

TONE, INTONATION AND THEIR INTERACTION IN LANGUAGE CONTACT:  
CASE STUDIES OF NANNING MANDARIN, NANNING CANTONESE,  
AND STANDARD MANDARIN

Xinzi Hou

Doctor of Philosophy

University of York

Language and Linguistic Science

November 2025

## ABSTRACT

In tonal languages, fundamental frequency (f<sub>0</sub>) simultaneously encodes lexical tone and paralinguistic intonation. This dual use of f<sub>0</sub> becomes more complex in language contact situations, particularly when the contacting varieties employ divergent prosodic strategies: Standard Mandarin extensively employs f<sub>0</sub> modulation, including on-focus expansion and out-of-focus compression of f<sub>0</sub> range (e.g., Y. Xu et al., 2012), whereas Cantonese relies less on f<sub>0</sub> and lacks post-focus compression (e.g., W. L. Wu & Xu, 2010). How these distinct pitch systems interact within a variety arising from contact therefore remains difficult to predict.

This thesis investigates tone, intonation, and their interplay within Nanning Mandarin, a stabilised intermediate variant arising from contact between Nanning Cantonese (source) and Standard Mandarin (target). Through comparison with the source and target languages, it aims to provide the first systematic, acoustic-phonetic investigation of the pitch systems of Nanning Mandarin, an under-researched variety, and to examine the interaction between lexical tone and intonational focus as reflected in tone-specific focus effects.

Two production studies investigated the f<sub>0</sub> realisation of lexical tone systems and intonational focus across the three Sinitic varieties, each recording over 60 native speakers. The tone study shows that Nanning Mandarin inherited the tone–contour mapping from Standard Mandarin, but exhibited greater similarities to Nanning Cantonese, taking similar-contoured Nanning Cantonese tones as main tonal templates. Regarding focus realisation, f<sub>0</sub> patterns diverged across all three varieties. Although Nanning Mandarin resembled Nanning Cantonese more than Standard Mandarin in focus marking, unique focus realisation and tonal variation were observed, with very restricted focus-driven f<sub>0</sub> modulation. These findings suggest that Nanning Mandarin, as a local variant of Standard Mandarin, may have developed distinct prosodic features, despite strong influence from Nanning Cantonese. The cross-linguistic divergence in tone-specific focus effects further suggests a complex tone–intonation interaction, requiring more fine-grained contour examination in continuous speech.

献给妈妈蓝柏坚、爸爸侯长明，和家乡南宁

感谢你们爱我，支持我，帮助我

事无巨细，始终如一

*Dedicated to my mother LAN Bojian, my father HOU Changming, and my hometown Nanning*

*Thank you for your love, support, and help*

*Unwaveringly, in every possible way*

## ACKNOWLEDGEMENT

Writing a non-academic piece in English is somehow even trickier than writing a thesis, because it requires a kind of writing ability that I most likely do not possess in my second language, unfortunately. But I will try my best. English is not my native language, yet here I am.

First of all, I want to thank my supervisor, Eleanor Chodroff, for being interested in my project even before I started my PhD. She made it seem feasible when I had no idea about quantitative methods or scripting. I am grateful for her constant patience, support, encouragement, care, and understanding—every time she managed to understand exactly what I meant through my rubbish expressions, it felt like there was nothing more I could possibly ask for. I also thank her for introducing me to acoustic analysis, statistics, and scripting, and more importantly, for making them so accessible and engaging. I had previously tried and failed to teach myself these more than once, but she helped me get past the initial barrier, enabling me to face further learning independently. I have repeatedly said this to different people, and I will say it again here: it is my greatest fortune to have Eleanor as my supervisor, and I dare to claim that she is the best supervisor ever (statistical evidence missing). The only regret is that I barely spent any time with her in person—I can only imagine how much more enjoyable my PhD experience would have been if I had.

I would also like to thank all those who provided academic guidance and advice during my PhD. My Thesis Advisory Panel deserves special thanks: Dominic Watt provided feedback during the early stages of the project, while Sam Hellmuth and George Bailey offered very useful guidance on experimental design, acoustic measurement, and statistical analysis of the intonation study. My examiners also contributed valuable insights: Shayne Slogett gave helpful advice on experiment design and shared perspectives from the perception side, while Yiya Chen raised many critical questions and offered insightful comments as an expert in prosody. Moreover, they both made the viva an unexpectedly enjoyable and memorable experience. I also benefited from support beyond York. The regular reading groups and discussions with my supervisor's lab members at the University of Zurich inspired me greatly in improving my methods and interpretations. I am also

grateful to Jonathan F. Kominsky for the timely help in solving a PsychoPy problem that had been troubling me for a long time, and to Kechun Li for insightful discussions on material design.

My PhD experience was somewhat unusual, as I spent much more time at home than at the university. As a result, I owe a great deal to the support of my family and friends.

Above all, my deepest gratitude goes to my mother, LAN Bojian, and my father, HOU Changming. The fact that I have come this far is closely tied to them. They raised me in an ocean of languages and music: one natively speaks Qin-Lian Cantonese and the other Henan Mandarin; one is an editor and the other a director, both working with words, the latter also with music. More importantly, they have always given me all the support they could every time I set out to pursue my dreams and follow my passions, whatever they might be. They were also deeply involved in my PhD through most stages, from material and experiment design to participant recruitment, recording, and validation... Especially my mother—I bet she could now confidently give a brief introduction to my research when people ask what exactly her daughter has been doing, which happened quite often when she tried to interest people in taking part in my study.

Another family member I would like to thank is my cousin, CHEN Xiaorui. As a bilingual native speaker of Nanning Cantonese and Nanning Mandarin, she was involved in many aspects of this study. Apart from recruiting participants for me, she served as a pilot participant, a participant (not in the same experiment!), an interlocutor, and a validator. She also provided invaluable suggestions, native-speaker intuition, and useful resources. Throughout the project, she has always been enthusiastic, patient, and genuinely interested in the research.

This research was also made possible by the kindness, generosity, and encouragement of many others around me. It could not have been realised without their continuous contributions, nor without their love for Nanning and its languages, which leaves me with a long list of people I would like to and need to thank.

For advice and discussion, I would like to thank my parents, my cousin CHEN Xiaorui, my friends

FENG Xin, QIN Haoyi, ZHONG Zhenghao, and CHEN Meilin for their thoughtful advice during the design, pilot, and validation stages of this project. I am also grateful to my family member HOU Xiulian, my friends YU Ning and LIU Junliang, my parents' friends (LU Mingyan, PANG Junrong, SHI Qiuyuan, XU Haibin, and ZENG Menxin), and my cousin's friends (Yuki, LUO Bin, and LIANG Haizhu) for sharing their native-speaker intuitions and perspectives. In addition, SUN Minghan patiently helped me work through a series of technical problems. I was also lucky to have my sister, LIU Yi, as my occasional mathematical rescue.

I am also deeply grateful to all those who helped recruit participants for this study—my parents, my family, my family friends, and my friends. Special thanks go to ZENG Menxin, DENG Xin, ZHENG Hao, CHEN Xiaorui, XU Haibin, FENG Xin, LEI Yutian, Master LEI, YU Zicheng, LAO Huan, SHI Qiuyuan, HOU Xiulian, HUANG Ling, HUANG Yufei, LI Ou, LI Yuanwei, TANG Ping, TANG Qian, WU Ya, WU Ziqian, and Cathy. This study simply would not have been possible without your help.

I would also like to thank all the participants who volunteered for this study. Many Nanning speakers, and even some Northern Mandarin speakers, told me that they found this research on languages of Nanning interesting, meaningful, and even amazing, especially at a time when fewer and fewer people speak the traditional local varieties. Some participants volunteered to recruit additional speakers after taking part themselves. Their enthusiasm and support made me feel that this research mattered not only to me, but also to many people who care about Nanning and its languages. Their encouragement was a constant source of motivation throughout this project.

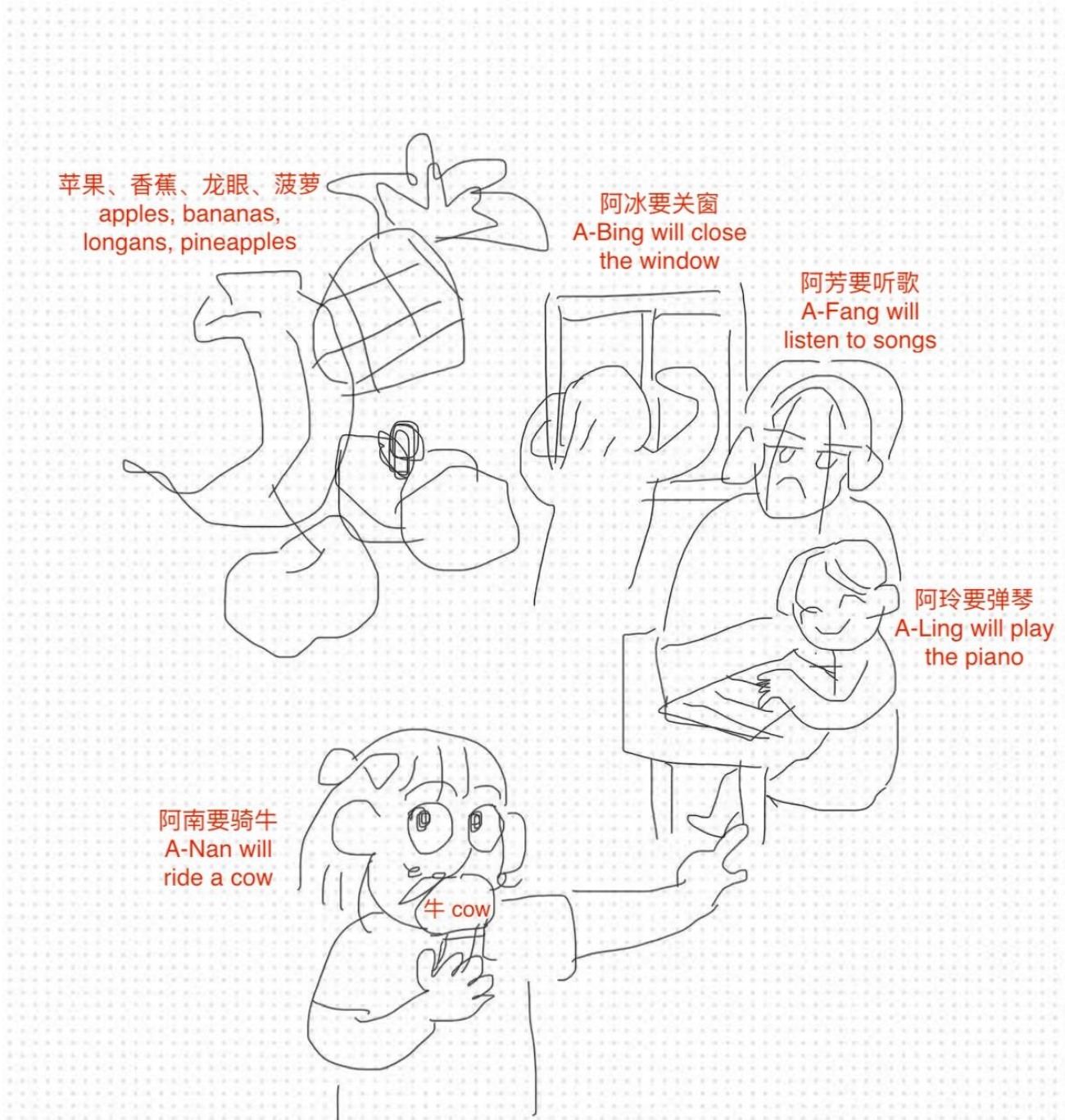
Beyond research, I was also fortunate to have meaningful connections throughout these years. Much of my PhD journey was spent without the companionship of fellow students and friends. In this journey, I would first like to thank FENG Xin, who is not only my closest friend, but also someone with whom I have shared remarkably similar academic paths. We supported each other through the difficult periods and were fortunate enough to welcome the dawn together. I am also grateful to my friends in Nanning, as this project brought us back into more frequent and closer contact. I also thank

my friends and colleagues at LLS and HRC. Getting to know them took time, but the conversations, lunches, and occasional outings made my time in York much less lonely. Among them, my friendship with GUO Buhan grew slowly but became a genuinely valued part of my time in York. I also treasure the friends I met through conferences—those precious like-minded people from around the world. And of course, Oreo, one of the kindest, gentlest, and most generous little cats in the world, who warmed my poor little life in York.

I also need to thank those things beyond PhD that brought me joy and kept me going. Daily piano practice and hip-hop dancing a few times a week were among the few things that could still give me positive feedback, especially when my research did not go well. They gave me the satisfaction of seeing myself improve, a still relatively healthy body (at least so far), and a sense of achievement beyond research. Stargazing and skywatching rewarded my eyes and helped me recharge during work, especially when the first two could not be guaranteed on a regular basis in York. Birdwatching in York and the gorgeous landscapes of Campus West offered me many moments of peace and wonder.

And finally, my home sweet home, Nanning, with its tropical scenery, amazing food, energetic atmosphere, and of course its fascinating languages—it has always been where I turned for comfort. I did not realise how deeply rooted I was there, even when we were thousands of miles apart.

I cannot say that I never imagined it, but I genuinely did not expect my PhD to become one of the happiest times of my life.



A funny doodle of my intonation stimuli (spoiler alert for the experiment!), drawn during the recording. Credit to my talented artist friend LAO Huan.

# TABLE OF CONTENT

<b>ABSTRACT</b> .....	<b>II</b>
<b>ACKNOWLEDGEMENT</b> .....	<b>IV</b>
<b>TABLE OF CONTENT</b> .....	<b>IX</b>
<b>LIST OF TABLES</b> .....	<b>XV</b>
<b>LIST OF FIGURES</b> .....	<b>XVII</b>
<b>DECLARATION</b> .....	<b>XIX</b>
<b>Chapter 1: INTRODUCTION AND LITERATURE</b> .....	<b>1</b>
1.1. LANGUAGE BACKGROUND.....	2
1.1.1. Varieties in China and the lingua franca.....	2
1.1.1.1. <i>Chinese and non-Chinese languages</i> .....	3
1.1.1.2. <i>Regional varieties: categorisation and hierarchy</i> .....	3
1.1.1.3. <i>Debates on terminology: “language” or “dialect”?</i> .....	7
1.1.1.4. <i>Lingua franca</i> .....	8
1.1.2. Varieties in Nanning.....	8
1.1.2.1. <i>Nanning, the city: geography and population</i> .....	8
1.1.2.2. <i>Local traditional varieties: history and distribution</i> .....	10
1.1.2.3. <i>Nanning Mandarin</i> .....	13
1.1.3. Existing literature on Nanning Mandarin and Nanning Cantonese.....	18
1.1.3.1. <i>Nanning Mandarin</i> .....	18
1.1.3.2. <i>Nanning Cantonese</i> .....	19
1.2. TONES OF CHINESE VARIETIES.....	20
1.2.1. Historical tone categories.....	21
1.2.2. Tone systems of Standard Mandarin, Nanning Cantonese, and Nanning Mandarin.....	23
1.2.2.1. <i>Tonal systems</i> .....	23
1.2.2.2. <i>Correspondence between tone categories</i> .....	25
1.2.2.3. <i>Nanning Mandarin tones</i> .....	27

1.2.3. Tones in contexts.....	27
1.2.3.1. <i>Neutral tones</i> .....	28
1.2.3.2. <i>Tone sandhi and tone changes</i> .....	29
1.2.3.3. <i>Alternative pronunciations</i> .....	31
1.2.4. Previous studies on tones in Nanning Mandarin and Nanning Cantonese.....	32
1.2.4.1. <i>Nanning Cantonese</i> .....	32
1.2.4.2. <i>Nanning Mandarin</i> .....	36
1.2.5. Research gaps in tone research.....	40
1.3. INTONATION OF CHINESE VARIETIES.....	41
1.3.1. Focus.....	41
1.3.1.1. <i>Definition and classification</i> .....	41
1.3.1.2. <i>Cross-linguistic variations in f0 realisations</i> .....	43
1.3.2. Previous studies on intonation in Nanning Cantonese and Nanning Mandarin.....	48
1.3.2.1. <i>Nanning Cantonese</i> .....	49
1.3.2.2. <i>Nanning Mandarin</i> .....	49
1.3.2.3. <i>Particles</i> .....	51
1.3.3. Research gaps in intonation research.....	52
1.4. PRESENT STUDY.....	52
1.4.1. Research gaps.....	52
1.4.2. Present study and aims.....	54
1.4.3. Research questions.....	54
<b>Chapter 2: PRODUCTION OF LEXICAL TONES.....</b>	<b>56</b>
2.1. CORPUS BUILD-UP.....	56
2.1.1. Wordlists.....	56
2.1.1.1. <i>Tone mapping</i> .....	56
2.1.1.2. <i>Word selection</i> .....	60
2.1.2. Participants.....	61
2.1.3. Experiment.....	63

2.1.3.1. <i>Before recording</i> .....	63
2.1.3.2. <i>During recording</i> .....	64
2.2. DATA PROCESSING AND ANALYSIS.....	65
2.2.1. Acoustic processing.....	65
2.2.1.1. <i>Segmentation and alignment</i> .....	65
2.2.1.2. <i>Validation of the NM accent</i> .....	66
2.2.2. Standard Mandarin data.....	69
2.2.3. Acoustic analysis.....	69
2.2.3.1. <i>F0 contour extraction</i> .....	69
2.2.3.2. <i>Cleaning data</i> .....	70
2.2.3.3. <i>Final datasets</i> .....	70
2.2.3.4. <i>F0 standardisation</i> .....	72
2.2.4. Contour modelling.....	73
2.2.4.1. <i>Description of the final model</i> .....	74
2.2.4.2. <i>(1) Basic model: f0 trajectory ~ tone</i> .....	75
2.2.4.3. <i>(2) Add random structures</i> .....	76
2.2.4.4. <i>(3) Add an AR1 error model</i> .....	78
2.2.4.5. <i>(4) Add another fixed effect: f0 trajectory ~ tone + preceding consonant</i> .....	80
2.2.4.6. <i>(5) Remaining discussions</i> .....	80
2.3. RESULTS AND DISCUSSIONS.....	82
2.3.1. Phonetic tone inventories.....	82
2.3.1.1. <i>Tone systems</i> .....	82
2.3.1.2. <i>Age difference</i> .....	91
2.3.1.3. <i>Individual variation</i> .....	93
2.3.2. Phonetic similarity of similar-contoured tones.....	94
2.3.2.1. <i>Overall phonetic distance</i> .....	94
2.3.2.2. <i>Contour similarity with f0 height differences</i> .....	96
2.3.2.3. <i>Contour similarity with f0 heights standardised</i> .....	106

