop “。” at the end of the utterances, and call the names out when seeing a Chinese exclamation mark “!” . For the calls, they were told to imagine that the person with the name on the PowerPoint is at the other side of a valley in a mountain, and the informants needed their attention. All the materials were randomised. To avoid any potential list effect, each utterance was placed on a separate slide. The experimenter controlled the pace of each slide, leaving a short pause after each utterance. The words

were placed in the middle of the screen to avoid head movement. Each item was non-consecutively repeated twice.

In many similar studies of the CC tune, a Discourse Completion Task is often used. However, many of those studies are of non-tonal languages. In Tianjin Mandarin, to investigate the performances of all tones, four times as much data as a non-tonal language would be needed. A Discourse Completion Task would make the recording too long, especially for the CC tune which requires high energy output.

The original instructions followed the majority of existing studies to ask the participants to call someone downstairs to come upstairs for lunch. However, that did not create any melodious chanted call, instead the informants solely relied on the change of amplitude. Once the instruction was changed to the mountain scenario, their prosody changed. Borrás-Comes et al. (2015) found physical distance to be an influencing factor in their study. In Tianjin Mandarin, the distance is enlarged. This indicates that for a tonal language, the alternation of the intonational tune may be an additional strategy and would not be necessary unless more intensive methods are needed. Another factor may be the small space of the recording booth which limited the imagination of the informants, since in daily life or in a drama, chanted calls are not particularly rare.

6.3.3 Data Selection

Since the data was collected in a laboratory setting and did not give the context before every utterance, some of the instances were not exaggerated enough to be regarded as a CC tune. To eliminate this factor from such instances, a pre-selection test was conducted by the researcher. All the utterances – a mixture of statements and CCs –

were presented in a random order to the native-speaking researcher. The researcher made forced choices between “statement” and “call”, and categorised them. Only the ones that were correctly identified were kept for further measurements and analyses.

6.4 Production Study 3: Monosyllabic Chanted Calls

The goal of this study is to investigate the CC tune in Tianjin Mandarin when only one syllable is involved. By examining monosyllabic calls, we can identify the interaction between lexical tones and post-lexical intonation without the distraction of lexical stress.

The organisation of the current section (Section 6.4) is as follows: Section 6.4.1 introduces the new methods used in this study which have not been covered in Section 6.3. Section 6.4.2 reports the results for the monosyllabic chanted calls: the results observed from visual inspection are displayed in §6.4.2.1; then the following aspects are presented in the same order as in Chapter 4: duration (§6.4.2.2), register (§6.4.2.3), pitch accent (§6.4.2.4), and edge and boundary tone (§6.4.2.5). Section 6.4.3 analyses and discusses the results. Finally, Section 6.4.4 concludes the whole of Section 6.4 with a summary of all results and analyses for the monosyllabic CC tune.

6.4.1 Methods

6.4.1.1 Materials

Eight monosyllabic words, as presented in Table 6.3, were used as materials to elicit the CC tune. Four of them were kinship terms, such as “mom” and “grandpa”. The other four were common given names. They were recorded in both the statement tune and CC tune. All the words had sonorant consonants since F0 contours can be better detected and extracted.

Table 6.3 Monosyllabic Calls

CATEGORIES	L TONE (TONE 1)	H TONE (TONE 2)	LH TONE (TONE 3)	HL TONE (TONE 4)
Kinship Terms	妈	爷	奶	妹
	/ma/	/je/	/nai/	/mei/
	‘mom’	‘grandpa’	‘grandma’	‘sister’
Given Names	薇	雷	蕊	瑞
	/wei/	/lei/	/zwei/	/zwei/
	‘Wei’	‘Lei’	‘Rui’	‘Rui’

6.4.1.2 Procedures and Informants

Six native Tianjin speakers participated in this study. Detailed information about the informants and procedures has already been provided in Section 6.3.1 and 6.3.2.

6.4.1.3 Statistical Analyses

After the data selection process, 14 out of 96 CC instances – 14.6% of the total number of instances – were discarded. Statistical analyses in this section were conducted using R (R Core Team, 2015), and a mixed-effect model was used. TYPE (Statement, Question) and TONE (L-Tone, H-Tone, LH-Tone, HL-Tone) were taken as fixed-effect factors. SUBJECT (6 different informants) and CATEGORY (kinship terms, given names) were held as random-effect factors. The factors that are important to this study are TYPE, TONE, and their interactions – TYPE*TONE. The focus of this section therefore is mainly on reporting these results. The error bars shown in all figures under Section 6.4.2 are ± 1 standard error.

6.4.2 Results

6.4.2.1 Visual Inspection

The first step when analysing Tianjin Mandarin intonation data (after the data have been labelled) is to make a visual inspection, as stated in Section 4.3.3. Unlike the IntQ data in Chapter 4, the CC tune appeared to be visibly different from the statement tune, as shown in Figure 4.6 to Figure 4.9. The F0 maxima seemed to be much higher, and the register was clearly lifted. A L% boundary tone could be observed at the end of the CC contours for all tones – the L% was especially easy to spot at the end of rising tones.

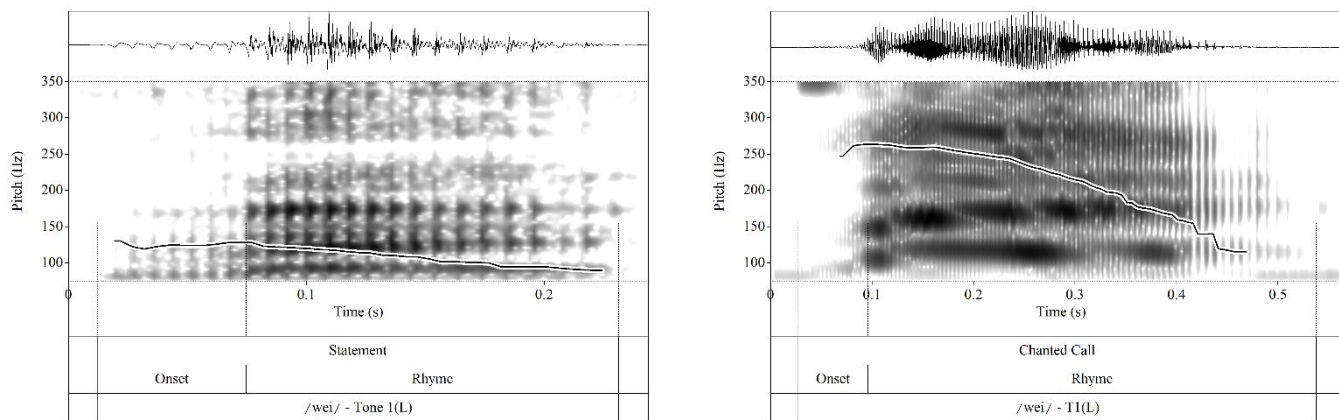


Figure 6.1 Statement (left) and CC (right): Monosyllabic – Tone 1 (L Tone)

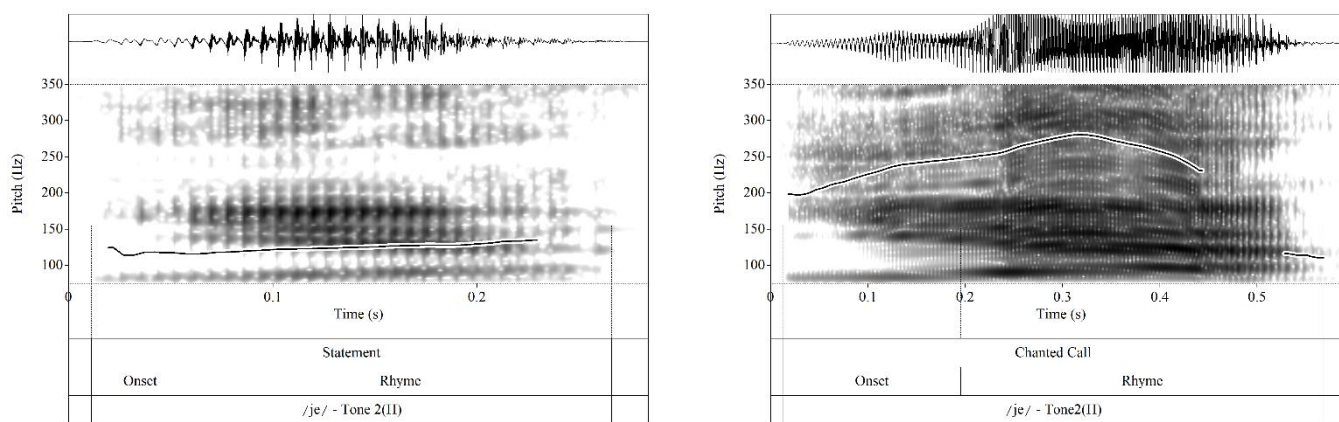


Figure 6.2 Statement (left) and CC (right): Monosyllabic – Tone 2 (H Tone)

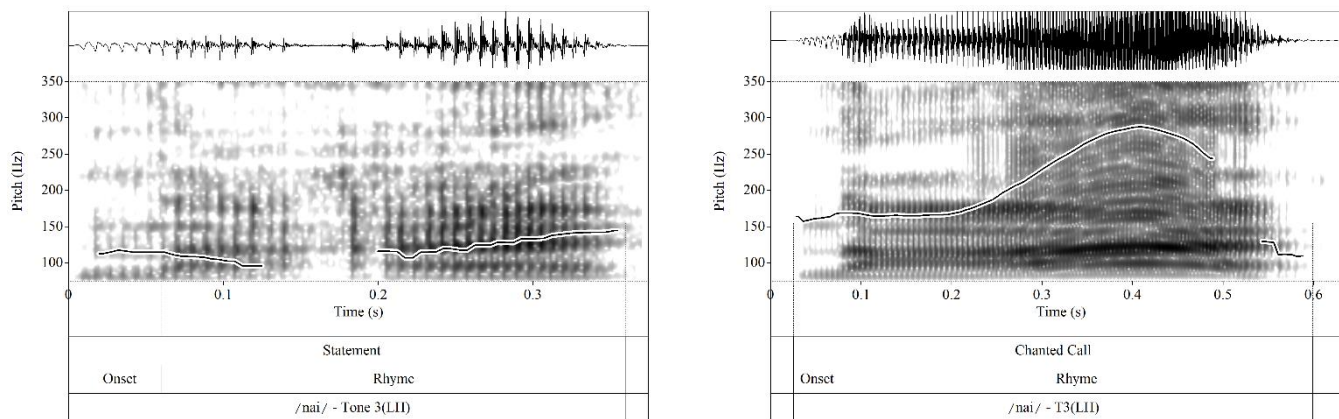


Figure 6.3 Statement (left) and CC (right): Monosyllabic – Tone 3 (LH Tone)

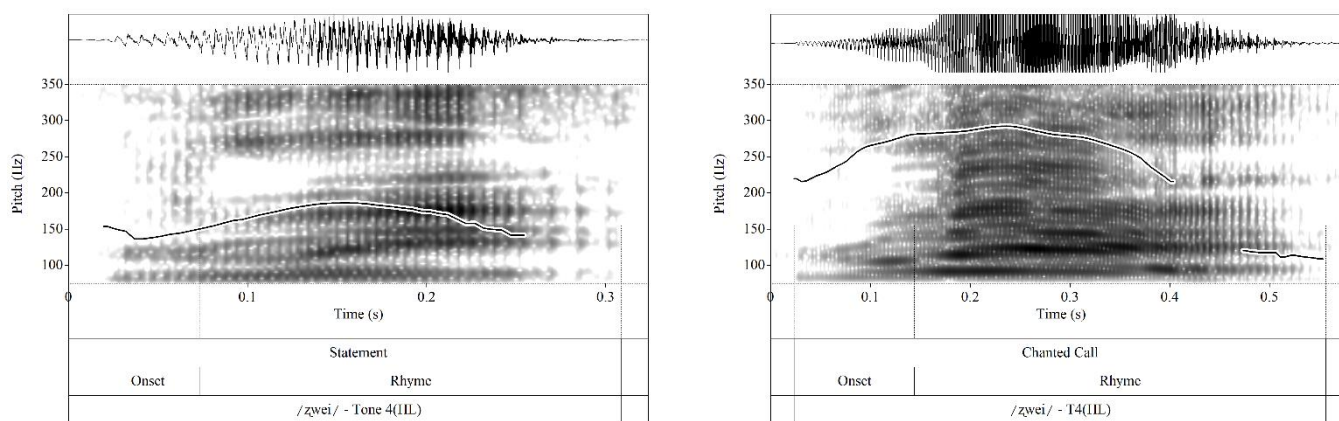


Figure 6.4 Statement (left) and CC (right): Monosyllabic – Tone 4 (HL Tone)

6.4.2.2 Duration

The average duration of the monosyllabic words on the whole was longer in the CC tune than in the statement tune. Figure 6.5 shows the duration of the **monosyllabic onsets**. The duration of the onsets in the CC tune was slightly longer than in the statement tune. The duration of the onsets was influenced by TYPE ($\chi^2(1) = 3.98$, $p = 0.046$ *).

The **monosyllabic rhymes** (Figure 6.6) were much longer in the CC tune (321.1ms on average, $\overline{SD} = 115.2$ ms) than in the statement tune (192.2ms on average, $\overline{SD} = 55.2$ ms). The average length of the CC rhymes was 1.67 times the average length of the statement rhymes. The length of monosyllabic rhymes was also influenced by TYPE ($\chi^2(1) = 87.38$, $p < 0.001$ ***), as well as TONE ($\chi^2(3) = 32.87$, $p < 0.001$ ***).

The difference between the duration of statements and CCs was much larger than the difference between statements and IntQs reported in Chapter 4. The durational difference between the IntQ tune and statement tune was 32.4ms on average. The difference between the CC tune and the statement tune was nearly three times more than the difference between the IntQ tune and the statement tune.

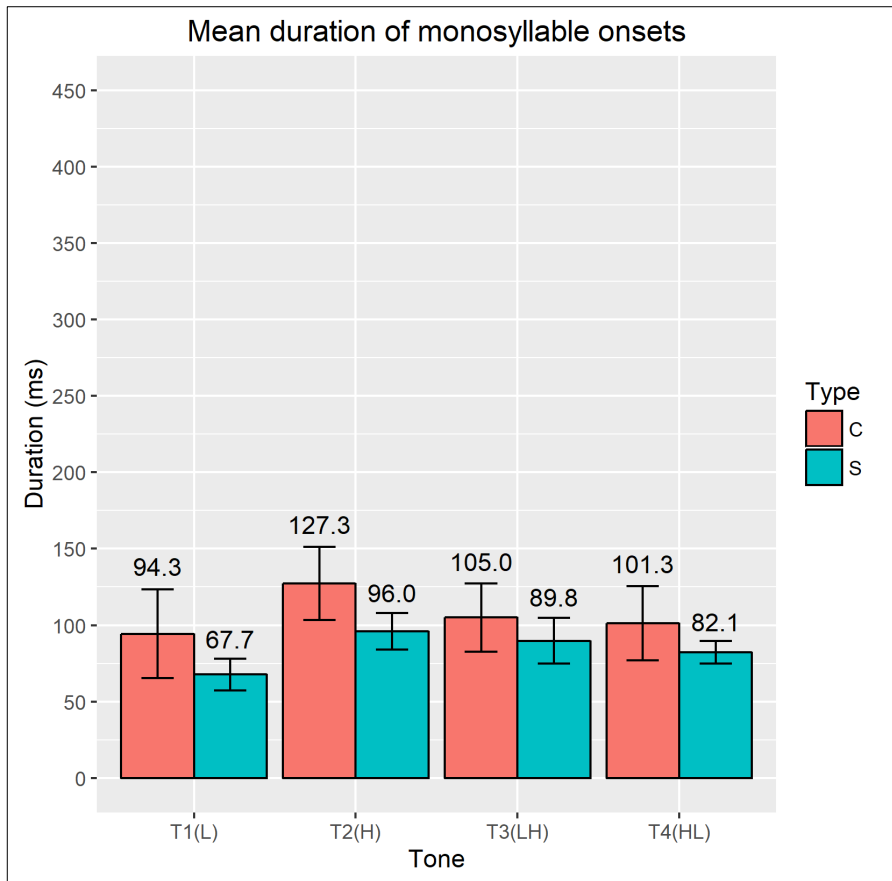


Figure 6.5 Duration: Monosyllabic onsets

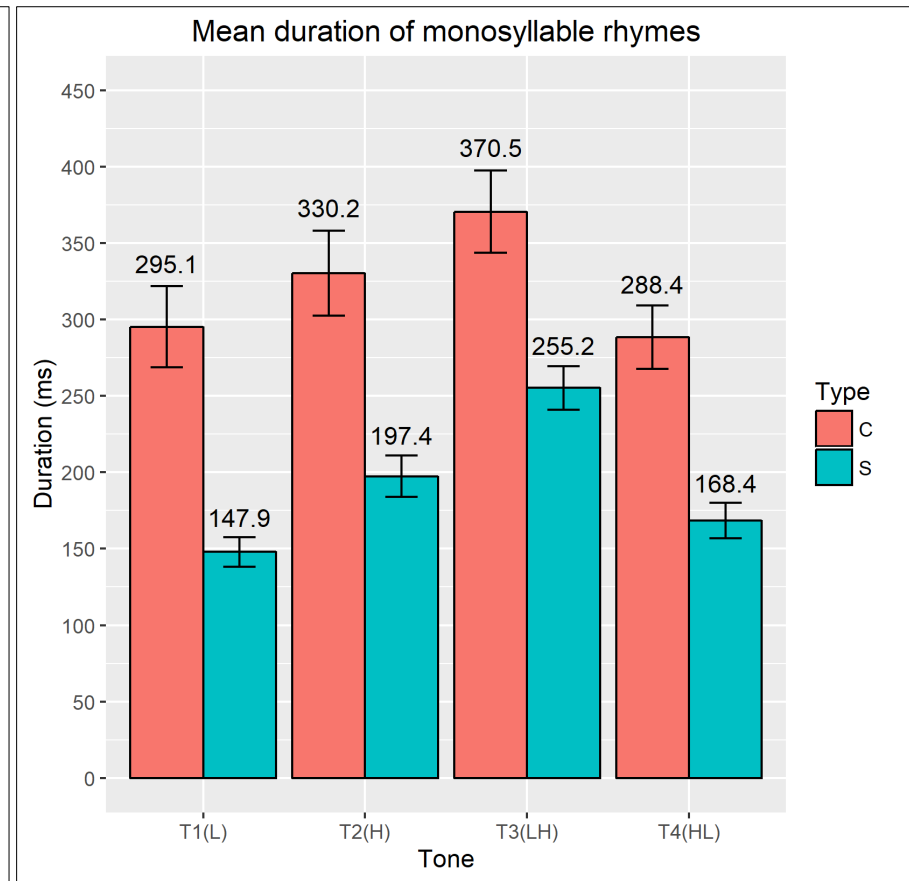


Figure 6.6 Duration: Monosyllabic rhymes

6.4.2.3 Register

Mean Pitch

The mean pitch of both onsets and rhymes of the CC tune was higher than that of the statement tune. The mean pitch of **monosyllabic onsets** (Figure 6.7) was affected by TYPE ($\chi^2(1) = 157.47$, $p < 0.001^{***}$), TONE ($\chi^2(3) = 71.97$, $p < 0.001^{***}$). Besides, onset had a marginally significant interaction between TYPE and TONE – TYPE*TONE ($\chi^2(3) = 8.9$, $p = 0.03^*$). The mean pitch of the **monosyllabic rhymes** (Figure 6.8) was also affected by TYPE and TONE separately, TYPE ($\chi^2(1) = 209.54$, $p < 0.001^{***}$), TONE ($\chi^2(3) = 59.69$, $p < 0.001^{***}$).

Compared with the mean pitch differences between IntQs and statements in Mono_(ISO) (Section 4.4.2.2), the differences between the mean pitch of CC tune and that of the statement tune were much larger. The average difference between the IntQs and the statements was 0.54ERB²³, while it was 1.5ERB between CCs and statements. The register of CC was expanded even more than that of the IntQ tune.

²³ ERB was calculated using the same formula as used in Chapter 4: $ERB = 16.7 \log_{10}(1 + f/165.4)$.

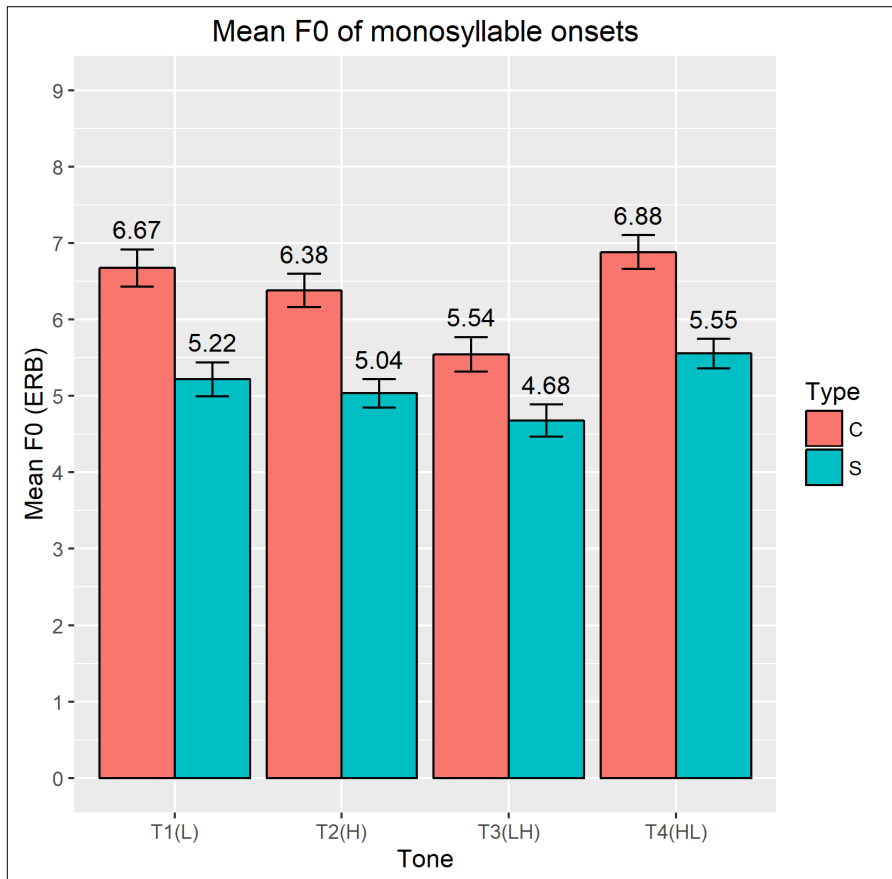


Figure 6.7 Mean Pitch: Monosyllabic onsets

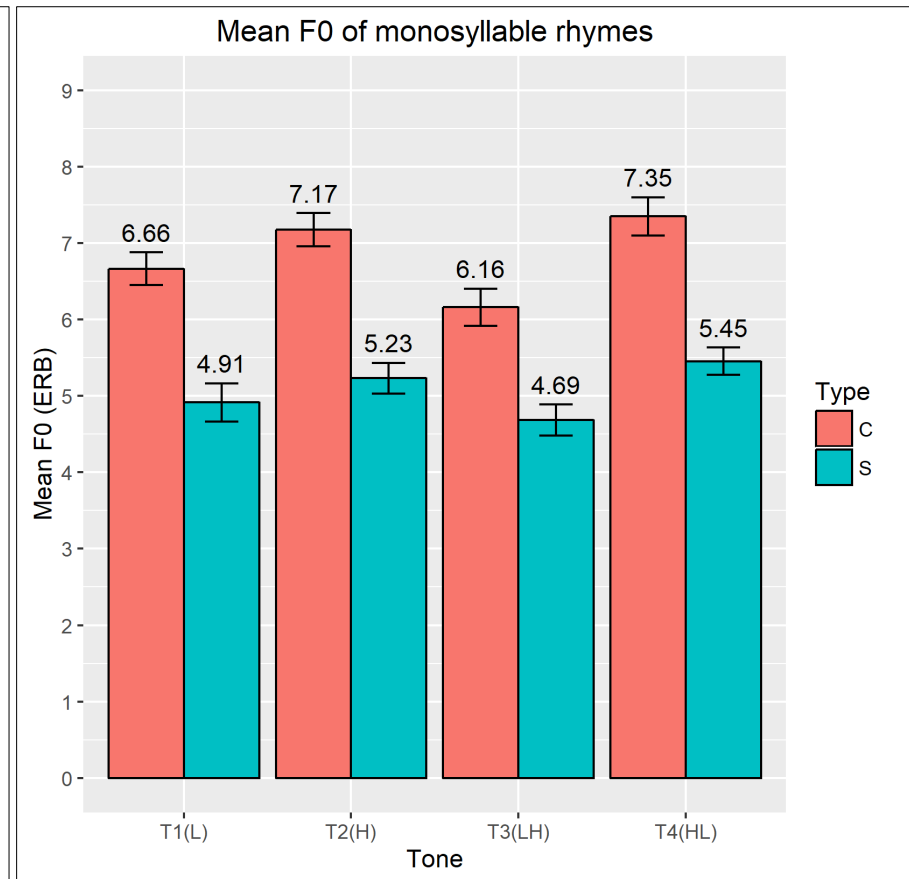


Figure 6.8 Mean Pitch: Monosyllabic rhymes

Tonal Scaling: Values of F0 Maxima and Minima

Mean pitch data alone could only suggest that the register was expanded. F0 extrema data provided further useful insight into whether the register was lifted. In Table 6.4, both maximum F0 and minimum F0 are higher in a CC tune than in a statement tune. If the register were only expanded or the mean pitch raised solely due to a H* pitch accent, the F0 minimum would not be raised together with the maximum. Therefore, the mean pitch data, together with the F0 extrema data, indicate that the register on the whole was lifted in the CC tune.

Table 6.4 F0 Scaling: Monosyllabic CC (Unit: ERB)

Tone	Type	Max F0	Min F0
T1 (L)	S	5.40	4.41
	CC	7.33	5.06
T2 (H)	S	5.90	4.90
	CC	8.32	5.83
T3 (LH)	S	5.85	3.58
	CC	6.98	5.14
T4 (HL)	S	6.20	4.83
	CC	7.84	6.00

6.4.2.4 Pitch Accent

The difference between the F0 maxima locations of IntQ tune and of statement tune in Mono_(ISO) did not differ by more than 7%. However, as shown in Figure 6.9, the F0 maxima locations all changed substantially from a statement tune to a CC tune. In this figure, the orange bar represents the CC tune, and the blue bar, the statement tune. Apart from Tone 1, all the other three tones had earlier alignment of F0 maxima. This is likely to be a result of the extra L% boundary tone. The L% boundary tone added length to

the segments and therefore the alignment of pitch maxima was brought back to near the beginning of the monosyllabic rhymes with a CC tune.