2.3.2.4. <i>Interim summary</i> .....	116
2.3.3. Formation of Nanning Mandarin citation tones.....	117
2.3.3.1. <i>Source language as main template</i> .....	117
2.3.3.2. <i>Hybrid features under joint influence</i> .....	118
2.3.3.3. <i>Target language as acoustic reference</i> .....	118
2.4. SUMMARY.....	121
<b>Chapter 3: PRODUCTION OF INTONATIONAL FOCUS</b> .....	<b>123</b>
3.1. CORPUS BUILD-UP.....	123
3.1.1. Materials.....	125
3.1.1.1. <i>Meaning conditions</i> .....	125
3.1.1.2. <i>Sentence lists</i> .....	128
3.1.1.3. <i>Word selection</i> .....	132
3.1.1.4. <i>Interim summary of materials</i> .....	135
3.1.2. Experiment design.....	135
3.1.2.1. <i>Pseudo-randomisation of trial order</i> .....	136
3.1.2.2. <i>Filler trials</i> .....	137
3.1.2.3. <i>Visual prompts: screen display</i> .....	139
3.1.2.4. <i>Audio prompts: pre-recorded elicitations</i> .....	140
3.1.2.5. <i>Validation and pilot study</i> .....	143
3.1.3. Participants.....	144
3.1.3.1. <i>Nanning Mandarin and Nanning Cantonese speakers</i> .....	144
3.1.3.2. <i>Standard Mandarin speakers</i> .....	147
3.1.4. Recording procedure.....	148
3.1.4.1. <i>Before recording</i> .....	148
3.1.4.2. <i>During recording</i> .....	150
3.2. DATA PROCESSING AND ANALYSIS.....	151
3.2.1. Annotation and segmentation.....	151
3.2.2. Validation of accents and naturalness.....	153

3.2.3. Data selection.....	156
3.2.4. Extracting f0 min, max, and range.....	157
3.2.4.1. F0 extraction.....	158
3.2.4.2. Manual f0 correction and re-extraction.....	164
3.2.5. Cleaning and standardisation.....	167
3.3. RESULTS OF F0 MIN, MAX, AND RANGE.....	168
3.3.1. Focus effects.....	171
3.3.1.1. Nanning Mandarin.....	172
3.3.1.2. Nanning Cantonese.....	173
3.3.1.3. Standard Mandarin.....	174
3.3.2. Tone effects.....	175
3.3.2.1. Differential focus effects: tone × meaning interactions.....	180
3.3.2.2. Post-hoc comparisons: EMMs results.....	182
3.3.3. Cross-linguistic comparisons of main focus effect.....	185
3.3.3.1. Nanning Cantonese vs. Nanning Mandarin.....	186
3.3.3.2. Standard Mandarin vs. Nanning Mandarin.....	187
3.3.3.3. Numerical distance.....	188
3.4. DISCUSSIONS.....	188
3.4.1. Focus effects.....	189
3.4.1.1. Focus marking in the three varieties.....	189
3.4.1.2. Cross-linguistic comparison.....	190
3.4.2. Tone effects.....	193
<b>Chapter 4: CONCLUDING REMARKS.....</b>	<b>198</b>
4.1. SUMMARY OF THE KEY FINDINGS.....	198
4.1.1. RQ 1: hybrid tone system with greater source influence in Nanning Mandarin.....	198
4.1.2. RQ 2 and 3: unique focus marking with restricted f0 modulation in Nanning Mandarin.....	199
4.2. FURTHER DISCUSSIONS: INTERPRETATION AND THEORETICAL IMPLICATIONS	200

4.2.1. Limitations of synchronic data.....	201
4.2.2. Focus-marking strategy and prosodic variation in language contact.....	202
4.2.3. Tone–intonation interaction: additive model?.....	204
4.3. CONCLUSION.....	205
<b>APPENDIX A: TONE STUDY.....</b>	<b>206</b>
A.1. SPEECH MATERIALS AND EXPERIMENTS.....	206
A.1.1. Monosyllabic wordlists.....	206
A.1.2. Disyllabic wordlists.....	209
A.1.3. Screenshot of trial display.....	212
A.2. ACOUSTIC RESULTS.....	213
A.2.1. Raw f0 values in Log-T Chao tone numbers.....	213
<b>APPENDIX B: INTONATION STUDY.....</b>	<b>214</b>
B.1. SPEECH MATERIALS AND EXPERIMENTS.....	214
B.1.1. Target sentences.....	214
B.1.2. Filler trials.....	215
B.1.3. Full set of question-answer pairs.....	217
B.1.4. Screenshot of trial display.....	223
B.2. STATISTICAL MODELS AND RESULTS.....	223
B.2.1. Analysis of focus effect (§ 3.3.1 and § 3.3.2.1).....	223
B.2.2. Post-hoc pairwise comparisons of tone-specific focus effect (§ 3.3.2.2).....	242
B.2.3. Cross-linguistic comparisons (§ 3.3.3).....	249
<b>REFERENCES.....</b>	<b>252</b>

## LIST OF TABLES

Table 1.1.1 : Language use of Nanning Mandarin and Nanning Cantonese by participants of this thesis.....	14
Table 1.2.1 : Middle Chinese tone categories; adapted from M. Y. Chen (2000).....	21
Table 1.2.2 : Lexical tone systems of Standard Mandarin and Nanning Mandarin. Tone values are from Lee and Zee (2003) and Lin (2008), respectively.....	24
Table 1.2.3 : Lexical tone system of Nanning Cantonese. Tone values are from Lin and Qin (2008).....	25
Table 1.2.4 : Correspondence between tone categories of the two Mandarin varieties and Nanning Cantonese.....	26
Table 1.2.5 : Summary of the tone values and descriptions (if tone values were not provided) of each of the four NM tones in the aforementioned previous studies. Bold indicates inconsistencies with the majority.....	33
Table 1.2.6 : Summary of the tone values and descriptions (if tone values were not provided) of each of the four NM tones in the aforementioned previous studies.....	39
Table 2.1.1 : Correspondence between the non-checked tone categories of Mandarin and Cantonese.....	57
Table 2.1.2 : Number of items in the wordlists summarised by tone.....	59
Table 2.1.3 : Word samples for different syllabic structures in Nanning Mandarin and Nanning Cantonese.....	61
Table 2.1.4 : Age and gender distribution of the recruited participants of Nanning Mandarin and Nanning Cantonese.....	63
Table 2.2.1 : Grouping of NM speakers. Group (a) and (b) (in bold) served as the reference points for rating.....	67
Table 2.2.2 : Detailed rating criteria. The rating of 4–5 (in bold) indicates an acceptable, authentic NM accent.....	68
Table 2.2.3 : Summary of the age and gender information of the selected speakers.....	72
Table 2.2.4 : Models containing different combinations of focus effects and their corresponding AIC values. Checkmarks indicate the contained random effects.....	78
Table 2.3.1 : Chao tone numbers for each phonological tone in Nanning Mandarin, Standard Mandarin, and Nanning Cantonese. Tones are aligned according to their historical correspondence.....	83
Table 2.3.2 : Comparison of the SM tone systems from different sources.....	85

Table 2.3.3 : Comparison of the NC tone systems from different sources. Bold indicates the tone values different from all the previous transcriptions.....	86
Table 2.3.4 : Summary of tonal contrasts in Nanning Cantonese from different sources.....	88
Table 2.3.5 : Comparison of the NM tone systems from different sources.....	89
Table 2.3.6 : RMSD values on the Chao tone numbers listed first in each cell, and on the z-scored values, second. Each row presents a comparison of an NM tone to tones with a similar contour in Nanning Cantonese and Standard Mandarin. Underlining indicates the smallest RMSE numbers, that is, the highest similarity to NM contours.....	95
Table 2.3.7 : Summary of the contour comparisons between the NM tones and the corresponding NC and SM tones.....	116
Table 3.1.1 : Four meaning conditions with elicitations and targets. Underlining indicates target sentences; bold indicates focused items.....	126
Table 3.1.2 : Examples of five-syllable target sentences.....	129
Table 3.1.3 : Example of question–answer pairs across meaning conditions.....	131
Table 3.1.4 : Lexical items inconsistent across the three varieties.....	135
Table 3.1.5 : Factors under investigation of intonation.....	135
Table 3.1.6 : Cross-linguistically inconsistent lexical items in filler trials.....	138
Table 3.1.7 : Age and gender distribution of the recruited participants of Nanning Mandarin and Nanning Cantonese.....	146
Table 3.1.8 : Age and gender distribution of the recruited participants of Standard Mandarin.....	148
Table 3.2.1 : Summary of the age and gender information of the selected speakers.....	157
Table 3.3.1 : Differences of the main focus effect between Nanning Mandarin and its source (top) and target (bottom) languages, compared to Broad Focus. Significance levels: $p < 0.001$ , ***, $p < 0.01$ , **, $p < 0.05$ , *.....	186
Table 3.4.1 : Summary of the main effects of the narrow focus in f0 range, min and max. ↑ indicates expanded f0 range or higher f0 min/max; ↓ indicates compressed f0 range or lower f0 min/max.....	189
Table 3.4.2 : Summary of the tone-specific focus effects in f0 range, min and max. ↑ indicates expanded f0 range or higher f0 min/max; ↓ indicates compressed f0 range or lower f0 min/max.....	194

## LIST OF FIGURES

Figure 1.1.1 : Geographical distribution of Chinese varieties; the right figure zooms in Southeast China. Different colours indicate different language groups.....	4
Figure 1.1.2 : Geographical distribution of languages in Guangxi. Different colours indicate different language families.....	5
Figure 1.1.3 : Positions of language branching tree for Beijing Mandarin and language varieties spoken in Nanning.....	6
Figure 1.1.4 : The location of Nanning, Guangxi Zhuang Autonomous Region, China.....	9
Figure 1.1.5 : Geographical distribution of languages in Nanning. Blue texts are Chinese varieties, and green texts are non-Chinese varieties. Chinese varieties: G (Southwest Mandarin), Y (Cantonese), K (Hakka), P (Pinghua), M (Min). Non-Chinese varieties: Zh (Zhuang), Ya (Mienic), Bn (Bunu).....	12
Figure 2.2.1 : Autocorrelation graph for the so-far best model, overall (left) and by trajectory (right). The height of the second line indicates the amount of autocorrelation at lag 1.....	79
Figure 2.2.2 : Autocorrelation graph for the so-far best model after adding an AR1 model ( $\rho = 0.550$ ), overall (left) and by trajectory (right). The height of the second line indicates the amount of autocorrelation at lag 1.....	79
Figure 2.2.3 : Estimated $f_0$ contours for the four level tones with four types of syllable-onset consonants.....	81
Figure 2.3.1 : $F_0$ contours in Chao tone numbers of Nanning Mandarin, Standard Mandarin, and Nanning Cantonese for each phonological tone. Each ribbon represents 1 standard error from the mean.....	83
Figure 2.3.2 : $F_0$ contours in Chao tone numbers of Nanning Mandarin and Nanning Cantonese by each of the four age groups for each phonological tone. Each ribbon represents 1 standard error from the mean.....	92
Figure 2.3.3 : $F_0$ contours in Chao tone numbers of Nanning Mandarin and Nanning Cantonese, plotted for individual speakers across phonological tones.....	93
Figure 2.3.4 : Overall $f_0$ contours and difference curves for level tones. Shaded bands represent pointwise 95%-confidence intervals.....	98
Figure 2.3.5 : Overall $f_0$ contours and difference curves for rising tones. Shaded bands represent pointwise 95%-confidence intervals.....	100
Figure 2.3.6 : Overall $f_0$ contours and difference curves for mid-falling tones. Shaded bands represent pointwise 95%-confidence intervals.....	102

Figure 2.3.7 : Overall f0 contours and difference curves for high-falling tones. Shaded bands represent pointwise 95%-confidence intervals.....	104
Figure 2.3.8 : Difference curves between the two pairs of NC tone (near-) mergers. Left: difference between NC T2 and NC T6 (falling tones); difference not significant. Right: difference between NC T3 and NC T4 (rising tones); difference significant at points 5.5–10. Shaded bands represent pointwise 95%-confidence intervals.....	105
Figure 2.3.9 : Overall f0 contours and difference curves for level tones after <i>z-score</i> standardisation. Shaded bands represent pointwise 95%-confidence intervals.....	109
Figure 2.3.10 : Overall f0 contours and difference curves for rising tones after <i>z-score</i> standardisation. Shaded bands represent pointwise 95%-confidence intervals.....	111
Figure 2.3.11 : Overall f0 contours and difference curves for mid-falling tones after <i>z-score</i> standardisation. Shaded bands represent pointwise 95%-confidence intervals.....	113
Figure 2.3.12 : Overall f0 contours and difference curves for high-falling tones after <i>z-score</i> standardisation. Shaded bands represent pointwise 95%-confidence intervals.....	115
Figure 2.3.13 : Difference curves between SM T1 and the corresponding NC level tones. Difference between SM T1 and: NC T1 (significant at points 1–9.8); NC T5 (significant at points 7–10). Shaded bands represent pointwise 95%-confidence intervals.....	119
Figure 2.3.14 : Difference curves between SM T2 and the corresponding NC rising tones. Difference between SM T2 and: NC T3 (significant at points 1–8.6); NC T4 (significant at points 1–7.5 and 9.8–10). Shaded bands represent pointwise 95%-confidence intervals.....	120
Figure 3.2.1 : Example of carryover effect in the subject with NM Tone 2.....	166
Figure 3.3.1 : Average f0 range across each focus condition.....	172
Figure 3.3.2 : Model-estimated average f0 range across each focus condition. Arrows indicate significant expansion (↑) or compression (↓) compared to Broad Focus.....	172
Figure 3.3.3 : Average f0 range (in LT Chao tone) in Nanning Mandarin, across each focus condition and tone. Left plot is the raw acoustics; right plot is the model estimation.....	177
Figure 3.3.4 : Average f0 range (in LT Chao tone) in Nanning Cantonese, across each focus condition and tone. Left plot is the raw acoustics; right plot is the model estimation..	178
Figure 3.3.5 : Average f0 range (in LT Chao tone) in Standard Mandarin, across each focus condition and tone. Left plot is the raw acoustics; right plot is the model estimation.....	179

## DECLARATION

I declare that this thesis is a presentation of original work and that I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. Parts of the work presented in this thesis have been disseminated in the publications listed below. I confirm that I am the first author of these publications and that the work included in this thesis represents my own major contribution to the research. All sources are acknowledged as references.

Hou, X., Zhao, L., & Chodroff, E. (2023). Intermingling tone systems: The relationship of Nanning Mandarin to Nanning Cantonese and Standard Mandarin. *Proceedings of the 20th International Congress of Phonetic Sciences, 1935–1939.*

Hou, X., & Chodroff, E. (2026). Focus marking in a language contact context: A comparison of focus-induced f<sub>0</sub> effects in Nanning Mandarin and Standard Mandarin. *Proceedings of Speech Prosody 2026*, 358–362. <https://doi.org/10.21437/SpeechProsody.2026-72>

The 2026 conference paper was published after the initial submission of this thesis and before the submission of the final version.

## Chapter 1: Introduction and literature

The melody of speech can convey multiple functions or roles from speaker emotion or affect to aspects of sentential or lexical meaning. For example, people in anger usually raise their pitch, while in depression, people tend to speak in a lower pitch. Other than emotion, speech melody can also express different sentential meanings: a rising pitch in “cat” may imply a question, while a falling one indicates a statement. These pitch patterns, or phrasal melodies, are the intonation.

In many Asian and African languages, word meanings are also distinguished by pitch patterns, unlike in English, where the words “kit” and “cat” differ only in the vowel sound. These pitch patterns conveying lexical-level contrast are known as lexical tone. For instance, in Standard Mandarin, the syllable /ma/ with a rising pitch means “hemp”, but with a falling pitch, it means “to scold”. In tonal languages, lexical tone and intonation both operate on the pitch and coexist over entire phrases. This presents a profound conflict: when pitch variation is already used obligatorily to distinguish word meaning, how do speakers simultaneously use the same pitch signal to realise intonation, such as signalling a question or marking emphasis?

Furthermore, even between tonal languages, both tone systems and intonation patterns can differ significantly. Standard Mandarin, for example, has four lexical tones, while Cantonese has six, and their respective pitch patterns are largely dissimilar. The divergence in intonation is even more pronounced. To signal focus, Standard Mandarin extensively employs fundamental frequency ( $f_0$ ), typically through  $f_0$  range expansion and/or  $f_0$  raising on the focal item, as well as  $f_0$  range compression and/or  $f_0$  lowering in the post-focus region (e.g., Y. Xu et al., 2012). In contrast, Cantonese does not reliably use  $f_0$  to mark focus, nor does it exhibit post focus-compression (e.g., W. L. Wu & Xu, 2010). The cross-linguistic divergence also exists in the interaction between tone and intonation. While sentential pitch in Standard Mandarin is often accounted for by ‘additive’ or ‘superimposed’ intonation onto tone proposed by Chao (1933), this may not hold true for other language varieties. More cross-linguistic comparisons are needed to understand the not-yet-clear true relationship between tone and intonation.

The pitch systems can involve even greater complexity in language contact situations. When two language varieties, such as Cantonese and Standard Mandarin, have vastly different tone systems and divergent (perhaps even conflicting) strategies for using  $f_0$  to realise intonation, it becomes difficult to predict how the pitch systems of a variety emerging from contact will be realised and interact with each other. Nanning Mandarin, a fossilised intermediate variant spoken in a multilingual city in China, presents a special case of tone and intonation developing from language contact. With Nanning Cantonese as its primary source language and Standard Mandarin as its target, Nanning Mandarin offers a window through which to investigate the compromises and innovations that arise from this tone–intonation interaction in a language contact context.

The core aim of this thesis is to systematically investigate the realisations of the two pitch systems that function together in Nanning Mandarin, using acoustic-phonetic approaches and cross-linguistic comparison. Through microvariation-like analysis of three related language varieties, this research explores how the pitch systems of Nanning Mandarin are similar to or different from those of Nanning Cantonese and Standard Mandarin. At the same time, this research aims to initially establish the prosodic profiles of two understudied Chinese varieties: Nanning Mandarin and Nanning Cantonese.

Following this introduction, the remainder of Chapter 1 will provide the relevant linguistic background, review the literature on tone and intonation, and detail the specific research questions and hypotheses of the study. Chapter 2 presents the methodology and results of the tone systems in Nanning Mandarin and their relation to Nanning Cantonese and Standard Mandarin tone systems. Chapter 3 details the methodology and results of the intonational focus study in Nanning Mandarin, again comparing to Nanning Cantonese and Standard Mandarin. Finally, Chapter 4 discusses how the findings advance our understanding of tone, intonation and their interaction in language contact.

## **1.1. Language background**

### **1.1.1. Varieties in China and the lingua franca**

### **1.1.1.1. Chinese and non-Chinese languages**

China is a multi-ethnic nation, where languages from different language families are spoken. The languages spoken by the major ethnic, Han people (汉族 Hànzú), is called Han languages (汉语 Hànyǔ); in language family classification, it is often referred to by Sinitic languages or Chinese languages from the Sino–Tibetan language family. The languages spoken by the other 55 ethnic minorities span several language families, including the non-Chinese branches of the Sino–Tibetan language family, the Altaic<sup>1</sup>, Austroasiatic, Austronesian, Indo–European, Koreanic<sup>2</sup> languages, and mixed languages (Institute of Linguistics CASS et al., 2012). There are debating views about the classification of language families, for example, whether Kra–Dai (also known as Tai–Kadai) languages is a member of the Sino–Tibetan language family<sup>3</sup>. Such debates, though, are not the focus of this study.

The main language varieties traditionally spoken in Nanning involve Sinitic (Yue, Pinghua) and Kra–Dai (Zhuang) languages. Mandarin is also a Sinitic language.

### **1.1.1.2. Regional varieties: categorisation and hierarchy**

Chinese languages form the Sinitic languages, which is the major branch of the Sino–Tibetan language family. Chinese languages comprise a rich set of regional varieties, which have traditionally been divided into various major groups (区 qū). Since 1985, the two most widely accepted categorisations are seven groups and ten groups (X. Li & Xiang, 2020), although this is still controversial. The seven-group categorisation was proposed by Yuan (1960): Guanhua (Mandarin; 官话), Wu (吴), Xiang (湘), Gan (赣), Hakka (客家), Yue (Cantonese; 粤), Min (闽). Later in 1987,

---

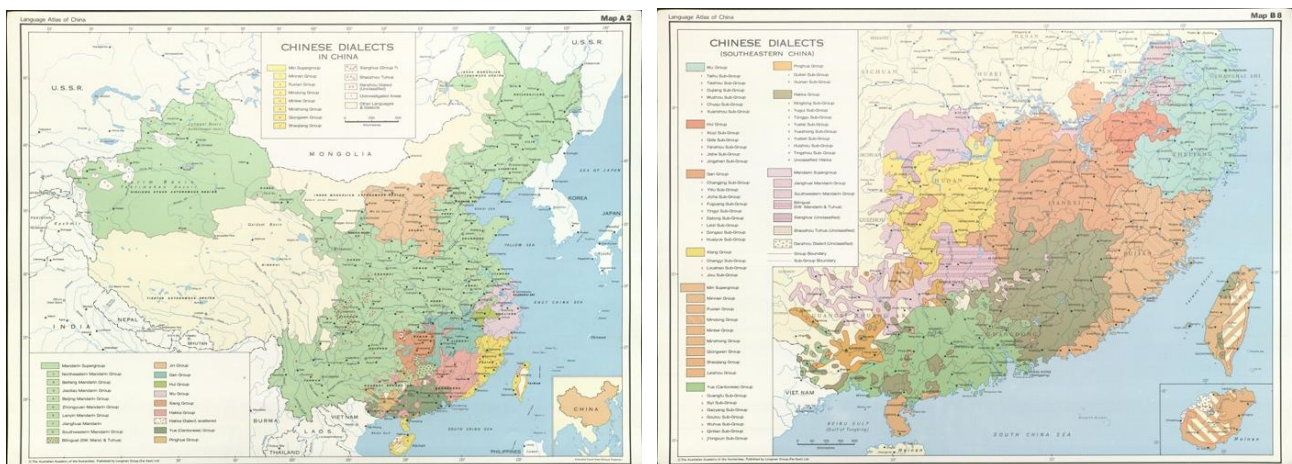
<sup>1</sup> Altaic is a hypothetical language family including the Mongolic(-Khitan), Turkic and Tungusic languages.

<sup>2</sup> Koreanic languages are uncategorised due to the controversies in whether it is a member of the Altaic languages.

<sup>3</sup> Language Atlas of China (2012) associates Kra–Dai languages to Sino–Tibetan languages, while Glottolog (<https://glottolog.org/>) does not.

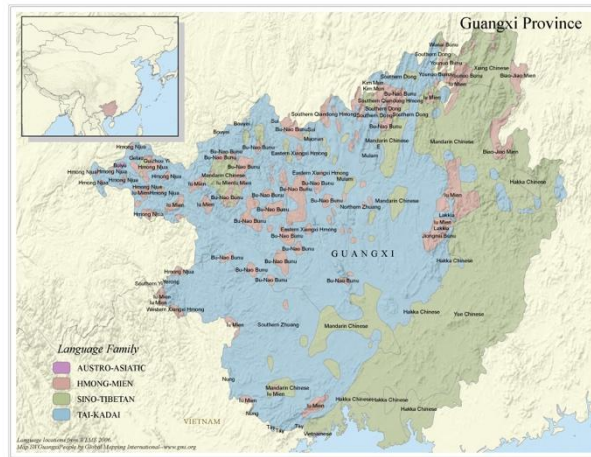
the first edition of the *Language Atlas of China* (2012), Jin (晋) and Hui (徽) were separated from Northern Mandarin, and an additional group “Pinghua (平话) and Tuhua (土话)” was added for some unclassified varieties in Southern China.

From the maps in Figure 1.1.1, we can see that northern China shares more homogeneity as a Mandarin-speaking region, while the other variety groups (except for Jin) are all located in the south. In particular, the Guangxi Zhuang Autonomous Region, where Nanning is located in and Nanning Mandarin is spoken, displays an even more complex linguistic environment, as shown in Figure 1.1.2. Guangxi is on the border of China and Vietnam and the transition between the Northern Mandarin and other Southern groups of Chinese varieties. Within the region, it contains a melange of Chinese subgroups and non-Chinese languages.



Source: Left figure from <https://i.redd.it/c783tzubtwh71.jpg>; right figure from [https://miro.medium.com/v2/resize:fit:8480/1\\*buRjS-v-D\\_eAZrymM3\\_7uA.jpeg](https://miro.medium.com/v2/resize:fit:8480/1*buRjS-v-D_eAZrymM3_7uA.jpeg).

**Figure 1.1.1:** Geographical distribution of Chinese varieties; the right figure zooms in Southeast China. Different colours indicate different language groups.



Source:

<https://www.missioninfobank.org/mib/images/gmms2007/SWGuangxiPeople.gmms07.jpg>

**Figure 1.1.2:** Geographical distribution of languages in Guangxi. Different colours indicate different language families.

These major groups are further divided into a hierarchy of subgroups. For example, Mandarin is first divided into eight smaller groups, each containing 2–13 subgroups (片 piàn). The Mandarin variety spoken in Beijing, which serves as the base pronunciation for Standard Mandarin, is a local dialect of the Jing–Cheng subgroup from the Beijing Mandarin group (one of the eight Mandarin groups). Yue (Cantonese) is divided into seven subgroups: Guangfu (广府), Siyi (四邑), Gao–Yang (高阳), Wu–Hua (吴化), Goulou (勾漏), Yong–Xun (邕浔), Qin–Lian (钦廉).

The main language varieties traditionally spoken in Nanning include: a local **Cantonese** dialect of the Yong–Xun subgroup from Yue (Cantonese) Chinese; local **Zhuang** dialects of both the Yongbei variety from the Northern group and the Yongnan variety from the Southern group of Zhuang (Kra–Dai); and a local **Pinghua** dialect of the Guinan subgroup from Pinghua (Chinese). Figure 1.1.3 below illustrates the position of branching where language varieties spoken in Nanning are located.



Source: Institute of Linguistics CASS et al. (2012), Institute of Linguistics CASS (2012), Chen & Li (2005)

**Figure 1.1.3:** Positions of language branching tree for Beijing Mandarin and language varieties spoken in Nanning.

These categorisations of Chinese varieties, though have not reached a consensus, were all based on the phonology and vocabulary; essentially, it was based on phonological sound change with respect to Middle Chinese (circa AD500–AD1200). The primary phonological criterion was the evolution of the syllabic initial consonants (声母 *shēngmǔ*) that were voiced obstruents (全浊 *quánzhuó*) in Middle Chinese (X. Li & Xiang, 2020). For example, below are the phonological criteria for the categorisation of Mandarin and Cantonese; detailed introductions of the historical tone categories are in § 1.2.

Mandarin corresponds to the modern varieties where among the historical voiced obstruent initials,

- 1) all of them became voiceless;
- 2) stops and affricatives
  - a) for historical *píng* (Tone I; ‘level’) tones, merged to aspirated voiceless
  - b) for historical *shǎng* (Tone II; ‘rising’), *qù* (Tone III; ‘departing’) and *rù* (Tone IV; ‘entering’) tones, merged to unaspirated voiceless

Cantonese corresponds to the modern varieties where among the historical voiced obstruent

initials,

- 1) all of them became voiceless
- 2) stops and affricatives
  - a) for historical *píng* (Tone I) tones, merged to aspirated voiceless
  - b) for historical *qù* (Tone III) and *rù* (Tone IV) tones, merged to unaspirated voiceless
  - c) for historical *shǎng* (Tone II) tones,
    - i. merged to aspirated voiceless if the tone is now *yáng shǎng* (Tone IIb)
    - ii. merged to unaspirated voiceless if the tone is merged to *yáng rù* (Tone IIIb)

### **1.1.1.3. Debates on terminology: “language” or “dialect”?**

In Chinese, regional varieties are referred to by the term *fāngyán* (方言), literally “regional speech”. The term *fāngyán* originated from the name of the book *Fāngyán*, a dictionary compiled by Yang Xiong in the early 1st century CE to document the regional vocabulary of its era.

This Chinese term has been traditionally translated as “**dialect**” in English by Chinese linguists (e.g., Y. R. Chao, 1968). However, the Chinese term *fāngyán* and the English term “dialect” actually have divergent references, and thus, the mismatched references are misleading. To define a dialect, the essential linguistic criterion is mutual intelligibility: two language varieties that are mutually intelligible are different dialects of one language, otherwise, they are different languages (Comrie, 2018; Mair, 1991).

This is obviously not the case when it comes to, for example, Mandarin and Cantonese. The Chinese varieties are distinct to an extent that they are largely mutually unintelligible, mainly due to the phonological and lexical differences. Therefore, Chinese varieties are often analogically compared to a language group such as Romance. Moreover, the linguistic distance between some Chinese “dialects” is even larger than between European “languages” (C. Tang et al., 2025).

However, speakers of Mandarin and Cantonese seldom claim that they speak different languages. This is attributed to both linguistic and social (cultural and political) factors: the various speech

communities of Chinese varieties have been using a shared logographic written system (i.e., 汉字 *Hànzì* ‘Chinese characters’), resident in a united nation with a bonding culture and history, not to mention the same historical origin and synchronic generic link among the varieties. In this case, Chinese varieties are in a unique situation different from European languages, making “language” a less appropriate term. In addition, the criterion of mutual intelligibility is challenged in the case of a dialect continuum, which is common in Chinese varieties (Comrie, 2018).

To handle the misunderstanding caused by mistranslation, “**regionalects**” (DeFrancis, 1986) and “**topolects**” (Mair, 1991) were adopted as alternative terms to depict the mutually unintelligible forms of Chinese varieties; the naming of these two terms only differs in the space size of the area. These debates on terminology, though, are not the main focus of this study. Therefore, we will use a more inclusive, generic term, “**variety**”, to refer to the regional varieties of Chinese languages.

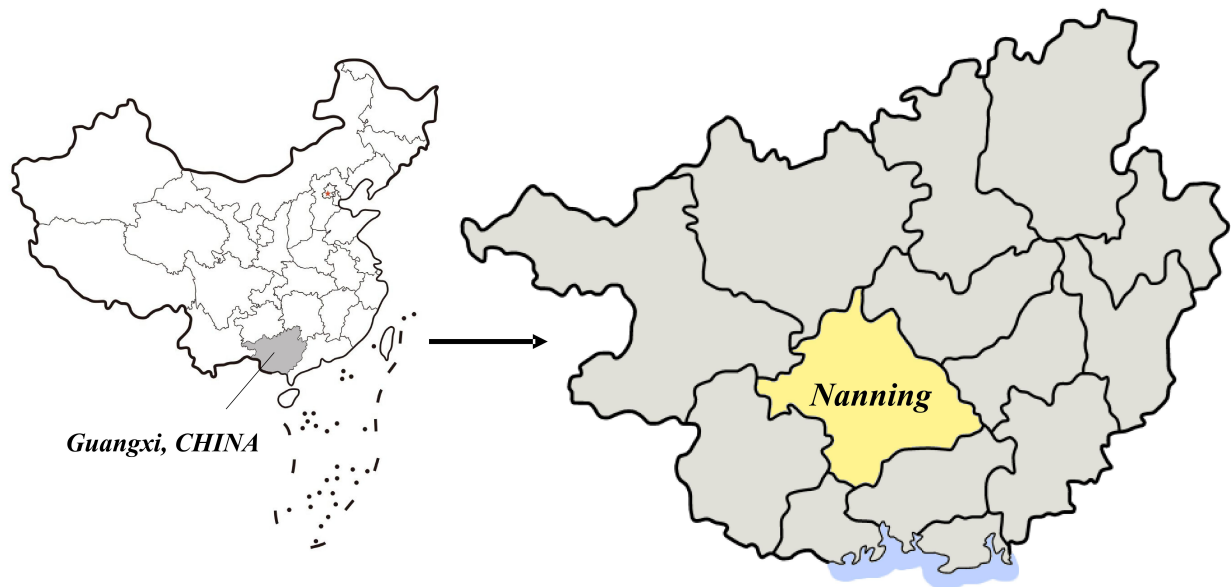
#### **1.1.1.4. Lingua franca**

The variability and mutual unintelligibility of the Chinese varieties necessitated the emergence of a domestic lingua franca to facilitate communication. Despite earlier common languages, the current lingua franca in China is Putonghua (普通话, ‘common speech’). Putonghua was established with Northern Mandarin as the principal base dialect and Beijing pronunciation as the reference for the standard.

Putonghua is not only the lingua franca, but also a standardised variety with legitimate prestige. Putonghua has been promoted throughout China required by the language policy since 1956, and has been set as the domestic lingua franca by law since 2001. Therefore, Putonghua is equivalent to the term Modern Chinese (现代汉语 *xiàndài Hànyǔ*) and is usually referred to as Standard Mandarin in English.

#### **1.1.2. Varieties in Nanning**

##### **1.1.2.1. Nanning, the city: geography and population**



Source: Adapted from two source maps:  
<https://img.uptt.com/uploads/allimg/250425/1-25042510532T491.jpg> and  
[https://upload.wikimedia.org/wikipedia/commons/4/43/Location\\_of\\_Nanning\\_Prefecture\\_within\\_Guangxi\\_%28China%29.png](https://upload.wikimedia.org/wikipedia/commons/4/43/Location_of_Nanning_Prefecture_within_Guangxi_%28China%29.png).

**Figure 1.1.4:** The location of Nanning, Guangxi Zhuang Autonomous Region, China.

Nanning (南宁 Nánning)<sup>4</sup> is the capital city of the Guangxi Zhuang Autonomous Region on the border of southern China. Historically, Nanning has served as a key gateway, facilitating trade between the inland regions of southwestern China and the coastal areas, as well as with the Mainland and Maritime Southeast Asia, since the Tang Dynasty (618AD–907AD). After the establishment of the People’s Republic of China (PRC), the rapid development of transport (including water, road, rail, and air travel) and communications has further established Nanning as a major hub city (Y. Lin & Qin, 2008).

Nanning’s linguistic complexity is rooted in its long history as a multi-ethnic and multi-lingual region. For centuries, it has been receiving migrants from within and beyond Guangxi, fostering a

<sup>4</sup> Also known as Yong (邕 Yōng).

high degree of linguistic inclusiveness. The Zhuang population is the indigenous ethnic group, while the Han population consists of several historical migration waves. Pinghua (Sinitic) speakers are largely the descendants of migrants from the Tang to Song Dynasties<sup>5</sup> (618–1279 AD), while Cantonese (Yue; Sinitic) speakers are mainly linked to migration since the mid-1800s (de Sousa, 2015; Y. Lin & Qin, 2008).

This historical layering has resulted in a high degree of contemporary multilingualism. According to an early questionnaire of language use (H. Chen & Li, 2005), the vast majority of respondents were bilingual (54.83%) or tri-lingual (38.88%), with monolingual speakers accounting for only 6.29%. These high percentages of multilingualism are situated within a large population: by the end of 2024, Nanning's total resident population was approximately 8.97 million people, with about 6.45 million residing in urban areas (Nanning Municipal Bureau of Statistics & Nanning Survey Office of the National Bureau of Statistics, 2025).

### **1.1.2.2. Local traditional varieties: history and distribution**

Currently, three main traditional varieties are still spoken in Nanning: the Chinese varieties of Nanning Cantonese and Pinghua, and the indigenous Tai variety of Zhuang. Nanning Cantonese is primarily spoken in the city centre, Pinghua is primarily spoken in the surrounding suburbs, nearby rural areas and urban villages, while Zhuang is mainly spoken in the rural villages. There is also a small population of speakers of other language varieties in Nanning.

**Nanning Cantonese** (南宁白话 Nánning Báihuà) was the predominant variety traditionally spoken in Nanning. It is a member of the Yong–Xun<sup>6</sup> subgroup of Yue (Cantonese) Chinese. Nanning Cantonese was brought to Nanning by migrants from Guangdong Province and remained the predominant variety in the city centre from the 1930s or 1940s until Nanning Mandarin became the

---

<sup>5</sup> Pinghua is the official language (guānhuà官话) of northern China during Tang and Song Dynasties.

<sup>6</sup> Yong and Xun are the historical name of places. Yong area includes the current Nanning city.

de facto lingua franca. In recent decades, the “old” (老派) Nanning Cantonese has been gradually replaced by the “new” (新派) one (Y. Lin & Qin, 2008); more detailed discussions on this can refer to § 1.2.4.1.

**Nanning Pinghua** (南宁平话 Nánning Píng huà) was one of the historically predominant varieties prior to Nanning Cantonese. It is a local dialect affiliated with the Guinan<sup>7</sup> subgroup of Pinghua Chinese. After Cantonese entered Nanning and became predominant, the distribution of Nanning Pinghua stepped out of the city and down to the suburbs and urban villages.

Nanning Cantonese, same as the other Cantonese varieties, is not mutually intelligible with Standard Mandarin. However, Nanning Cantonese shares many phonological similarities with Pinghua (de Sousa, 2015), leading to higher mutual intelligibility than that with other local varieties (Y. Lin & Qin, 2008). This is reflected in the debates on Pinghua’s categorical affiliation, i.e., whether Pinghua should be categorised as a branch of Cantonese (Mai, 2012; Wu W., 2001).

The **Zhuang varieties** (壮语 Zhuàngyǔ) spoken in Nanning include Yongnan variety and Yongbei variety. The Yong River (邕江 Yōngjiāng), the river that Nanning is situated on, divides the city into the north and south; Yongnan and Yongbei literally mean south or north of the Yong River, respectively. The distribution of Yongnan and Yongbei Zhuang approximately follows the division of the river. Yongnan Zhuang has a more complex sound inventory than Yongbei Zhuang.

All these language varieties spoken in Nanning have been influencing each other with considerable contact. Nanning Cantonese has imposed great influence on the local Pinghua; meanwhile, Nanning Cantonese has received influence from Pinghua and the surrounding Chinese varieties, including Cantonese dialects. Zhuang varieties have also been in deep contact with the Chinese varieties; especially, Nanning Cantonese has a strong substratum from and has been influenced by Zhuang during evolution (de Sousa, 2015; W. Huang, 2018; Y. Lin & Qin, 2008; Z. Wang, 2023). Under the

---

<sup>7</sup> Guinan, literally “southern Guangxi”

language policy of SM promotion, a trend of regional varieties converging towards the standard variety is expected.



**Figure 1.1.5:** Geographical distribution of languages in Nanning. Blue texts are Chinese varieties, and green texts are non-Chinese varieties. Chinese varieties: G (Southwest Mandarin), Y (Cantonese), K (Hakka), P (Pinghua), M (Min). Non-Chinese varieties: Zh (Zhuang), Ya (Mienic), Bn (Bunu<sup>8</sup>).