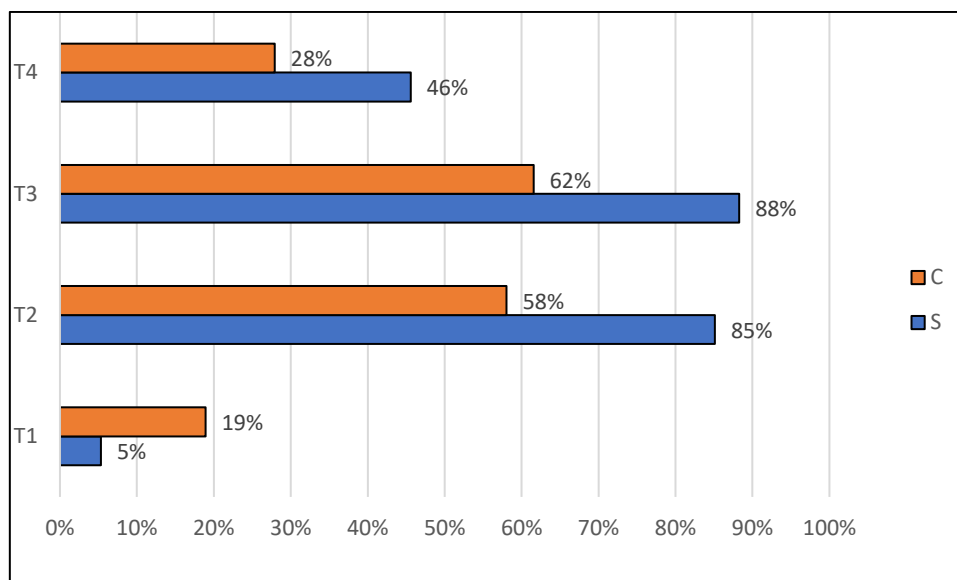


Figure 6.9 F0 Maxima Alignment in CC tune and Statement Tune

Another possibility is that there is a H* pitch accent. This explanation can account for the pitch alignment change in Tone 1. The statement in Figure 4.6 (p. 77) clearly depicts an extra H* aligned with the boundary between the onset and the rhyme of Tone 1. A high plateau can also be found in Figure 4.9 (p. 78) for Tone 4. In Figure 6.2 (Tone 2) and Figure 6.3 (Tone 3), a H* pitch accent can also be observed to align with the medial position of the rhyme in CC tune. The results presented in Figure 6.9 attest the observations. Figure 6.10 provides a more detailed view of where each maximum F0 point is. The statements are again shown in blue and the CCs are in orange. In this figure, as for Tone 2 and Tone 3, which are both rising tones, the statements have the maximum F0 points mostly at the end of the rhymes. However, in CCs the maximum F0 points moved significantly towards the middle of the rhymes, which is resulted from an extra L% boundary tone attached to the CC utterances. The maximum F0 points in

T4 gather at the early portion of the rhymes, which is approximately the same as where the Tone 1's maximum points occur. Since both T4 and T1 are falling tones, their F0 maxima alignment is with the H of the lexical tones. Yet, even though the alignment seems to be with the high portion of the lexical tones, the possibility still exists that the H* pitch accent is associated with the middle portion of the tone-bearing rhyme. Therefore, whether the H* pitch accent is associated with the H tone in the lexical tones or the middle portion of the tone-bearing unit needs to be tested in longer words and sentences. In the disyllabic chanted call tune study in Section 6.5, the association of pitch accent will be revisited.

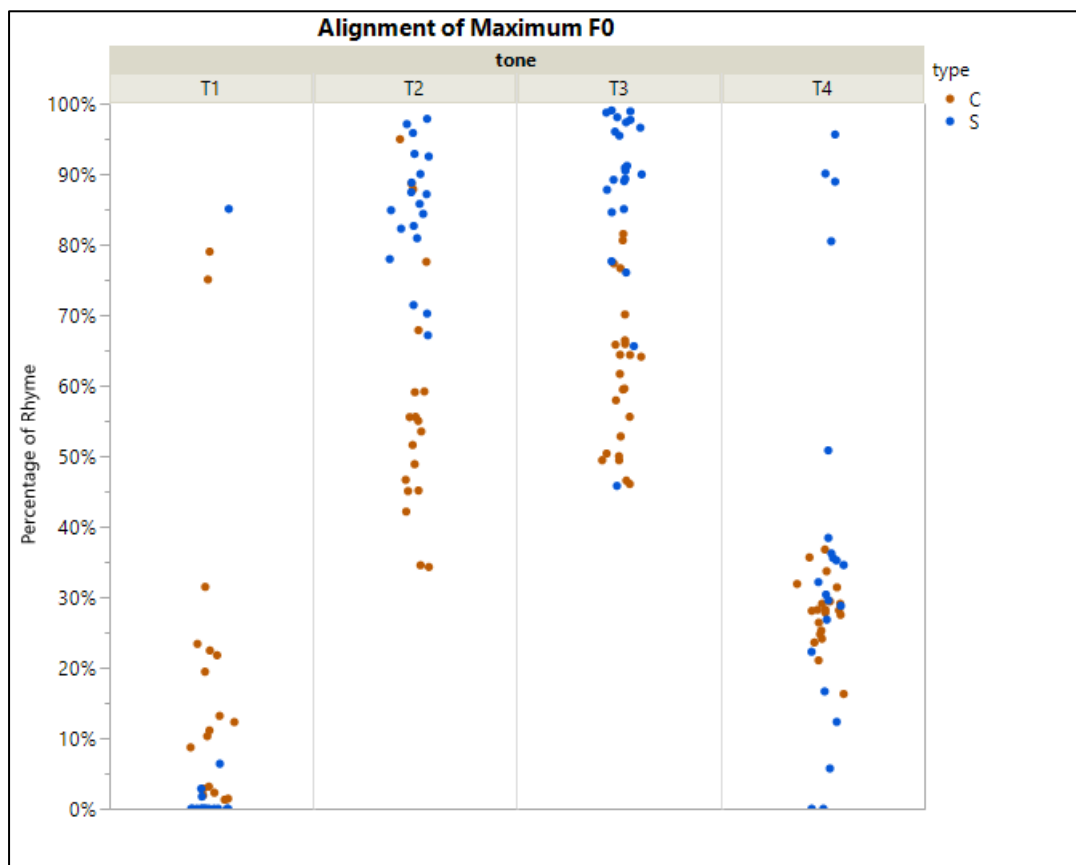


Figure 6.10 Alignment of Maximum F0 in CC Tune and Statement Tune

6.4.2.5 Edge/Boundary Tone

As observed previously in Section 6.4.2.1, all four tones ended in an actual L% boundary tone. To verify this observation with quantitative data, I examine the alignment of minimum F0 in Figure 6.11. In this figure, the orange points represent CC tune while the blue ones are of statement tune.

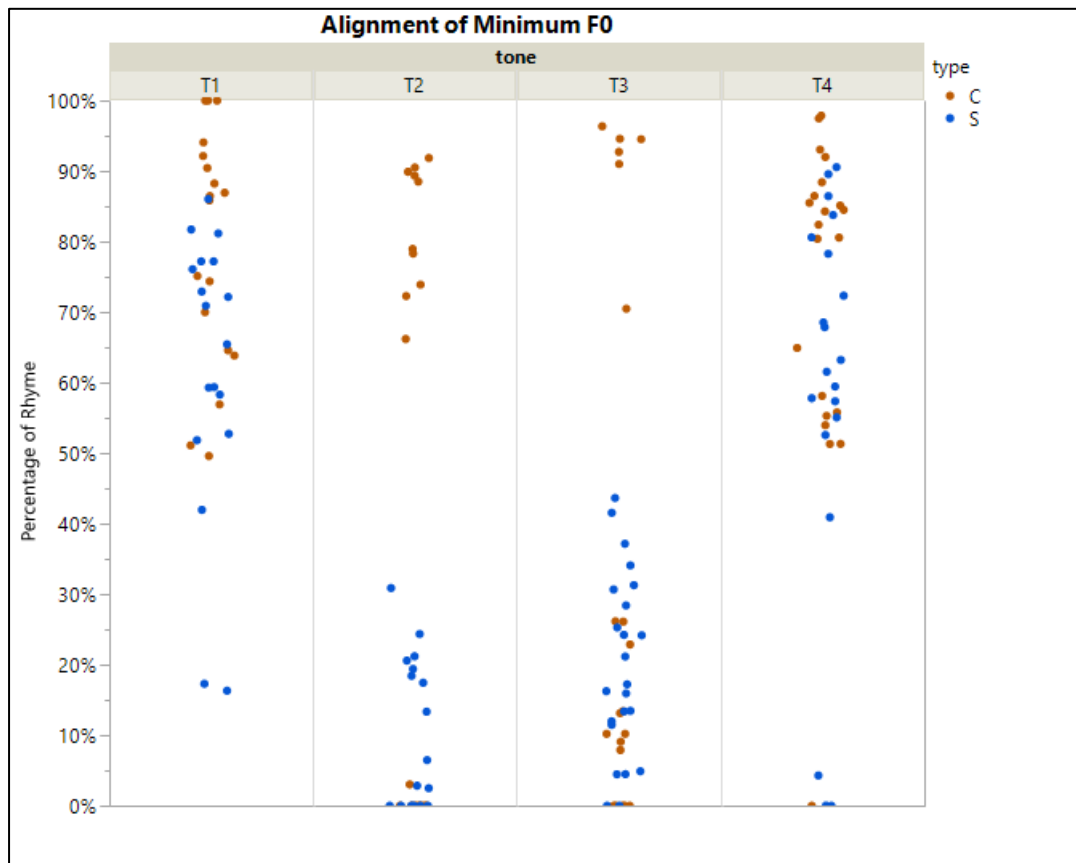


Figure 6.11 Alignment of Minimum F0 in CC Tune and Statement Tune

Figure 6.11 displays the alignment of F0 minima of the rhyme in the CC tune and in the statement tune respectively. Tone 2 and Tone 3 in the figure, both being rising tones, have the minimum F0 points in statement tune mainly at the beginning of the rhymes. On the other hand, in the CC tune, the minimum F0 points moved to the end of the rhymes. In the two falling tones, Tone 1 and Tone 4, the minimum F0 points come even

later than the minimum F0 in the original statement tune. The fact that the CC tune has more minimum F0 points towards the end of the rhyme than the statements further supports that there is a L% for the CC tune in Tianjin Mandarin.

6.4.3 Discussion

6.4.3.1 Lengthening

The lengthening in the CC tune is substantially longer than the difference between IntQ tune and statement tune. Although the L% boundary tone partially contributed to the addition of duration, the lengthening in the current study of CC tune is likely to be a different type of lengthening from the boundary-induced lengthening in the IntQ study.

The first reason for this claim is that the durational difference in the current study of CC tune is much larger than the lengthening in the IntQ study. As discussed in Section 4.5.3.1, a number of studies investigated the perceptual value of the durational cue in IntQ tune, but found it to be insignificant for the perception and identification of IntQ tune. However, in the CC tune, even if judged by the duration alone, the longer tune was easily identified not to be the default statement tune.

Many studies on the CC tune in non-tonal languages address the duration issue. Hayes and Lahiri (1992) analysed the relationship between duration and stress in both English and Bengali and proposed a full account of the interaction. In many other languages, the CC tune also has a sustained H pitch at the end – the “sustainment” refers to final lengthening.

If the final boundary-induced lengthening in IntQ tune is a side product of the floating H% boundary tone, then the lengthening in the current study of CC tune is a major component that plays a key role in this tune.

6.4.3.2 Pitch Accent

The current study focuses on the CC tune of monosyllabic calls. In Chapter 4, when the words were also monosyllabic but produced in isolation as IntQs, it was not possible to see any pitch accent. In the current study, a H* pitch accent can be visibly detected from the F0 contours and is also shown in the pitch alignment data. This suggests that it is possible for a tonal language to have a pitch accent that is not created by the sentence focus. The smallest unit for such a pitch accent can be as small as a monosyllabic word. Further investigations are needed to ascertain the association of this pitch accent. This will be reported and discussed in Section 6.5 in the current chapter.

6.4.3.3 Boundary Tone

Chapter 4 investigated the IntQ tune and found a floating H% boundary tone. The floating boundary tone serves as a phonological device that facilitates rising and deters falling according to each individual lexical tone. The “benefit” of the floating boundary tone is that it both alters the post-lexical intonation and also keeps the lexical tone contour unchanged. This gives us the impression that, in such a tonal language as Tianjin Mandarin, lexical tones cannot be overridden by intonation. However, results from the current study of chanted call tune have challenged this idea. The L% boundary tone is an interpolation of an extra L tone at the IP boundary and does not modify the neighbouring tones, which is against the principles of floating boundary tone (see p.

134). The L tone thus is not a “floating” boundary tone but a real boundary tone. The fact that a real boundary tone exists in the CC tune, in turn, supports the “floating boundary tone” analysis for the IntQ tune in Chapter 4, in that it is essential to distinguish these two different types of boundary tones – not only cross-linguistically but also within a tonal language itself (i.e. Tianjin Mandarin in this case).

6.4.4 Summary for Monosyllabic Chanted Calls

In Section 6.4, the chanted call tune was investigated with monosyllabic calls. The results showed that CC tune in Tianjin Mandarin has a sustained H* pitch accent on top of a lifted register, as well as a L% boundary tone: $[[H^*]_{\text{sustained}}]_{\text{higher register}} + L\%$.

6.5 Production Study 4: Disyllabic Chanted Calls

Following the examination of the CC tune of monosyllabic names, this section will investigate the CC tune of disyllabic names, which will provide further insights into (i) whether the CC tune which is derived from the monosyllabic study is able to hold in utterances that consist of more than one syllable; (ii) how an intonation tune manifests on a toneless syllable.

The organisation of the current section (Section 6.5) is as follows: Section 6.5.1 presents the specific methods for this production study of **disyllabic chanted calls**; Section 6.5.2 reports the results for the disyllabic chanted calls: visual inspection (§6.5.2.1), duration (§6.5.2.2), mean pitch (§6.5.2.3), tonal scaling (§6.5.2.4), and tonal alignment (§6.5.2.5). Section 6.5.3 analyses and discusses the results, and finally, Section 6.5.4 concludes with a summary of all results and analyses for the chanted call tune for disyllabic utterances.

6.5.1 Methods

6.5.1.1 Materials

40 tokens of disyllabic words were used as materials to elicit the CC tune in disyllabic utterances. These tokens were divided into two sets with different lexical tone formations and word prominences: Set 1 was a full array of all possible non-neutral tonal combinations, in which the word prominences were on the second syllable. Each combination had two different tokens, as shown in Table 6.5. Set 2 used the same words as in the monosyllabic CC tune study (see Table 6.3), but repeated the monosyllable to create kinship terms and disyllabic names, which is common in creating nicknames in Mandarin. In Set 2, the second syllables carried neutral tones as shown in Table 6.6, which also led to the word prominence being on the first syllable. The numbers beside the phonetic transcriptions in Table 6.5 and Table 6.6 are their lexical tones. For example, /je² je⁰/ is the disyllabic sequence of /je/ with Tone 2 (H Tone) and /je/ with Tone 0, i.e. the neutral tone. Set 1 and Set 2 were randomly mixed and presented to the informants in the same study. The division here is merely to facilitate the clarity of the result presentation.

Set 1 in Table 6.5 consists of 32 common names. The first syllables of the disyllabic names are all common family names, which are followed by common given names in the second syllables. Tone A represents the tone of the first syllable and Tone B stands for the tone of the second syllable. Set 2 in Table 6.6 have the same composition as the monosyllabic names. Four of the disyllabic words in Set 2 are kinship terms, while the other four are given names, usually used as nicknames among family members or close

friends. All the materials in both Set 1 and Set 2 used sonorant onsets for better detection of F0 contours and extraction of F0 values.

Table 6.5 Disyllabic Calls (Set 1)

		Tone A			
		L Tone (Tone 1)	H Tone (Tone 2)	LH Tone (Tone 3)	HL Tone (Tone 4)
Tone B		殷薇	王薇	李薇	魏薇
	L Tone (Tone 1)	/jin1 wei1/ 安伊 /an1 ji1/	/waŋ2 wei1/ 吴伊 /wu2 ji1/	/li3 wei1/ 米伊 /mi3 ji1/	/wei4 wei1/ 万伊 /wan4 ji1/
	H Tone (Tone 2)	/jin1 lei2/ 安茹 /an1 zu2/	/waŋ2 lei2/ 吴茹 /wu2 zu2/	/li3 lei2/ 米茹 /mi3 zu2/	/wei4 lei2/ 万茹 /wan4 zu2/
	LH Tone (Tone 3)	/jin1 zwei3/ 安武 /an1 wu3/	/waŋ2 zwei3/ 吴武 /wu2 wu3/	/li3 zwei3/ 米武 /mi3 wu3/	/wei4 zwei3/ 万武 /wan4 wu3/
	HL Tone (Tone 4)	/jin1 zwei4/ 安璐 /an1 lu4/	/waŋ2 zwei4/ 吴璐 /wu2 lu4/	/li3 zwei4/ 米璐 /mi3 lu4/	/wei4 zwei4/ 万璐 /wan4 lu4/

Table 6.6 Disyllabic Calls (Set 2)

		Tone A			
Categories		L Tone (Tone 1)	H Tone (Tone 2)	LH Tone (Tone 3)	HL Tone (Tone 4)
Tone B = Tone 0	Kinship Terms	妈妈 /ma1 ma0/ ‘mom’	爷爷 /je2 je0/ ‘grandpa’	奶奶 /nai3 nai0/ ‘grandma’	妹妹 /mei4 mei0/ ‘sister’
	Given Names	薇薇 /wei1 wei0/ ‘Weiwei’	雷雷 /lei2 lei0/ ‘Leilei’	蕊蕊 /zwei3 zwei0/ ‘Ruirui’	瑞瑞 /zwei4 zwei0/ ‘Ruirui’

6.5.1.2 Procedures and Informants

6 native Tianjin speakers participated in this study. They also participated in the previous CC tune study of monosyllabic calls. Detailed information about the informants and procedures has already been provided in Section 6.3.1 and 6.3.2.

6.5.1.3 Statistical Analyses

After the data selection process, 94 out of 480 CC instances – 19.6% of the total number of instances – were excluded. Statistical analyses in this section were conducted using R (RStudio-Team, 2017), and a mixed-effect model was used. TYPE (Statement, 8 Question), TONE (L-Tone, H-Tone, LH-Tone, HL-Tone), and SYLLABLE (First syllable, Second syllable) were taken as fixed-effect factors. SUBJECT (six different informants) and ITEM (two different tokens) were held as random-effect factors. The factors that are important to this study are: (i) TYPE – to examine whether the differences between different types are statistically significant; (ii) TONE – to examine whether the differences between different lexical tones are statistically significant; (iii) TYPE * TONE interaction – to examine whether the differences in the dependent variables (duration, mean pitch, maximum F0, minimum F0) are jointly affected by both the type and the tones in a categorical manner; (iv) SYLLABLE – to examine whether the differences between the two syllables of the disyllabic utterances are statistically significant. The error bars shown in the bar graphs under Section 6.5.2 are ± 1 standard error.

6.5.2 Results

Since T0, i.e. the neutral tone or toneless tone, differs from the other four tones in both temporal and pitch values, the results are presented separately in this section. The

results for the data that do not contain T0 are Set-1 data as introduced in Table 6.5; the ones that contain T0 are presented as Set-2 data (Table 6.6). For clarity, the results for the two sets of data are reported separately in the following sections.

The tone pairs are represented as TxTx in the result section, e.g. T1T2. Four tone pairs undergoing tonal dissimilation. The tone names in Set-1 materials, as well as their underlying and surface height of the tone combinations, are listed in Table 6.7. The shaded combinations undergo tonal dissimilation. Set 2 contains the underlying tones with a neutral tone, i.e. T0.

Table 6.7 Tone Name vs Tone Height

Tone Name	Underlying Tone Height	Surface Tone Height (if different)
T1T1	L+L	H+L
T1T2	L+H	
T1T3	L+LH	
T1T4	L+HL	
T2T1	H+L	
T2T2	H+H	
T2T3	H+LH	
T2T4	H+HL	
T3T1	LH+L	
T3T2	LH+H	
T3T3	LH+LH	H+LH
T3T4	LH+HL	
T4T1	HL+L	H+L
T4T2	HL+H	
T4T3	HL+LH	
T4T4	HL+HL	L+HL

6.5.2.1 Visual Inspection

Figure 6.12 to Figure 6.27 display instances of the disyllabic chanted call contours from the Set-1 data. Set-2 data are shown in Figure 6.28 to Figure 6.31. The instances with the least creakiness and comparatively smoother contours were selected as representatives in the figures. All were produced by male speakers. These figures are also useful during the process of understanding the results from phonetic details reported in the next few sections.

Set 1: without T0

An initial visual inspection of the contours for Set 1 data (Figure 6.12 to Figure 6.27) confirms the conclusions drawn from the monosyllabic calls. The **register** is evidently higher for the utterances with the CC tune than for those with the statement tune. Further descriptive data and statistical analyses will be presented by way of the mean pitch, values of maximum F0 and minimum F0 in Section 6.5.2.3 and Section 6.5.2.4.

The contours of the CCs are much more exaggerated than the statements, owing to the **pitch accents**. A pitch accent of H* can be seen for the CC tune in the figures. Moreover, the H* pitch accent is associated with the H of each syllable for all of the lexical tones in Set 1 data. For example, within the T1Tx utterances which have T1 as the initial tone, T1T1 undergoes tonal dissimilation and becomes a rising H tone and a falling L tone, while the other three T1Tx combinations keep the falling L tone for the first syllables. If the argument of the association between H* and the H of T1 in the CC tune was valid, the H* should align with the ending part of the T1 syllable in T1T1 in the CC tune, and the beginning portion of the T1 syllable in other three T1Tx combinations in the CC tune. Figure 6.12 compares the statement tune and the CC tune of the T1T1 combination.

In the CC tune figure on the right, the contour of the first syllable, which is almost flat in the statement figure on the left, turns into a distinct rising tone. In the other three figures (Figure 6.13, Figure 6.14, Figure 6.15), the local H* was found towards the beginning of the first syllables, which was particularly clear in T1T3 since the second syllable, T3, starts with a L tone too, so that it does not create a high transition between the lexical tones like T1T2 and T1T4.

The T4Tx utterances also lent support for the association with H*. The peak of T4 is towards the middle of the T4 syllables in its citation form. T4T1 becomes H+L and T4T4 becomes L+HL due to tonal dissimilation rules. Therefore, if the H* is indeed associated with the high parts of the lexical tones, the first syllable of the CC tune of T4T1 would have a H* towards the end, and that of T4T4 would have a H* towards the beginning, while T4T2 and T4T3 would have a H* in the middle of the first syllables. Figure 6.24 displays the statement and CC tunes of T4T1. In the statement tune, the T4 syllable rises subtly, while the rise in the CC tune is much more exaggerated. The H* can be found towards the end of the T4 syllable. In contrast, T4T2 (Figure 6.25) and T4T3 (Figure 6.26) both have their local maxima at the middle portion of the first syllables in the CC tune. In T4T4 (Figure 6.27), the highest point is at the beginning of the syllable rhyme and the falling slope is not as sharp as the T4s in T4T2 and T4T3, which is exactly what distinguishes a T1 (L Tone) from a T4 (HL Tone).

T2Tx and T3Tx (shown in Figure 6.16 to Figure 6.23) had lexical tones that end with H tones in the first syllables, including the T3T3 sequence after tonal dissimilation. It is notable that when these two rising tones (T2 and T3) are succeeded by lexical tones that start with a L tone, the local maxima for the rising tones shifts towards the middle

of the syllable. This is a pure phonetic transition effect and is not a part of the phonological tune of the CCs.

A **L% boundary tone** can be observed at the end of all utterances with the CC tune. It is particularly evident in the utterances that end with rising or high tones (see Figure 6.13, Figure 6.14, Figure 6.17, Figure 6.18, Figure 6.21, Figure 6.22, Figure 6.25, Figure 6.26). For instance, both Figure 6.17 (T2T2) and Figure 6.22 (T3T3) only contain rising tones in the statement tune, however, at the end of the CC tune, L% boundary tones clearly exist. The tonal sequences that end low in statement tune – the ones ending in T1 and T4 – either end in a lower tone in the CC tune than in statement tune, or the falls are sharper in the CC tune than in the statement tune. Despite being unable to serve as decisive evidence, these data support the L% tone analysis.