Figure 1.1.5 is an old map of Nanning in 2012 adjusted from the *Language Atlas of China* (2012). The map displays the city proper of Nanning and its then jurisdiction of one district (Yongning) and four (rural) counties (Mashan, Long'an, Wuming, Shanglin, Binyang, Heng). The administrative levels of some regions have changed, but the city territory has remained largely unchanged. The

---

<sup>8</sup> A Hmongic language.

current Nanning city covers the areas of Nanning, Yongning County and Wuming County on Figure 1.1.5.

### **1.1.2.3. Nanning Mandarin**

#### **Emergence of Nanning Mandarin**

In addition to the traditional varieties, another primary variety in Nanning is the local variant of Standard Mandarin. Following the nationwide promotion of Standard Mandarin since the 1950s, the standardised lingua franca is officially required in education, news media, government affairs, and other formal public situations. The complex multilingual environment in Nanning strengthens the need for a lingua franca, a situation further intensified by significant migration to the capital city. Apart from migrants moving for economic reasons and urbanisation (mainly from within Guangxi and neighbouring regions), a large influx of military officials and cadres from northern China brought northern Mandarin to Nanning. The language spoken by this prestigious social group in turn boosted the prestige of Standard Mandarin (Yu & Qin, 2008). Facilitated by the influxes of migrants with different language backgrounds from within and beyond Guangxi, and further accelerated by the growth of mass media and digital communication since the 1990s, the Standard Mandarin promotion in Nanning has been dramatically rapid, widespread, and profound, at a pace that leads within Guangxi and exceeds other Cantonese-speaking areas (Y. Lin & Qin, 2008).

In the 1990s, the strong nationwide promotion resulted in the widespread prevalence of Standard Mandarin. According to early questionnaires on language use, 88.31% of the sample population in Nanning had Standard Mandarin competence, ranked second in Guangxi (H. Chen & Li, 2005; Yu & Qin, 2008); among them, the competence rate for speakers aged 45 or below exceeded 90%. A later investigation (Huang N., 2018) of primary to middle school students (mainly aged 9–16) further claimed a younger trend of (Nanning) Mandarin use. In their report, although only 6.1% of the participants reported Mandarin as their L1, over half of them used Nanning Mandarin with peers and parents, and 39% of them had their parents using Nanning Mandarin with each other. In the local

public space, 80%–90% of participants used Nanning Mandarin.

While Nanning Cantonese was the primary L1 for older generations, Nanning Mandarin has increasingly become the native language for younger speakers; moreover, younger speakers are now usually monolingual Nanning Mandarin speakers (Y. Lin & Qin, 2008). This trend is also reflected in the demographic distribution of the recruited participants in this thesis, as shown in Table 1.1.1 below. In our tone and intonation studies, participants were required to have acquired Nanning Mandarin or Nanning Cantonese before the age of twelve. Speakers in the mid to old age group (age > 40) mostly acquired Nanning Cantonese before Nanning Mandarin or at least simultaneously, whereas younger speakers (age < 40) mostly acquired Nanning Mandarin as their native language, with some of them not acquiring proficient Nanning Cantonese.

	<b>Tone study</b> (30 usable speakers)		<b>Intonation study</b> (52 speakers; 31 selected for analysis)	
	<b>Age &gt; 40</b>	<b>Age &lt; 40</b>	<b>Age &gt; 40</b>	<b>age&lt;40</b>
<b>NC earlier</b>	16	1	27	4
<b>NC = NM</b>	3	3	5	3
<b>NM earlier</b>	1	2	1	2
<b>NM speaker with low/no NC proficiency</b>	0	4	1	9

**Table 1.1.1:** Language use of Nanning Mandarin and Nanning Cantonese by participants of this thesis.

However, “competence of Standard Mandarin” does not only refer to fluent, standard form, but also includes the language variety when spoken with various levels of local accents and fluency (H. Chen & Li, 2005). The less standard variety is, in fact, what the vast majority of local speakers use in real practice. This is because Standard Mandarin is a variety that has been officially standardised, rather than one that has evolved naturally as a regional variety. Therefore, this distinct Mandarin dialect,

namely Nanning Mandarin (南宁普通话 Nánning pǔtōnghuà), is effectively the current lingua franca of Nanning.

### **Classification for non-standard Mandarin varieties**

Chinese linguistics started to realise the prevalence and research value of such new, non-standard Mandarin varieties emerging from language contact. However, the definition and according terminology were found to be various and vague in the previous literature. Existing literature usually categorised Nanning Mandarin as a type of ‘interlanguage’ (Yu, 2008) or ‘Local Mandarin’ (Lan, 2007), most of which used both terms interchangeably (H. Chen & Li, 2005; Fan, 2020; Y. Li, 2012; Y. Liang, 2005; Lu, 2012; Y. Wu, 2017).

The term ‘**interlanguage**’ was proposed by Selinker (1969, 1972) for second language (L2) learning. An interlanguage is defined as a separate, intermediate linguistic system resulting from a learner’s attempt to produce the norm of the target second language. The notion, originally defined specifically for individual non-native speakers, was later adopted and extended by Chinese linguists (e.g., Y. Chen, 1991; Z. Chen, 1990; R. Li, 1988) to describe collective linguistic behaviour that emerged during the SM promotion. The new term specifically for this case, ‘**Local Mandarin**’ (literally, ‘local Standard Mandarin’ or ‘local Putonghua’; 方言普通话/地方普通话), was proposed by Yao (1989). Local Mandarin includes any non-standard variants of Standard Mandarin that meet the criteria of communication between various regional varieties. For any specific Local Mandarin variety, the word ‘local’ can be replaced by the name of the place, e.g., Nanning Putonghua.

However, the term ‘interlanguage’ does not seem to be completely accurate. It is widely agreed that most of the Standard Mandarin learners actually acquire the non-standard variants in the first place, especially when the teaching resources are restricted. Moreover, an increasing number of speakers, especially in the younger generations, now acquire such non-standard variants as their native language, and thus are monolingual. This situation, therefore, does not seem to tightly align with L2 acquisition. Furthermore, Lyu and Yu (2021) believed that both ‘interlanguage’ and ‘Local

Mandarin' are limited in static description of the language systems, lacking in revealing the competition and variation in language contact. They therefore proposed to integrate the theories of **dialect levelling** (Trudgill, 1986) and **koineisation** (Siegel, 1985) to better account for Nanning Mandarin. In their opinion, the formation and variation of Nanning Mandarin basically conformed to the three processes of koineisation, i.e., mixing, levelling and simplification.

It should be noted that, koineisation applies to the contact between similar, mutually intelligible social or regional dialects (Siegel, 1985), and the three processes happen in what is called **immigrant koineisation** by Siegel (1985). However, migration is not the primary factor in the formation of Nanning Mandarin (Lyu & Yu, 2021). Moreover, the inherent continuity across generations of Nanning Mandarin speakers further argues against the situation of new dialect formation (Kerswill, 2010). Moreover, koineisation does not involve L2 acquisition, code-switching, imperfect language transmission or transfer, and does not have a target variety (Siegel, 2001), which is akin to the situation of Nanning Mandarin.

In terms of the **levelling** mechanism, that is, the reduction or attrition of marked, unusual variants, it leads to the reduction of differences between varieties in contact (Trudgill, 1986). This convergence is mutual (Kerswill, 2003), and can happen in both the cross-dialectal and dialect–standard dimension (Hinskens, 1998). Although the persistence of local Nanning Cantonese features in Nanning Mandarin inevitably results in a certain degree of “loss” of Standard Mandarin features, it is difficult to fully disentangle Nanning Cantonese and Standard Mandarin in the exact ways they influence Nanning Mandarin from observed language phenomena or behaviour. Meanwhile, the convergence between varieties is hardly mutual: it is generally the speakers of local varieties who accommodate to Standard Mandarin, not the other way around. We acknowledge the contact between Nanning Cantonese and Standard Mandarin through SM promotion, media communication and migration, but these existing terminologies arising from the literature of language contact do not seem to accurately apply to Nanning Mandarin either.

Regardless of the undetermined terminology, among the controversies on the nature of such

non-standard variants, the widely accepted view emphasises two key points: 1) The direction is local varieties approximating Standard Mandarin. 2) Such varieties are variants of Standard Mandarin (Y. Chen, 1991; Z. Chen, 1990). Therefore, Jin (2004) concluded that the local variants of Standard Mandarin should apply to the “continuum model” of interlanguage, rather than the “hybrid model”. This claim clearly positions the two ends of interlanguage: the continuum model applies to a continuum from elementary L2 to standard L2, which can be graded by the SM proficiency test, while the hybrid model applies to a continuum from L1 to L2. This widely accepted view is reflected in the consensus among previous studies on Nanning Mandarin, although the implications of terms, either interlanguage or Local Mandarin, were often left implicit.

In this study, Nanning Mandarin (NM) refers to the non-standard local variant of Standard Mandarin (SM) spoken by native speakers of the traditional regional varieties, primarily Nanning Cantonese (NC). Critically, it is a fossilised intermediate variety of language acquisition, taking Nanning Cantonese as the source language and Standard Mandarin as the target language. The terms ‘interlanguage’ and ‘Local Mandarin’ can still be used as long as they clearly follow the classification and definition specified in this study.

### **Characteristics**

Interlanguages are systematic and highly structured, but also unstable and vague at the same time.

**Systematicity and fossilisation.** Interlanguages usually have a relatively stable system within a synchronic time phase. It is a relatively independent language system that systematically differs from its native and target languages by rule. This brings up a related concept, ‘fossilisation’, proposed by (Selinker, 1972). Fossilisation refers to the group phenomenon that an interlanguage pauses its approximation to the target language norm, with interlanguage items remaining in the resultant variety. Fossilisation manifests primarily in phonetics, to a lesser extent in the lexicon and least in grammar (Y. Chen, 1991). Nevertheless, the fossilisation of Nanning Mandarin has been mitigated among the younger generation: an increasing number of younger speakers is able to adopt a more

standard variety given the language policy and language contact, and are in favour of switching to Standard Mandarin in formal situations, or even more, in expanding this usage to daily speech.

**Instability and variability.** On the other hand, such a language system is also dynamic and highly variable. Local Mandarin in itself is a language system that is always under approximation to Standard Mandarin (R. Li, 1988). In a beginner–standard continuum, any intermediate variety could be counted as ‘Local Mandarin’, which contains various developmental phases and SM proficiency levels (Yao, 1989). Moreover, the specific constitutions of regional varieties and Standard Mandarin vary by place, and even more, by speaker (R. Li, 1988). Local Mandarin is such a phenomenon that is shaped by numerous social and sociolinguistic factors, including regional varieties, register, idiolect, age, gender, education, addressee, and language context (Yao, 1989). The high degree of variability results in an unstable language system. Since pitch dynamics (tone and intonation) are the main focus of this study, a more detailed introduction to the pitch systems of Nanning Mandarin will be provided in the following sections.

Nanning Mandarin, rather than an official term, is what the Nanning locals refer to the variety. It in fact illustrates people’s self-awareness of the uniqueness of these varieties. With the internet developing, Nanning Mandarin is also increasingly appearing on social media and has been recognised by people from other parts of China.

### **1.1.3. Existing literature on Nanning Mandarin and Nanning Cantonese**

Despite the large population of speakers and the unique linguistic features, Nanning Mandarin and Nanning Cantonese have been understudied and underrepresented, especially compared to the plentiful studies on Standard Mandarin and standard Cantonese. Although Nanning Cantonese is rich in traditional documentation and Nanning Mandarin has been drawing more academic attention as a new variety, existing studies are mostly no more than descriptive accounts of linguistic phenomena, lacking systematic and robust linguistic investigations and analyses.

#### **1.1.3.1. Nanning Mandarin**

In recent years, Nanning Mandarin has gradually drawn attention, but has often been treated as more of a cultural or regional phenomenon than an object of linguistic research. Overall, previous studies are rare.

Chapters in two edited volumes on languages in Guangxi—one on language use (F. Qin, 2005) and the other on SM promotion (Y. Lin, 2008)—have described and discussed the sound, vocabulary and grammar of Nanning Mandarin. Other individual studies have discussed various aspects of Nanning Mandarin, including its sound system and prosody (Fan, 2020; Flaming, 2020; Y. Wu, 2017; Q. Tang & Liu, 2016; T. Li, 2012; Y. Liang, 2005; H. Yang, 2003), lexicon (Lu, 2012; Shen F., 2005; Shen X., 2000; L. Zhang, 2022), sentence-final particles (Lan, 2007; Y. Li, 2012), as well as language formation and attitudes (Huang N., 2018; X. Li, 2015; Y. Li, 2012; C. Lin, 2015; Z. Wang, 2023; Yu, 2008).

The majority of the studies focus on phonetics and phonology, or report on language use. A timeline of previous research shows that whilst Nanning Mandarin began to draw academic attention in the early 2000s, concentrated investigations on specific linguistic aspects did not appear until the 2010s. Even so, these investigations were typically based on comparisons with Standard Mandarin, and their findings were often framed as ‘failures’ to acquire the standardised variety. In other words, Nanning Mandarin was measured by its deviation from the benchmark of Standard Mandarin within a comparative framework.

In 2021, Lyu and Yu (2021) published the first review paper on existing literature on Nanning Mandarin, proposed the limitations and according recommendations. Since 2015, the establishment of the Chinese Language Resources Protection Project Collection and Display Platform also provided audio and video recordings of the Local Mandarin of Nanning Cantonese and Nanning Pinghua, effectively Nanning Mandarin.

### **1.1.3.2. Nanning Cantonese**

Nanning Cantonese, as the traditional variety that was numerically dominant in Nanning, has been well-documented in terms of sound system, lexicon and grammar (H. Yang, 1997; Y. Lin & Qin, 2008; Q. Tang, 2020; 广西壮族自治区地方志编纂委员会 [Guangxi Zhuang Autonomous Region Local Documentation Compilation Committee], 1998, GX Doc hereafter). Specifically, the sound system has been systematically documented by scholars and the government since the late 20th century; the rich literature includes (e.g., Hong, 1989; Xie, 1994; H. Yang, 1997; Xie, 2007; 南宁市地方志编纂委员会 [Nanning Municipal Local Records Compilation Commission], 1998, NN Doc hereafter). Audio and video recordings were also collected by Yang (1997) and the Chinese Language Resources Protection Project Collection and Display Platform 中国语言资源保护工程采录展示平台<sup>9</sup>. LeiMaau, a native Nanning personal contributor, organised, digitised, and visualised the existing literature, and developed a useful booklet and several tools for Nanning Cantonese<sup>10</sup>.

Apart from descriptive literature, analyses and investigations mainly focused on language use and attitude (H. Chen & Li, 2005; Huang N., 2018; C. Lin, 2015; Z. Wang, 2023), and on typology and language contact based on evidence of historical phonology, syntax and lexicon (e.g., de Sousa, 2015, 2022; de Sousa et al., 2015; Kwok, 2010, 2014; Kwok et al., 2011). As tone and intonation are the main focus of this thesis, previous studies on the tonal and intonational systems of Nanning Mandarin and Nanning Cantonese will be introduced in detail in § 1.2 and § 1.3.

## 1.2. Tones of Chinese varieties

This study investigates the pitch systems, that is, tone and intonation, of Chinese varieties. Therefore, above all, we will introduce the terms of tone and pitch first. In the book *Tone* (Yip, 2002), Yip made a clear discussion and distinction of the terminology. **Lexical tone** refers to a suprasegmental phonological category, mainly pitch, that can distinguish word meaning. The acoustic correlate of

---

<sup>9</sup> zhongguoyuyan.cn

<sup>10</sup> Booklet: <https://leimaau.github.io/book/>; online dictionary: <https://leimaau-webdict3.vercel.app/>.

tone is primarily **fundamental frequency (f0)**, i.e., the number of pulses per second produced by vocal fold vibration, along with voice quality (e.g., creaky voice). While f0 measures tone height in hertz (Hz) from the acoustic side, the term **pitch** refers to the tone height in the hearer’s perception, that is, the psychophysical correlate of f0 (Ladd, 2008). In phonetic research, raw f0 values (Hz) are typically **log-transformed** to better approximate perceived pitch. Accordingly, throughout this thesis, the term “pitch” is used largely interchangeably with log-transformed f0.

Languages that utilise lexical tone are therefore called tone languages (or tonal languages), which is one principal characteristic of Chinese varieties. The discussions and descriptions of Chinese tones usually involve two additional terms, tone categories and tone patterns. Tone categories refer to the phonological grouping, and tone patterns refer to the phonetic realisations. Due to language evolution, one tone category can be realised largely differently across modern Chinese varieties. To better understand these terms, § 1.2.1 will first introduce the historical tone categories in Middle Chinese, and § 1.2.2 will illustrate the tone systems of Standard Mandarin and Nanning Mandarin, and how these two modern varieties correspond to reflect the relationship between their tone categories.

### 1.2.1. Historical tone categories

Index	Category	Gloss	Derived tones	
			Voiceless (a)	Voiced (b)
I	<i>píng</i>	‘level’	<i>yīn píng</i>	<i>yáng píng</i>
II	<i>shǎng</i>	‘rising’	<i>yīn shǎng</i>	<i>yáng shǎng</i>
III	<i>qù</i>	‘departing’	<i>yīn qù</i>	<i>yáng qù</i>
IV	<i>rù</i>	‘entering’	<i>yīn rù</i>	<i>yáng rù</i>

**Table 1.2.1:** Middle Chinese tone categories; adapted from M. Y. Chen (2000).

The tone categories of modern Chinese varieties are inherited from and systematically correspond to the Middle Chinese tone categories.

Four main tone categories have been firmly established in Middle Chinese tonology, i.e., *píng* (I), *shǎng* (II), *qù* (III) and *rù* (IV). However, the specific tone values remain undocumented, and the reason for the traditional nomenclature is controversial, with only subjective descriptions of the pitch height and contour.

Two dichotomies are used to group the four primary tone categories from different aspects: *píng* and *zè*, *shū* and *cù*. *Píng* and *zè* describe the tones based on tonal patterns, where *píng shēng* ('level tones') refers to the level tone (I), while *zè shēng* ('oblique tones') refers to the non-level tones, including the contour tones (II, III) and the short, abrupt tone (IV). On the other hand, *shū* and *cù* describe the syllables, and thus the tones, based on the syllabic coda. Syllables with a stop coda (usually unreleased /p, t, k/ or /ʔ/) are called "checked" syllables, while the non-checked syllables (either an open syllable or with a nasal coda) are called "smooth" syllables. The tones carried by checked syllables are called "checked" tones or *cù shēng* (literally 'short and abrupt tones'), while the other three non-checked tones (I, II, III) are called "smooth" tones. The *rù* tone (IV) occurs exclusively with checked syllables and thus is often referred to as the "checked tone".

Along with the language evolution, the four primary tone categories have undergone a tone split by the voicing contrast in the syllabic onset. Each of the primary tones was further split into two registers: a high register with voiceless onset, and a low register with voiced onset, known as *yīn* (a) and *yáng* (b) in the traditional nomenclature. This yielded a systematic eight-tone system in Table 1.2.1, which is still sustained in varieties such as Cantonese. In some varieties, certain tones can be further split by the contrast on aspiration in the syllabic onset or on the length or quality of the vowel. For example, in Cantonese, the checked tone *yīn rù* (IVa) was further split into an 'upper' tone for short (lax) vowels and a 'lower' tone for long (tense) vowels, as the example in Table 1.2.3.

Major tone mergers during the evolution of many modern varieties include the merger of Tone IIb to

Tone III and the redistribution of Tone IV. Tone IIb with a voiced obstruent onset were merged with Tone III. The redistribution of checked tones (IV) is more complex. First, Tone IV was merged with the other three primary tones depending on the voicing in the syllabic onset: Tone Ib (*yáng píng*) if with a voiced obstruent onset, Tone III if with a sonorant onset, and Tone II if with a voiceless onset. The previously checked tones that are now in Tone II were further redistributed to Tone Ia, Ib and III, without explicit phonological rules. As a result, Tone IV has disappeared in many modern varieties (e.g., Standard Mandarin), but has been fully and systematically preserved in Cantonese.

## **1.2.2. Tone systems of Standard Mandarin, Nanning Cantonese, and Nanning Mandarin**

### **1.2.2.1. Tonal systems**

This section introduces the tone systems of Standard Mandarin (SM), Nanning Cantonese (NC), and Nanning Mandarin (NM) in the literature, followed by a summary of the systematic correspondence between the tone categories. Unlike the SM tone system, which has been well-established and extensively studied in literature (e.g., W.-S. Lee & Zee, 2003) and Modern Chinese textbooks by Peking University (2012; SM Textbook), studies on NC and NM tones are comparatively scarce (as discussed in § 1.2), consisting primarily of auditory-based phonological descriptions, with only a few small-scale acoustic analyses.

Table 1.2.2 displays the tone systems of Standard Mandarin and Nanning Mandarin; Table 1.2.3 for Nanning Cantonese. Standard Mandarin has four tone categories, while Nanning Cantonese has nine tone categories. The phonetic description for tones adopts the numeric stave tone system designed by Chao (1930). In this numeric system, the pitch space is divided into a five-degree scale, where 1 represents the lowest and 5 represents the highest. Multiple digits are used to depict a tonal contour, for example, [55] indicates a high-level tone, while [214] indicates a dipping tone starting low and ending high. Tone IV carried by the checked syllables are conventionally indicated by a single tone digit due to the short syllabic duration.

Based on the traditional nomenclature, contemporary varieties largely follow the historical tone categories (in the ‘Historical category’ column); adjustment is necessary only when a tone merger has happened to the majority of morphemes in a tone. As a result of the tonal evolution, it is common to see that tones named by the same historical category are realised in distinct tonal patterns across regional varieties, such as the *yáng píng* (Ib) category in Standard Mandarin and Nanning Cantonese. For easier reference, the tone categories are also labelled by Arabic numbering (in the ‘Tone index’ column), in accordance with the order of the historical categories.

For both language varieties, each tone is presented with an example syllable. The selected syllables are mostly /fu/. In Nanning Cantonese, there is a tonotactic gap in Tone IIb where no /fu/ syllable is available; a syllabic /mu/ was therefore chosen. For Tone IV associated with the checked syllables, the systematic mapping of the vowels (lax or tense), syllabic obstruent coda (/k/ or /t/) and the tone altogether determine the available options. The morpheme 阔 ‘vast’ was chosen primarily for Nanning Cantonese for consistent phonetic context; in Standard Mandarin, the pronunciation is less similar due to major sound change.

Historical category	Tone index	Tone value (SM)	Tone value (NM)	Example
<i>yīn píng</i> (Ia)	1	55	33 or 44	夫 /fu/ ‘man’
<i>yáng píng</i> (Ib)	2	35	24 to 34	扶 /fu/ ‘to hold’
				福 /fu/ ‘fortune’ 服 /fu/ ‘clothes’
<i>shǎng</i> (II)	3	214	21	府 /fu/ ‘mansion’ 母 /mu/ ‘mother’
<i>qù</i> (III)	4	51	53 or 55	富 /fu/ ‘rich’
				父 /fu/ ‘father’ 阔 /kuo/ ‘vast’

**Table 1.2.2:** Lexical tone systems of Standard Mandarin and Nanning Mandarin.

Tone values are from Lee and Zee (2003) and Lin (2008), respectively.

Historical category	Tone index	Tone value	Example
<i>yān píng</i> (Ia)	1	55	夫 /fu/ ‘man’
<i>yáng píng</i> (Ib)	2	21	扶 /fu/ ‘to hold’
<i>yān shǎng</i> (IIa)	3	35	府 /fu/ ‘mansion’
<i>yáng shǎng</i> (IIb)	4	24	母 /mu/ ‘mother’
<i>yān qù</i> (IIIa)	5	33	富 /fu/ ‘rich’
<i>yáng qù</i> (IIIb)	6	22	父 /fu/ ‘father’
<i>upper yān rù</i> (IVa-i)	7	5	福 /fɔk/ ‘fortune’
<i>lower yān rù</i> (IVa-ii)	8	3	阔 /fut/ ‘vast’
<i>yáng rù</i> (IVb)	9	2	服 /fɔk/ ‘clothes’

**Table 1.2.3:** Lexical tone system of Nanning Cantonese. Tone values are from Lin and Qin (2008).

### 1.2.2.2. Correspondence between tone categories

Table 1.2.4 displays the lexical tone systems of the two Mandarin varieties and Nanning Cantonese, with tone categories aligned according to the historical correspondence. Example words that are not consistent with the syllable /fu/ are due to sound change, similar to the cases in Table 1.2.2 and Table 1.1.3. Tone categories were divided by the solid line to separate the tones evolving from historically smooth tones (I to III) and checked tones (IV).

From Table 1.2.3 and Table 1.2.4, we can see that NC tone system is conservative with respect to retaining the full set of eight-tone categories (even with an additional split), whereas Standard Mandarin has undergone major mergers. Therefore, Nanning Cantonese still generally reflect the full tonal categories in Middle Chinese. Moreover, the systematic correspondence between tone categories of Standard Mandarin and Nanning Cantonese is largely consistent with the major traces of tonal development from Middle Chinese to Standard Mandarin. For example, the complex redistribution of historical Tone IV in modern varieties (details in § 1.2.1) can be observed in

Standard Mandarin: Tone IVa in Nanning Cantonese corresponds to each of the four SM tones, and Tone IVb in Nanning Cantonese corresponds to either Tone Ib or Tone III by syllabic onset (voiced obstruent or sonorant).

Standard Mandarin and Nanning Mandarin				Nanning Cantonese			
Historical category	Tone index	Tone value (SM)	Tone value (NM) Example	Historical category	Tone index	Tone value	Example
Ia	1	55	33/44 夫 /fu/ 'man'	Ia	1	55	夫 /fu/ 'man'
Ib	2	35	24/34 扶 /fu/ 'to hold'	Ib	2	21	扶 /fu/ 'to hold'
II	3	214	21 府 /fu/ 'mansion' 母 /mu/ 'mother'	IIa	3	35	府 /fu/ 'mansion'
				IIb	4	24	母 /mu/ 'mother'
III	4	51	53/55 富 /fu/ 'rich' 父 /fu/ 'father'	IIIa	5	33	富 /fu/ 'rich'
				IIIb	6	22	父 /fu/ 'father'
Ia	1	55	33/44 忽 /xu/ 'sudden'	IVa-i	7	5	忽 /fət/ 'sudden'
Ib	2	35	24/34 福 /fu/ 'fortune'				福 /fɔk/ 'fortune'
II	3	214	21 法 /fa/ 'law'	IVa-ii	8	3	法 /fat/ 'law'
III	4	51	53/55 阔 /kuo/ 'vast'				阔 /fut/ 'vast'
Ib	2	35	24/34 服 /fu/ 'clothes'	IVb	9	2	服 /fɔk/ 'clothes'
III	4	51	53/55 木 /mu/ 'wood'				木 /fɔk/ 'wood'

*Source:* Corresponding relationship according to Y. Lin & Qin (2008), Wan, (2012), and Z. Tang (2016).

**Table 1.2.4:** Correspondence between tone categories of the two Mandarin varieties and Nanning Cantonese.

### **1.2.2.3. Nanning Mandarin tones**

In this thesis, Nanning Mandarin, as a variant of Standard Mandarin, inherited the four tone categories from Standard Mandarin and thus applies the same correspondence to NC tones as Standard Mandarin does. Therefore, our investigation only focused on the phonetic realisations of tones. The reasons are as follows:

(1) The majority of the NM speakers do not show misalignment between the tone categories of Nanning Mandarin and Standard Mandarin, according to previous literature (F. Qin, 2005; Yu, 2008), which is consistent with the researcher's observation as a native speaker. Qin (2005) and Yu (2008) categorised the deviation of interlanguage from Standard Mandarin into "erroneous" (phonological) and "imperfect" (phonetic) realisations; this categorisation was borrowed from the rating system of the SM proficiency test. They proposed three developmental phases (low, median, high) with regard to the approximation of interlanguage to Standard Mandarin, that is, with low to high levels of SM proficiency. When with low SM proficiency, interlanguage speakers have phonological misalignment to the SM tone categories; while with mid to high SM proficiency, they were able to apply the SM tone categories, and only the phonetic patterns were different. According to the investigation of language use, the majority of interlanguage speakers in Guangxi were in their median phase.

(2) This also aligns with the observed features described in the proposed theoretical frameworks for non-standard Mandarin varieties. The literature has mostly agreed that such varieties usually focus more on the analogy of phonological categories from Standard Mandarin, whereas the phonetic realisations from regional varieties were borrowed and henceforth retained (Y. Chen, 1991; Chen Z., 2007; R. Li, 1988).

### **1.2.3. Tones in contexts**

In this section, we will also briefly introduce the contextual tones in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin. Although contextual tones are not the main focus and research

objects of this thesis, they are relevant to the material design of disyllables and sentences. Therefore, a brief introduction will help the readers to gain a better understanding of the materials used in this thesis.

### **1.2.3.1. Neutral tones**

In **Standard Mandarin**, a syllable can be realised with a short and light production, leading to the loss of its original citation tone. This is traditionally called *qīng shēng* (literally, ‘light tone’) or neutral tone (Y. R. Chao, 1968). Neutral tone is realised in weak stress and mostly occurs following a stressed syllable (i.e., enclitic) (Y. R. Chao, 1968). Acoustically, neutral tone syllables are correlated with largely shortened duration, compressed pitch range, and weakened vowels (neutralisation or loss), sometimes with voicing of voiceless consonants (SM Textbook, 2012).

Neutral tone does not have a fixed pitch; it has been considered as a contextual tone with a pitch heavily determined by its preceding tone (Chao, 1968; M. Y. Chen, 2000; M. Yip, 2002). In terms of pitch height, (Y. R. Chao, 1968) noted that the neutral tone is realised with the highest pitch after Tone 3 (half-high), a middle pitch after Tone 2, a half-low pitch after Tone 1, and a low pitch after Tone 4. In terms of tone contour, neutral tone is falling after SM Tone 1, Tone 2 and Tone 4, and level after SM Tone 3, according to auditory description (Y. R. Chao, 1968) and acoustic analysis (W.-S. Lee & Zee, 2008).

Neutral tone has been highly constrained by syntax and semantics. Certain syntactic constituents are obligatorily realised in neutral tones, such as particles and suffixes, which only contain a very small number of morphemes. For certain words, realisation in full citation tone or neutral tone will differentiate word meanings.

Cantonese, however, does not have such a neutral tone for unstressed syllables (Bauer & Benedict, 1997; Y. R. Chao, 1968), and the same is true for **Nanning Cantonese** (Hong, 1989). Similarly, **Nanning Mandarin** also does not have neutral tone (Fan, 2020; Liang G., 1994; Y. Lin, 2008; H. Yang, 2003, 2008), and existing research has rarely investigated the phonetic realisations of

morphemes that are neutral-toned in Standard Mandarin. An intonation study (H. Yang, 2003, 2008) investigated such non-particle morphemes in the sentence-final position, finding that they showed a high f<sub>0</sub> and level contour without weakening. The study, therefore, concluded that Nanning Mandarin realised these morphemes in a higher pitch than their neutral-toned counterparts in Standard Mandarin.

### 1.2.3.2. Tone sandhi and tone changes

A succession of tones will lead to allophonic and phoneme changes on certain tones, known as **tone sandhi** (Y. R. Chao, 1968). Tone sandhi in Standard Mandarin has been extensively established in literature and textbooks (M. Y. Chen, 2000; SM Textbook, 2012). In Standard Mandarin, all four tones have tone sandhi, among which Tone 3 sandhi is the most salient one, as it changes the canonical citation tone pattern. Tone 3 is realised in the full dipping form [214] only in pre-pausal positions, i.e., sentence final, phrase final, or monosyllables. When SM Tone 3 is followed by another Tone 3, it is realised as [35] (same as SM Tone 2); otherwise, it is realised as low-level [211] (or [21] for simpler notation).

- (1) a. T3 → 35 / \_\_\_ T3
- b. T3 → 21 / \_\_\_ T1/T2/T4

When the tone following SM Tone 3 is a neutral tone, the same rule applies depending on the underlying form of the neutral tone. When in a multisyllabic context, Tone 3 sandhi will also be governed by morphosyntactic rules (phonological phrase or prosodic phrase) and affected by complex factors such as pauses, prominence, syntactic structures and intonation.

An extended discussion of the phonological rule (1) is the contentious underlying target of SM Tone 3. Increasingly, researchers believe that the true nature of SM Tone 3 is a low-level tone, in terms of phonetic production, phonological contrast, and perception (F. Shi & Ran, 2011; Zhou, 2017). It was stated in the textbooks that the low-level portion is the main body of the tone, so the tonal

transcription [214] is actually simplified from more accurate [2114], which displays a fall–level–rise contour (SM Textbook, 2012). The rising tail, on the other hand, is conditional and thus predictable (F. Shi & Ran, 2011; Yip, 2002), making the dipping form [214] a variant rather than the underlying form.

In addition, a couple of morphemes (e.g., 一 /ji<sup>55</sup>/, 不 /bu<sup>51</sup>/) also undergo tone sandhi, some of which are in free variation (SM Textbook, 2012).

For Nanning Mandarin, existing studies have not explicitly mentioned tone sandhi. According to the researcher’s observation, Nanning Mandarin at least inherits the Tone 3 sandhi from Standard Mandarin.

Cantonese, however, only has a limited set of tonal changes (Féry, 2016b) and does not have systematic tone sandhi (Y. Lin & Qin, 2008). In Cantonese, **tonal variation or tone change** (变音 biànyīn; 变调 biàndiào) can happen to certain morphemes for pragmatic purpose. By changing to high-level Tone 1 [55] or high-rising Tone 3 [35], the realisation in high tones can show a diminutive (intimacy, affection, or softening) function. These instances are not very productive and thus can be exhaustively listed. As demonstrated in (2), the low-level Tone 2 [22] changes to high-level Tone 1 [55] (Y. Lin & Qin, 2008).

- (2) 妹 ‘younger sister’      妹仔 ‘girl’  
       [**mui**<sup>22</sup>]                      → [bui<sup>55</sup> tsei<sup>35</sup>]

In certain morphemes where segmental elision is permitted, the associated tones are preserved but undergoes tonal modification. As demonstrated in (3), the checked syllable [mæt<sup>5</sup>] first loses its stop coda and alters the vowel to assimilate to [jɛ<sup>24</sup>], the bearing tone thus changes from a short checked tone [5] to a smooth, non-checked tone [Tone]. This elision can progress further, where the segments of [jɛ<sup>24</sup>] are fully dropped and the tone [24] floats to the preceding syllable [55], resulting in a surface tone realisation of [535] (Y. Lin & Qin, 2008). This is accounted for by tone stability under the

autosegmental framework in (M. Y. Chen, 2000).

(3) 乜嘢 ‘*what*’

[mɛt<sup>5</sup> jɛ<sup>24</sup>] → [mɛ<sup>55</sup> jɛ<sup>24</sup>] → [mɛ<sup>535</sup>]

In addition, a few words display tone assimilation or dissimilation. As demonstrated in (4), the tone of [ny<sup>24</sup>] assimilates to [tsɛi<sup>35</sup>] and also becomes [35] (Y. Lin & Qin, 2008).

(4) 女 ‘*daughter*’          崽女 ‘*son and daughter*’

[ny<sup>24</sup>]                      → [tsɛi<sup>35</sup> ny<sup>35</sup>]

### 1.2.3.3. Alternative pronunciations

In Nanning Cantonese, there are characters associated with more than one set of pronunciations, traditionally referred to as *yì dú* 异读 (literally, ‘variant pronunciation’). These variations are mainly systematic, with a limited number of irregular ones.

The systematic variations are primarily the ones termed 文白异读 (literally, ‘variation in the literary and colloquial pronunciations’). The ‘literary’ pronunciation refers to the set typically used in reading texts and spoken phrases from the older literature, while the ‘colloquial’ pronunciation is the set used in everyday speech (Féry, 2016b). One major alternation rule for tone is that, tone originally associated with the historical low register (b, *yáng*) changes to its high register (a, *yīn*) counterpart in colloquial reading, while it remains its original tone in literal reading (Y. Lin & Qin, 2008). As demonstrated in (5), the citation tone of [mɛn<sup>21</sup>] (Tone IIb) is preserved in the literal reading, but changes to [55] (Tone IIa) in the colloquial reading.

(5) Literal reading          Colloquial reading

文化 ‘*culture*’          一文钱 ‘*one [classifier] money*’

[mɛn<sup>21</sup> fa<sup>33</sup>]          → [jɛt<sup>5</sup> mɛn<sup>55</sup> ts<sup>h</sup>in<sup>21</sup>]

However, the two sets are not so absolutely distinct in the two separate registers or linguistic domains, as it may be suggested by name; instead, they actually overlap with each other (Féry, 2016b; Y. Lin & Qin, 2008). In addition, the syllable (morpheme) bearing such alternative readings often presents the differences when forming different words, instead of forming contrasts within the same word, as in the minimal pairs. Such multiple readings sometimes even vary by individual (Y. Lin & Qin, 2008).

This variation in fact reflects the influence of Guangfu Cantonese, Zhuang, and other regional varieties on Nanning Cantonese. The literary reading is generally inherited from the modern Guangzhou Cantonese, while the colloquial reading is mainly inherited from early Guangfu Cantonese, along with some irregular items traceable to Zhuang and other language varieties.

#### **1.2.4. Previous studies on tones in Nanning Mandarin and Nanning Cantonese**

In this section, we will introduce the previous findings on the tones of Nanning Mandarin and Nanning Cantonese.

##### **1.2.4.1. Nanning Cantonese**

Compared to the newly emerged Nanning Mandarin, the tone system of Nanning Cantonese was relatively richer in traditional documentation using auditory transcriptions, along with initial empirical research using acoustic-phonetic methods.

NC reference	T1	T2	T3	T4	T5	T6	T7	T8	T9	Comments
Xie (1994)	55	21	35	<b>13</b>	33	22	5	3	2	Old NC
Hong (1989)							5	3/33	2/22	
H. Yang (1997)										n=2, old NC
GX Doc (1998)							55	33	22	
NN Doc (1998)										Old NC
S. Li (2002)	55	21	35	24	33	22				
Y. Lin & Qin (2008)										
H. Chen & Lin (2009)							5	3	2	
Teng (2018)										n=1
Tang (2020)										fieldwork
Xie (2007)	55	21	35	<b>23</b>	33	22	55	33	22	fieldwork
Shi S. (2007)	55	<b>31</b>	35	<b>34</b>	<b>44</b>	<b>32</b>	5	<b>4</b>	<u><b>32</b></u>	n=2, acoustic old NC
S. Wang (2015)	55	21	35	24	33	<b>31</b>	55	33	<b>21</b>	n=1, acoustic
T. Luo (2015)	55	<b>31</b>	35	24	33	<b>21</b>	<u>55</u>	<u>33</u>	<u>21</u>	n=6, acoustic

**Table 1.2.5:** Summary of the tone values and descriptions (if tone values were not provided) of each of the four NM tones in the aforementioned previous studies. Bold indicates inconsistencies with the majority.