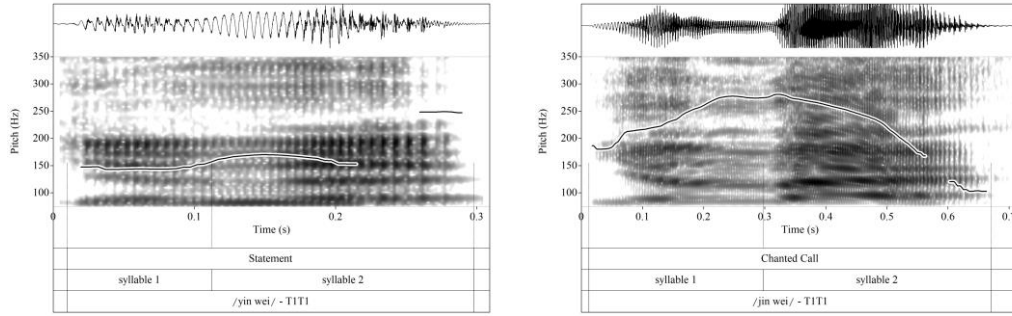


Figure 6.12 Statement (left) and CC (right): Disyllabic (Set1) – T1T1

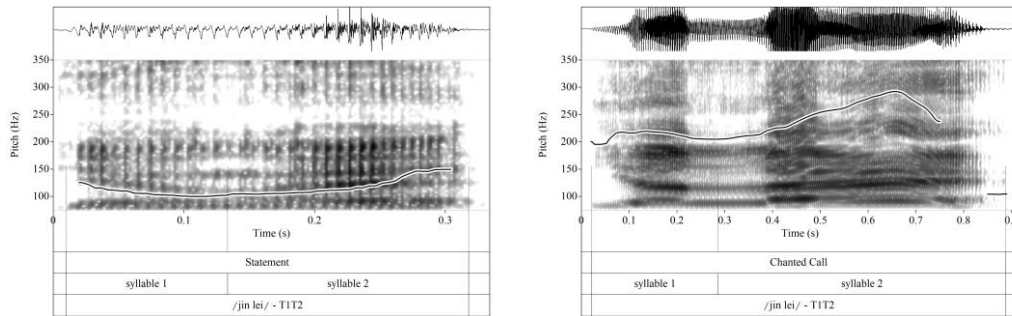


Figure 6.13 Statement (left) and CC (right): Disyllabic (Set1) – T1T2

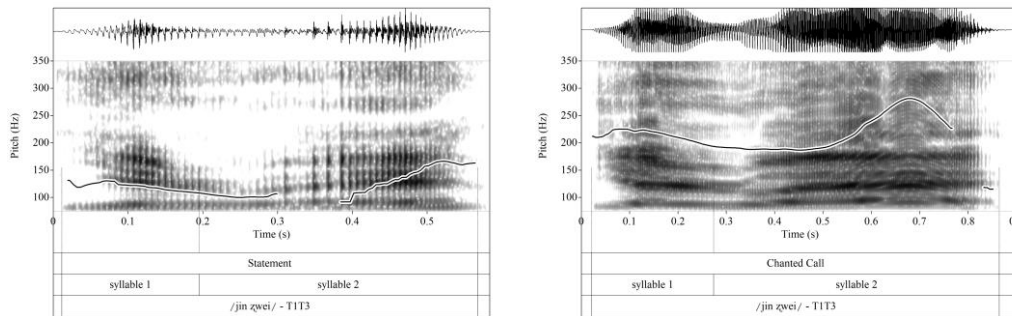


Figure 6.14 Statement (left) and CC (right): Disyllabic (Set1) – T1T3

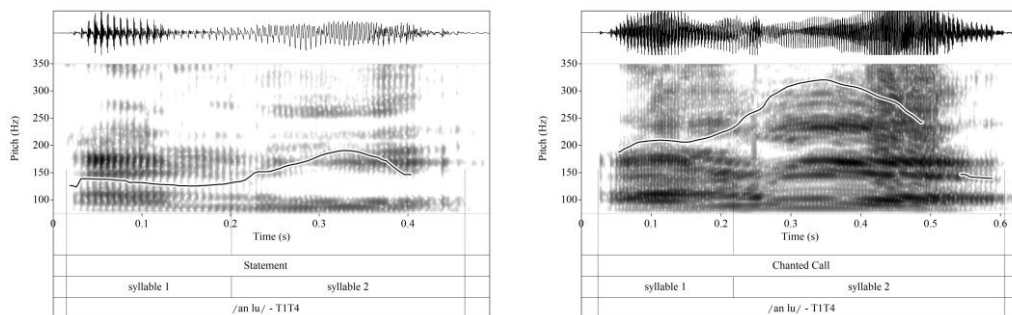


Figure 6.15 Statement (left) and CC (right): Disyllabic (Set1) – T1T4

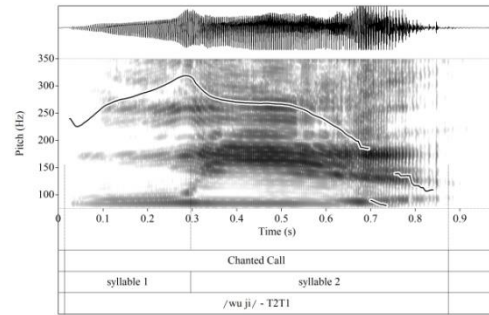
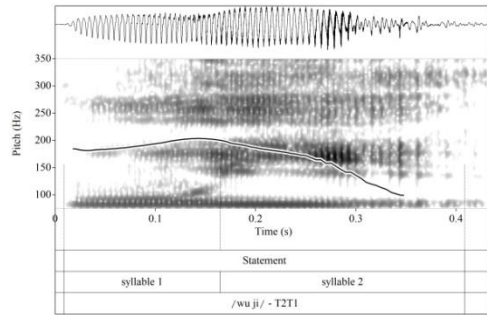


Figure 6.16 Statement (left) and CC (right): Disyllabic (Set1) – T2T1

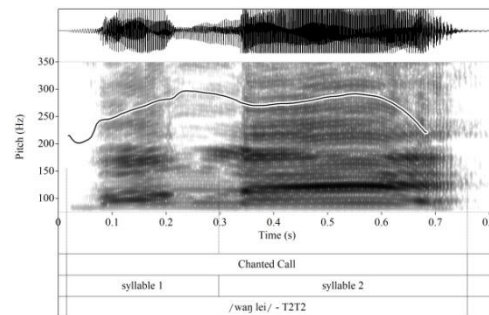
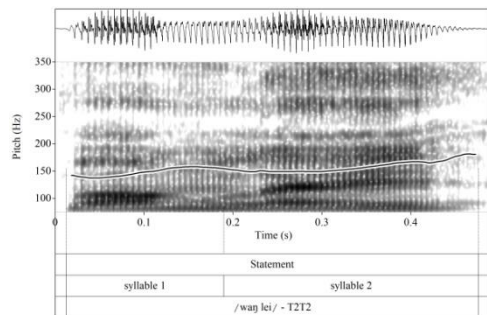


Figure 6.17 Statement (left) and CC (right): Disyllabic (Set1) – T2T2

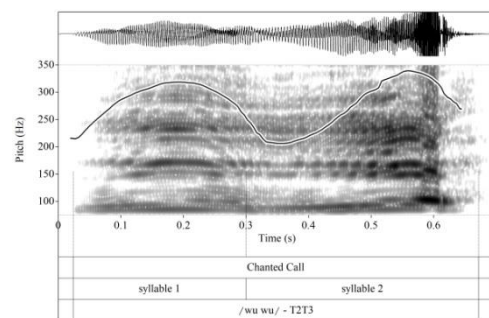
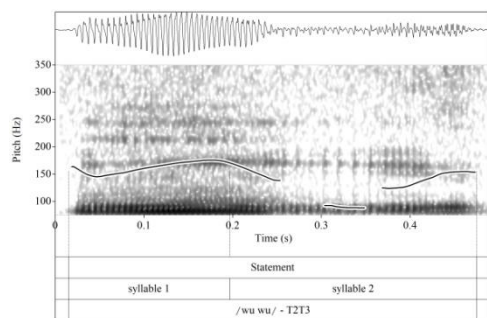


Figure 6.18 Statement (left) and CC (right): Disyllabic (Set1) – T2T3

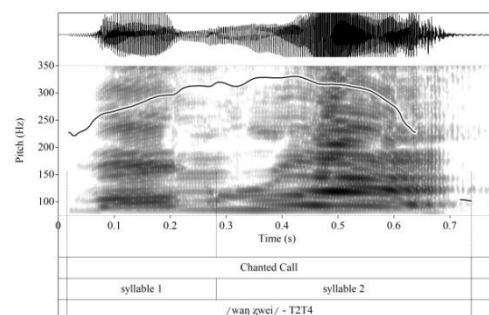
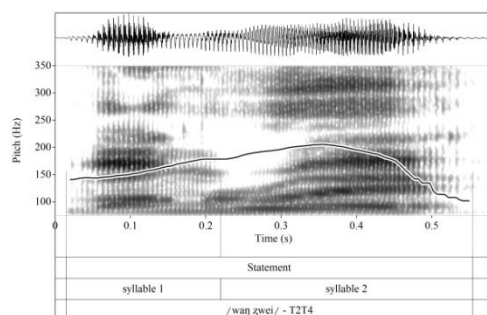


Figure 6.19 Statement (left) and CC (right): Disyllabic (Set1) – T2T4

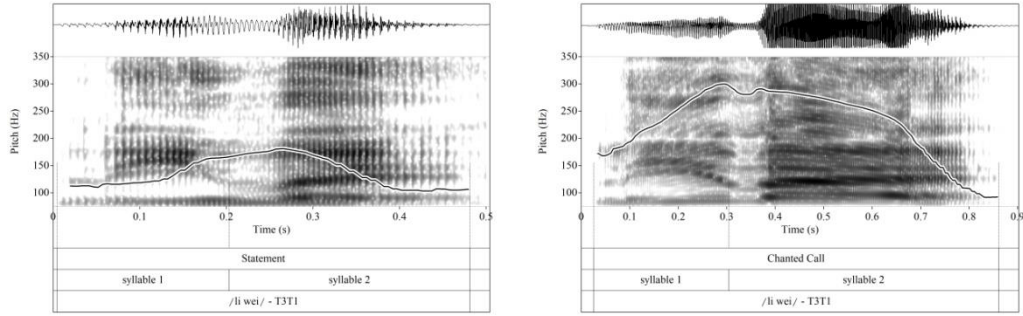


Figure 6.20 Statement (left) and CC (right): Disyllabic (Set1) – T3T1

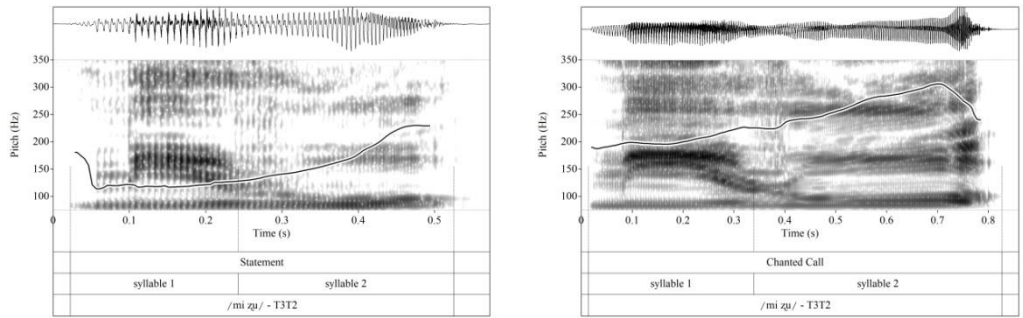


Figure 6.21 Statement (left) and CC (right): Disyllabic (Set1) – T3T2

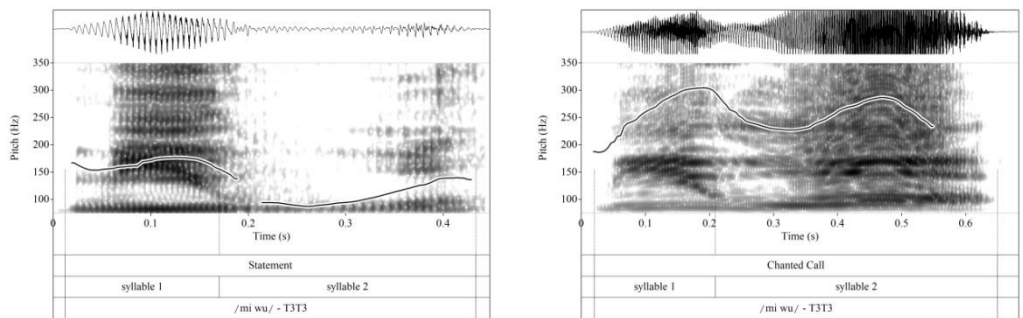


Figure 6.22 Statement (left) and CC (right): Disyllabic (Set1) – T3T3

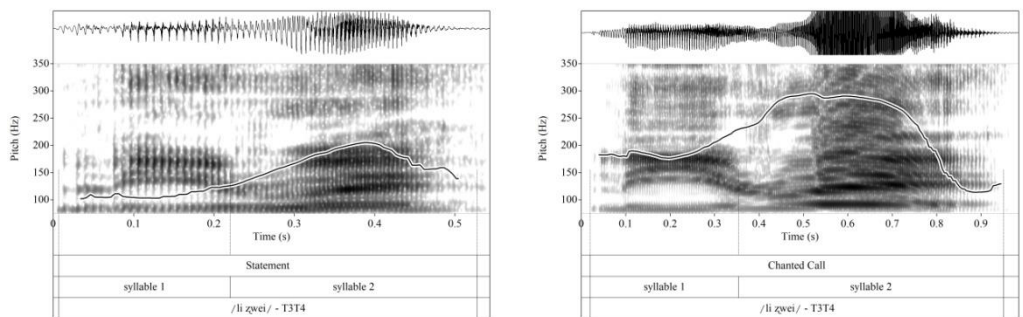


Figure 6.23 Statement (left) and CC (right): Disyllabic (Set1) – T3T4

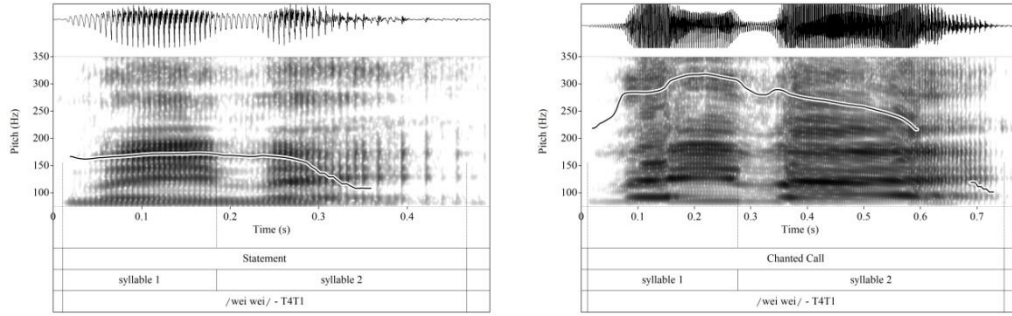


Figure 6.24 Statement (left) and CC (right): Disyllabic (Set1) – T4T1

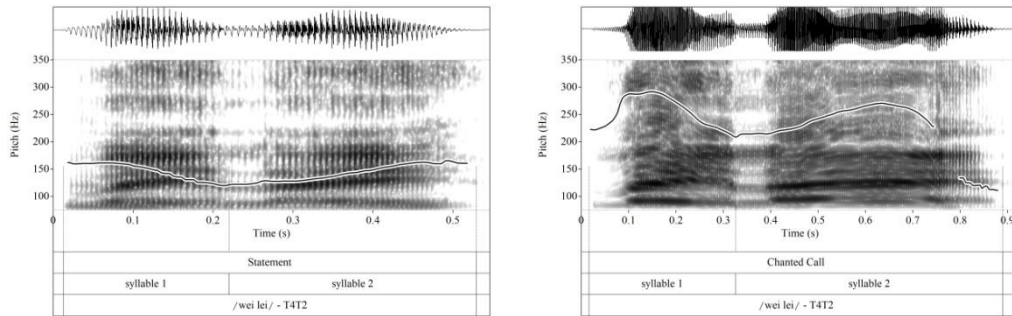


Figure 6.25 Statement (left) and CC (right): Disyllabic (Set1) – T4T2

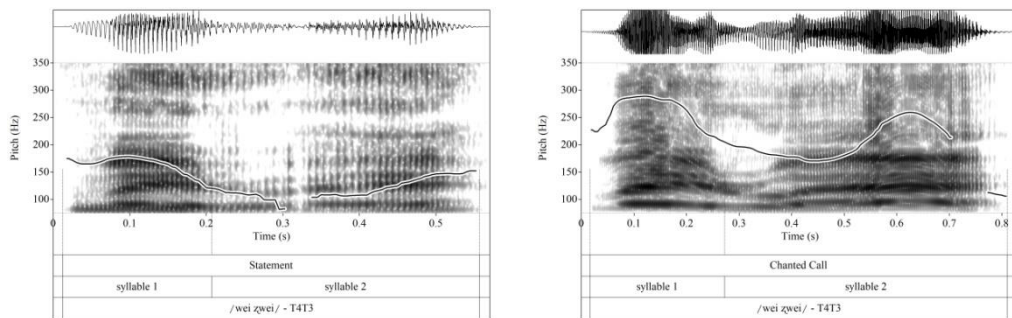


Figure 6.26 Statement (left) and CC (right): Disyllabic (Set1) – T4T3

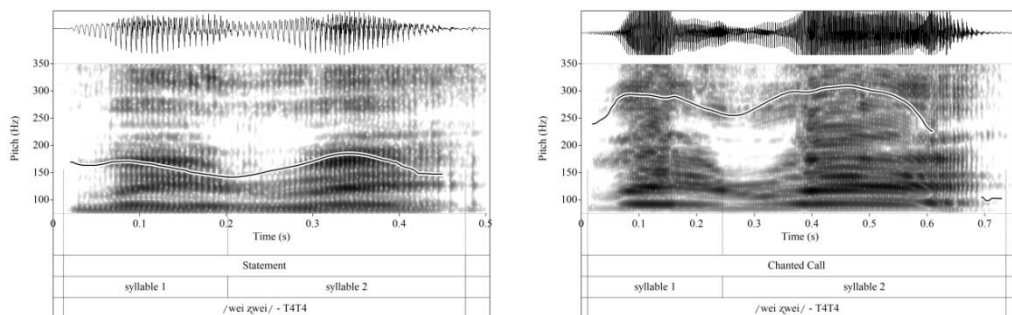


Figure 6.27 Statement (left) and CC (right): Disyllabic (Set1) – T4T4

Set 2: with T0

Set-2 data, as shown in Figure 6.28 to Figure 6.31, all end with T0. Their statement tune has low and flat contours for most of the T0 syllable. In a number of instances that have not been displayed here, T0 does not have an identifiable F0 contour or F0 value. The lexical tones of the first syllables in the statements have contours that bear more resemblance to the contours of their own citation tones.

A raise in the **register** of the CC tune in Set 2 can also be observed. The **pitch accent** in Set 2 is in accordance with the Set-1 results: a H* pitch accent is associated with the H of the lexical tones in each syllable. The H* is manifested in the first syllables of all tonal sequences visibly in the figures. The second syllable, i.e. the neutral tone T0, resembles a lower T1. Therefore, the H* alignment of the CC tune for all four tonal sequences in Set 2 are similar to the ones with T1 as the second tone in Set 1 (c.f. Figure 6.28 vs Figure 6.12, Figure 6.29 vs Figure 6.16, Figure 6.30 vs Figure 6.20, Figure 6.31 vs Figure 6.24). T4T0 differs from T4T1 the most among all the above combinations due to the tonal dissimilation in T4T1. A **L% boundary tone** can also be found at the right edge of the IP in the CC instances. Since T0 does not possess its own underlying tone, the H* is shown directly on the T0 syllable, aligned with the first half of the syllable. The L% boundary tone, together with the H* at the beginning of the second syllables, gives rise to the falling contour in T0 with the CC tune.

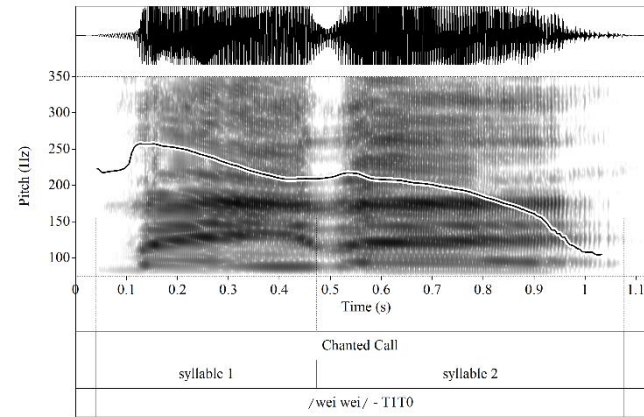
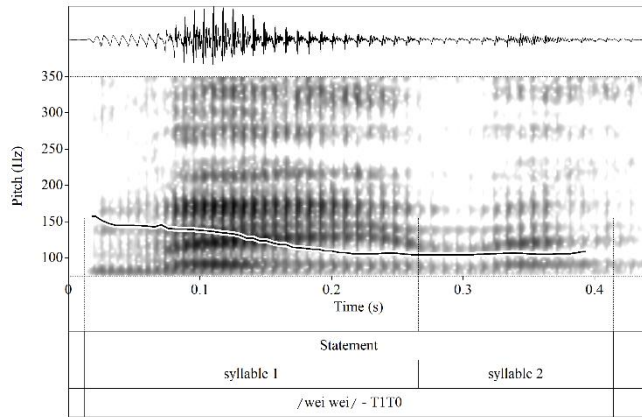


Figure 6.28 Statement (left) and CC (right): Disyllabic (Set2) – T1T0

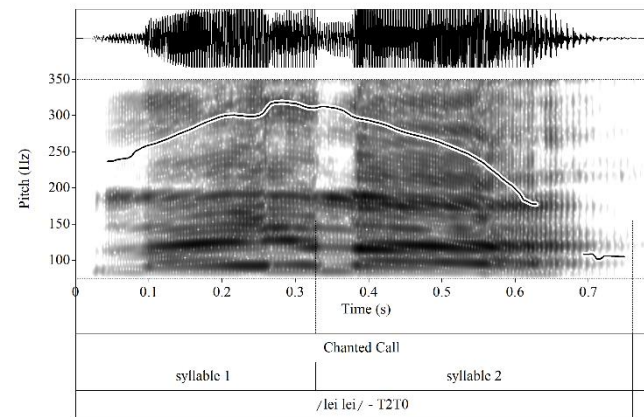
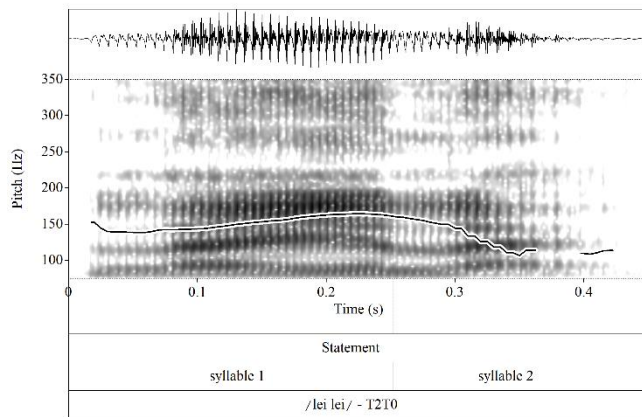


Figure 6.29 Statement (left) and CC (right): Disyllabic (Set2) – T2T0

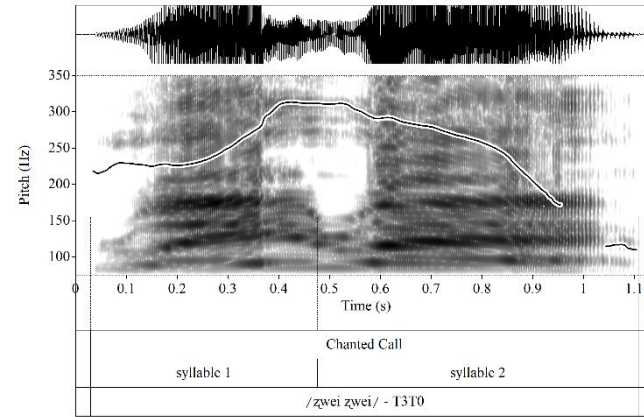
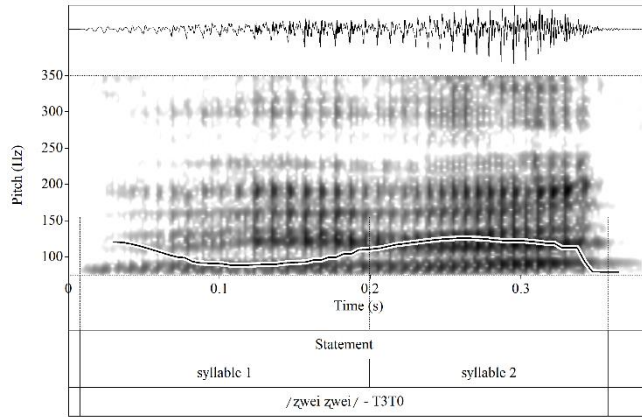


Figure 6.30 Statement (left) and CC (right): Disyllabic (Set2) – T3T0

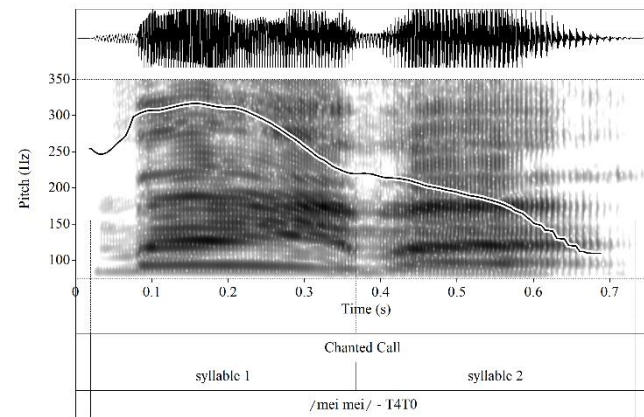
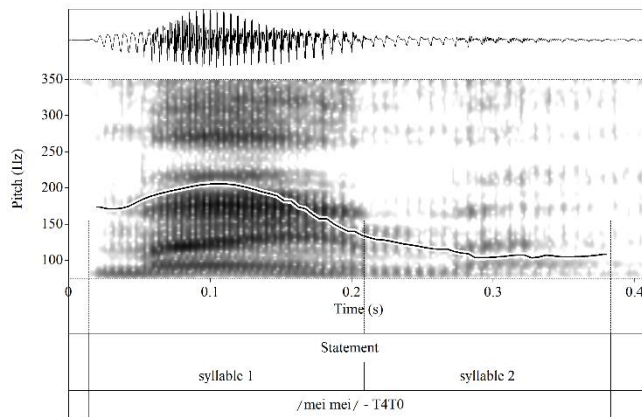


Figure 6.31 Statement (left) and CC (right): Disyllabic (Set2) – T4T0

6.5.2.2 Duration

Set 1: without T0

The average duration of the disyllabic words on the whole was longer in the CC tune than in the statement tune. Figure 6.32 shows the average duration of the **first syllables**. The duration of the first syllables in the CC tune (209.1ms on average, $\overline{SD} = 41.5$ ms) was longer than in the statement tune (191.3ms on average, $\overline{SD} = 34.6$ ms). The duration of the first syllables was influenced by TYPE ($\chi^2(1) = 42.79$, $p < 0.001$ ***) and TONE ($\chi^2(15) = 81.01$, $p < 0.001$ ***).

The **second syllables** (Figure 6.33) were much longer in the CC tune (381.2ms on average, $\overline{SD} = 67.8$ ms) than in the statement tune (264.5ms on average, $\overline{SD} = 51.6$ ms). The average length of the second syllables with CC tune was 1.44 times the average length of those in the statement tune. The length of the second syllables was also influenced by TYPE ($\chi^2(1) = 613.68$, $p < 0.001$ ***), as well as TONE ($\chi^2(15) = 178.37$, $p < 0.001$ ***).

The difference between the average duration of the first syllables and the second syllables resembles that which is between the onsets and rhymes in monosyllabic calls – the lengthening occurred at the latter half of the word, which was the rhyme of a monosyllabic word and the second syllable of a disyllabic word.

Set 2: T0 data

In this set of data, the second syllables all had T0 – the neutral tone. A neutral-tone syllable is a reduced syllable which has a much shorter duration than a full syllable that has one of the other four lexical tones.

Similar to the results of Set-1 data, the average duration of the disyllabic words on the whole was also longer in the CC tune than in the statement tune in Set-2 data. Figure 6.34 shows the average duration of the **first syllables** preceding the neutral-tone syllables. The duration of the first syllables in Set 2 that had the CC tune (240.7ms on average, $\overline{SD} = 55.2\text{ms}$) was longer than the ones with the statement tune (207.0ms on average, $\overline{SD} = 33.5\text{ms}$). The duration of the first syllables was influenced by TYPE ($\chi^2(1) = 33.15, p < 0.001$ ***), but not TONE ($\chi^2(3) = 81.01, p = 0.085$).

The **second syllables** (Figure 6.33) were much longer in the CC tune (349.6ms on average, $\overline{SD} = 80.2\text{ms}$) than in the statement tune (184.9ms on average, $\overline{SD} = 46.3\text{ms}$). The average length of the second syllables with CC tune was 1.89 times the average length of those in the statement tune. The length of the second syllables was also influenced by TYPE ($\chi^2(1) = 187.95, p < 0.001$ ***), but not TONE ($\chi^2(3) = 2.5, p = 0.47$). The T0 syllables were shorter than the preceding full syllable in the statement tune, but were nearly doubled in the CC tune and were even longer than the preceding full syllables.