### **Auditory descriptions: traditional documentation**

From Table 1.2.5, we can see that the majority (nine) of the auditory transcriptions agreed on the tone values of the non-checked tones (Tone 1 to 6), and all of them agreed on the transcriptions of the checked tones, with minor discrepancies in notation. Since the checked tones are not the main focus of this thesis, we will not discuss the checked tones in detail hereafter. Overall, all auditory studies agreed on the tonal descriptions of all NC tones except for NC Tone 4. Although all of them identified NC Tone 4 as a low-to-mid rising tone, the specific tone values showed discrepancy.

Most of these studies did not provide explicit methodological descriptions. Yang (1997) and Teng (2018) briefly mentioned the recruited speakers; Xie (2007) and Tang (2020) conducted the descriptions based on fieldwork. Among the nine studies that reported identical transcriptions, three are associated with the same researcher, Lin Yi (林亦): Lin was a co-author of Lin and Qin (2008), the contributor of the NC data in Chen & Lin (2009), and one of the two speakers in Yang (1997). This may have led to potential homogeneity in the NC data.

### **Acoustic analyses**

While auditory analysis is valuable in providing “a sensible heuristic to the perceived tone pattern” (C. Xu & Zhang, 2024), acoustic analysis is undoubtedly essential. Since the late 2000s, acoustic approaches, such as log-transform, have been applied to investigations on monosyllabic tones.

Using the existing audio recordings collected by Yang (1997) from two speakers, Shi (2007) obtained the phonetic transcriptions using two methods of  $f_0$  normalisation: Log T-value transforms (logarithm) and fraction of range (FOR; linear) transforms. Shi concluded that the acoustic transcriptions largely agreed with the auditory literature, although the reported tonal values for NC Tones 2, 4, 5 and 6 were different. Moreover, Shi stated that Log T-values were closer to perceived NC tones than FOR. Therefore, only the Log T-values were listed in Table 1.2.5.

Wang (2015) and Luo (2015) investigated the tonal acoustics based on their self-collected data. Wang (2015) analysed the tone production from one young speaker, while Luo (2015) examined the tone productions from six speakers that were balanced in gender and evenly distributed across age groups (young, mid, old). The tone patterns and values reported in both studies were largely consistent with the traditional transcriptions, except for Tone 6, which will be discussed later.

An additional study (X. Li, 2018) investigated the monosyllabic and disyllabic tones of Cantonese spoken in Yongning District of Nanning. The narrow reference of Nanning Cantonese (i.e., Cantonese spoken in the city centre) usually excludes Yongning, due to its suburban location and its historical administrative status. Previously, Yongning was administered as a county within

Nanning's jurisdiction, rather than being part of the city proper. Despite its subsequent reclassification as a district in Nanning, the narrower terminological scope of Nanning Cantonese persists. Nevertheless, Yongning Cantonese, also a member of Yong–Xun subgroup, shares the same tone categories of Nanning Cantonese with minor differences in tonal patterns. Li (2018)'s acoustic findings from six speakers also supported this traditional observation.

## **Summary**

Overall, all three acoustic studies on Nanning Mandarin reached an agreement on the tonal patterns and Chao tone numbers of Tone 1 [55] and Tone 3 [35], which was also consistent with the auditory studies. For Tones 2, 4, and 5, although the tone values showed discrepancies not only within the acoustic studies but also compared to the auditory studies, all of the previous studies agreed on the tonal patterns: low-falling for Tone 2, low rising for Tone 3, and mid-level for Tone 5. The most salient misalignment between the acoustic and auditory transcriptions occurs in Tone 6: acoustic studies identified a falling contour, different from the low-level tone unanimously reported in the auditory literature. This in fact introduces the well-noted merger of NC Tone 6 to Tone 2, which was first mentioned in Hong (1989) and has since been noted in younger speakers' speech by subsequent studies (i.e., “new” Nanning Cantonese) (Y. Lin & Qin, 2008; T. Luo, 2015; S. Wang, 2015; NN Doc, 1998).

The division of “old” and “new”, though mostly determined by speaker's age, is unclear and varied in previous studies. Therefore, this division is unreliable and ambiguous in describing the variation of Nanning Cantonese. Nevertheless, we can at least figure out what the researchers were referring to when they used the terms “old” or “new”. Xie (1994), Yang (1997), and NN Doc (1998) clearly stated that they transcribed the “old” Nanning Cantonese; the acoustic transcriptions in Shi (2007) are therefore also “old”. On the other hand, it was made explicit that Wang (2015) investigated the “new” Nanning Cantonese, while Luo (2015) and Li (2018) compared the “old” and “new” Nanning Cantonese. According to the participants' demographics provided in these studies, we can roughly summarise that a tone merger was not observed in speakers born before the 1980s, but did occur in

speakers born since the 1980s and continues to develop in the younger generation's speech.

In acoustic studies, Luo (2015) compared the pitch, duration and intensity between Tone 2 and 6 across the age groups (young, mid, old), finding that all three acoustic correlates showed the strongest distinction between tones in the old speakers and weakest in the young speakers. A similar finding was also reported in Wang (2015) from one of the young speakers in Luo (2015), and Li (2018) from four young speakers and two old speakers of Yongning Cantonese. Nevertheless, these comparisons were not supported statistically. Xian and Yang (2022) calculated the acoustic distances between Tone 2 and 6 from 20 young speakers; the closer distance indicated tone merger in production.

Luo (2015) and Xian and Yang (2022) also conducted perception experiments: Luo (2015) by asking participants to discriminate their own productions, and Xian and Yang (2022) by asking listeners to discriminate one speaker's productions. Results of both studies showed that tone merger in perception occurred in over half of the participants.

Apart from the tone merger of NC Tone 6 to Tone 2, Xian and Yang (2022) also examined all possible tone pairs in Nanning Cantonese for potential merger. They found that the Tone 3–4 pair (both rising) showed a higher extent of merger than the Tone 2–6 pair in both production and perception. The merger of Tone 4 to Tone 3 was also indicated in Shi (2007)'s transcriptions. Yongning Cantonese was also reported to have this tone merger, especially from young speakers Li (2018), although the tonal distinction remained in multi-syllabic word contexts (Y. Lin & Qin, 2008). Xian and Yang (2022) also found that Tone 1–5 (both level tones) showed merger in perception but not in production.

#### **1.2.4.2. Nanning Mandarin**

Research on Nanning Mandarin started in the 2000s, later than that on Nanning Cantonese. Compared to the traditional tone studies of regional varieties, the instability and variability of non-standard Mandarin are expected to introduce more complexity when investigating the tones.

## **Auditory descriptions**

Few studies have directly examined the lexical tone systems of Nanning Mandarin, and they mostly rely on impressionistic auditory transcription. Qin (2005) briefly mentioned how NM tones deviated from SM tones, but did not report the tone inventory. A chapter from an early edited volume about the SM promotion in Guangxi Lin (2008) reported the tone inventory of the interlanguage between Nanning Cantonese and Standard Mandarin, which was effectively Nanning Mandarin; however, the methodology was not made explicit. In addition, Fan (2020) conducted an auditory analysis of the tone inventory in Nanning Mandarin mainly from one younger speaker and one mid-age speaker out of 10 recruited speakers, and found that the differences between generations were minor. Two additional studies have indirectly examined Nanning Mandarin through an auditory comparison of Nanning Cantonese and Standard Mandarin in terms of similar contour shapes (T. Li, 2012; Y. Liang, 2005).

Overall, all four auditory studies (Fan, 2020; T. Li, 2012; Y. Liang, 2005; Y. Lin, 2008) identified a level tone for NM Tone 1, a rising tone for NM Tone 2, and a low falling tone for NM Tone 3; however, they differed in the description of NM Tone 4. They also differed in the exact Chao tone numbers assigned to many of these tones; detailed tonal transcriptions can be referred to in Table 1.2.6.

Compared to the corresponding SM tones, these auditory studies agreed that NM Tone 1 (level) was lower. Tone 2 (rising) presented a narrow pitch range with lower onset and offset, and thus was located in lower pitch space; moreover, Li (2012) added that the lengthened low portion at the onset of Tone 2 led to a curlier contour. Tone 3, which was dipping in Standard Mandarin, lacked the rising tail and thus merely falling; nevertheless, Fan (2020) observed occasional realisations of the full dipping form by the young speaker. Similar to Tone 2, Tone 4 also presented a narrower pitch range and thus a flatter contour, although the motivation were not agreed on: Liang (2005) and Lin (2008) attributed to a higher offset of falling than SM Tone 4; Li (2012) attributed to a lower onset, but meanwhile used tone values of [53] or [54], which indicated a tonal pattern effectively consistent

to that of Liang (2005) and Lin (2008).

### **Acoustic analyses**

However, the tone system of Nanning Mandarin has not yet been systematically investigated using acoustic methods. Existing acoustic studies only investigated the tones in contexts, either phrases or utterances, rather than the citation tone inventory. One early intonational study (H. Yang, 2003) investigated the  $f_0$  realisations of sentence-final tones in spontaneous conversations. By summarising the proportion of tone patterns within each NM tone from four speakers, the findings were generally consistent with the auditory results; for NM Tone 4, a lower onset was observed. The tonal pattern of NM Tone 3 was mostly in agreement with the auditory results, but occasional instances of the dipping contour were also observed. Flaming (2020) investigated the  $f_0$  range of NM Tone 4 in phrases and sentences for possible confusion with NM Tone 1. By comparing NM Tone 4 produced by four speakers to SM Tone 4 from existing data, Flaming found that NM Tone 4 was realised with a narrower  $f_0$  range and thus a flatter contour than SM Tone 4 in various tonal contexts (i.e., before T4, after T3, after T4, and utterance final). Specifically, two-thirds to three-fourths of the falls that were predicted to cover a pitch range over six semitones in SM were only realised with a pitch range less than six semitones. Flaming also confirmed that NM Tone 4 was acoustically distinct from NM Tone 1.

However, Flaming's study mainly focused on  $f_0$  range (along with onset and offset) in semitones. Although pitch contours (not time-normalised) of some non-final NM Tone 4 were provided, tonal values in Chao tone numerals for the entire citation tone system were not provided. Moreover, among the four speakers, only two were born and/or raised in Nanning, and three of them did not report any local Nanning varieties as their L1 (two Hakka, one Yulin Cantonese<sup>11</sup>). Given their atypical backgrounds, it was unclear how representative their productions were of Nanning Mandarin.

---

<sup>11</sup> A member of Goulou subgroup of Cantonese.

## Summary

Table 1.2.6 summarises the tone values in Chao tone numerals and descriptions (if tone values were not provided) of each of the four NM tones from previous studies.

NM reference	Tone 1	Tone 2	Tone 3	Tone 4	Comments on method
Y. Lin (2008)	33/44	24 to 34/35	21	53/55	<b>Auditory</b> ; method implicit
Fan (2020)	55 (sometimes 33/44)	35 (sometimes 24)	21 (sometimes 31/41/42; occasionally 213)	51	<b>Auditory</b> ; n=2 (primary) out of 10
Y. Liang (2005)	44	24/35	21	53/54	<b>Indirect examination</b>
T. Li (2012)	between 33 and 55	Narrower pitch range and lower f0 level than SM Tone 2	21	53/54	<b>Indirect examination</b>
H. Yang (2003)	33/44	Mostly 12/13	falling without dipping	Mostly 31/42	<b>Intonation study</b> ; n=4
Flaming (2020)				Falling, flatter with a narrower range than SM Tone 4	<b>Intonation study</b> ; n=2 (primary) out of 4

**Table 1.2.6:** Summary of the tone values and descriptions (if tone values were not provided) of each of the four NM tones in the aforementioned previous studies.

Overall, the consensus among previous studies is that Nanning Mandarin has a narrower pitch range and thus flatter contour than Standard Mandarin (Fan, 2020; Flaming, 2020; T. Li, 2012; Y. Liang, 2005; Y. Lin, 2008; F. Qin, 2005; H. Yang, 2003). Specifically, NM Tone 1 (level) and 2 (rising) have lower pitch height. NM Tone 3 (falling) is mostly non-dipping, while NM Tone 4 shows divergent descriptions regarding both pitch height and contour. A similar observation was reported on sentence-final tones in an intonational study of Nanning Mandarin (H. Yang, 2003). This convergence among previous findings suggests that the phonetic realisations of NM tones are

systematically different from those of Standard Mandarin, even though features from both the target and source languages were observable.

Nevertheless, discrepancies still remain regarding the precise characterisations of the NM tone system. From Table 1.2.6, we can see that the specific phonetic transcriptions of the NM tones have not reached an agreement. All four auditory studies with tonal transcriptions (Fan, 2020; T. Li, 2012; Y. Liang, 2005; Y. Lin, 2008) indicated intermediate and unstable phonetic realisations between Nanning Cantonese and Standard Mandarin.

These studies (Fan, 2020; T. Li, 2012; Y. Liang, 2005; Y. Lin, 2008; F. Qin, 2005) conjectured that the identified patterns in NM tones can be explained as phonetic variation caused by interference of the L1 tone realisation; the mechanism were usually described from the perspectives of (partial) L1 transfer and deviation from Standard Mandarin. However, this point of view overlooked the complicated factors during the development of Nanning Mandarin, such as complex language contact, community- and region-level commonality, and stratum in sound change (Lyu & Yu, 2021).

### **1.2.5. Research gaps in tone research**

Based on the review in this chapter, significant gaps remain in our understanding of the understudied tone systems of Nanning Cantonese and Nanning Mandarin. As detailed in § 1.2.4, while existing research on Nanning Cantonese offers rich auditory documentations, acoustic investigations are very few and based on very small sample sizes (one to six participants). Research on Nanning Mandarin is even more insufficient, featuring discrepancies in tonal descriptions, small sample sizes (two to four primary participants) and lacking direct acoustic examination of its tone system. Furthermore, there has been no systematic, cross-linguistic comparison of Nanning Mandarin, its source language (NC), and its target language (SM) under controlled experimental conditions.

Consequently, the tone systems of both Nanning varieties lack a systematic description grounded in solid acoustic-phonetic approaches and robust statistical analysis. This reliance on descriptive reports

and limited data effectively makes the conclusions lack statistical validity and generalisability. It therefore remains unclear and fragmented how the NM tone system truly functions and how it relates to its source language (NC) and target language (Standard Mandarin).

### **1.3. Intonation of Chinese varieties**

Pitch is not only used in tone to convey lexical contrast, but also used to convey utterance or discourse-level contrast, known as intonation. Intonation is the pitch variation of speech that conveys pragmatic meaning at the sentence level, in other words, speech melody (e.g., Y. Chen, 2022; Féry, 2016a). It is a core element of prosody, although the term is sometimes used more broadly to include lexical-level contrasts (Y. Xu, 2019) or other suprasegmental aspects (Féry, 2016a; Ladd, 2008); in the latter case, the term is used interchangeably with prosody.

Intonation fundamentally uses  $f_0$  or pitch in a linguistically-structured way to signal a range of linguistic (or, post-lexical) meanings, such as prominence and sentence modality. The ways in which  $f_0$  is adjusted by intonation, however, can be largely language-specific. For example, when marking focus, Standard Mandarin extensively uses  $f_0$  with compression in the post-focus region (e.g., Y. Xu et al., 2012), while Cantonese shows the opposite (e.g., W. L. Wu & Xu, 2010); when signalling question intonation, Standard Mandarin primarily displays global pitch rising and largely retains the tone contour (Gu, 2016; J. Zhang et al., 2020), while Cantonese shows more significant rising at the sentence-final position that alters the tone identity (Gu et al., 2006).

This section will mainly introduce focus marking, one of the linguistic functions of intonation, with a focus on highlighting the variations by language and tone categories in its phonetic realisations.

#### **1.3.1. Focus**

##### **1.3.1.1. Definition and classification**

Focus has been defined differently under different theoretical frameworks, usually associated with

highlighting new information (e.g., Halliday, 1967), alternatives (e.g., Krifka, 2008), and emphasis for attention (e.g., Shi F. & Xia, 2022; Y. Xu, 2005).

The definition from various perspectives is highly related to the categorisation of various focus types. Focus types are usually divided based on the meaning they express, especially on information structure. Early studies in English have examined a list of focus types (Gussenhoven, 2007; Krifka, 2008). Among these, a key distinction is between focus used to correct wrong information (Gussenhoven, 2007) and focus that introduces new information (Halliday, 1967).

The terminology referring to focus types, however, is fairly inconsistent across different studies. For example, the focus that denotes new information, usually known as information focus, was referred to as emphatic focus in Xu (2005). An even more problematic case concerns the term ‘contrastive focus’, which is commonly used to refer to ‘narrow focus’ (Gussenhoven, 2007), but is also used to denote corrective focus (Chafe, 1974; Y. Xu, 2005) or alternative focus (Selkirk, 2008). This in fact reflects the overlap pragmatic meaning between correcting a piece of wrong information and rejecting an alternative; Yang (2022) therefore treated corrective focus as a subtype of alternative focus. Therefore, scholars have proposed to investigate and classify focus type in terms of paralinguistic functions (Y. Xu, 2019) or a continuum of informativeness (Shi F. & Xia, 2022).

On the other hand, when a sentence contains no emphatic pattern, or when an entire sentence is emphasised, such a condition is also denoted by various terms, usually broad focus (Ladd, 1980), but also neutral focus (Y. Xu & Xu, 2005) or natural focus (Shi F. & Xia, 2022). The terminology, especially the term broad focus, is controversial in the accuracy in focus size (or scope) and information structure (Gussenhoven, 2007; Katz & Selkirk, 2011; Ladd, 2008; Y. Xu, 2019); Katz and Selkirk (2011) therefore proposed to use the term all (discourse-)new instead of broad focus.

Acknowledging the debates on terminology, we will state the terms used in this thesis for clarity. We will use a combination of the traditional terms (narrow vs. broad focus) and information structure (new vs. given) to refer to the various types of narrow focus: focus that occurred with given

information will be termed as **new information focus** (denotes new information), **corrective focus** (corrects wrong information), and **alternative focus** (points out alternatives); while the broad, sentential focus with all-new information will be referred to as **broad focus**.

### **1.3.1.2. Cross-linguistic variations in f0 realisations**

Languages can take an ensemble of acoustic cues to mark focus; the cue weighting and acoustic variations, however, can greatly differ across languages, even across closely related Chinese varieties, although they are all tonal languages. Standard Mandarin uses f0 as a major cue for focus marking (Shi F. & Xia, 2022), but also adjusts duration, intensity and hyper-articulation (Y. Chen & Gussenhoven, 2008; Shih, 1988; J. Zhang et al., 2020). Cantonese, on the other hand, mainly relies on duration and intensity to signal focus, while f0 is not a reliable acoustic cue (K. Li, 2024; W. L. Wu & Xu, 2010; Y. Yang, 2022). As this thesis focuses on f0 characteristics of intonational focus, only the acoustic correlate of f0 will be further discussed in the following sections, with particular emphasis on the cross-linguistic variation.

#### **Standard Mandarin**

Existing studies on focus in Chinese varieties are mostly on Standard Mandarin. Standard Mandarin has been repeatedly reported to make extensive use f0 as a major acoustic correlate to mark various types of focus (Jin, 1996; Y. Xu et al., 2012; Shi F. & Xia, 2022; Y. Chen, 2022; Y. Yang, 2022). Typical f0 manipulations under focus involve expanding the f0 range, raising the f0 mean, and/or exaggerating the f0 contour. Post-focus positions usually show compressed f0 range and/or lowered f0 mean, whereas pre-focus components show only slight or no compression (Y. Chen & Gussenhoven, 2008; Jin, 1996; Y. Xu, 1999; Y. Xu et al., 2012; Yuan, 2004; among others).

The on-focus expansion is usually driven by a raised f0 maximum (max) (Y. Chen & Gussenhoven, 2008; J. Shen, 1985; Y. Yang, 2022), whereas the f0 minimum (min) may be lowered or remain unchanged depending on the tone (Y. Chen & Gussenhoven, 2008). At sentence-final position, focus effects tend to be weaker than in non-final positions, often showing only mild or no f0 expansion or

raising (Jin, 1996; Y. Xu, 1999; Y. Yang, 2022). This is related to the similar f<sub>0</sub> realisation of final focus and broad focus, especially the wider f<sub>0</sub> range in the sentence-final position (Jin, 1996; F. Liu & Xu, 2005; F. Shi et al., 2009; Visceglia et al., 2012).

In the post-focus region, an early study also found a lowering top line (Jin, 1996), which may be one of the factors contributing to the f<sub>0</sub> range compression. Nevertheless, an early study reported no post-focus compression; instead, the post-focus f<sub>0</sub> range merely returns to the neutral baseline (Shih, 1988), although this finding has not been widely replicated in later work.

## **Cantonese**

Compared with the extensive studies on Standard Mandarin, Cantonese focus marking is relatively less studied. Existing work has concentrated exclusively on so-called “standard” Cantonese, namely the varieties spoken in Hong Kong (Vance, 1976; C. H. V. Man, 1999; Bauer et al., 2004; Gu & Lee, 2009; W. L. Wu & Xu, 2010; Y. Xu, 2011; Hsu et al., 2018; Y. Yang, 2022; among others; Han, 2013) and Guangzhou (Gu et al., 2006; K. Li, 2024; K. Li et al., 2023), while Nanning Cantonese has so far remained unexamined.

Unlike Standard Mandarin, Cantonese has been found not to rely heavily on f<sub>0</sub> for focus marking, with only limited or conditional f<sub>0</sub> modulation reported (K. Li, 2024; Vance, 1976; W. L. Wu & Xu, 2010; Y. Yang, 2022). Some work notes an absence of f<sub>0</sub> adjustment, or an overall f<sub>0</sub> raising or lowering relative to broad focus, indicating that Cantonese tends to maintain its global f<sub>0</sub> pattern rather than enhance local on-focus versus out-of-focus f<sub>0</sub> contrast (Gu & Lee, 2009; K. Li, 2024; Vance, 1976; Y. Yang, 2022). A broad consensus also indicates a lack of post-focus compression (Gu & Lee, 2009; W. L. Wu & Xu, 2010; Y. Yang, 2022; Hsu et al., 2018; K. Li, 2024). In some cases, f<sub>0</sub> range expansion (Gu & Lee, 2009) and f<sub>0</sub> raising (Gu & Lee, 2009; W. L. Wu & Xu, 2010) have been observed, although the latter was restricted to the low-level Tone 6 in Wu and Xu (2010). On the other hand, on-focus f<sub>0</sub> realisation is more variable across tones and across studies; these tonal effects will be discussed later.

A small number of studies reported f0 manipulation resembling Standard Mandarin (C. H. V. Man, 1999), although these findings are generally attributed to differences in experimental design and are not directly comparable to the majority of work.

### **Post-focus compression and focus marking under language contact**

Based on the fact that post-focus compression (PFC) can be present or absent in closely related tonal languages, along with findings from a range of world languages, Xu et al.(2012) proposed that PFC may be a typological feature and have a single historical origin, although this hypothesis is still speculative, requiring more large-scale cross-linguistic investigation. Moreover, under language contact, focus marking often involves greater complexity, particularly in contexts of L1 influence and L2 production. One related discussion is about the transferability of PFC; particularly, cases simultaneously involving languages with and without PFC are expected to be more complicated.

Previous findings suggested that PFC was difficult to gain in L2 production. When acquiring a PFC language, learners usually presented no, weak or inconsistent PFC in their L2 productions, whether their native language had PFC or not. When the learners' native language had no PFC, PFC in Mandarin was not acquired by L1 Tsat speakers (B. Wang et al., 2012) and Taiwan Mandarin<sup>12</sup> speakers (Y. Xu et al., 2012), although later research observed native-like PFC in the L2 Mandarin by younger speakers of L1 Quanzhou Min, who gained more exposure to Mandarin (Y. Chen, 2014; Y. Chen et al., 2015). In addition, PFC in English was absent in f0 range and weaker in f0 contour by L1 Taiwan Mandarin speakers (Visceglia et al., 2012).

Moreover, even when both their native and target languages have PFC, PFC was found to be weaker and inconsistent in speakers' L2 production. PFC in English was found to be weaker or not native-like by L1 Mandarin speakers (Y. Chen, 2014, 2015; Visceglia et al., 2012). Furthermore, similar to the case of L2 Mandarin by L1 Quanzhou Min speakers, PFC in L2 productions showed

---

<sup>12</sup> Taiwan Mandarin is effectively the non-standard variety of Standard Mandarin spoken in Taiwan. In Xu et al. (2012), the recruited Taiwan Mandarin speakers included monolingual speakers and bilingual Taiwan Mandarin-Taiwan Min speakers.

correlation to learners' L2 proficiency. For example, L1 English speakers who had earlier exposure to Mandarin produced some PFC in their L2 Mandarin, while later learners produced none; L1 Mandarin learners with longer residencies in the United States produced more PFC in their L2 English (Y. Chen, 2014, 2015). Likewise, PFC in English was acquired only by advanced learners speaking L1 Korean (J. Liu et al., 2019). The acquisition challenge of PFC was also found in simultaneous English-Cantonese bilingual speakers, who inconsistently produced PFC in English but never produced PFC in Cantonese (W. L. Wu, 2013; W. L. Wu & Chung, 2011).

However, one case of consistent PFC production was reported in L2 Cantonese by L1 Mandarin speakers (Y. Yang, 2022), where the speakers' native language had PFC but the target language did not. Nevertheless, the author agreed with the correlation between PFC learnability and L2 proficiency, pointing out the need for data from speakers with low Cantonese proficiency to validate this account.

These findings altogether indicate that PFC is easy to lose, but difficult to gain from either the native language or the target language (Y. Xu et al., 2012); more generally, PFC is not easy to transfer between languages (Y. Chen, 2015). The acquisition of PFC, however, is not entirely impossible, alongside the increasing L2 experience and proficiency. These findings suggest the unclear transferability of PFC in L2 acquisition.

Apart from the transferability of PFC, researchers also discussed the potential L1 and L2 influence on intonation realisation. Chen (2014, 2015) proposed that, when one of the L1 and L2 is tonal and the other is non-tonal, lexical tones as a novel phonological feature will block the acquisition of prosody focus, including PFC. However, this account does not apply to the cases where L1 and L2 are both tonal (e.g., Mandarin–Cantonese) or both non-tonal (e.g., Korean–English).

In terms of L1 influence, for example, researchers unanimously attributed the loss of PFC in Mandarin by L1 Min speakers to the long contact with Min (Y. Chen, 2014, p. 2; Y. Chen et al., 2015; Y. Xu et al., 2012). Yang (2022) found that Cantonese by L1 Mandarin speakers made use of

f<sub>0</sub> more than native Cantonese, especially in the non-final focal region; PFC was also presented. Regarding L2 influence, in addition to the aforementioned effect of L2 proficiency, Yang (2022) observed that Cantonese learners' native Mandarin used less f<sub>0</sub> in focus marking, presumably due to the influence of their L2 Cantonese; bidirectional influence between speakers' L1 and L2 was therefore concluded.

Despite the complexity of focus realisations in L2 acquisition, few studies have investigated the intonation realisations in L2, let alone tonal languages and interlanguages, especially when tonal variations have displayed complicated, inconsistent patterns in the realisation of focus intonation within and across language varieties. Specifically, the case studies on Taiwan Min, Taiwan Mandarin and Standard Mandarin (Y. Xu et al., 2012) provided a situation similar to the case of Nanning Mandarin, raising our curiosity about the focus realisation of Nanning Mandarin.

### **Tonal variations in focus realisation**

While lexical tone and prosodic focus both influence pitch realisation, the variation in post-focus f<sub>0</sub> realisation across tonal languages suggests that focus may be largely independent of lexical tone. Despite this, evidence also reveals interaction between the f<sub>0</sub> realisations of lexical tone and prosodic focus.

Above all, the magnitude of on-focus f<sub>0</sub> change varies by tone category. In Standard Mandarin, focus marking patterns vary across the high, low and dynamic (rising and falling) tones. Chen & Gussenhoven (2008) stated that the falling Tone 4 [51] exhibited greater on-focus expansion than the rising Tone 2 [35] and high Tone 1 [55], caused by different mechanisms. In addition, multiple studies claimed that high tones are subject to greater focus effect, such as displaying more visible f<sub>0</sub> raising; while the focus realisation on low tones is still debatable, as data showed that low tones are lowered or largely unaffected under focus (Jia et al., 2006; Shih, 1988; Y. Xu, 1999).

Although Cantonese makes relatively little use of f<sub>0</sub> adjustments to signal focus, it also displays tone-dependent patterns, with even more inconsistent variations. Regarding the f<sub>0</sub> range expansion,

Gu and Lee (2009) and Wu and Xu (2010) both observed the clearest expansion on the rising T3 and T4, while the low-to-mid level or falling tones, i.e., T2, T5 and T6, also showed expansion in some of the focus positions (initial, median or final). Nevertheless, Li (2024) observed only weak f0 range expansion. In terms of the mean f0 contour, f0 adjustment similar to Standard Mandarin was found, but more inconsistent cases were also reported. Similar to Standard Mandarin, Gu and Lee (2009) and Li (2024) found greater f0 raising on high register tones; while on low tones, Li (2024) observed greater f0 lowering, and Gu and Lee (2009) found the lowest pitch targets may remain unchanged under focus. However, more controversial, inconsistent patterns were also reported. Wu and Xu (2010) also observed f0 raising on the low-to-mid level or falling tones in some focus position, specifically, T5 (medial position), T2 and T6 (final position). In addition, while Li (2024) observed no f0 lowering on the high-level T1 and high checked T7, Hsu et al. (2018) found on-focus f0 lowering on T1.

Summarising from these findings, Standard Mandarin and standard Cantonese seem to show some shared patterns of on-focus tonal variation. For example, dynamic tones involve greater f0 range expansion, high tones involve greater f0 raising, and low tones are disputed between f0 lowering or remaining unchanged, although many exceptional cases still exist. Despite these f0 changes, the tonal identities are generally unaffected by the focus conditions (Y. Xu, 2019), especially in Cantonese (K. Li, 2024; V. C. H. Man, 2002), although focus realisations are often associated with enhanced distinctiveness on-focus tonal contours. In addition to the tone-bearing units, the tonal contexts can also interact with focus, further complicating the realisation of sentential pitch (Y. Chen & Gussenhoven, 2008; Shih, 1988). These tonal variations are also observed in PFC. Even for the PFC languages, PFC were found absent in specific tones (Y. Yang et al., 2018) or tonal contexts (Y. Chen, 2010; Duan et al., 2013; C. Shen & Xu, 2016; Y. Xu, 1999). Yang et al. (2018) therefore suggested that PFC, rather than an all-or-nothing feature of languages, need to be investigated considering tone and tonal contexts.

### **1.3.2. Previous studies on intonation in Nanning Cantonese and Nanning Mandarin**

Unlike Standard Mandarin and Guangfu Cantonese, previous studies have rarely investigated the intonation in Nanning Mandarin and Nanning Cantonese; among those, none have investigated the focus intonation.

### **1.3.2.1. Nanning Cantonese**

Regarding question intonation, one study examined the pitch of question particles (X. Liu, 2024). Following the view that the pitch realisations of question-final particles are affected by question intonation, this study described the pitch values of these particles, and also mentioned that one of those particle questions can be optionally realised as intonation question (i.e., question without any interrogative markers, including particles) with final pitch rising. However, this study is mainly descriptive with a focus on syntax and comparison to Guangzhou Cantonese, without any acoustic investigation.

### **1.3.2.2. Nanning Mandarin**

Compared to Nanning Cantonese, a few studies have acoustically investigated the statement and question intonation of Nanning Mandarin. An early study (H. Yang, 2003) investigated sentential-final tones using spontaneous in-person conversations. Yang concluded that Nanning Mandarin extensively uses rising pitch in the sentence types that end with a falling pitch in Standard Mandarin, including statements, *wh*- questions and face-to-face questions. However, no quantitative evidence was provided, and the example spectrograms for questions in fact presented a falling final pitch. Similarly, Qin (2005) also mentioned the rising and lengthening in sentence-final intonation observed in NC–SM interlanguage.

Tang and Liu (2016) and Wu (2017) employed the methodology (‘intonation patterns’ 语调格局) proposed by Shi (1999) to investigate the  $f_0$  range of each prosodic phrase in a sentence. Pitch contour and range for each syllable (in semitone) were obtained, from which a measure termed ‘undulation’ (起伏度, *qǐfúdù*) was calculated for each prosodic phrase (F. Shi, 2008). This measure presents the difference between phrasal pitch ranges in the maximum, minimum and mean in

percentage.

Tang and Liu (2016) examined statement intonation in Nanning Mandarin with comparison to existing Standard Mandarin data. Findings showed that Nanning Mandarin presented narrower sentential and phrasal pitch ranges as well as a greater declination than Standard Mandarin, presumably due to the narrow pitch range of NM tones. In addition, pitch range expansion was observed in both initial and final phrases in Nanning Mandarin, unlike Standard Mandarin, where only the final pitch range was expanded; Shi et al. (2009) thus considered Standard Mandarin showed final focus in statement intonation.

Wu (2017) investigated the intonation of four types of questions: yes-no questions, alternative questions (A-or-B disjunctive questions, A-not-A questions), and *wh*- questions; alternative and *wh*- questions contained grammatical interrogative markers as suggested. Each question type was further presented with and without final question particles, and compared to the corresponding statements. Statement intonation generally showed a downtrend, but instances were observed to have rising intonation. The sentential position of the widest phrasal pitch range was not consistent. Regarding question intonation, results showed that yes-no questions that contained no interrogative markers always presented a final pitch rise, with or without particles. For question types containing interrogative markers, particle questions were always associated with final pitch fall; when without particles, alternative questions allowed for both final rise and fall, while *wh*- questions always presented final fall. According to Wu's observation, NM speakers preferred final rise when final fall was optional in alternative questions, because rising intonation was more common in Nanning Mandarin, presumably due to the influence of Nanning Cantonese.

To summarise, findings on statement intonation showed discrepancies: Yang (2003) observed extensive use of rising pitch, Wu (2017) found mainly downtrend with rising instances, while Tang and Liu (2016) only observed downtrend. This may be the consequence of the naturalness of materials and sample sizes. Regardless, previous studies generally agreed that Nanning Mandarin tended to use rising intonation in speech (F. Qin, 2005; Y. Wu, 2017; H. Yang, 2003), even for the

sentence types that were realised as downtrend in Standard Mandarin. The sentential position of the widest phrasal pitch range also did not show a consistent pattern in Tang and Liu (2016) and Wu (2017). Regarding question intonation, Yang (2003)'s finding on *wh*- questions effectively aligns with Wu (2017)'s finding, that is, *wh*- questions with particle(s) were realised with final pitch fall.

Although existing studies did not systematically investigate the interaction between tone and intonation, some of them have described the tonal realisations in utterances. As mentioned in § 1.2.4, the only two acoustic studies on NM tones in fact investigated the tonal realisations in phrases or utterances. Yang (2003) reported a lower f0 level, narrower f0 range, and a flatter f0 contour in the sentence-final tone, as well as the absence of neutral tones. Relatedly, Flaming (2020) found that NM Tone 4 was realised with a narrower f0 range and thus a flatter contour than SM Tone 4 in various tonal contexts. These findings are in line with the existing auditory descriptions of the NM tone system. In addition, Wu (2017) briefly mentioned occasional pitch variation in intonation caused by tonal variation in connected speech. Particularly, the pitch maximum usually occurred on the high-falling NM Tone 4.

### **1.3.2.3. Particles**

Compared to research on sentential intonation, more studies focused on the pitch realisation of sentence-final particles. Cantonese is widely known for having a rich set of particles as is Nanning Cantonese. Cantonese particles are typically realised with full pitch contour; therefore, existing research generally included the pitch transcriptions of particles (Huo, 2020; X. Liu, 2024; NN Doc, 1998). However, the research focus is mainly on pragmatic meanings and sentence types. Huo (2020) summarised the phonological distribution of particles by main vowel and tone categories, concluding that one particle can usually be realised in different pitch patterns to encode various intonation.

Nanning Mandarin inherited the rich set of particles from Nanning Cantonese, making it a remarkable characteristic distinct from Standard Mandarin. Previous studies have discussed NM particles through auditory transcriptions (Lan, 2007; Y. Lin, 2008) and acoustic analyses (Y. Wu,

2017; H. Yang, 2003, 2008). Yang (2003, 2008) and Lin (2008) both agreed that NM particles are generally not unstressed with a relatively long duration and often a rising pitch, unlike Standard Mandarin. Similar to Nanning Cantonese, NM particles also vary in pitch contours to express different meanings (H. Yang, 2003, 2008). On the other hand, Lan (2007) is mainly a description of the pragmatic meaning of particles; Wu (2017) investigated the intonation of particle questions, but the research focus is not on the particles.

### **1.3.3. Research gaps in intonation research**

In summary, the intonation systems of Nanning Mandarin and Nanning Cantonese have not been examined using systematic experimental approaches or robust methodologies (as detailed in § 1.3.2). While preliminary work on Nanning Mandarin has addressed statement and question intonation, other key aspects of intonation remain largely unexplored, including focus marking.

Methodologically, existing studies were based on small sample sizes (two to six participants) and lacked inferential statistical analysis. Although the measure of undulation aims to provide a set of descriptive, numerical values to objectively measure the degree of change between phrasal  $f_0$  ranges, no statistical significance tests were employed to validate the reported differences. In addition, experimental materials were either read sentences strictly in control and without contexts (Q. Tang & Liu, 2016; Y. Wu, 2017) or free conversation lacking control (H. Yang, 2003). Crucially, these experiments were not designed to account for systematic tonal variation.

Moreover, no study has investigated the resulting  $f_0$  patterns arising from the contact between language varieties that use vastly different  $f_0$  strategies to mark focus, concerning the cases of Standard Mandarin and standard Cantonese reviewed earlier.