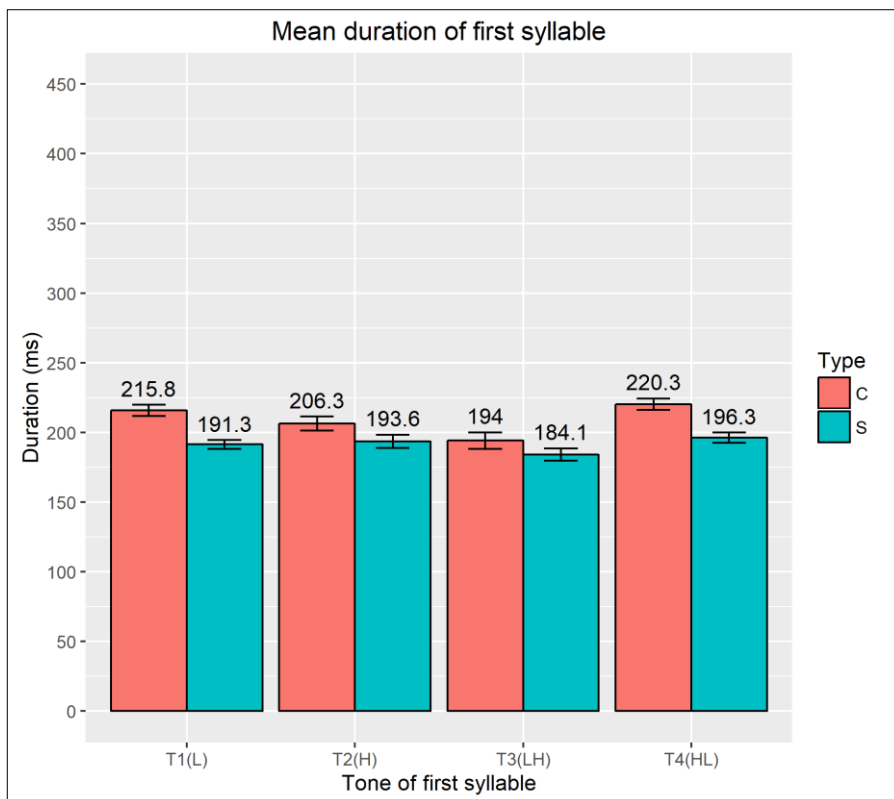


Figure 6.32 Duration: First syllable (Set 1)

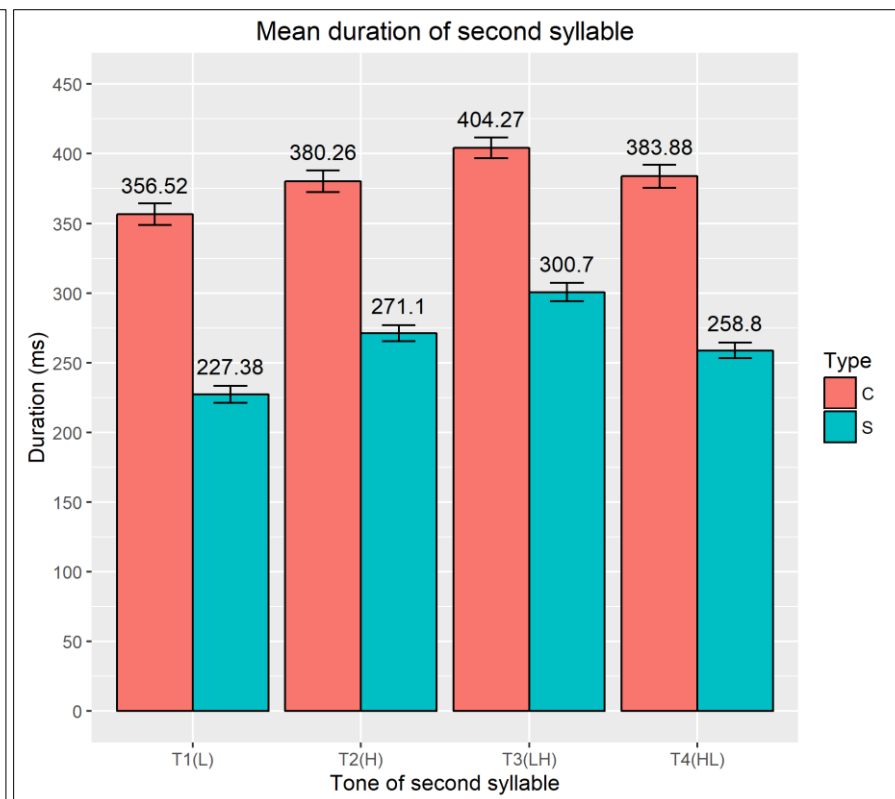


Figure 6.33 Duration: Second syllable (Set 1)

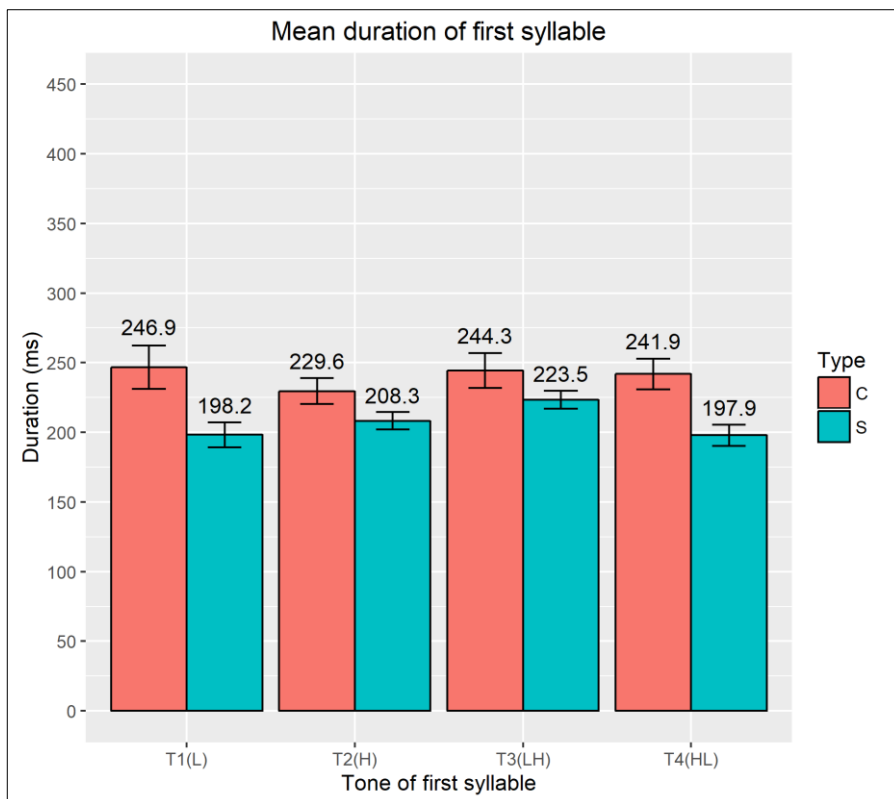


Figure 6.34 Duration: First syllable (Set 2)

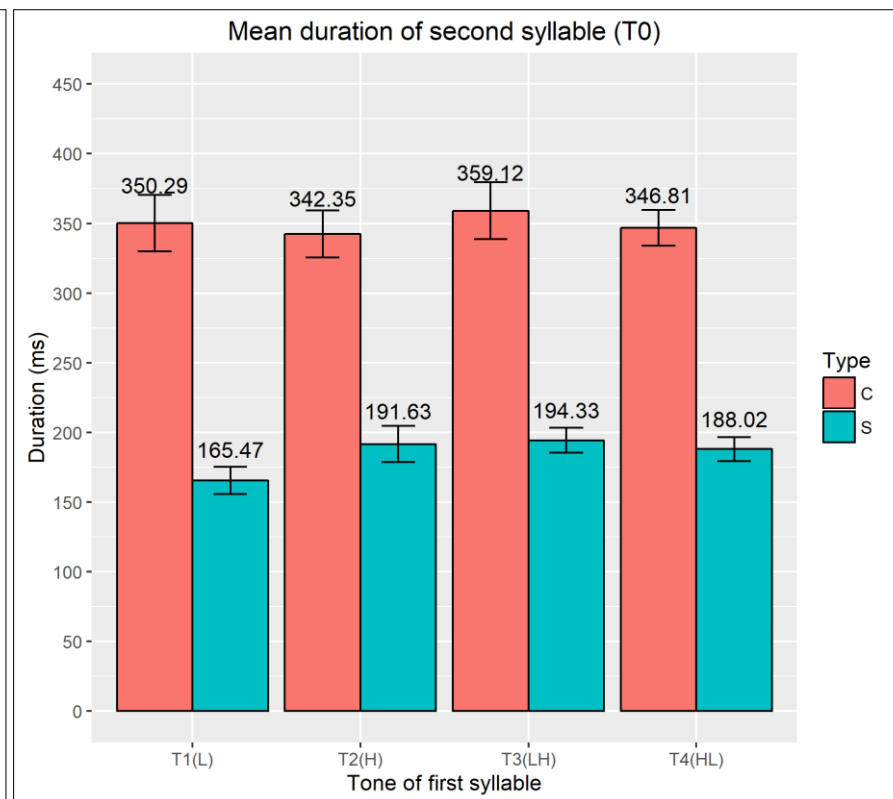


Figure 6.35 Duration: Second syllable – T0 (Set 2)

6.5.2.3 Mean Pitch

Mean pitch is one of the indicators for the register, which answers the question of whether the register of the CC tune is raised or expanded on the basis of the statement, and provides information on which specific parts of the tunes change in the register. Moreover, the mean pitch data facilitates the analysis of pitch accents. The mean pitch results of Set 1 and Set 2 data are presented in this section. Further discussions will continue in Section 0.

Set 1: without T0

The fact that the CC tune has a higher register than the statement tune can be ascertained by way of visual inspection, since the height of the whole pitch contours are evidently higher in the CC tune (TYPE ($\chi^2(1) = 1716.5$, $p < 0.001^{***}$), TONE ($\chi^2(15) = 314.77$, $p < 0.001^{***}$)). The register difference between CC tune and statement tune is substantially larger than the difference between the IntQ tune and statement tune in Chapter 4.

Comparing the two syllables in the disyllabic words, the second syllable was higher than the first syllable, both in terms of the F0 values and in terms of the difference between the average mean pitches of the CC tune and the statement tune. The average mean F0 value of the first syllables was 5.20ERB ($\overline{SD} = 0.90\text{ERB}$) for the statement tune, and 7.02ERB ($\overline{SD} = 1.09\text{ERB}$) for the CC tune. The average mean F0 of the second syllables were 5.30ERB ($\overline{SD} = 0.93\text{ERB}$) for the statement tune and 7.18ERB ($\overline{SD} = 1.06\text{ERB}$). The second syllables were slightly higher than the first syllables: the average difference between the two syllables was 0.16ERB in the CCs, and 0.10ERB in the statements. The mean F0 of the CCs was higher than the statements by 1.82ERB

on average in the first syllables, and 1.88ERB in the second syllables. A mix-effect model was used to examine the main effect of syllable position. The difference between the two syllables was statistically significant, SYLLABLE: $\chi^2(1) = 19.5$, $p < 0.001$ ***.

Table 6.8 shows the mean pitch of the disyllabic words for Set 1. Tone A is the tone of the first syllable and Tone B represents the tone of the second syllable. The intersection of Tone A and Tone B contains the mean pitch in ERB for the tonal sequence of TATB, as well as the standard deviation in brackets; for example, the mean pitch for T1T2 with CC tune is 6.45ERB and the standard deviation is 1.2ERB, while the mean pitch for T1T2 with statement tune is 4.67ERB with a standard deviation of 0.8ERB.

Table 6.8 Mean Pitch of Disyllabic Words: Set1 (Unit: ERB)

		Tone A								
		T1		T2		T3		T4		
		Syl. 1	Syl.2	Syl. 1	Syl.2	Syl. 1	Syl.2	Syl. 1	Syl.2	
Tone B	T1	C	7.26(0.9)	7.06(1.2)	7.39(0.9)	7.05(1.1)	6.69(1.2)	7.15(1.3)	7.87(1.0)	7.29(1.1)
		S	5.26(0.9)	5.44(1.0)	5.39(0.9)	5.30(1.1)	4.88(0.9)	5.31(1.0)	5.73(0.9)	5.47(1.0)
	T2	C	6.45(1.2)	7.16(1.0)	7.62(1.0)	7.83(0.9)	6.24(1.1)	7.38(1.0)	7.09(1.2)	7.03(1.1)
		S	4.67(0.8)	5.03(0.7)	5.50(0.9)	5.67(0.9)	4.81(0.8)	5.18(0.8)	5.37(0.9)	5.13(0.7)
	T3	C	6.45(1.1)	6.54(0.8)	7.52(1.0)	7.04(1.0)	7.27(1.1)	6.83(1.0)	7.32(1.3)	6.42(1.3)
		S	4.67(0.9)	4.83(0.9)	5.47(1.0)	4.86(0.8)	5.45(1.0)	5.20(1.2)	5.36(0.9)	4.78(1.1)
	T4	C	6.39(1.0)	7.21(1.1)	7.33(1.0)	7.89(1.0)	6.15(1.2)	7.37(1.2)	7.31(1.1)	7.58(1.0)
		S	4.89(1.0)	5.28(1.0)	5.54(0.8)	6.05(1.0)	4.59(0.9)	5.51(0.8)	5.58(0.9)	5.81(0.9)

(Syl. 1: the first syllable; Syl. 2: the second syllable.)

Set 2: T0 data

Table 6.9 shows the mean pitch results for Set-2 data. The CC tune was also significantly higher than the statement tune (TYPE ($\chi^2(1) = 354.48$, $p < 0.001$ ***), TONE ($\chi^2(3) = 46.01$, $p < 0.001$ ***)). The TYPE*TONE interaction was not significant, due to

the fact that all the second syllables carry the same neutral tone T0. Figure 6.36 visualises the mean pitch of the T0 syllable when preceded by different lexical tones (x-axis). In statements, T0 was higher when preceded by T2 (H Tone) and T3 (LH Tone) than by T1 (L Tone) and T4 (HL Tone), which means that T0 was strongly influenced by the preceding tones' ending pitch height.

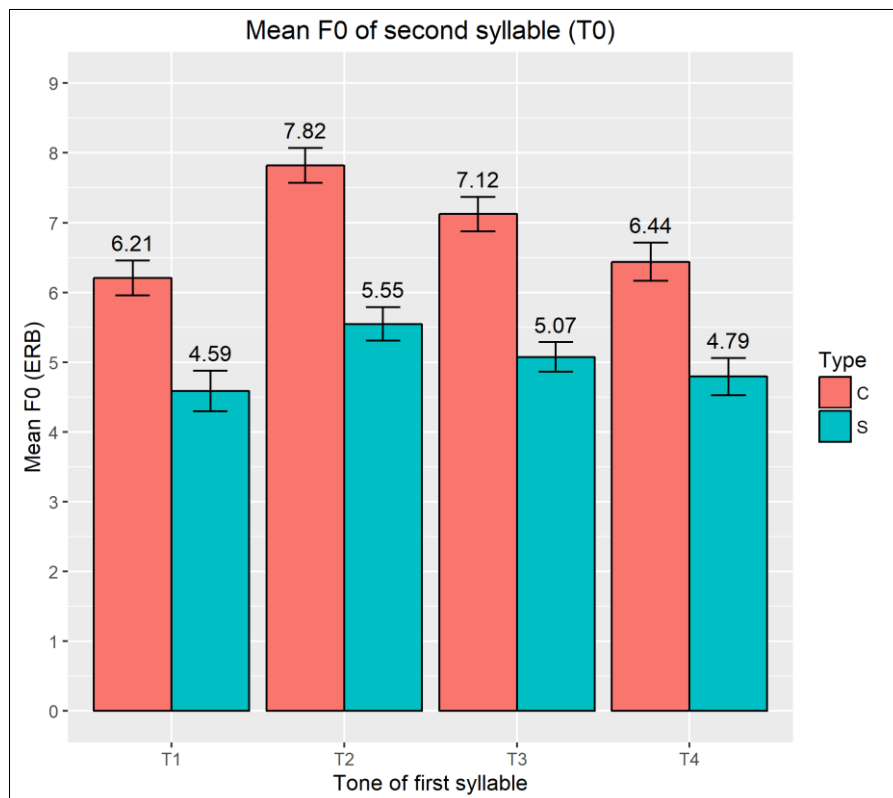


Figure 6.36 Mean F0: Second syllable – T0 (Set 2)

The difference between the first syllables and the second syllables in terms of the mean pitch exhibited a slightly different pattern from the Set-1 results: the second syllables were lower than the first syllables on average. The average mean F0 value of the first syllable was 5.21ERB ($\overline{SD} = 1.05\text{ERB}$) for the statement tune, and 7.15ERB ($\overline{SD} = 1.10\text{ERB}$) for the CC tune. The mean pitch of the first syllable in the statement tune was approximately the same with that of the Set-1 results, with the difference being

merely 0.1ERB. The average for the second syllables, i.e. the T0 syllables, in the statement tune was 5.00ERB ($\overline{SD} = 1.13\text{ERB}$), and 6.90ERB ($\overline{SD} = 1.17\text{ERB}$) in the CC tune. The second syllables were lower than the first syllables by 0.25ERB in the CCs, and by 0.21ERB in the statements. The mean F0 of the CCs was higher than the statements by 1.94ERB on average in the first syllables; the second syllables were also higher in the CCs than the statements by 1.90ERB. However, the second syllables were lower than the first syllables on average. A mix-effect model was used to examine the main effect of syllable position. The difference between the two syllables was statistically significant, SYLLABLE: $\chi^2(1) = 9.82$, $p = 0.0017$ **.

Table 6.9 Mean Pitch of Disyllabic Words: Set2 (Unit: ERB)

		Tone A							
		T1		T2		T3		T4	
		Syl. 1	Syl.2	Syl. 1	Syl.2	Syl. 1	Syl.2	Syl. 1	Syl.2
ToneB = T0	C	7.05(1.1)	6.21(1.1)	7.53(1.0)	7.82(1.1)	6.26(1.0)	7.12(1.2)	7.76(1.2)	6.44(1.3)
	S	4.89(1.0)	4.59(1.2)	5.29(0.8)	5.55(1.1)	4.89(1.6)	5.07(1.0)	5.75(1.0)	4.79(1.3)

6.5.2.4 Tonal Scaling

Tonal scaling (F0 maxima and F0 minima) in the current study of disyllabic calls not only provides insight into the register change, but also facilitates with the understanding of pitch accents in the CC tune. Therefore, the aim of the presentation of tonal scaling data is to answer the following three questions:

- (i) Is the register raised or expanded?
- (ii) Is the H* associated with the first syllable, the second syllable, or the entire disyllabic word?
- (iii) Is the H* associated with the H of the lexical tones, or the L?

The major contents of the current section are the values of F0 maxima and F0 minima, and Section 6.5.3.3 and Section 6.5.3.4 will further discuss how they are relevant to the analyses of register and pitch accents and how these data help to answer the above questions. The results for both Set-1 data and Set-2 data are presented in the following order: firstly, the general description data and statistical analyses of the maximum F0 and minimum F0 of the CC tune and the statement tune are reported; secondly, the average values of F0 maxima of the two syllables within each tune are compared, and the values of the two tunes within each syllable are also compared; finally, the average values of F0 minima are cross compared in the same manner: first between the two syllables of each tune, and then between the two tunes within each syllable. Table 6.10 to Table 6.13 present the complete results of the F0 maxima and minima for the two sets of data separately. The numbers in brackets are the standard deviations for each category. “CC” represents the chanted call tune, and “S” represents the statement tune.

Set 1: without T0

The average **maximum F0** of the CCs was significantly higher than that of the statements, TYPE ($\chi^2(1) = 811.02, p < 0.001^{***}$), TONE ($\chi^2(15) = 187.84, p < 0.001^{***}$). The average maximum F0 in the CC tune was 7.80ERB ($\overline{SD} = 1.10\text{ERB}$), while that in the statement tune was 5.99ERB ($\overline{SD} = 1.29\text{ERB}$). The average **minimum F0** was also higher in the CC tune (5.73ERB, $\overline{SD} = 1.34\text{ERB}$) than in the statement tune (5.04ERB, $\overline{SD} = 0.94\text{ERB}$), TYPE ($\chi^2(1) = 596.74, p < 0.001^{***}$), TONE ($\chi^2(15) = 236.86, p < 0.001^{***}$), TONE*TYPE ($\chi^2(15) = 33.87, p = 0.0035^{**}$).

Another significant main effect that influenced the **F0 maxima** was SYLLABLE ($\chi^2(1) = 208.52, p < 0.001^{***}$). The first syllables were not as high as the second syllables on

average, in the statement tune and the CC tune respectively: the first syllables averaged to 6.61ERB ($\overline{SD} = 1.07\text{ERB}$), and the second syllable to 7.19ERB ($\overline{SD} = 1.32\text{ERB}$). In the CC tune, the second syllables were higher than the first syllables by 0.56ERB, while the difference was 0.59ERB in statements, with the seconding syllables being the higher ones. The differences between the two tunes in each syllable also supported that the second syllables had bigger change. The first syllables in the CC-tune utterances were higher than the corresponding statements by 1.83ERB on average, while the second syllables were higher in the CC utterances than the statements by 1.79ERB. The maximum pitch difference between the two tunes was higher in the second syllables than in the first syllables.

F0 minima also significantly differed between CC tune and statement tune in terms of SYLLABLE ($\chi^2(1) = 207.86$, $p < 0.001^{***}$). However, contrary to the values of F0 maxima, the minimum F0 of the first syllables were higher than the second syllables on average and in each tune. The average minimum F0 of the first syllables was 6.06ERB ($\overline{SD} = 1.05\text{ERB}$), and that of the second syllables was 4.72ERB ($\overline{SD} = 1.23\text{ERB}$). The first syllables were higher than the second syllables by 1.40ERB in the CCs, while they were higher than the second syllables by 1.28ERB in the statements. The difference between the CC tune and the statement tune was manifested by a larger margin in the first syllables too. The difference between the minimum F0 of the CC tune and the statement tune in the first syllables was 0.75ERB, while that in the second syllables was 0.63ERB.

Set 2: T0 data

The data in Set 2 obtained similar results to those in Set 1. The major difference lay in the difference between the minimum F0 of the CC tune and the statement tune in the second syllables, which can be ascribed to the uniqueness of the lexical tone of the second syllables, i.e. T0. Further details will be presented below.

The **maximum F0** of the CCs was significantly higher than that of the statements on average: TYPE ($\chi^2(1) = 231.06$, $p < 0.001^{***}$), TONE ($\chi^2(3) = 60.84$, $p < 0.001^{***}$), TONE*TYPE ($\chi^2(3) = 10.32$, $p = 0.016^*$). The average maximum F0 in the CC tune was 7.55ERB ($\overline{SD} = 1.16\text{ERB}$), while that in the statement tune was 5.69ERB ($\overline{SD} = 2.74\text{ERB}$). The average **minimum F0** was also higher in the CC tune (5.76ERB, $\overline{SD} = 1.52\text{ERB}$) than in the statement tune (4.47ERB, $\overline{SD} = 0.96\text{ERB}$): TYPE ($\chi^2(1) = 94.02$, $p < 0.001^{***}$), TONE ($\chi^2(3) = 27.70$, $p < 0.001^{***}$). The minimum F0 result in the CC tune is similar to that in the Set-1 data (5.73ERB).

Another significant main effect that influenced the **F0 maxima** was SYLLABLE ($\chi^2(1) = 4.03$, $p = 0.045^*$). Similar to the maximum F0 results in Set 1, the first syllables were not as high as the second syllables in both the statement tune and the CC tune: the first syllables averaged to 6.76ERB ($\overline{SD} = 1.06\text{ERB}$), and the second syllables, 6.72ERB ($\overline{SD} = 1.48\text{ERB}$). However, despite the difference of the values of F0 maxima between the two syllables, the changes between the two syllables are the same in both tunes: in the CC tune, the average maximum F0 of the second syllables was higher than the first syllables by 0.56ERB; in the statement tune, the difference between the two syllables was 0.59ERB, with the second syllables being higher. Moreover, the difference between the maximum F0 of the CC tune and the statement tune in the first syllables

and the second syllables respectively were 1.83ERB and 1.79ERB, with the CC tune being higher than the statement tune in both syllables.

The values of **F0 minima** also significantly differed between the CC tune and the statement tune in terms of SYLLABLE ($\chi^2(1) = 53.71, p < 0.001^{***}$). Similar to the results of minimum F0 in Set 1, the minimum F0 of the first syllables were higher than the second syllables on average in the Set-2 data as well. The average minimum F0 of the first syllables was 6.06ERB ($\overline{SD} = 1.05\text{ERB}$), and that of the second syllables was 4.72ERB ($\overline{SD} = 1.23\text{ERB}$). In the CCs, the first syllables were higher than the second syllables by 1.20ERB in terms of the minimum F0 results, while the difference was only 0.60ERB in the statement tune, with the first syllables doubling that of the second syllables. In the results reported before – including all maximum F0 data in Set 1 and Set 2, as well as the minimum F0 data in Set 1 – all had comparable differences across syllables or tunes. However, this comparison between the first syllable and the second syllable in the CCs and statements differ substantially. This discrepancy will be further discussed in Section 6.5.3.4 on boundary tones. The difference between the CC tune and the statement tune was manifested by a much larger margin in the first syllables, which was in conformity with the Set-1 results. The difference between the minimum F0 of the CC tune and the statement tune in the first syllables was 1.59ERB, while that in the second syllables was 0.99ERB.

Table 6.10 Average Maximum F0 Values (Set 1)

	Syllable 1		Syllable 2		Average by Syllable		Average by Tune	
	CC	S	CC	S	Syllable 1	Syllable 2	CC	S
T1T1	7.87(0.95)	5.74(0.88)	7.93(1.15)	6.35(1.68)	6.81(0.91)	7.14(1.41)	7.90(1.05)	6.04(1.28)
T1T2	6.95(1.57)	5.25(1.09)	8.03(0.92)	5.91(0.84)	6.10(1.33)	6.97(0.88)	7.49(1.24)	5.58(0.97)
T1T3	6.77(1.19)	5.34(1.17)	7.57(0.89)	6.33(1.95)	6.05(1.18)	6.95(1.42)	7.17(1.04)	5.83(1.56)
T1T4	6.74(1.04)	5.75(1.45)	8.12(1.16)	6.13(1.58)	6.24(1.25)	7.13(1.37)	7.43(1.10)	5.94(1.52)
T2T1	8.05(0.88)	5.94(0.98)	7.96(0.83)	6.25(1.66)	6.99(0.93)	7.10(1.25)	8.00(0.85)	6.09(1.32)
T2T2	8.16(0.99)	6.02(1.11)	8.34(0.95)	6.59(1.40)	7.09(1.05)	7.47(1.18)	8.25(0.97)	6.30(1.25)
T2T3	8.06(0.92)	6.06(1.13)	8.24(1.25)	6.15(1.72)	7.06(1.03)	7.20(1.49)	8.15(1.09)	6.10(1.42)
T2T4	7.94(0.94)	6.02(0.79)	8.64(1.13)	6.90(1.72)	6.98(0.87)	7.77(1.42)	8.29(1.04)	6.46(1.26)
T3T1	7.59(1.12)	5.23(0.93)	7.91(1.14)	6.01(1.37)	6.41(1.03)	6.96(1.25)	7.75(1.13)	5.62(1.15)
T3T2	6.58(1.08)	5.03(1.04)	8.17(0.87)	6.35(1.29)	5.81(1.06)	7.26(1.08)	7.37(0.98)	5.69(1.17)
T3T3	7.99(1.06)	5.82(1.02)	8.07(1.34)	6.34(1.77)	6.90(1.04)	7.20(1.55)	8.03(1.20)	6.08(1.39)
T3T4	6.56(1.26)	4.52(0.91)	8.34(1.28)	6.09(0.86)	5.54(1.09)	7.22(1.07)	7.45(1.27)	5.30(0.89)
T4T1	8.27(0.99)	6.12(0.94)	8.22(1.09)	6.46(1.61)	7.19(0.96)	7.34(1.35)	8.25(1.04)	6.29(1.27)
T4T2	7.45(1.22)	6.09(1.25)	7.83(1.18)	6.04(1.23)	6.77(1.23)	6.93(1.20)	7.64(1.20)	6.06(1.24)
T4T3	7.82(1.20)	6.07(1.00)	7.69(1.43)	6.00(2.08)	6.94(1.10)	6.85(1.76)	7.76(1.31)	6.03(1.54)
T4T4	7.62(1.05)	6.17(1.20)	8.26(1.17)	6.72(1.75)	6.89(1.12)	7.49(1.46)	7.94(1.11)	6.44(1.48)
Average by Tone	7.53(1.09)	5.70(1.06)	8.08(1.11)	6.29(1.53)	6.61(1.07)	7.19(1.32)	7.80(1.10)	5.99(1.29)

Table 6.11 Average Minimum F0 Values (Set 1)

	Syllable 1		Syllable 2		Average by Syllable		Average by Tune	
	CC	S	CC	S	Syllable 1	Syllable 2	CC	S
T1T1	6.39(0.98)	4.90(0.99)	5.15(1.06)	4.37(1.06)	5.65(0.99)	4.76(1.48)	5.77(1.02)	4.64(1.03)
T1T2	6.11(1.11)	4.31(0.78)	5.84(0.78)	4.42(0.78)	5.21(0.94)	5.13(1.12)	5.97(0.94)	4.36(0.78)
T1T3	5.85(1.08)	4.17(0.85)	5.62(0.85)	3.62(0.85)	5.01(0.97)	4.62(0.93)	5.73(0.97)	3.90(0.85)
T1T4	5.95(1.13)	4.55(0.91)	5.38(0.88)	4.35(0.88)	5.25(1.02)	4.87(1.23)	5.67(1.01)	4.45(0.90)
T2T1	6.44(1.08)	5.03(0.85)	5.12(0.93)	4.22(0.93)	5.74(0.97)	4.67(1.34)	5.78(1.00)	4.63(0.89)
T2T2	6.75(1.14)	5.10(0.96)	6.89(0.88)	5.29(0.88)	5.92(1.05)	6.09(1.21)	6.82(1.01)	5.20(0.92)
T2T3	6.61(1.14)	5.12(0.93)	5.78(0.90)	3.73(0.90)	5.87(1.03)	4.76(1.19)	6.20(1.02)	4.42(0.91)
T2T4	6.60(1.03)	5.27(0.84)	5.79(0.95)	5.39(0.95)	5.94(0.93)	5.59(1.45)	6.19(0.99)	5.33(0.89)
T3T1	5.63(1.51)	4.57(0.89)	5.45(1.01)	4.44(1.01)	5.10(1.20)	4.94(1.39)	5.54(1.26)	4.50(0.95)
T3T2	5.80(1.01)	4.38(0.86)	6.31(0.88)	4.59(0.88)	5.09(0.94)	5.45(1.12)	6.06(0.95)	4.48(0.87)
T3T3	6.20(1.30)	4.94(1.12)	5.62(0.99)	3.95(0.99)	5.57(1.21)	4.78(1.09)	5.91(1.14)	4.45(1.06)
T3T4	5.76(1.07)	4.17(0.88)	5.49(0.95)	4.42(0.95)	4.97(0.98)	4.95(1.39)	5.62(1.01)	4.30(0.92)
T4T1	7.99(1.03)	5.99(0.90)	5.22(0.91)	4.28(0.91)	6.99(0.96)	4.75(1.38)	6.60(0.97)	5.14(0.90)
T4T2	6.81(1.26)	6.01(1.31)	6.03(0.74)	4.67(0.74)	6.41(1.28)	5.35(0.95)	6.42(1.00)	5.34(1.02)
T4T3	6.92(1.34)	5.99(1.05)	4.99(0.88)	3.70(0.88)	6.45(1.19)	4.34(1.16)	5.95(1.11)	4.84(0.96)
T4T4	7.15(1.10)	6.02(1.22)	6.28(1.00)	4.97(1.00)	6.59(1.16)	5.62(1.21)	6.71(1.05)	5.50(1.11)
Average by Tone	6.43(1.14)	5.03(0.96)	5.68(0.91)	4.40(0.91)	5.73(1.05)	5.04(1.23)	6.06(1.03)	4.72(0.94)

Table 6.12 Average Maximum F0 Values (Set 2)

	Syllable 1		Syllable 2		Average by Syllable		Average by Tune	
	CC	S	CC	S	syl1	syl2	C	S
T1T0	7.47(1.19)	5.75(1.18)	6.77(1.18)	5.46(2.24)	6.61(1.19)	6.12(1.71)	7.12(1.19)	5.61(1.71)
T2T0	8.32(0.95)	5.92(0.84)	8.51(1.01)	6.12(1.22)	7.12(0.89)	7.31(1.11)	8.41(0.98)	6.02(1.03)
T3T0	7.45(1.25)	4.42(0.88)	7.86(1.26)	5.75(1.68)	6.07(1.07)	6.80(1.47)	7.66(1.25)	5.08(1.28)
T4T0	8.14(1.14)	6.32(1.04)	7.44(1.33)	5.83(1.89)	7.23(1.09)	6.63(1.61)	7.79(1.23)	6.07(1.47)
Average by Tone	7.85(1.13)	5.60(0.99)	7.65(1.19)	5.79(1.76)	6.76(1.06)	6.72(1.48)	7.75(1.16)	5.69(1.37)

Table 6.13 Average Minimum F0 Values (Set 2)

	Syllable 1		Syllable 2		Average by Syllable		Average by Tune	
	CC	S	CC	S	syl1	syl2	C	S
T1T0	6.39(1.24)	4.94(1.38)	4.11(0.82)	3.78(0.93)	5.25(1.03)	4.36(1.16)	5.25(1.03)	4.36(1.16)
T2T0	6.09(1.67)	5.56(2.00)	4.82(0.90)	4.85(1.13)	5.45(1.28)	5.21(1.57)	5.45(1.28)	5.21(1.57)
T3T0	5.53(1.16)	5.29(1.84)	3.97(0.99)	4.27(1.03)	4.75(1.08)	4.78(1.44)	4.75(1.08)	4.78(1.44)
T4T0	7.43(1.26)	4.84(1.57)	6.19(1.11)	3.78(0.78)	6.81(1.19)	4.31(1.17)	6.81(1.19)	4.31(1.17)
Average by Tone	6.36(1.33)	5.16(1.70)	4.77(0.96)	4.17(0.97)	5.56(1.14)	4.66(1.33)	5.56(1.14)	4.66(1.33)

6.5.2.5 Tonal Alignment

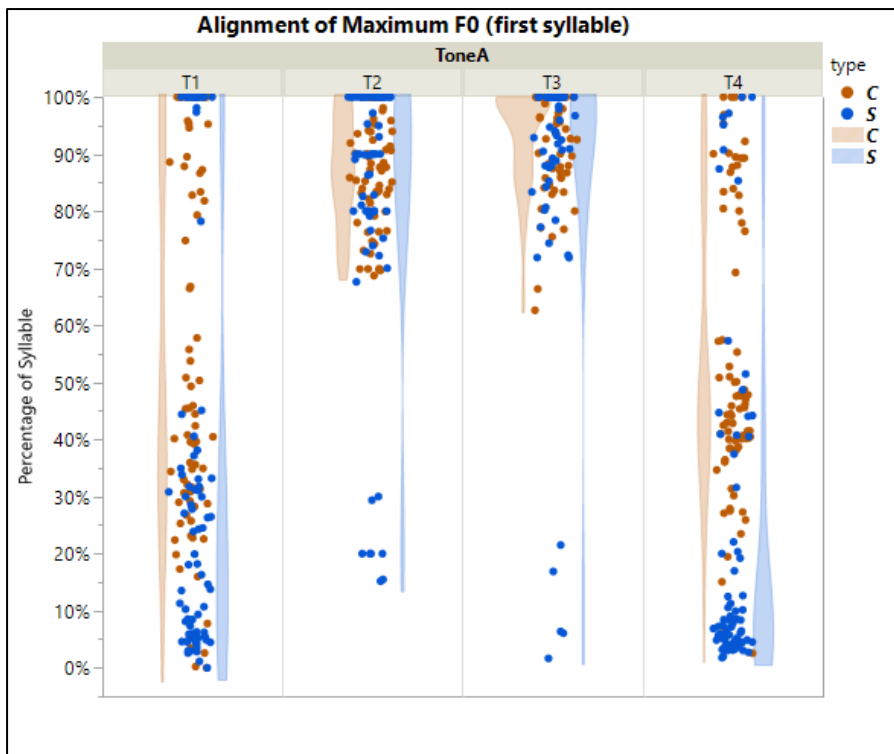
As in the monosyllabic chanted calls reported in Section 6.4, the tonal alignments of the F0 maxima and minima equip the analyses of pitch accents and boundary tones in the CC tune with qualitative data. Based on the observations from visual inspection (Section 6.5.2.1), the H* pitch accent and the L% boundary tone are of interest in the current section. Therefore, the current section aims at elaborating on these two aspects by way of tonal alignment: from the results of F0 maxima alignment, the alignment and association of the H* pitch accent can be further examined; the existence of L% can also be further attested by employing the values of F0 minima alignment.

A simple presentation of the average percentage will not be sufficient since the lexical tones go through tonal dissimilation and are affected by tonal coarticulation. Therefore, a number of figures (Figure 6.37 to Figure 6.43) are used in the current section to illustrate the location of extrema in relation to the syllables in percentage. The figures present the results by syllables – the results for the first syllables of the disyllabic words (Tone A) are grouped by Tone A; the ones representing the second syllables (Tone B in Set 1, and Tone 0 in Set 2) are grouped by Tone B. The dots in the figures depict the F0 extrema alignment, while the area plots sketch the density and distribution of the F0 extrema location. Different types of tunes are shown in different colours – the colour orange symbolises the CC tune, while blue embodies the statement tune. These figures are used for the presentation of all instances in the data, in order to supplement the examples quoted in the visual inspection section (Section 6.5.2.1). In this section, the F0 maxima alignment will be presented first for Set-1 data, followed by Set-2 data; then, the F0 minima alignment results will be presented respectively for Set 1 and Set 2.

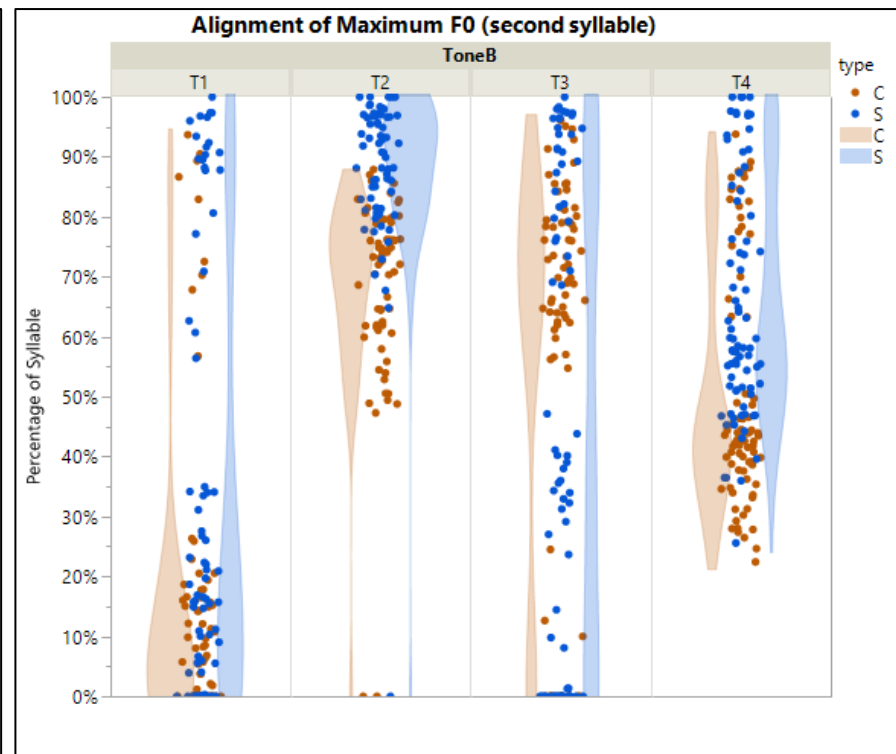
F0 Maxima Alignment: Set-1 Data

Through visual inspection (Section 6.5.2.1), a H* was overserved to be associated with the H part of each lexical tone. Figure 6.37 and Figure 6.38 further examine whether this association applies to most instances instead of merely to the example figures by means of displaying the maximum F0 alignment points for all instances. In the first syllables (Figure 6.37), most of T2 and T3's F0 maxima were aligned at the end of the first syllable, since they were both rising tones. T1 and T4 had more F0 maxima points towards the middle of the syllable, which was also where the H in the rhymes starts; however, due to tonal dissimilation, the T1 (falling L) becomes T2 (rising H) in instances in the first syllables of T1T1 and T4T1. For example, the F0 maxima alignment in the first syllable of T1T1 is shown in Figure 6.39. This strongly suggests that the H* is associated with the H of the lexical tone.

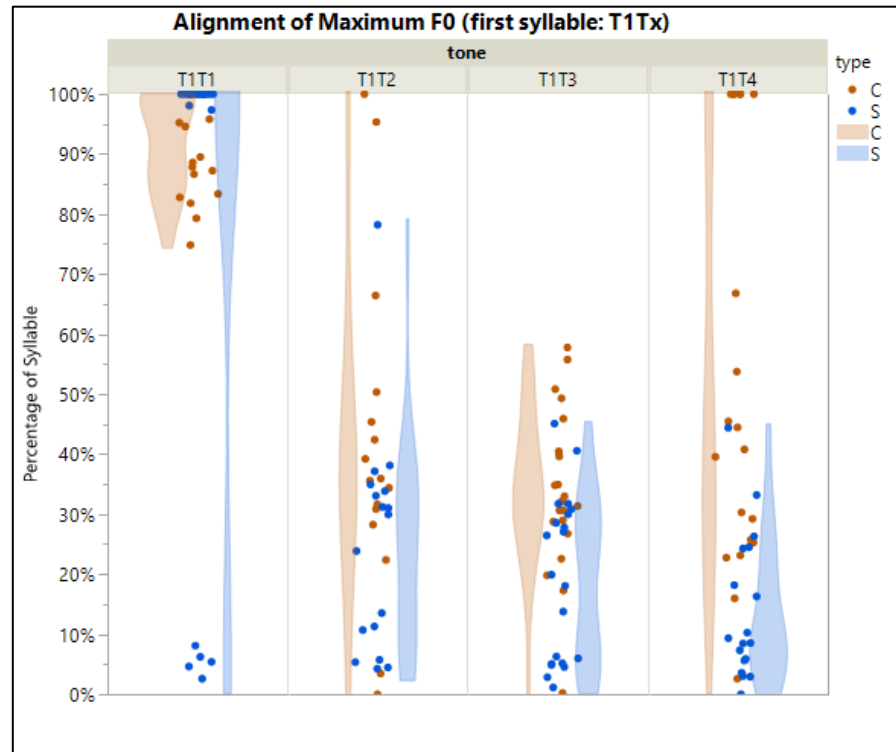
In the second syllables (Figure 6.38), the alignment of maximum F0 in TxT1 and TxT4 with the CC tune were similar to their counterparts in statements –T1 in TxT1 aligned at the beginning of the syllable, and T4 in TxT4, at the middle of the syllable, which were consistent with the F0 maxima alignment of the lexical tones in their forms. In the instances ending in rising tones, i.e. T2 and T3, however, the F0 maxima alignment moved away from the end of the syllable by a small but noticeable margin. The change was due to the assertion of the L% boundary tone, which will be further elaborated in the F0 minima alignment data in Figure 6.42.



**Figure 6.37 Set 1: Alignment of Maximum F0
in the First Syllables**



**Figure 6.38 Set 1: Alignment of Maximum F0
in the Second Syllables**

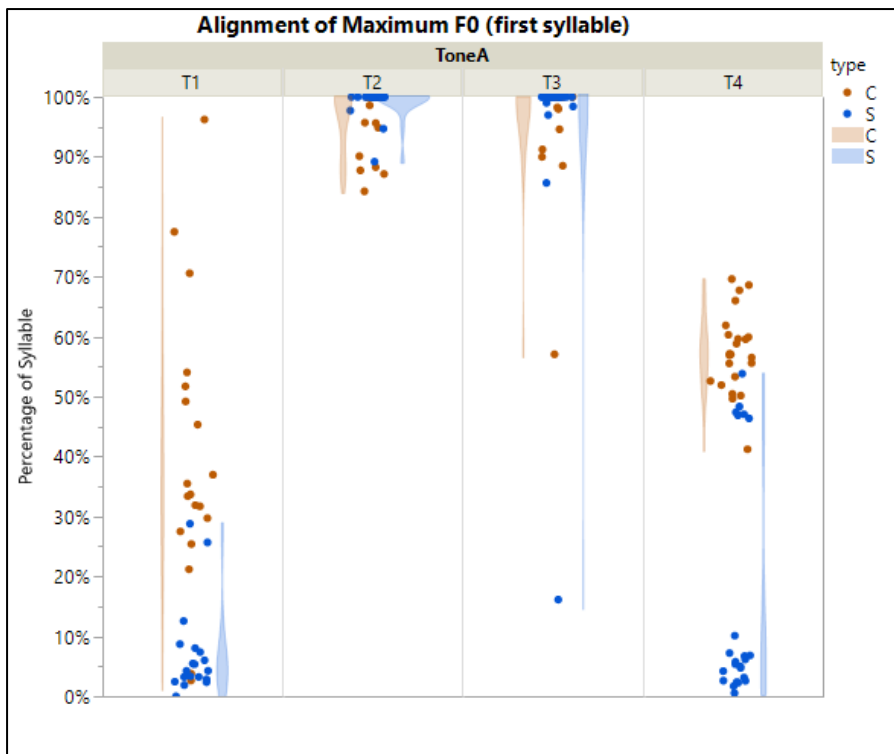


**Figure 6.39 Set 1: Alignment of Maximum F0
(First Syllables of T1Tx)**

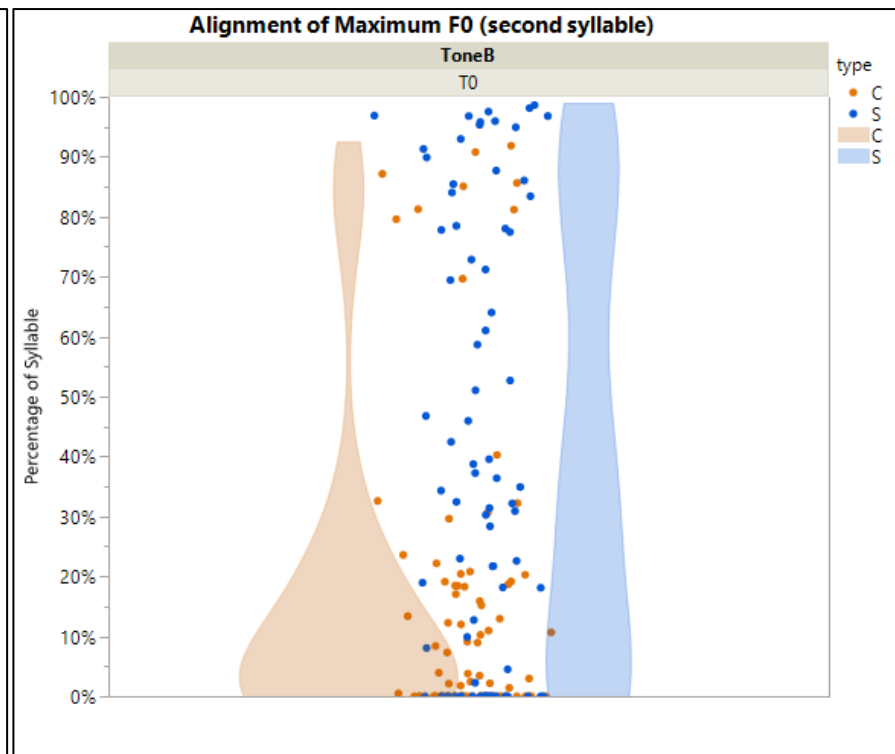
F0 Maxima Alignment: Set-2 Data

In Set-2 data, since the second syllables all carried T0, the first syllables were able to receive less influence from the second syllables. Therefore, the results of Set 2 are presented separately from the Set-1 data. Figure 6.40, a display of the F0 maxima alignment in the first syllables in Set 2, further supports and elaborates on the analyses in Set 1. The F0 maxima of T1 with the CC tune was distributed sparsely, with the densest area being aligned to around 30% of the syllable. The dispersal was partly due to tonal dissimilation. As introduced in Chapter 2, T0 in Tianjin Mandarin has no underlying tone and often realises as a low and falling contour, which is similar to T1. The fact that the maximum F0 points appeared towards the end of the syllable was due to a similar tonal dissimilation as in T1T1. The F0 maxima aligned more densely at the middle of the syllable in the CC tune than in the statement tune, which clearly indicated that the H* was associated with the H of the Tone 4, which is around the middle in its citation form. The fact that the F0 maxima of T2 and T3 mostly gathered at the end of the first syllable also lends support for the H* association with the H in lexical tones.

The F0 maxima of the second syllables, i.e. the ones with T0, are shown in Figure 6.41. The distribution of the maximum F0 points was more disperse in the statement tune than in the CC tune, due to the fact that T0 does not have any underlying tones. In the CC tune, the majority of the F0 maxima were aligned with the H beginning of the T0 syllable; however, the H* pitch accent is associated with the H tone in the previous lexical tones.



**Figure 6.40 Set 2: Alignment of Maximum F0
in the First Syllables**



**Figure 6.41 Set 2: Alignment of Maximum F0
in the Second Syllables (T0)**

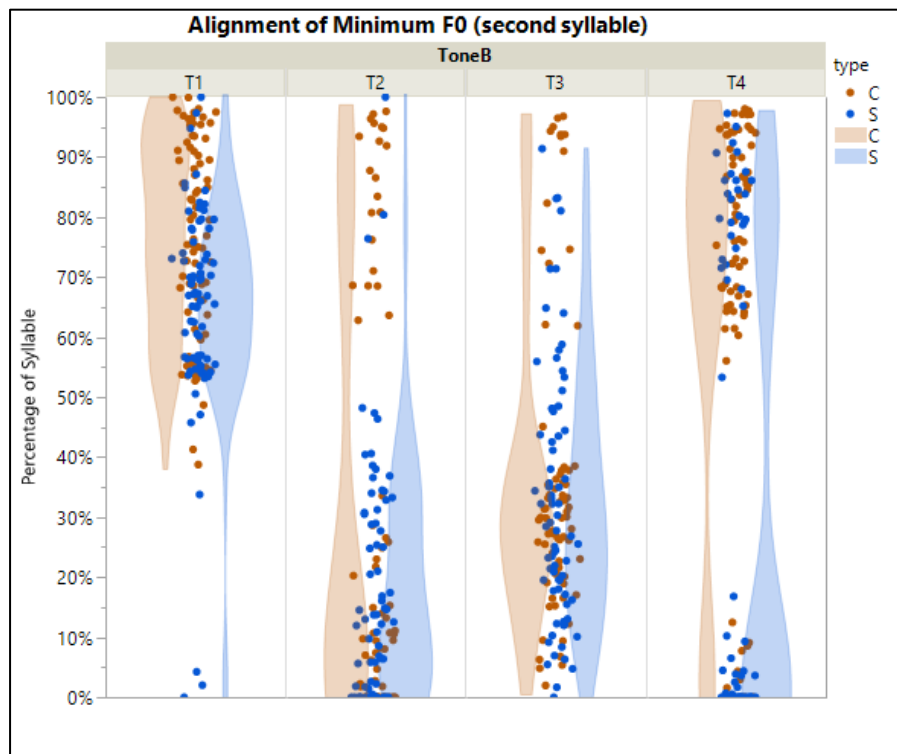
F0 Minima Alignment: Set-1 Data

Since F0 minima alignment only serves as an indicator of the L% boundary tone, only the data for the second syllables are displayed in this part. Figure 6.42 shows the alignment of F0 minima in the second syllables of Set-1 data. All four tones saw the major part of their F0 minima aligning with the portion in which their citation tones would – early in the rising tones (T2 and T3) and late in the falling tones (T1 and T4). However, what I am trying to find in the figures is evidence for the existence of L%. We therefore need to examine whether there were more F0 minima aligned to the end of the syllables in the CC tune than their statement counterparts; to be specific, in Figure 6.42, if towards the end of the four tones, there were more dots in orange than in blue, we could conclude that the L% existed in most CC utterances. The results shown in Figure 6.42 confirm the expectation. In T1 and T4, the density of F0 minima was higher in the CC tune than in the statement tune. In T2 and T3, the statement instances hardly had any F0 minima that were aligned with the end of the utterances; however, there were many more dots in orange than in blue, which suggested that some utterances that ended with a rising tone had lower endings than their beginnings. These results attest the analysis that a L% boundary tone can be found in the CC tune.

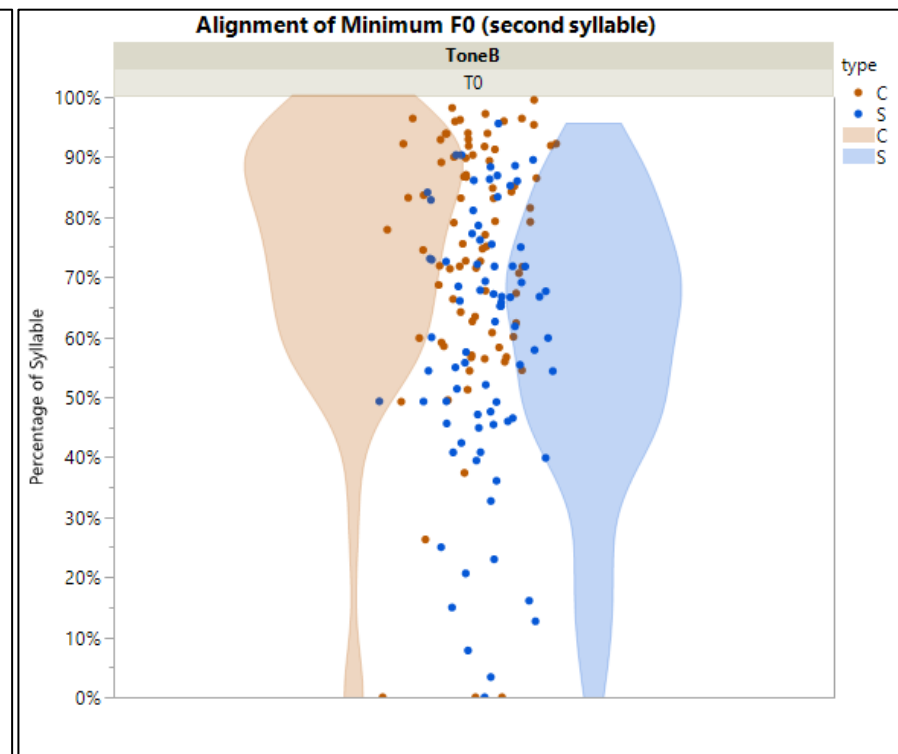
F0 Minima Alignment: Set 2 data

Figure 6.43 presents the F0 minima alignment for Set-2 data, which only contains T0 syllables. In the statement tune, the majority of the F0 minima were at approximate 60% of the T0 syllable, while the CC tune had much later F0 minima alignment, most of which were at around 90% of the syllable. Since T0 does not have an underlying tonal contour, the F0 minima points directly depicts the L% boundary tone. The fact that the

L% is aligned more to the end of the T0 syllables in the CC tune than in the statement tune has provided strong evidence for the L% boundary tone analysis.



**Figure 6.42 Set 1: Alignment of Minimum F0
in the Second Syllables**



**Figure 6.43 Set 2: Alignment of Minimum F0
in the Second Syllables (T0)**

6.5.3 Discussion

6.5.3.1 Lengthening

The results of the current study show that the duration of both syllables in the disyllabic words are significantly lengthened in the CC tune, with the second syllable being lengthened by a much larger margin. The sources of the lengthening are two: one is the side product of the L% boundary tone; the other one is lengthening solely for the purpose of chanting and calling.

The evidence for the first source, i.e. the L% boundary tone, not only can be found in the duration section (Section 6.5.2.2, p. 214), but also can be gained from the analyses of F0 maxima alignment (Section 6.5.2.5, p. 229). As shown in Figure 6.38 (p. 231), the F0 maxima in T2 and T3 are aligned towards an earlier stage of the syllables with the CC tune than with the statement tune. The only explanation is that another tone was interpolated at the IP boundary.

The second source is from the tune itself. In the discussion in the IntQ chapter (Section 4.4.3.1, p. 93), as well as the discussion of the results for the monosyllabic CC tune (Section 6.4.3.1, p. 196), the notion of boundary-induced lengthening has been mentioned, with the boundary-induced lengthening and the lengthening in African languages compared in the IntQ discussion, and the differences between the boundary-induced lengthening versus the “sustainment” in a tune compared in the monosyllabic CC tune discussed. I address this issue again here, since this is one of the key points for delineating the principles that we should follow to determine whether lengthening is a component of a specific tune or not. The type of boundary-induced lengthening in question is a side product of the boundary tone, regardless of whether it is a floating

boundary tone or an actual boundary tone. To identify such lengthening, both production data and perception data are useful. From the perspective of production, the durational differences do not take up a huge percentage, and removing the temporal difference will not create a distinctively different tune. From the perspective of perception, native listeners cannot consciously identify the temporal difference, irrespective of whether they are capable of distinguishing them subconsciously or not. The IntQ tune in Tianjin Mandarin is a typical case of such boundary-induced lengthening. If the durational difference is of this type, then it should not be included in its tune.

However, if the temporal aspect serves the main purpose of forming a tune, then it should be included in the phonological description of the tune. For example, in the CC tune discussed in the current chapter, if it had the same length as a statement tune, the H* pitch accent would not sustain and thus perceptually could be regarded as a separate tune that served a different function.

It is regrettable that AM theory does not involve any temporal expression, which is nevertheless much needed. Only in the CC tune alone, due to the lack of temporal feature notation system, at least four different methods (using the boundary tone sign “%” alone, using “0%”, labelling with the text “(sust)”, using boundary tone in conjunction with a phrase tone) have been used to depict the “sustainment” of the boundary tone, as reviewed in 6.2. Some languages change their intonation by means of changing the temporal aspect of the utterances. Rialland (2007) lists a number of African languages that use lengthening as the question marker. The integration of the temporal descriptors into AM framework also prevents missing important information

in depicting different pitch accents. For instance, Arvaniti and Garding (2007) found significant temporal difference between L+H* and L*+H. Although the association was successfully captured, the temporal difference was not included in the tune.

6.5.3.2 Register

Similar to all three previous production studies (two on IntQ tune and one on monosyllabic CC tune) in this thesis, the register change is the clearest difference between the tunes in question and their statement counterparts, and is directly detectable through visual inspection. The current study of the disyllabic chanted call tune again saw a raised register, based on phonetic evidence from the mean pitch, values of pitch maxima and minima. All of these properties were consistently higher in the CC tune than in the statement tune.

In Set 1 data, the second syllables, which carry the lexical prominence, were higher than the first syllables on average. In Set 2 data, since the second syllables were all T0, the first syllables carry the prominence. The first syllables were higher than the second syllables on average. In both sets of data, the maximum F0 was higher in the second syllables and the minimum F0 was lower in the second syllable, which means the range of the first syllables was much smaller than the second syllables.

The three following possibilities can be inferred from these results. The first one is that the register is lifted on the entire utterance, with the lexical prominence being raised the most. The second possibility is that there is a pitch accent aligned with the lexical prominence. Further discussion on the pitch accents will be made in the next section. The third possibility is that the lexical prominence itself creates a higher pitch. This is

possible since many studies have shown (e.g. Fry, 1955, 1958) that higher pitch is one of the properties of lexical stress. It is most likely to be a combination of all three possibilities that created the register difference between prominence and non-prominent part of the utterance.

The register is another feature that needs to be included in the transcription system of AM framework. For a contour tone language, the register is an important feature that distinguishes tunes both in production and in perception. Pitch excursion, which shows the extent of the Hs and Ls, is also desperately needed in the transcription of tunes. Not only the results of the current study highlight this need, other studies of non-tonal languages also suggest such absence. For instance, Sadat-Tehrani (2007) reported the only difference between a default chanted call and a chanted call in anger was that the H* of the anger CC tune was higher in the register (see Appendix 2 for graphic information).

6.5.3.3 Pitch Accent

In the study of monosyllabic CC tune, I discovered a H* pitch accent, whose association was yet to be ascertained. The current study provides new insight into the association. Based on visual inspection in Section 6.5.2.1 and the F0 maxima alignment results in Section 6.5.2.5, there is strong evidence for the association of H* pitch accent with the H of every lexical tone, including the high parts of the level lexical tones. In the case of T0, which does not have an underlying tone, the H* is only associated with the first syllable, but spreads to the second syllable, as illustrated in (47):

(47) Examples of H* association rules:



Comparing the results cross-linguistically, the results are directly in contrast with the CC tune in Thai, another contour tone language. The H* pitch accent of the CC tune in Tianjin Mandarin magnifies the contours of the lexical tones, while in Thai, the CC tune levels the lexical tones, as reviewed in Section 6.2 (p. 176): for example, a high lexical tone, which has a continuous rise and is followed by a fall, becomes a mid-high level tone in the CC tune.

6.5.3.4 Boundary Tone

As reported in the results section, the current study of the disyllabic chanted call tune confirms the observation of the L% boundary tone depicted in the previous study (Section 6.4.3.3). Clear evidence from visual inspection, together with the results of duration, F0 maxima alignment, and F0 minima alignment in particular, suggests that there is a L% boundary tone.

The discovery of the L% boundary tone in Tianjin Mandarin is noteworthy, in that it is an actual boundary tone, instead of a floating boundary tone as in the IntQ tune reported in Chapter 4. This shows that it is entirely possible to have an extra tone interpolated at the IP boundary in Tianjin Mandarin – even for rising lexical tones, a L% boundary tone can be attached to the end. Such boundary tones have already been reported in Cantonese (J. K.-Y. Ma et al., 2006), this thesis is the first to report such phenomenon in Tianjin Mandarin.

Although both L% and sustainment occur in the CC tune in Tianjin Mandarin as they do in the non-tonal languages reviewed in Section 6.2, the CC tune differs from those languages. In the reviewed languages, they had sustained boundary tone, rather than a lengthened pitch accent. As discussed in the lengthening section (6.5.3.1), the durational differences are not solely brought by the boundary tone in Tianjin Mandarin. Evidence from F0 extrema alignment can also serve as evidence that the boundary tone only took a small percentage of the whole tune. This suggests that even though a boundary tone is interpolated in this tonal language, the contours of the lexical tones have to be kept in their original form as much as possible. Thus, inserting a new boundary tone which takes a substantially large proportion of the tune is not likely.

6.5.4 Summary for Disyllabic Chanted Calls

In Section 6.5, the results of disyllabic chanted calls were reported in terms of visual inspection, duration, mean pitch, tonal scaling, and tonal alignment, and were subsequently discussed in terms of lengthening, register, pitch accent, and boundary tone. The quantifiable results are summarised in Table 6.14. In the table, the first row compares whether CC tune or the statement tune, is higher in the parameters in different columns (duration, mean F0, Maximum F0, and minimum F0): C represents the chanted call tune, while S stands for the statement tune. The second row compares which syllable of the disyllabic utterances produces higher values of the examined factors. The following rows are comparisons between the CC tune and statement tunes in the same syllables, and between the two syllables within the same tunes.

Table 6.14 Summary of Key Results

	Duration (ms)		Mean F0 (ERB)		MaxF0 (ERB)		MinF0 (ERB)	
	SET1	SET2	SET1	SET2	SET1	SET2	SET1	SET2
C vs S	C	C	C	C	C	C	C	C
Syll.1 vs Syll.2	2nd	2nd	2nd	1st	2nd	2nd	1st	1st
Syll.1: C-S	17.8	33.7	1.82	1.94	1.83	2.18	0.75	1.59
Syll.2: C-S	116.7	164.8	1.88	1.90	1.79	1.86	0.63	0.99
C: Syll.2-Syll.1	172.1	109.0	0.16	-0.25	0.56	-0.20	-1.40	-1.20
S: Syll.2-Syll.1	73.2	-22.1	0.10	-0.21	0.59	0.12	-1.28	-0.60

Based on all the results and discussions, the following process can be observed to apply to derive the CC tune of Tianjin Mandarin:

- a. Raise the register;
- b. Lengthen every syllable, with the biggest amount on the last syllable;
- c. Associate H* to every H in each lexical tone;
- d. Spread H* to syllable without lexical tone, if any;
- e. Attach L% at IP boundary.

Having studied the disyllabic chanted calls, the two questions raised at the beginning of Section 6.4 can now be answered.

Q1: Can the CC tune derived from the monosyllabic study be applied in utterances which consist of more than one syllable?

A1: Yes. The three major components – register lift, sustained H* pitch accent, and L% boundary tone all exist in disyllabic chanted calls. Moreover, disyllabic chanted calls provided more evidence for the association of the H* to the H of each lexical tone.

Q2: How does an intonation tune manifest on a toneless syllable?

A2: Despite having no underlying form, the T0 syllable is lengthened by a large margin, as in a full syllable. The H* pitch accent does not directly associate with the T0 but spreads to the T0 syllable. The L% is also attached to the end of the syllable, if T0 was the last syllable.

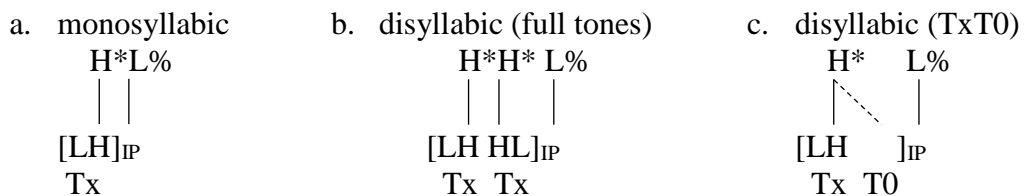
6.6 Conclusion

This chapter contains two production studies: one on the chanted call tune by examining monosyllabic utterances; and one on the same tune by examining disyllabic utterances. The three major findings of the chanted call tune are raised register, sustained H* which associates with the H of each lexical tone, and L%.

It goes through the following processes and can be analysed as (48):

- a. Raise the register;
- b. Lengthen every syllable, with the biggest amount on the last syllable;
- c. Associate H* to every H in each lexical tone;
- d. Spread H* to syllable without lexical tone, if any;
- e. Attach L% at IP boundary.

(48) Association of H* and L% in monosyllabic and disyllabic CC tunes:



Following the notation from the monosyllabic chanted call study in Section 6.4.4, the CC tune in Tianjin Mandarin is: $[[H^*]_{\text{sustained}}]_{\text{higher register}} + L\%$.

Chapter 7 Final Discussion and Conclusions

This chapter starts with a summary of the key findings from the previous chapters of this thesis (Section 7.1), followed by further discussions on some general issues (Section 7.2), including the applicability of AM theory in tonal languages, tone-tune interaction, as well as the necessity of keeping the notations in AM framework consistent. Section 7.3 points out some directions for future research. Finally, in Section 7.4, this thesis ends with a presentation of how this thesis managed to meet the three research goals (theoretical, documentation, and methodological) set out in the first chapters.

7.1 Summary of Key Findings

Chapter 4 presented two production studies of the intonational Yes/No question tune. Compared with the statement tune, the IntQ tune has a lifted register, and a H% boundary tone. The IntQ tune also has longer duration than its statement counterpart. However, the lengthening is likely to be boundary-induced rather than an actual part of the tune. In longer sentences, a H* is associated with the sentence prominence, i.e. focus, regardless of the tune, and a post-focus compression is also present in both the statement tune and the IntQ tune.

Chapter 5 reported three perception experiments of the identification of the tunes of intonational Yes/No question and statements. The results show that native listeners utilise both the register information and the boundary information – for statements, the L register information is used, and for IntQs, the H boundary tone is sought.

Chapter 6 investigated two production studies of the tune of the chanted calls. The CC tune again has a much higher register – even higher than the IntQ tune. The CC tune also has a sustained H* associated with the H part of each lexical tone, followed by a L% boundary tone. Similar to that in the production of the IntQ tune, the results for the CC tune also exhibited lifted register and lengthening. The lengthening in this tune, however, is not merely induced by the boundary. It is an essential part of the tune, since part of the tune is reliant on the sustainment of the pitch. It is therefore noteworthy that the AM framework needs more notation devices for temporal and register changes, which are phonologically meaningful intonational tune components.

7.2 Further General Discussion

7.2.1 Application of AM Theory in a Tonal Language

One of the central goals of this thesis is to apply AM theory to a tonal language, especially a contour tone language, since studies of intonation in contour tone languages are still very rare. The rarity can be ascribed to the fact that many aspects of AM theory sound entirely impossible in a contour tone language; for example, “sequential tonal structure” alone is shocking in the first instance. In this section, the relationship between AM theory and contour tone languages is scrutinised from the perspectives of the four tenets of AM theory, drawing on the analyses in the studies presented in this thesis. The four tenets (Ladd, 2008) are repeated here, in the order of discussion in the current section.

- (49) Four tenets of AM theory:
- (a) Distinction between pitch accent and stress;
 - (b) Sequential tonal structure;
 - (c) Analysis of pitch accents in terms of level tones;
 - (d) Local sources for global trends.

Distinction between pitch accent and stress

Stress, as defined in Section 2.3.2.2 Chapter 2, refers to the lexical prosody that provides a phonologically deducible docking point for a potential high pitch to be associated with, and thus becomes the lexical prominence of a word; sentence prominence consequently can be associated with it. Pitch accent is the sentence-level pitch component that is associated with the sentence prominence. Since stress is an abstract phonological slot, and pitch accent is an actual pitch performance, the distinction between stress and pitch accent is straightforward. The real distinction this tenet is trying to draw is between lexical prosody and sentence prosody, i.e. between the phonetic realisation of the lexical stress and sentence-level pitch accent. In a tonal language, this aspect manifests much more transparently: lexical prosody – including lexical tones, tonal alterations (including but not limited to tone sandhi), and lexical prominence (or more commonly, stress) – needs to be separated from the sentence prosody. Separating lexical tones from sentence prosody seems to be presumed; however, the following tenet of AM theory – the sequential tonal structure – often creates the impression that all pitch performances should be taken as one linear string of tones.

Sequential tonal structure

This tenet refers to the fact that AM theory only addresses the phonologically meaningful “events” and ignores the “transitions”. It emphasises the non-hierarchical

structure of the intonational tunes at the same time. In an intonation language, this tenet is intuitive and easy to abide by, since the only pitch modulation is for intonation and its related features. Yet, in a tonal language like Tianjin Mandarin, describing the pitch contours with sequences of tones does not seem to be sensible, since many aspects of prosody, including lexical tones, tonal alterations, lexical prominence, sentential prominence, etc., all manifest through pitch. For example, in Tianjin Mandarin, it would be impossible to have one intonation if the pitch for lexical tones is not excluded from the intonation. A L tone would then have a L-L% boundary for a polar question, and a H tone would have a H-H% intonation. An analysis like this is neither representative nor predictive, and does not provide us with any insight into how a question tune is represented in the brain at all.

In order to solve this paradox, the first tenet I discussed above about the distinction between lexical prosody and sentential prosody then plays its role: lexical prosody is to be considered separately from sentence prosody. Lexical tones therefore are not limited by this tenet and can keep their original descriptions, regardless of whether they are described with their contours (e.g. rising tone, falling tone) or level tones (L, H, M, etc.). The linearity only refers to the structure of sentence-level prosody, including focus-induced pitch accent, boundary tones, etc. This way, the lexical tones can be fully described according to the needs of the specific languages, while the intonational tunes still only keep the phonologically meaningful events, in the form of sequences of tones.

With regards to teasing apart the lexical tones from sentence-level prosody, this thesis has provided an illustration of how to exclude lexical tones from sentential prosody in the methods sections of Chapter 4 and Chapter 6: by comparing the F0 contours of

citation tones and altered tones in different tune types through mean pitch, F0 range, tonal scaling, tonal alignment, and so on. Although the current methods are sufficient for the phonological analysis of the tunes, it is beneficial if more precisely defined phonetic correlates can be acquired through future studies.

Analysis of pitch accents in terms of level tones

This tenet limits the description of intonational tunes to two level tones: H and L. In order to discuss whether this tenet holds valid in a tonal language, two separate questions need to be asked – are two tones enough in tonal languages? Are two tones enough in any languages?

The answer of the first question needs to be based on the condition that the two-tone system is sufficient for the languages that have already been studied and reported. These languages consist of intonational languages mainly, some pitch-accent languages, and a few tonal languages. If it is true that in these languages, the two-tone system is sufficient for the description of all their tunes, then there is no reason why it would not be sufficient in a tonal language. In fact, since tonal languages, especially contour tone languages, have assigned many roles for pitch (lexical stress, lexical tone, focus marking, etc.), other uses of pitch are very limited. The fact that many people believe pitch cannot take up so many roles in one language also leads to this misconception that “tonal languages cannot have intonation”. Although tonal languages do have intonation, it is not surprising how much more restricted than non-tonal languages they are. Therefore, if two tones are enough to describe intonational tunes in non-tonal languages, it should be enough in tonal languages too. Nevertheless, human languages are full of unexpected phenomena, therefore it still holds possible for a tonal language that has yet

to be studied to have richer intonation than any non-tonal language, since the speakers are probably more sensitive to every aspect of pitch. So far, based on our current understanding of intonation in tonal languages, AM theory's two-level system can be applied and is sufficient in all known tonal languages.

We shall return to see whether the premise of the above discussion is valid and pose the question: are two levels adequate in intonational languages, or in the languages that have been studied under the AM framework? The short answer would be that this coin has two sides. Take the chanted call tune as an example. In many languages, a downstepped !H is used to transcribe a tone that is neither very high nor very low. On the one hand, the merit of this notation is that a value can be assigned to the H tone as the highest level in a speaker's range, and !H can change accordingly. Imagine that in the future some form of phonetic modelling or speech synthesis needs to be done based on the phonological description of a certain pitch contour, a !H% is clearly a dependent value on the H tone. If another level, say M tone, is used in the notation, then the value of the M tone needs to be reasoned and assigned separately and with more caution. The point is that H% and M are essentially different in nature although the phonetic values may be the same. On the other hand, the drawback of this method is that in many analyses, although the downstep sign is used, there is no phonologically downstepping process involved. The downstepping sign is only used to indicate that this new tone is not as high as the highest tone in a tune. The motivation for the two-tone system was partially to prevent the excessive use of many levels, and two is the minimal number that are proven to be useful. If an extra level is indeed needed, then it should be used instead of using other expressions that may be confused with other phonological

processes. The consistency of notations in AM theory will be further discussed in Section 7.2.3.

Local sources for global trends

This tenet is mainly about how to decide on the height of the pitch accents. This part is not disputable since it is the only way that tonal sequences can be abstracted, and many previous theoretical works have explained the motivation of the local sources (see Ladd, 2008; Pierrehumbert, 1980). The real question is, does it mean global trend is not important at all? Admittedly, AM theory never completely denies global trend – it believes that if a global trend is phonologically meaningful, then it should be able to be analysed as an upstep or a downstep; if it is a pure phonetic performance such as declination, then it is not noteworthy. It is in a way undeniable, but, at the same time, insufficient. The production studies in this thesis showed clear global register change when there is a tune change. It is also perceptually important since, in the perception study, register serves as the cue for statement identification. However, because of the neglect of global trend, register change is not represented in AM theory. Therefore, although AM theory is theoretically sound, it needs further discussions and development in terms of the global trends.

With tonal languages in consideration, I adjust the four tenets as follows:

- (50) Four tenets of AM theory:
- (a) Distinction between sentence prosody and lexical prosody;
 - (b) Sequential tonal structure of sentence prosody;
 - (c) Analysis of pitch accents in terms of level tones;
 - (d) Local sources for global trends first; if insufficient, then phonologically meaningful global trends.

7.2.2 Tone-Tune Interaction

A central question of this thesis is how to tease apart tone and intonation. As discussed in the previous section, the key is to distinguish lexical prosody from sentence prosody. Lexical tone and lexical prominence evidently belong to lexical prosody. Tonal alteration is also lexical, though it is not specified in the mental lexicon. Sentential prosody includes the focus of the sentence, i.e. sentence prominence, the pitch accent associated with it, and the boundary tones (including edge tones at ip boundary and IP boundary).

Another question that needs to be answered is, why can lexical tones be overridden in some tunes (e.g. CC tune), while they cannot in some others (e.g. IntQ tune). One plausible explanation is that the major function of the CC tune is to arouse the attention of the addressee. The following factors – people are more sensitive to their own names; names are comparatively unique strings of syllables; kinship terms are high-frequency words – may jointly decide that it is not extremely important to keep the contours of the lexical tones. However, in an intonational Yes/No question, whether it is to seek information or to gain confirmation, it is important for the lexical information to be kept so that the meaning can get across to the listener. A similar phenomenon has been reported in Thai (Luksaneeyanawin, 1983): the rising and falling lexical tones become levelled in extreme cases where the meaning of the words are not important, such as professional callings or recitations.

7.2.3 Consistency of Terms and Notation in AM Framework

In this thesis, I proposed the usage of “floating boundary tone” for representing the type of boundary tone that (1) is not an interpolation of an extra tone; (2) makes a functional

difference in the sentence level tune; (3) has no phonetic realisation itself but modifies neighbouring tones. I also pointed out the lack of notation devices for temporal representation and register change in the AM framework. The misuse of !H in CC tune when it is not a downstep was discussed as well. These all seem to be trivial notational issues that do not affect the nature of research, and emphasising such issues seems to be pedantic. I nonetheless regard this matter as an important issue that has roots in the nature of linguistics studies. Similar to mathematics, linguistics tries to use a collection of formal methods to analyse languages. To realise, or at least jointly contribute to, this goal, we need to make collective efforts of keeping the notations consistent

7.3 Future Research

For the production study of IntQ tune, there are two future studies that can be done. The first one is to investigate the phonological analysis of post-focus compression by way of alternating the last lexical tone of the carrier sentence. Further examination of the interaction between lexical tone and post-focus compression can provide further information on how the compression occurs – is it caused by the deletion of a part of the lexical tone, or the addition of a new tone? Moreover, by changing the last lexical tone, stronger evidence can also be provided for the floating H% boundary tone analysis. The second study is to examine disyllabic words. In disyllabic words, different types of lexical prominence can be designed, in order to investigate how the lexical tones, lexical prominence, and IntQ tune interact. If the disyllabic words are also placed in the sentence prominence position, an extra element of sentence prominence can also be added.

The perception study in the current thesis is an initial exploration of how people perceive and identify intonational tunes in a tonal language. Our current understanding of how people perceive and retrieve intonational tunes are extremely limited – not only in tonal languages, but in all languages. Further investigations on all aspects of the perception, encoding, retrieval of intonational tunes are much needed.

To further investigate the production of the tune of chanted calls, new materials with extra neutral-tone syllables can be studied. For example, in (51), calling someone who is named *xiao zhang zi*, with an exclamation particle at the end, the last two syllables both have neutral tones. How are the L% boundary tone and the H* pitch accent associate with the text in such cases?

- (51) 小 张 子 哟!
- xiao zhang zi yo!
- T3 T1 T0 T0

7.4 Concluding Remarks

The three research goals set out in the first chapter were the theoretical goal, the documentation goal, and the methodological goal. This section summarises how these three goals are met.

This chapter has contributed to the AM theory with data and analyses from a tonal language, Tianjin Mandarin. The results of Chapter 4, Chapter 5, and Chapter 6 have provided empirical bases for the new notations devices that are needed in the AM framework, i.e. the representations for floating boundary tone, register change, and temporal change. Together with the discussion in Section 7.2.1 and Section 7.2.3, this

thesis discussed how AM theory is applicable in tonal languages, the improvements that are needed, and the necessity for the improvements.

The documentation goal is successfully reached since the tunes are successfully analysed, as presented in this thesis. The tune for intonation Yes/No questions is:

[H* pitch accent + (post-focus compression) + floating H% boundary tone] higher register ;

And the tune for the chanted calls is:

[[H*]_{sustained}]_{higher register} + L%.

Native Tianjin speakers distinguish the IntQ tune from the statement tune using both the register information and the boundary information. The register and boundary of the lexical tones thus become distractors during the identification.

This thesis employed a laboratory phonology approach throughout, integrating phonology and phonetics, in order to meet the methodology goal. The phonological analyses in the current thesis are all based on phonetic data. The approaches used in this thesis can be taken as examples of how to tease apart intonation and lexical tones in a contour tone language.

All in all, the findings in this thesis enrich our knowledge of prosodic phonology with concrete evidence and abstract interpretations from a tonal language, Tianjin Mandarin. To fully understand the intricate interaction between tones and tunes in both production and perception, further investigations are needed in Tianjin Mandarin and other tonal languages alike.

References

- Abe, I. (1962). Call-contours. In *Proceedings of the Fourth International Congress of Phonetic Sciences. The Hague: Mouton* (pp. 519–523).
- Abe, I. (1998). Intonation in Japanese. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 363–378). Cambridge, UK: Cambridge University Press.
- Armstrong, M. E. (2012). *The development of yes-no question intonation in Puerto Rican Spanish* (Doctoral thesis). The Ohio State University.
- Arvaniti, A. (2009). Greek intonation and the phonology of prosody: polar questions revisited. *Proceedings of the 8th International Conference on Greek Linguistics*, 14–29.
- Arvaniti, A. (2011). Representation of Intonation. In *Blackwell Companion to Phonology* (pp. 757–780). Blackwell Publishing.
- Arvaniti, A., & Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 1–52). Oxford: Oxford University Press.
- Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. In J. Cole & J. H. Hualde (Eds.), *Papers in Laboratory Phonology 9* (pp. 547–576). Berlin, New York: Mouton de Gruyter.
- Arvaniti, A., & Ladd, D. R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, 26(01), 43. <https://doi.org/10.1017/S0952675709001717>
- Arvaniti, A., & Ladd, D. R. (2015). Underspecification in intonation revisited: A reply

- to Xu, Lee, Prom- on and Liu. *Phonology*, 32(3), 537–541.
<https://doi.org/10.1017/S0952675715000305>
- Arvaniti, A., Żygis, M., & Jaskuła, M. (2016). The Phonetics and Phonology of the Polish Calling Melodies. In *Phonetica* (pp. 338–361).
<https://doi.org/10.1159/000446001>
- Bao, Z. (2003). Tone, accent and stress in Chinese. *Journal of Linguistics*, 39(1), 147–166. <https://doi.org/10.1017/S0022226702001962>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software; Vol 1, Issue 1 (2015)* .
- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.
- Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer. Version 5.3.51. Retrieved September 3, 2014, from <http://www.praat.org/>
- Bolinger, D. (1978). Intonation across languages. *Universals of Human Language*, 2, 471–524.
- Borràs-Comes, J., Sichel-Bazin, R., & Prieto, P. (2015). Vocative Intonation Preferences are Sensitive to Politeness Factors. *Language and Speech*, 58(1), 68–83. <https://doi.org/10.1177/0023830914565441>
- Bruce, G. (1982). Textual Aspects of Prosody in Swedish. *Phonetica*, 39, 274–287.
<https://doi.org/10.1017/CBO9781107415324.004>
- Bruce, G., & Gårding, E. (1978). A prosodic typology for Swedish dialects. *Nordic*

Prosody, 1.

Cahill, M. C. (2013). Polar question intonation in five Ghanaian languages. In *LSA Annual Meeting Extended Abstracts* (Vol. 4).

Cangemi, F., & D'Imperio, M. (2013). Tempo and the perception of sentence modality. *Laboratory Phonology, 4*(1), 191–219. <https://doi.org/10.1515/lp-2013-0008>

Cao, J. (2001). Hanyu yunlv qiefen de yuyinxue he yuyanxue xiansuo [Phonetic and linguistic clues in Chinese prosodic segmentation and grouping]. In *Modern Phonetics in the New Century: Proceeding of the Fifth National Modern Phonetics Conference*. Beijing, China: Tsinghua University Press.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique, 45*, 24–27.

Chao, Y. R. (1933). Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology, 121–134*.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Univ of California Press.

Chen, M. Y. (1979). Metrical structure: evidence from Chinese poetry. *Linguistic Inquiry, 10*(3), 371–420.

Chen, M. Y. (1986). The paradox of Tianjin tone sandhi. *CLS, 22*(98), 114.

Chen, M. Y. (1987). Introductory remarks to a symposium on Tianjin tone sandhi. *Journal of Chinese Linguistics, 15*(2), 203–227.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge University Press Cambridge, UK:

Chen, Y. (2010). Post-focus F0 compression—Now you see it, now you don't. *Journal*

of Phonetics, 38(4), 517–525. <https://doi.org/10.1016/j.wocn.2010.06.004>

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 36(4), 724–746. <https://doi.org/10.1016/j.wocn.2008.06.003>

Cohen, A., & 't Hart, J. (1965). *Perceptual analysis of intonation patterns*. Instituut voor Perceptie Onderzoek.

Cui, J., & Li, Y. (1995). Tianjinhua de yuliu yinbian [Phonological rules in Tianjin Mandarin]. *Zhongguo Yuwen*, 1, 36–44.

Daly, N., & Warren, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics*, 5(1), 85–96. <https://doi.org/10.1111/1467-9481.00139>

Davison, D. S. (1991a). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics*, 50–57.

Davison, D. S. (1991b). Stress and tonal targets in Tianjin Mandarin. *UCLA Working Papers in Phonetics*, 78, 58–77.

Dilley, L. C., & Brown, M. (2007). Effects of pitch range variation on f₀ extrema in an imitation task. *Journal of Phonetics*, 35(4), 523–551. <https://doi.org/10.1016/j.wocn.2007.01.003>

Duan, W., & Jia, Y. (2014). The typology of focus realization of Northern Mandarin. *Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium On*. <https://doi.org/10.1109/ISCSLP.2014.6936589>

- Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages* (Doctoral thesis). MIT.
- Duanmu, S. (2000). *The Phonology of Standard Chinese* (2nd ed.). Oxford University Press.
- Duanmu, S. (2011). Chinese syllable structure. In M. van Oostendorp, E. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (Vol. 5, pp. 2151–2777). Malden, MA & Oxford: Wiley-Blackwell.
- Dung, Đ. T., Huong, T. T., & Boulakia, G. (1998). Intonation in Vietnamese. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 398–420). Cambridge, UK: Cambridge University Press.
- Fagyal, Z. (1997). Chanting intonation in French. *U. Penn Working Papers in Linguistics*, 4(2), 77–90.
- Feng, H. (2015). *Chengshihua jinchengzhong chengshi fangyan bianyi tezheng yanjiu: yi tianjin fangyan weili* [On linguistic variation in urban dialects during urbanisation: A case study of Tianjin Mandarin]. Tianjin University Press.
- Feng, S. (1995). *Prosodic structure and prosodically constrained syntax in Chinese*. University of Pennsylvania, Ann Arbor.
- Feng, S. (1996). Lun Hanyu de “Yunlvci” [On Mandarin prosodic word]. *Zhongguo Shehui Kexue*, 1, 161–176.
- Feng, S. (1997). *Interactions between morphology syntax and prosody in chinese*. Beijing, China: Peking University Press.

- Feng, S. (1998). Lun Hanyu de “Ziran Yinbu”[On “Natural Feet” in Mandarin]. *Zhongguo Yuwen*, (1), 40–47.
- Flemming, E., & Nie, H. (2016). Are there boundary tones in Mandarin Chinese echo questions? In *Linguistics Association of Great Britain Annual Meeting 2016*. York, UK: University of York.
- Frota, S. (2014). The intonational phonology of European Portuguese. In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 6–42). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof>
- Frota, S., Cruz, M., Svartman, F., Collischonn, G., Fonseca, A., Serra, C., ... Vigário, M. (2015). Intonational variation in Portuguese: European and Brazilian varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 235–283). Oxford: Oxford University Press.
- Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, 27(4), 765–768.
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(2), 126–152.
- Gårding, E. (1984). Chinese and Swedish in a generative model of intonation. *Nordic Prosody*, 111, 79–91.
- Gårding, E. (1987). Speech Act and Tonal Pattern in Standard Chinese: Constancy and Variation. *Phonetica*, 44(1), 13–29.
- Gibbon, D. (1976). *Perspectives of intonation analysis* (Vol. 9). Bern: Lang.

- Goldsmith, J. A. (1976). *Autosegmental phonology* (Doctoral thesis). MIT.
- Grice, M., Baumann, S., & Benzmüller, R. (2005a). German Intonation in Autosegmental-Metrical Phonology. In *Prosodic Typology*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199249633.003.0003>
- Grice, M., Baumann, S., & Benzmüller, R. (2005b). German Intonation in Autosegmental Metrical Phonology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 55–83). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof>
- Grice, M., Roettger, T. B., Ridouane, R., & Fougeron, C. (2011). The association of tones in Tashlhiyt Berber. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)* (pp. 775–778). Hong Kong.
- Grice, M., & Savino, M. (2003). Question type and information structure in Italian. *Proceedings of Prosodic Interface 2003*.
- Grønnum, N. (1998). Intonation in Danish. In *Intonation Systems: A Survey of Twenty Languages* (pp. 131–151). Cambridge, UK: Cambridge University Press.
- Gu, X. (2008). Tianjinhua shengmu jiqi yanbian [Consonants and their evolution in Tianjin Mandarin]. *Yuyan Yingyong Yanjiu*, 111–113.
- Guo, S. (1938). Zhongguo yuci zhi tanxing zuoyong [The elasticity of Chinese words]. *Yanjing Xuebao*, 24.
- Gussenhoven, C. (1983). Testing the Reality of Focus Domains. *Language and Speech*, 26(1), 61–80. <https://doi.org/10.1177/002383098302600104>

- Gussenhoven, C. (1993). The Dutch foot and the chanted call. *Journal of Linguistics*, 29(1), 37–63. <https://doi.org/10.1017/S0022226700000049>
- Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and Phonology. In *Speech Prosody 2002: Proceedings of the First International Conference on Speech Prosody* (pp. 47–57).
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, C., Rietveld, T., Kerkhoff, J., & Terken, J. (2012). ToDI second edition. Retrieved December 31, 2017, from <http://todi.let.kun.nl/ToDI/home.htm>
- Hayes, B., & Lahiri, A. (1991). Bengali intonational phonology. *Natural Language & Linguistic Theory*, 47–96.
- Hayes, B., & Lahiri, A. (1992). Durationally specified intonation in English and Bengali. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Proceedings of the 1990 Wenner-Gren Center Conference on Music, Language, Speech and the Brain* (pp. 78–91). Wenner-Gren Center International Symposium Series. Palgrave, London.
- Heeren, W. F. L., Bibyk, S. A., Gunlogson, C., & Tanenhaus, M. K. (2015). Asking or Telling – Real-time Processing of Prosodically Distinguished Questions and Statements. *Language and Speech*, 58(4), 474–501. <https://doi.org/10.1177/0023830914564452>
- Hermans, D., & van Gestel, J. C. (1991). The frequency scale of speech intonation. *The Journal of the Acoustical Society of America*, 90(1), 97–102.
- Ho, A. T. (1977). Intonation Variation in a Mandarin Sentence for Three Expressions:

- Interrogative, Exclamatory and Declarative. *Phonetica*, 34(6), 446–457.
<https://doi.org/10.1159/000259916>
- Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, 23(02), 225–257.
<https://doi.org/10.1017/S0952675706000893>
- Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch-accent. *Language Sciences*, 31, 213–238. <https://doi.org/10.1016/j.langsci.2008.12.007>
- Hyman, L. M. (2010). Markedness and the phonological typology of two-height tone systems. *UC Berkeley Phonology Lab Annual Report*, 376–394.
https://doi.org/10.1163/9789004303218_010
- Hyman, L. M. (2011). The Representation of Tone. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology*.
- Hyman, L. M., & Tadadjeu, M. (1976). Floating tones in Mbam-Nkam. *Studies in Bantu Tonology: Southern California Occasional Papers in Linguistics*, (3), 59–111.
- Jansche, M. (1998). A two-level take on Tianjin tone. In *Proceedings of the Third ESSLLI Student Session, 10th European Summer School on Logic, Language and Information* (pp. 162–174). Saarbrücken, Germany.
- Jones, J., Louw, J., & Roux, J. C. (1998). Perceptual Experiments on Queclaratives in Xhosa. *South African Journal of Linguistics*, 16(sup36), 19–32.
<https://doi.org/10.1080/10118063.1998.9724400>
- Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody* (Doctoral thesis). The Ohio State University. <https://doi.org/10.16953/deusbed.74839>

- Jun, S.-A., & Fougeron, C. (2000). A Phonological Model of French Intonation. In A. Bontinis (Ed.), *Intonation: Analysis, Modeling and Technology*. (pp. 209–242). Dordrecht: Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-011-4317-2>
- Jun, S.-A., & Oh, M. (2000). Acquisition of second language intonation. In *INTERSPEECH* (pp. 73–76).
- Kennard, H. (2013). *Breton Morphosyntax in Two Generations of Speakers: Evidence from Word Order and Mutation* (Doctoral thesis). University of Oxford.
- Kochanski, G., & Shih, C. (2003). Prosody modeling with soft templates. *Speech Communication*, 39(3–4), 311–352. [https://doi.org/10.1016/S0167-6393\(02\)00047-X](https://doi.org/10.1016/S0167-6393(02)00047-X)
- Ladd, D. R. (1978). Stylized intonation. *Language*. <https://doi.org/10.1353/lan.1978.0056>
- Ladd, D. R. (1984). Declination.: a review and some hypotheses. *Phonology*, 1, 53–74. <https://doi.org/10.1017/S0952675700000294>
- Ladd, D. R. (1992). An introduction to intonational phonology. *Docherty, Ladd, Papers in Laboratory Phonology. II: Gesture, Segment, Prosody*, 321–334.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge, MA: Cambridge University Press.
- Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge University Press.
- Ladd, D. R., & Cutler, A. (1983). Introduction: Models and measurements in the study

of prosody. *Prosody, Models and Measurements*, 1–10.

Leben, W. R. (1973). *Suprasegmental phonology* (Doctoral thesis). Massachusetts Institute of Technology.

Leben, W. R. (2011). Autosegments. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *Blackwell Companion to Phonology*.

Lee, O. J. (2005). *The prosody of questions in Beijing Mandarin* (Doctoral thesis). Ohio State University, Columbus, Ohio.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185–199.
<https://doi.org/10.1017/S0305000900001598>

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Li, D., & Cheng, M. (2008). *Waiguo ren shiyong hanyu yufa [A practical Chinese Grammar for Foreigners]* (Revised ed). Beijing Language and Culture University Press.

Li, Q., & Chen, Y. (2012). Trisyllabic tone sandhi in Tianjin Mandarin. In *Proceedings of The Third International Symposium on Tonal Aspects of Languages*. Nanjing, China.

Li, Q., & Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics*, 54, 123–150.
<https://doi.org/10.1016/j.wocn.2015.10.002>

- Li, Q., Chen, Y., & Xiong, Z. (2017). Tianjin Mandarin. *Journal of the International Phonetic Association*, 1–20. <https://doi.org/10.1017/S0025100317000287>
- Li, Shiyu. (1956). Shuo tianjinhua de ren zenyang xuexi putonghua [How do Tianjin Mandarin speakers learn standard Mandarin]. *Zhongguo Yuwen*, 46(4), 24.
- Li, Shiyu, & Han, G. (1991). Luelun tianjin fangyandao [On Tianjin Dialect Island]. *Journal of Tianjin Normal University*, 2, 71–76.
- Li, Sijing. (1986). *Hanyu “er”yinshi yanjiu [A Study of the history of /ə/ in Chinese]*. Beijing: Commercial Press.
- Li, X., & Liu, S. (1985). Tianjin fangyan de liandu biandiao [The tone sandhi in Tianjin Dialect]. *Zhongguo Yuwen*, 1, 76–80.
- Liberman, M. Y. (1975). *The intonational system of English*. Massachusetts Institute of Technology.
- Liberman, M. Y., & Prince, A. S. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.
- Lin, H. (2008). Variable directional applications in Tianjin tone sandhi. *Journal of East Asian Linguistics*, 17(3), 181–226. <https://doi.org/10.1007/s10831-008-9024-x>
- Lin, M. (2004). On production and perception of boundary tone in Chinese intonation. In *Proceedings of the international symposium on tonal aspects of languages: with emphasis on tone languages* (pp. 125–130). Beijing.
- Lin, M. (2006). Yiwen he chenshu yuqi yu bianjiediao [Interrogative vs. declarative and the boundary tone in Standard Chinese]. *Zhongguo Yuwen*, (4).

- Lin, M., & Yan, J. (2000). Putonghua yunlvci he zhongyin wenti [Prosodic word and its stress in standard Chinese]. *Report of Phonetic Research*, 56–77.
- Lindau, M. (1986). Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America*, 80(3), 757–764. <https://doi.org/10.1121/1.393950>
- Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, 62(2–4), 70–87. <https://doi.org/10.1159/000090090>
- Luksaneeyanawin, S. (1983). *Intonation in Thai* (Doctoral thesis). University of Edinburgh.
- Luksaneeyanawin, S. (1998). Intonation in Thai. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 376–394). Cambridge, UK: Cambridge University Press.
- Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *Journal of the Acoustical Society of America*, 120(6), 3978–3987. <https://doi.org/10.1121/1.2363927>
- Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *Journal of the Acoustical Society of America*, 129(2), 1012–1023. <https://doi.org/10.1121/1.3531840>
- Ma, Q. (2005). Zailun “tianjinhua liandu biandiao zhi mi” [More on the paradox of Tianjin tone sandhi]. *Dangdai Yuyanxue*, 7(2), 297–106.
- Ma, Q. (2012). Bidirectional tone sandhi in Tianjin dialect: Problem and analysis. In *The Sixth International Conference on Speech Prosody*. Shanghai, China.

- Ma, Q., & Jia, Y. (2006). Tianjinhua shangsheng de liangtiao “biandiao guize” bianxi [Two New Third Tone Sandhi Rules in Tianjin Dialect: A Critical Re-analysis]. *Journal of Tianjin Normal University (Social Science)*, 1(1), 53–59.
- Makarova, V. (2001). Perceptual correlates of sentence-type intonation in Russian and Japanese. *Journal of Phonetics*, 29, 137–154. <https://doi.org/10.1006/jpho.2001.0137>
- Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers Association*, 24(3), 49–65.
- Muñiz Cachón, C., Rodríguez, R. G., Gómez, L. D., & Pedrero, M. A. (2006). Prosodia gallego-asturiana en enunciados SVO. *Revista de Filología Asturiana*, (6), 335–349.
- Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.
- Ni, J., & Kawai, H. (2004). Pitch targets anchor Chinese tone and intonation patterns. In *Speech Prosody* (pp. 95–98). Nara, Japan.
- Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2–5.
- Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, 40(1), 1–18.
- Ohala, J. J. (1984). An ethological perspective on common cross-Language utilization of F 0 of voice. *Phonetica*, 41(1), 1–16. <https://doi.org/10.1159/000261706>
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*

- (Doctoral thesis). MIT. Distributed 1988, Indiana University Linguistics Club.
- Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor, Michigan: University of Michigan Press.
- R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Reetz, H., & Kleinmann, A. (2008). Multi-subject hardware for experiment control and precise reaction time measurement. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1489–1492). Barcelona, Spain.
- Rialland, A. (2007). Question prosody: an African perspective. In T. Riad & C. Gussenhoven (Eds.), *Tones and tunes* (Vol. 1). Walter de Gruyter.
- Rialland, A. (2009). The African lax question prosody: Its realisation and geographical distribution. *Lingua*, *119*(6), 928–949.
<https://doi.org/10.1016/j.lingua.2007.09.014>
- Rialland, A., & Robert, S. (2001). The Intonational System of Wolof. *Linguistics*, *39*(375), 893–940.
- Roettger, T. B. (2017). *Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments*. Berlin: Language Science Press.
- Sadat-Tehrani, N. (2007). *The intonational grammar of Persian*. University of Manitoba.

- Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, 33(1), 3–11. <https://doi.org/10.1037/0012-1649.33.1.3>
- Selkirk, E. O. (1972). *The phrase phonology of English and French*. (Doctoral thesis). MIT.
- Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11, 563–605.
- Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, 371–405.
- Shen, J. (1985). Beijinghua shengdiao de yinyu he yudiao [The register and intonation of Beijing Mandarin]. In T. Lin & L. Wang (Eds.), *Beijing Yuyin Shiyanlu [Reports of Beijing Phonetic Experiments]* (Vol. 73, pp. 1–21). Beijing: Peking University Press.
- Shen, J. (1992). Hanyu yudiao moxing chuyi (On Chinese intonation model). *Yuwen Yanjiu*, 45, 16–24.
- Shen, J. (1994). Hanyu yudiao gouzao he yudiao leixing [The structure and types of Mandarin intonation]. *Fangyan*, (3), 221–228.
- Shen, X. S. (1990). *The Prosody of Mandarin Chinese* (Vol. 118). Univ of California Press.
- Shi, F. (2009). *Shiyan yinxixue tansuo [Exploration of Experimental Phonology]*. Beijing, China: Peking University Press.

- Shi, P. (1980). Sizhong juzi de yudiao bianhua [The intonational change of four types of sentences]. *Yuyan Jiaoxue Yu Yanjiu*, 2, 71–81.
- Shih, C. (1986). *The prosodic domain of tone sandhi in Chinese* (Doctoral thesis). University of California, San Diego.
- Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83–109.
- Shih, C. (1997). Declination in Mandarin. In *Intonation: Theory, Models and Applications. Proceedings of an ESCA Workshop*. (pp. 293–296). Athens, Greece.
- Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 243--268). Kluwer Academic Publishers.
- Tianjin Municipal Bureau of Statistics. (2010). Tianjinshi 2010nian diliuci quanguo renkou pucha zhuyao shuju gongbao [Major figures of the 6th National Census: Tianjin]. Retrieved January 5, 2018, from <http://www.stats-tj.gov.cn/Item/16734.aspx>
- Tianjin Municipal Bureau of Statistics. (2015). 2015nian Tianjinshi renkou zhuyao shuju gongbao [Major data from 2015 population census]. Retrieved January 5, 2018, from <http://www.stats-tj.gov.cn/Item/26015.aspx>
- Trager, G. L., & Smith, H. L. (1951). *An Outline of English Structure (Studies in Linguistics, Occasional Papers 3)*. Norman, Oklah: Battenburg Press.
- Tseng, C. (1981). *An acoustic phonetic study on tones in Mandarin Chinese*. Brown University.

- Van der Hulst, H. (2012). Deconstructing stress. *Lingua*, 122(13), 1494–1521.
<https://doi.org/10.1016/j.lingua.2012.08.011>
- Varga, L. (2008). The calling contour in Hungarian and English. *Phonology*, 25(03), 469–497. <https://doi.org/10.1017/S0952675708001607>
- Wang, A., Chen, M., & Lu, S. (2004). The study of declination in speech database in standard Chinese. *Shengxue Xuebao*, 29(4), 353–358.
- Wang, B. (2002). *The Research on Perception of Prosody in Mandarin* (Doctoral thesis). Institute of Psychology, China Academy of Science, Beijing, China.
- Wang, H. (2000). Hanyu de yunlvci yu yunlv duanyu [the prosodic word and the prosodic phrase of Chinese]. *Zhongguo Yuwen*, 6(10), 2007.
- Wang, L., Jiang, Z., & Tang, A. (2009). Guanyu tianjin fangyan yuyin yanbian de jige wenti de taolun — jianlun tianjin fangyan de yuanliu guanxi [The phonetic evolution of Tianjin dialect and the origin of the Tianjin dialect]. *Yuwen Yanjiu*, 3, 45–50.
- Wang, M., & Lin, M. (2003). Putonghua ziran huayu zhong de xiaqing [Declination in Chinese spontaneous speech]. *Report of Phonetic Research*, 71–75.
- Wang, P., & Shi, F. (2010). Hanyu Beijinghua yiwenju yudiao de qifu du (The degree of rising and falling of Beijing Mandarin questions). *Nankai Linguistics*, 2.
- Wang, Z., & Feng, S. (2006). Shengdiao duibifa yu beijinghua shuangyinzu de zhongyin leixing [Tonal Contrast and Disyllabic Stress Patterns in Beijing Mandarin]. *Yuyan Kexue*, 5(1), 3–22.

- Wee, L.-H. (2005). Elision and tone sandhi in the Tianjin dialect. In *Department of English Language and Literature Conference Paper* (pp. 422–431). Monterey, California: Defense Language Institute: National University of Singapore.
- Wee, L.-H., Yan, X., & Lu, J. (2005). Tianjin fangyan de tunyin xianxiang [Phenomena of medial syllable swallowing in Tianjin dialect]. *Yuyan Kexue*, 4(4), 66–75.
- Wells, R. S. (1945). The pitch phonemes of English. *Language*, 21(1), 27–39.
- Welmers, W. E. (1959). Tonemics, Morphotonemics, and Tonal Morphemes. *General Linguistics*, 4, 1–9.
- Williams, E. S. (1976). Underlying tone in Margi and Igbo. *Linguistic Inquiry*, 7(3), 463–484.
- Wu, L., & Wang, X. (2014). Luanping Fangyan Yuyin Xitong Diaocha Baogao [Investigation Report on Phonetic System of Luanping County Dialect -- On the Initial Consonant System]. *Hebei Minzu Shifan Xueyuan Xuebao*, 34(4), 56–61. <https://doi.org/10.16729/j.cnki.jhnun.2014.04.007>
- Wu, W. L., & Xu, Y. (2010). Prosodic Focus in Hong Kong Cantonese without Post-focus Compression. *Proceedings Speech Prosody 2010*, 1–4. <https://doi.org/10.1007/s10831-011-9087-y>
- Wurm, S. A., Li, R., Baumann, T., & Lee, M. W. (1987). *Language Atlas of China*. Longman.
- Wyunhe. (2011). Map of sinitic dialect - English version. Retrieved January 3, 2018, from https://commons.wikimedia.org/wiki/File:Map_of_sinitic_languages_full-en.svg

- Xu, B. R., & Mok, P. (2012). Cross-linguistic Perception of Intonation by Mandarin and Cantonese Listeners. In *Proceedings of Speech Prosody* (pp. 99–102). Shanghai.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f₀ contours. *Journal of Phonetics*, 27(1), 55–105. <https://doi.org/10.1006/jpho.1999.0086>
- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46, 220–251. <https://doi.org/10.1016/j.specom.2005.02.014>
- Xu, Y. (2015). Speech Prosody — Theories, models and analysis. In A. R. Meireles (Ed.), *Courses on Speech Prosody* (pp. 146–177). Cambridge Scholars Publishing.
- Xu, Y., Chen, S., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review*, 29(1), 131–147. <https://doi.org/10.1515/tlr-2012-0006>
- Xu, Y., Lee, A., Prom-on, S., & Liu, F. (2015). Explaining the PENTA model: A reply to Arvaniti & Ladd. *Phonology*, 32(3), 505–535. <https://doi.org/10.1017/S0952675715000299>
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33(2), 159–197. <https://doi.org/10.1016/j.wocn.2004.11.001>
- Yip, M. (1980). *The tonal phonology of Chinese*. MIT.
- Yip, M. (1989). Contour tones. *Phonology*, 6(01), 149–174. <https://doi.org/10.1017/S095267570000097X>

- Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, 130(1), 4063–4069. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Yuan, J. (2012). The effects of speaking rate and intonation on the duration of tones in Mandarin Chinese. *Proceedings of the 6th International Conference on Speech Prosody (SP2012)*, 532–534.
- Yuan, J., & Liberman, M. Y. (2010). F0 declination in English and Mandarin broadcast news speech. In *INTERSPEECH* (pp. 134–137).
- Yuan, J., & Shih, C. (2004). Confusability of Chinese intonation. In *Speech Prosody 2004, International Conference*.
- Zeng, Xiaoyu. (2010). Lun tianjinhua de yuanliu [On the source and course of Tianjin dialect]. *Nankai Linguistics*, 2, 1–13.
- Zeng, XL, Martin, P., & Boulakia, G. (2004). Tones and intonation in declarative and interrogative sentences in Mandarin. In *TAL-2004* (pp. 235-238.). Beijing, China.
- Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(01), 153–201. <https://doi.org/10.1017/S0952675710000060>
- Zhang, J., & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, 68(3), 161–191. <https://doi.org/10.1159/000333387>
- Zhang, X. (1987). Lun xiandai Tianjinhua li de chiyinzi [On dentalisation in Tianjin Mandarin]. *Tianjin Shangxueyuan Xuebao*, 3, 55. <https://doi.org/DOI:10.15963/j.cnki.cn12-1401/f.1987.03.009>

Zheng, B., Wang, B., Yang, Y., Lu, S., & Cao, J. (2000). The regular accent in Chinese sentences. In *ICSLP-2000* (pp. 86–89).

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1), 3–42.

Appendices

Appendix 1: Figures from Production Study 2: Monosyllabic Words in Sentence Prominence (Mono_(SEN))

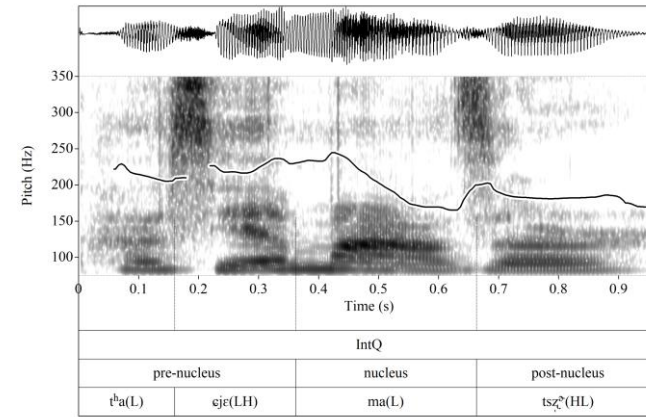
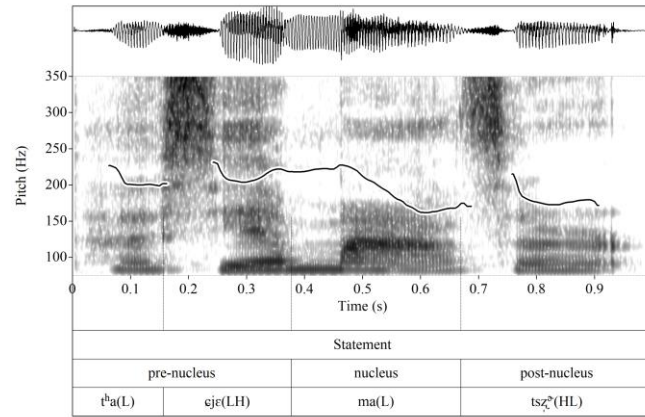


Figure A1 Statement (left) and IntQ (right): P3 + Tone 1 (L Tone)

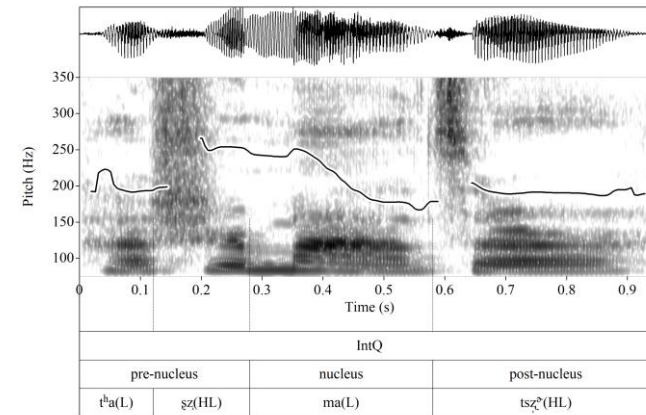
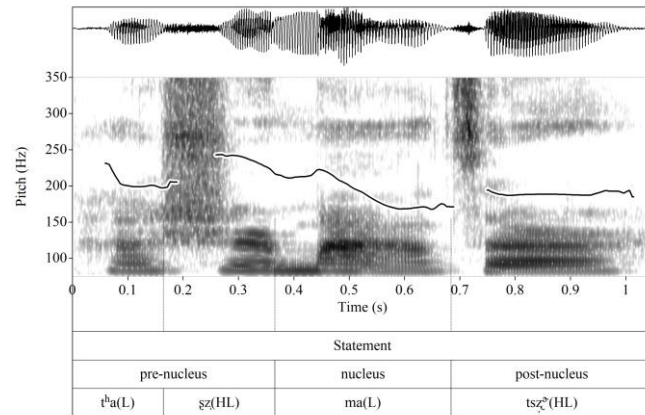


Figure A2 Statement (left) and IntQ (right): P4 + Tone 1 (L Tone)

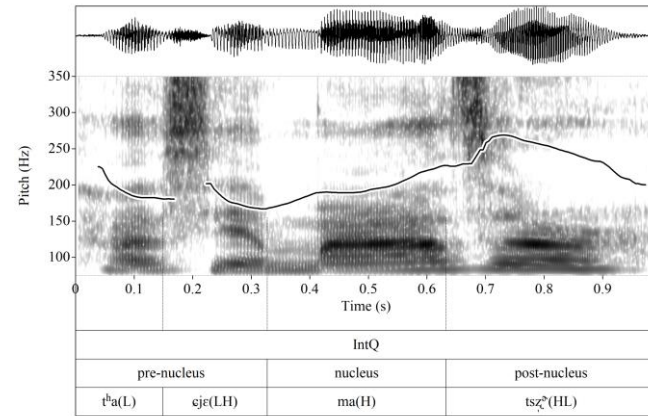
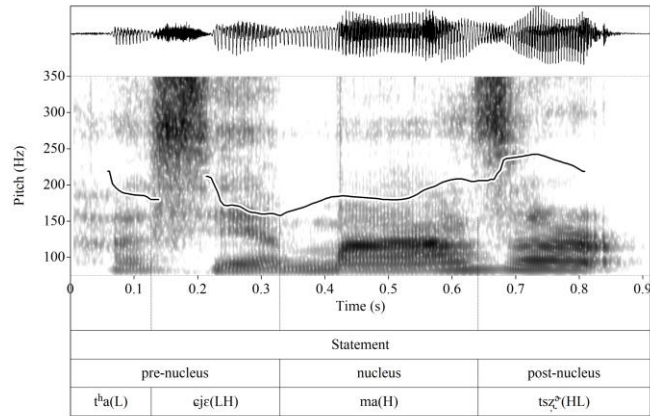


Figure A3 Statement (left) and IntQ (right): P3 – Tone 2 (H Tone)

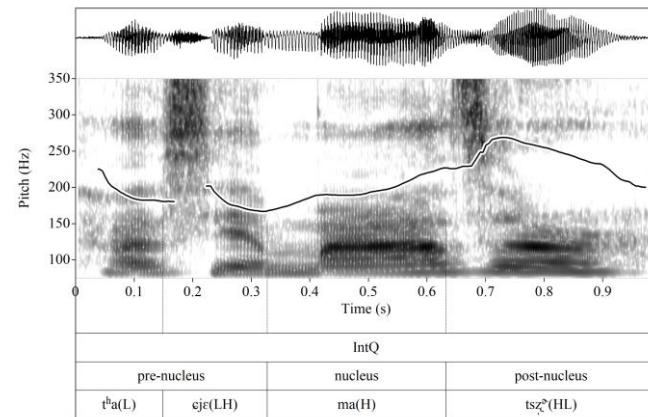
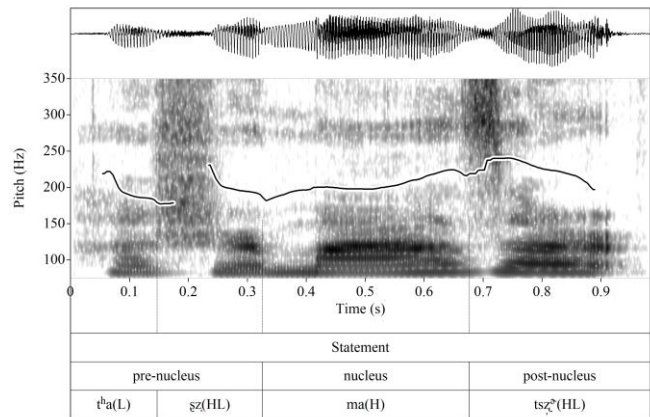


Figure A4 Statement (left) and IntQ (right): P4 – Tone 2 (H Tone)

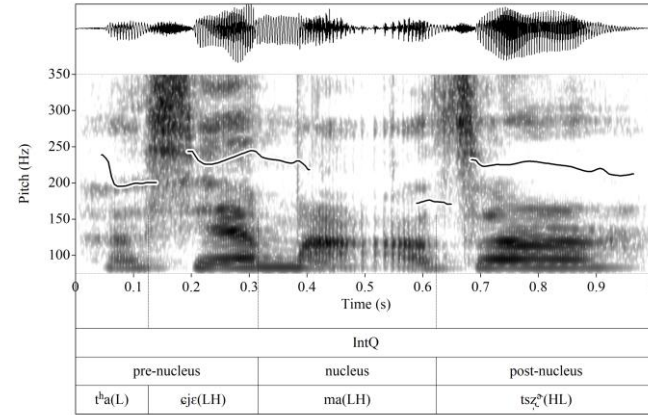
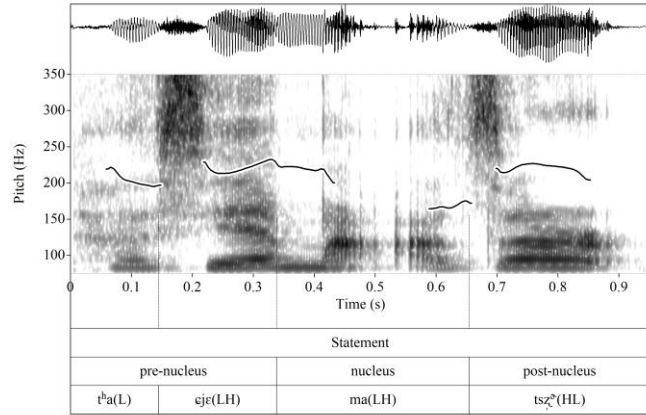


Figure A5 Statement (left) and IntQ (right): P3 – Tone 3 (LH Tone)

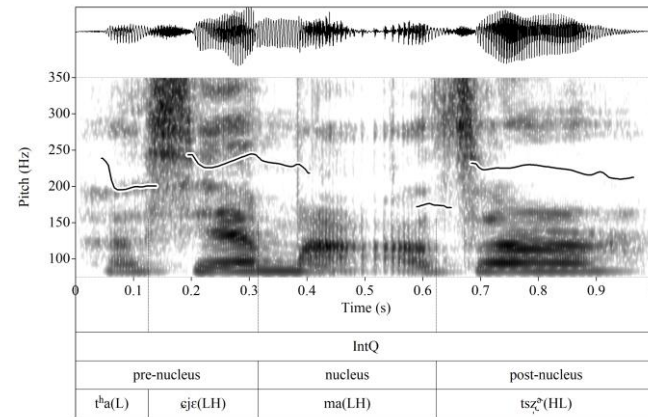
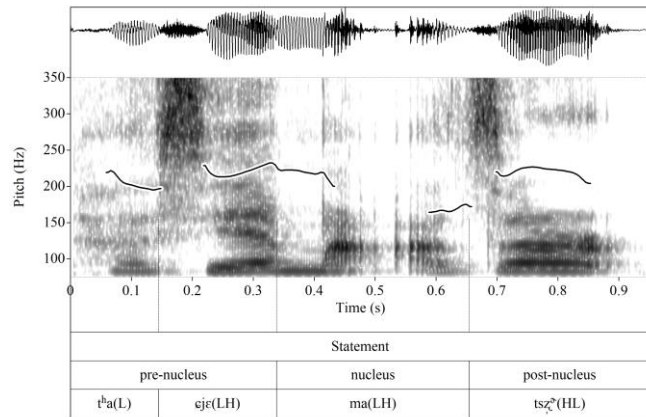


Figure A6 Statement (left) and IntQ (right): P4 – Tone 3 (LH Tone)

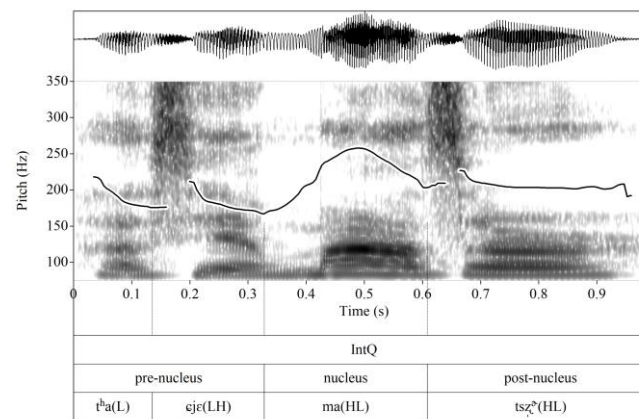
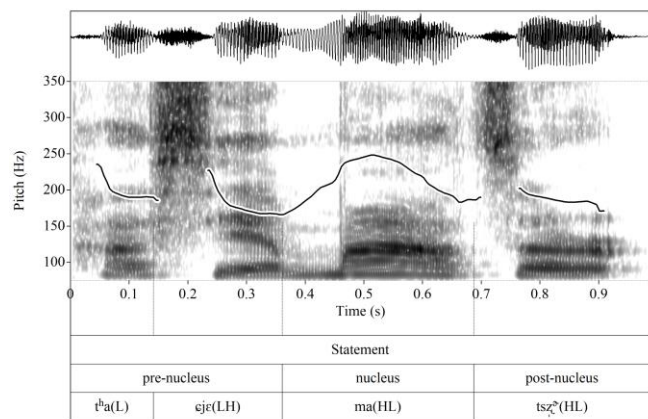


Figure A7 Statement (left) and IntQ (right): P3 – Tone 4 (HL Tone)

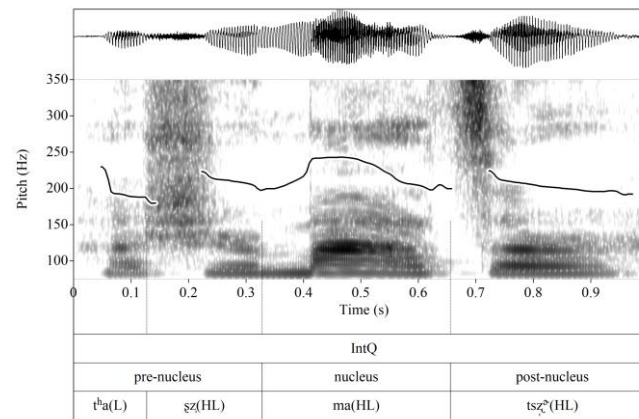
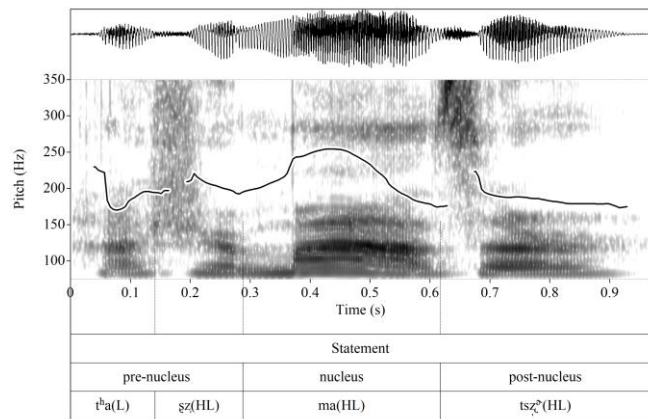
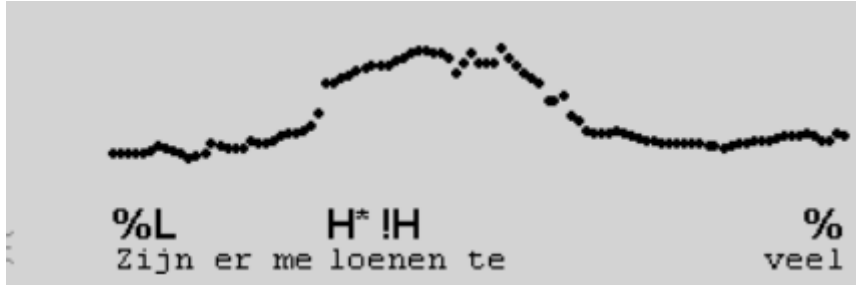



Figure A8 Statement (left) and IntQ (right): P4 – Tone 4 (HL Tone)

Appendix 2: Figures of Chanted Call Tunes in World Languages

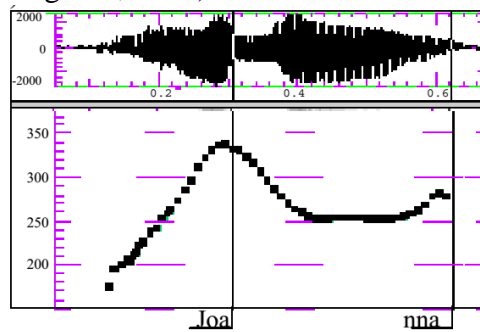
(1) Dutch (Gussenhoven, Rietveld, Kerkhoff, & Terken, 2012):



(2) German (Grice, Baumann, & Benzmueller, 2005):

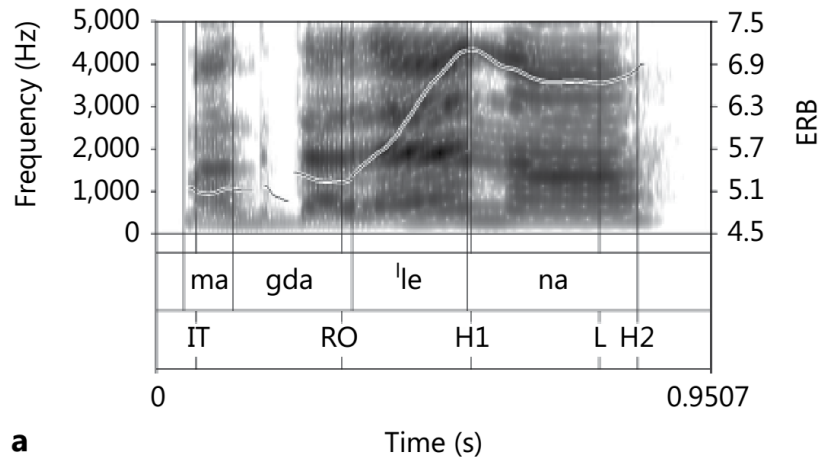
Stylised Step Down	(L+)H* !H-%		Calling contour	BECKenBAUer!
--------------------------	----------------	--	-----------------	---------------------

(3) French (Jun & Fougeron, 2000):

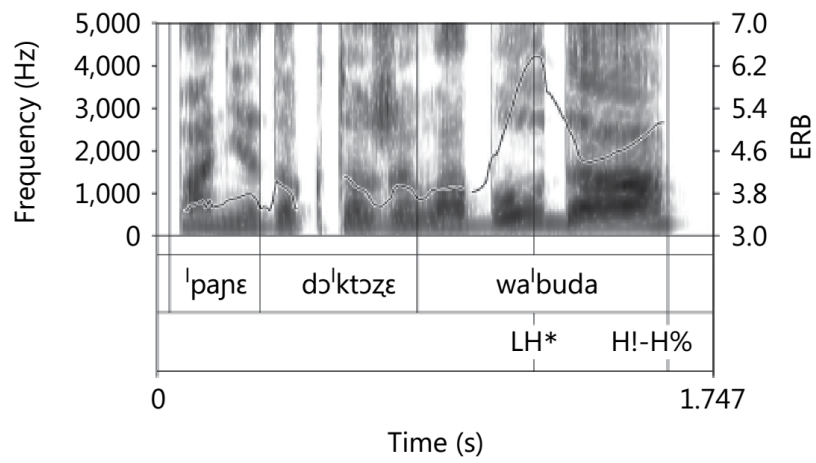


(4) Polish (Arvaniti, Żygis, & Jaskuła, 2016)

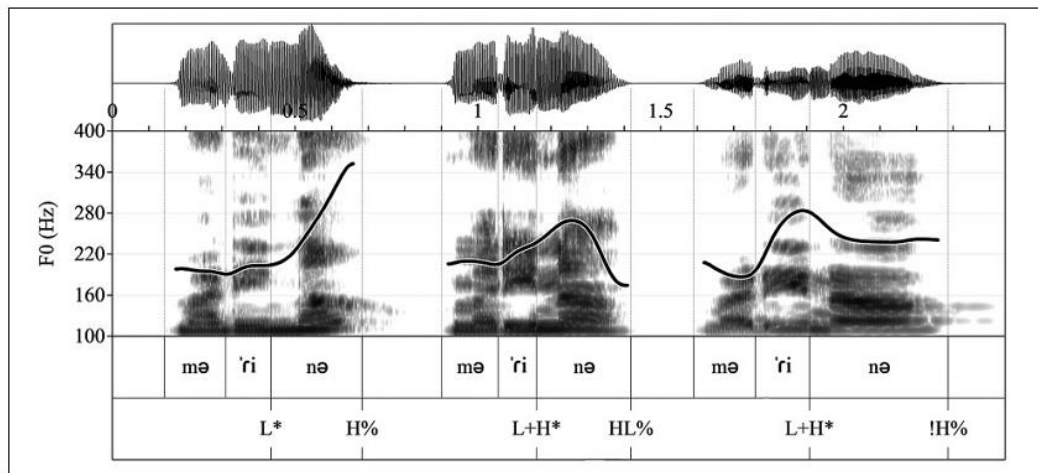
Routine Call:



Urgent Call:



(5) Catalan (Borràs-Comes, Sichel-Bazin, & Prieto, 2015)



(6) Portuguese (Frota, 2014):

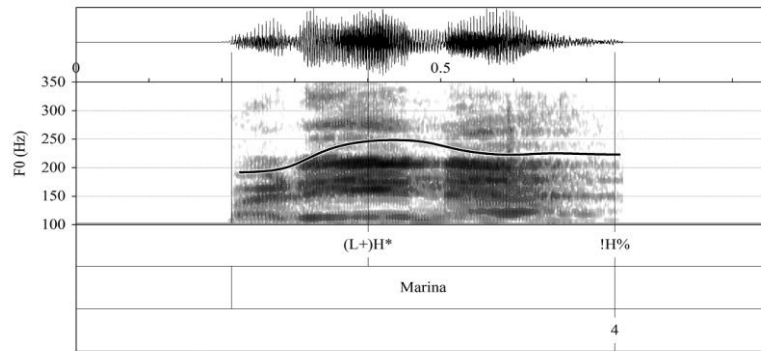


FIG. 7.35 Waveform, spectrogram, and F0 contour of the greeting call *Marina*, produced by a speaker of POR (EP, Northern variety, coast)

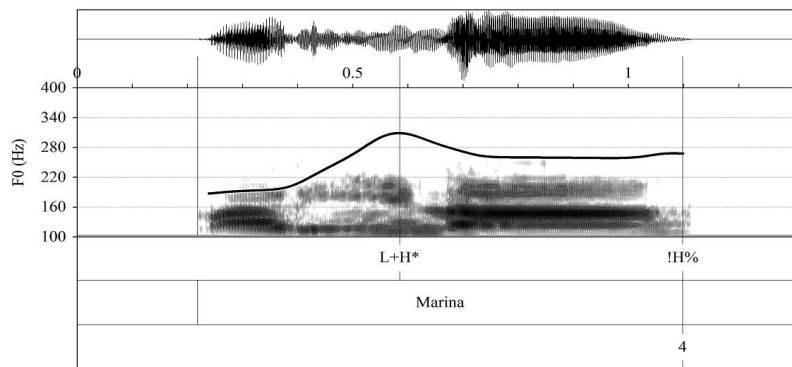
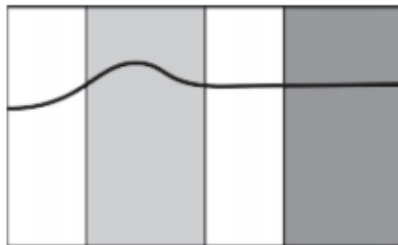


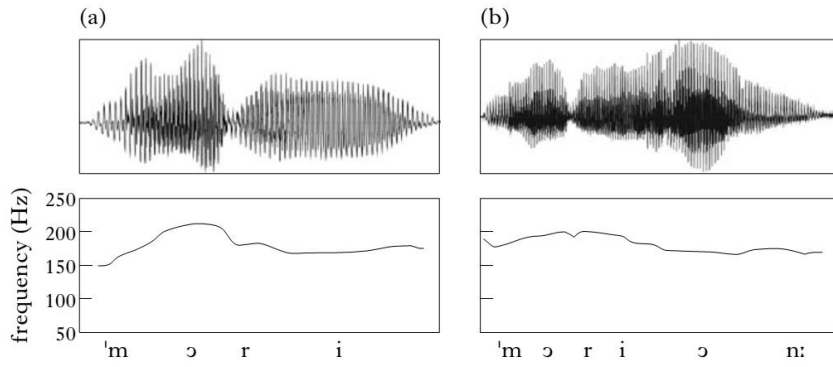
FIG. 7.37 Waveform, spectrogram, and F0 contour of the greeting call *Marina*, produced by a speaker of BH (BP, Baiano)



(L+)H* !H%

(52)

(7) Hungarian (Varga, 2008);



(8) Persian (Sadat-Tehrani, 2007):

Fig. 87a: The representation of the name *ebrahim* in default call.

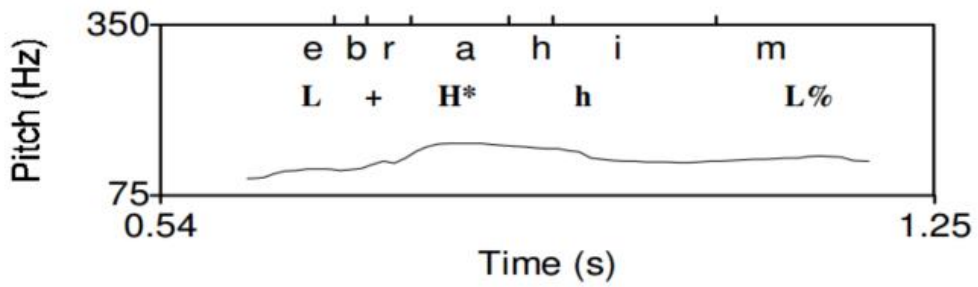


Fig. 87b: The representation of the name *ebrahim* in anger call.

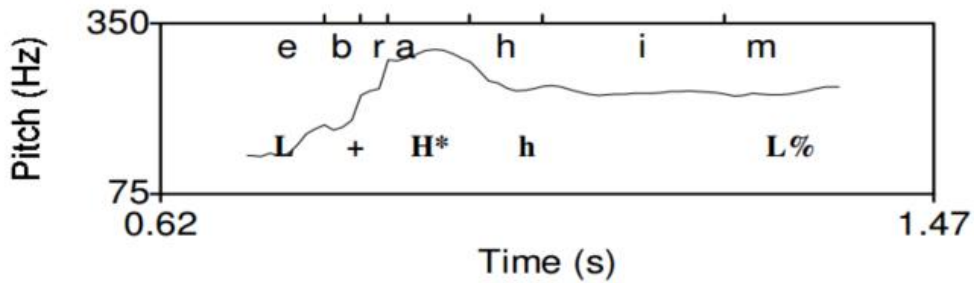
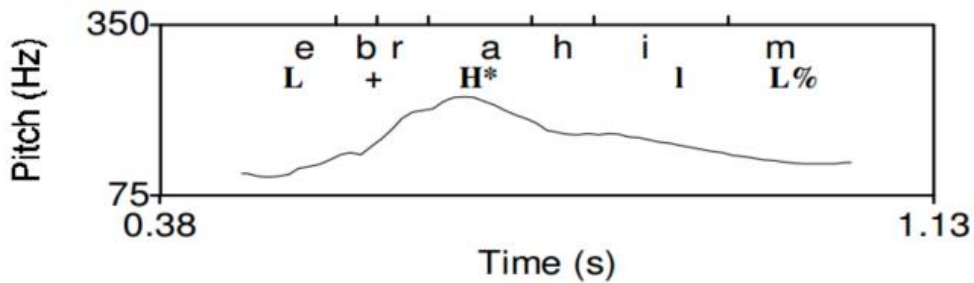
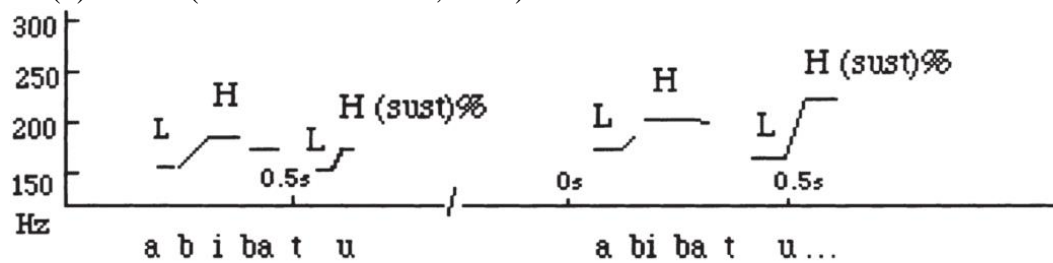


Fig. 87c: The representation of the name *ebrahim* in surprise call.



(9) Wolof (Rialland & Robert, 2001):



- (36) a. Abibatu! (N) b. Abibatu! (N)
 Abibatu! Abibatu!