## **1.4. Present study**

### **1.4.1. Research gaps**

First, as shown in the literature review, the current understanding of the tone systems and  $f_0$  realisations of Nanning Mandarin and Nanning Cantonese remains fragmented. In addition, inconsistent or discrepant descriptions still remain regarding the precise characterisations of these tones (as discussed in § 1.2). This is largely because previous research on NM and NC tones is mainly preliminary descriptive case reports. Due to the lack of systematic acoustic methods, sufficient sample sizes, and robust statistical testing, the conclusions of these studies are often empirically weak. This gap is critical, as a proper investigation and clear understanding of the NM and NC tone systems serves as an essential foundation for any subsequent study of prosodic systems and tone–intonation interaction in contact.

A more central gap is that the intonation of Nanning Mandarin—a special case arising from contact between two divergent prosodic systems (Standard Mandarin and Nanning Cantonese)—has not been investigated, particularly concerning focus. This is not merely a descriptive gap (i.e., the lack of a case study), but one that hinders progress on several key theoretical issues. As previous research has indicated, our understanding of PFC typology and transferability, the phonological account for the focus mechanism, and the nature of tone–intonation interaction require validation from a wider range of cross-linguistic data and the systematic consideration of tonal variation (Y. Chen, 2022; Y. Xu et al., 2012; Y. Yang, 2022; J. Zhang et al., 2020).

Finally, although non-standard Local Mandarin varieties are spoken by the vast majority of the population in China, the field of research is insufficient. Even in phonetics, which has received relatively more attention than other linguistic subfields, research is still in an emerging stage (Hu & Xu, 2020; Qie, 2015). Specifically, previous studies on Nanning Mandarin have often been framed as cultural or language descriptions. Moreover, the prevailing (theoretical) perspective in its phonetic research has tended to focus on ‘failure of acquisition’ or ‘deviations’ from Standard Mandarin, with an implicit goal of accent correction. Such a ‘deficit perspective’ has overlooked the value of studying Nanning Mandarin as an independent linguistic system. As an intermediate variety with systematicity and sociolinguistic functions, Nanning Mandarin has developed beyond individual-level imperfect L2 acquisition and is emerging as a community-level phenomenon that is

gaining L1 status. Moreover, Nanning Mandarin also manifests the complex processes of language variation and contact that are not fully captured within this perspective.

### **1.4.2. Present study and aims**

To address the multiple gaps identified in the previous section, which range from fundamental descriptive needs and key theoretical questions to broader perspectival limitations, the central aim of this thesis is to demonstrate the acoustic characterisations of tone, intonation, and their interaction in Nanning Mandarin. This research provides the first systematic attempt at an acoustic investigation of the Nanning Mandarin pitch systems, with a direct comparison to the historically dominant source language, Nanning Cantonese, and the target language, Standard Mandarin.

In contrast to previous studies, the present study assesses these  $f_0$  realisations and cross-linguistic relationships using robust acoustic-phonetic and statistical analysis. These analyses are based on data from carefully designed production experiments from a large number of NM, NC, and SM speakers.

Specifically, the tone study extends previous research by directly examining the relationship of the NM tone system to the NC and SM tone systems. The intonation study provides the first systematic attempt to investigate focus marking in Nanning Mandarin and Nanning Cantonese, using Standard Mandarin data as a well-established reference. At the same time, this research aims to establish initial prosodic profiles of two understudied Sinitic varieties, Nanning Mandarin and Nanning Cantonese.

### **1.4.3. Research questions**

The central aim of this thesis is specified by the following research questions (RQs). The questions are structured to systematically investigate three core components: (1) lexical tone, (2) intonational focus, and (3) their interaction.

#### **RQ 1: Lexical tone**

**How are the lexical tone systems of Nanning Mandarin, Nanning Cantonese, and Standard Mandarin acoustically realised and compared?**

- (1a) What are the primary  $f_0$  realisations (e.g., height and contour) of the NM and NC tone systems?
- (1b) In what ways are  $f_0$  realisations of the NM tone system similar to or different from those of Nanning Cantonese and Standard Mandarin?

**RQ 2: Intonational focus**

**How are intonational focus in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin acoustically realised and compared?**

- (2a) What are the primary  $f_0$  range modulation used to mark focus (compared to the broad focus condition) in Nanning Mandarin and Nanning Cantonese? For example, does Nanning Mandarin have post-focus compression (PFC), a feature present in Standard Mandarin but absent in Cantonese?
- (2b) In what ways are the  $f_0$  range modulation in NM focus marking similar to or different from those of Nanning Cantonese and Standard Mandarin?

**RQ 3: Tone–intonation interaction**

**How do the patterns of tone–intonation interaction in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin manifest acoustically, and how do they compare?**

- (3a) How does focus marking modulate the  $f_0$  range of different tone categories? Can specific patterns be identified within language varieties?
- (3b) In what ways are these tone–intonation interaction patterns in Nanning Mandarin similar to or different from those of Nanning Cantonese and Standard Mandarin? Can language-specific patterns or cross-linguistically consistent patterns be identified?

## **Chapter 2: Production of lexical tones**

### **2.1. Corpus build-up**

Data for all three language varieties—Nanning Mandarin, Nanning Cantonese, and Standard Mandarin—were collected for this study. However, due to the extensive time required to process the new recordings, the SM data analysed in the tone study was sourced from the existing ManDi corpus (Zhao, 2024). Consequently, the following methodology sections (i.e., materials, participants, experiment, and data processing) will focus on the procedures for the newly collected and processed NM and NC data. In the subsequent acoustic analysis section, all three language varieties were subjected to the same procedure and will therefore be described together. Readers are referred to the original publications for full details on the ManDi corpus.

#### **2.1.1. Wordlists**

##### **2.1.1.1. Tone mapping**

For Nanning Mandarin and Nanning Cantonese, a wordlist containing 72 monosyllabic words and 72 disyllabic words was designed based on the diachronic correspondence between Mandarin and Cantonese tone categories. As a variety of Standard Mandarin, Nanning Mandarin inherited the SM tone system. Nanning Mandarin and Standard Mandarin have four citation tones, while Nanning Cantonese has six citation tones that have historical correspondences with Mandarin tones (X. Li & Xiang, 2020; Y. Lin & Qin, 2008).

Table 2.1.1 shows the correspondence between Cantonese and Mandarin non-checked (smooth) tones adapted from Y. Lin & Qin (2008), Wan, (2012), and Z. Tang (2016). Tone values in the brackets are the canonical tone values from the phonological documentations for Nanning Mandarin (Yu, 2008), Nanning Cantonese (Y. Lin & Qin, 2008) and Standard Mandarin (SM Textbook, 2012).

In the table below, the checked tones (IV) were not included, as they are not involved in this study, although they still exist in Cantonese. This is because the three checked tones are usually underlyingly level tones (Yip, 2002), with similar pitch contour and range as the three non-checked level tones, but with substantially shorter duration (Gu & Lee, 2007; W.-S. Lee, 2004; S. Wang, 2015; Yip, 2002). That says, the difference between the checked and non-checked tones is primarily segmental (L. Zhang, 2013). Therefore, the checked tones are generally treated as the allophones of their non-checked counterparts (Mok & Zuo, 2012). In addition, in Mandarin, the checked tones had already merged into the other tone categories, making it non-parallel to include the checked tones when comparing Mandarin with Cantonese. To maintain cross-linguistic consistency and reduce workload, we decided not to include the checked tones in the present study.

Tone	Historical category	Nanning Mandarin	Standard Mandarin	Tone	Historical category	Nanning Cantonese
1	yīn píng (Ia)	33 or 44	55	1	yīn píng (Ia)	55
2	yáng píng (Ib)	24 to 34	35	2	yáng píng (Ib)	21
3	shǎng (II)	21	214	3	yīn shǎng (IIa)	35
				4	yáng shǎng (IIb)	24
4	qù (III)	53 or 55	51	5	yīn qù (IIIa)	33
				6	yáng qù (IIIb)	22

**Table 2.1.1:** Correspondence between the non-checked tone categories of Mandarin and Cantonese.

Table 2.1.2 below summarises the word counts of the monosyllabic and disyllabic wordlists across the three language varieties. The number of items was determined primarily with reference to the NC tone system, as Nanning Cantonese has more tones than the two Mandarin varieties. Specifically, each NC tone was assigned 12 monosyllabic words, and each NC tone combination was assigned two disyllabic words. The same set of words was also used for both Mandarin varieties. As a result

of the tone correspondence illustrated in Table 2.1.1, the word count associated with Mandarin Tones 3 and 4 was at least doubled. The number of words for each tone category and each tone combination reflects the main rules of tonal correspondence between Mandarin and Cantonese excluding exceptional cases.

To summarise, the 72 monosyllabic words cover all four tone categories in Nanning Mandarin with 12 or 24 words for each NM tone, and all six tone categories in Nanning Cantonese with 12 words for each NC tone. The 72 disyllabic words cover all 16<sup>13</sup> tone combinations in Nanning Mandarin with two, four or eight words for each NM tone combination and all 36 tone combinations in Nanning Cantonese with two words for each NC tone combination. The full wordlists with phonetic transcriptions and English glosses are attached in Appendix A.1.1 for monosyllables and A.1.2 for disyllables.

Due to time constraints, the present study focuses on the monosyllabic words alone. Citation tones form the foundation for establishing the prosodic profile of a language variety, especially for varieties that have not yet undergone systematic acoustic analysis. Therefore, monosyllabic data, which provide a necessary first step for characterising the tonal system, will be described in detail hereafter; disyllabic words are reserved for further investigation.

Wordlist	Nanning Mandarin and Standard Mandarin		Nanning Cantonese	
	Tone category	Count	Tone category	Count
Monosyllabic	M1	12	C1	12
	M2	12	C2	12
	M3	24	C3	12
			C4	12
	M4	24	C5	12
			C6	12

<sup>13</sup> There are different explanations about the third tone sandhi in Mandarin: some treat it as M3 + M3 → M2 + M3, and some argue the sandhi Tone 3 is not equivalent to Tone 2. Regardless the discrepancies, we will include it as a separate combination await for further explanations.

	<b>Tone combination</b>	<b>Count</b>	<b>Tone combination</b>	<b>Count</b>
	<b>Disyllabic</b>	M1+M1	2	C1+C1
M1+M2		2	C1+C2	2
M1+M3		4	C1+C3	2
			C1+C4	2
M1+M4		4	C1+C5	2
			C1+C6	2
M2+M1		2	C2+C1	2
M2+M2		2	C2+C2	2
M2+M3		4	C2+C3	2
			C2+C4	2
M2+M4		4	C2+C5	2
			C2+C6	2
M3+M1		4	C3+C1	2
			C4+C1	2
M3+M2		4	C3+C2	2
			C4+C2	2
M3+M3		8	C3+C3	2
			C3+C4	2
			C4+C3	2
			C4+C4	2
M3+M4		8	C3+C5	2
			C3+C6	2
			C4+C5	2
			C4+C6	2
M4+M1		4	C5+C1	2
			C6+C1	2
M4+M2		4	C5+C2	2
			C6+C2	2
M4+M3	8	C5+C3	2	
		C5+C4	2	
		C6+C3	2	
		C6+C4	2	
M4+M4	8	C5+C5	2	
		C5+C6	2	
		C6+C5	2	
		C6+C6	2	

**Table 2.1.2:** Number of items in the wordlists summarised by tone.

### 2.1.1.2. Word selection

The word selection was guided by two primary criteria: **high frequency** and **shared phonetic context** cross-linguistically. That says, the ideal items should optimally differ only in tone. Word frequency was determined by the author and validated by native speakers, and similar phonetic contexts are indicated by the IPA transcriptions.

All items in the wordlist were recognised cognates in both daily spoken Nanning Mandarin and Nanning Cantonese; a few might sound old-fashioned in Mandarin, as they are now more commonly used as bound morphemes in disyllabic or multisyllabic words in Modern Chinese. According to participants' feedback, a few NC words (e.g., 毯 'blanket', 害怕 'fear') felt Mandarin-like and thus less authentic. Nevertheless, as these words had been increasingly adopted by NC speakers under the influence of Standard Mandarin, they were retained for the NC tone categories lacking words. To enhance acoustic clarity and minimise potential errors, items with certain phonetic and lexical features were less preferred, unless a given tone category lacked sufficient items. These included: syllables with null (no initials<sup>14</sup>) or sonorant consonant onsets, to facilitate segmentation; characters with multiple readings (heteronyms), to minimise pronunciation errors; and syllables containing compound vowels (diphthongs or triphthongs), to simplify the syllabic structure and reduce variations in voice quality.

Finally, it should be noted that the study did not use a single syllable across all tone categories for most cases, although it may be an ideal practice in tonal research. This is because we prioritised cross-linguistic consistency and word frequency, resulting in very limited remaining options. Furthermore, using real words instead of non-words was considered essential for the naturalness of the materials, particularly when eliciting the NM accent. The selected words primarily feature simple V, CV, or CVN syllable structures, as illustrated by the examples in Table 2.1.3.

---

<sup>14</sup> Zero ( $\emptyset$ ) onset/initial refers to when the initial consonant is absent and phonetically realised as corresponding glides or glottal, which is an unstable status.

Syllabic structure	Nanning Mandarin	Nanning Cantonese	Gloss
V	/y <sup>21</sup> /	/jy <sup>24</sup> /	雨 ‘rain’
	/jou <sup>21</sup> /	/jɛu <sup>24</sup> /	有 ‘have’
CV	/ku <sup>33</sup> /	/ku <sup>55</sup> /	姑 ‘aunt’
	/kua <sup>33</sup> /	/k <sup>wa</sup> 55/	瓜 ‘melon’
	/ma <sup>33</sup> /	/ma <sup>55</sup> /	妈 ‘mother’
	/mai <sup>21</sup> /	/mai <sup>24</sup> /	买 ‘buy’
CVN	/fan <sup>33</sup> /	/fan <sup>55</sup> /	翻 ‘flip’
	/nan <sup>24</sup> /	/nam <sup>21</sup> /	南 ‘south’

**Table 2.1.3:** Word samples for different syllabic structures in Nanning Mandarin and Nanning Cantonese.

All words in the wordlist are chosen from authentic documentations and phonological dictionaries (Ding, 1981; Y. Lin & Qin, 2008; GX Doc, 1998) according to the author’s intuition as a native NM speaker and validated by native speakers of Nanning Mandarin and Nanning Cantonese. IPA transcriptions mainly refer to (Y. Lin & Qin, 2008; GX Doc, 1998) for Nanning Cantonese and the broad transcriptions of Standard Mandarin in (W.-S. Lee & Zee, 2003) for Nanning Mandarin with manual check.

### 2.1.2. Participants

For Nanning Mandarin and Nanning Cantonese, 52 native speakers were recruited. Among them, 48 were bilingual speakers, who were preferred for convenience; four were NM monolingual speakers. Acquaintances of the researcher were preferred for the convenience sampling. We aimed at collecting data from 30 usable speakers to reach a relatively large sample size. The eligibility of participation was determined based on a demographic interview concerning language background.

With a few exceptions, participants met the following criteria:

- They had been residing in Nanning before the age of 12 (1 exception);
- They had acquired Nanning Mandarin or Nanning Cantonese before the age of 12 (NM: 1 exception; NC: 4 exceptions);
- They did not have close family members or friends who spoke Northern Mandarin dialects (7 exceptions);
- They had not received formal training in linguistics.

Additional criteria were preferred:

- They had not stayed outside of Nanning for more than one year within the past ten years (7 exceptions);
- They had not received any explicit training or test in Standard Mandarin (12 exceptions).

Participants who did not meet the above criteria went through additional validation of their NM accent for further exclusion (see § 2.2.1.2). Participants also needed to be proficient in written Standard Simplified Chinese and have no auditory or speaking impairment.

Unlike traditional Chinese dialectological studies, which often require strict standards of linguistic purity (e.g., multi-generational native ancestry), our recruitment criteria were designed to reflect the actual sociolinguistic reality of Nanning. The complex linguistic landscape of Nanning is characterised by significant migration and language contact. Moreover, it has become a practical necessity to relocate for higher education and to obtain an SM proficiency certificate for employment, especially for the younger generation. Given that Nanning Mandarin is a recently formed interlanguage and Nanning Cantonese is itself a hybrid variety, imposing such strict criteria would be both unrealistic and unrepresentative of the local speaker community. Instead, we primarily ensured that participants' families of origin were all long-term residents of Guangxi. The additional Chinese languages spoken by participants' family of origin were mainly other local Guangxi varieties, including various Cantonese dialects, Zhuang languages, Southwest Mandarin, and Pinghua.

Furthermore, to avoid a homogenous sample, participants were recruited from diverse social backgrounds, varying in their residential district, occupation, and education.

Participants were roughly balanced for gender (27 females and 25 males) and evenly distributed from young to old for age (12 each for age 20–39, 40–49 and 50–59, 16 for age 60–69), with some leniency given the convenience sampling. Speakers aged 26 to 67 (median age = 51); among them, speakers aged 40 to 69 formed the majority. This is because they were less affected by other language varieties than the younger speakers, as they were less likely to have settled elsewhere in recent years and have had less exposure to Standard Mandarin at school; on the other hand, they articulated more clearly than the older speakers. Table 2.1.4 displays the detailed age and gender distribution of the recruited participants.

Gender \ Age	20–39		40–49	50–59	60–69	Total
	20–29	30–39				
Female	4	2	5	8	8	27
Male	5	1	7	4	8	25
SUM	9	3	12	12	16	52

**Table 2.1.4:** Age and gender distribution of the recruited participants of Nanning Mandarin and Nanning Cantonese.

### 2.1.3. Experiment

#### 2.1.3.1. Before recording

**Recording settings.** For Nanning Mandarin and Nanning Cantonese, recordings were conducted in person locally in Nanning. The entire process took place at a desk in a quiet room or booth in a location convenient for the participant. The speaker sat in front of the laptop, and the researcher sat to the right of the speaker. The smartphone was mounted on a tripod at approximately mouth height,

positioned 25 to 40 cm from and slightly to the right side of the speaker's mouth, to prevent recording air bursts.

The instructions and wordlist were presented in Simplified Chinese on a MacBook Pro (Retina, 13-inch, Early 2015) via PsychoPy v.2021.2.3 (Peirce et al., 2019). The production was recorded using Awesome Voice Recorder X v2.0.4 (by Newkline Co., Ltd.) on an iPhone 7 smartphone. Such a lossless format phone recording was proven to accurately track  $f_0$  (C. Zhang et al., 2021). The audio was recorded in WAV format, mono channel, with a sampling rate of 22,050 Hz, a bit rate of 320 kbps.

**Demographic interview.** Before the study, participants received a recorded demographic interview about their language and social backgrounds, including the age and approaches of Nanning Cantonese and Nanning Mandarin acquisition, the occasions and frequency of language use, SM and other language proficiency, educational backgrounds, occupations, age and gender.

**Instructions.** In the instructions, participants were familiarised with the requirements and procedure on the slides with oral instructions by the researcher, followed by an overview of the wordlist. When they felt prepared, the PsychoPy experiment started, and the entire progression was controlled by the participant by pressing any button on the keyboard.

The participants could raise relevant questions or provide feedback during the introductions, the practice phases and after the experiment.

### **2.1.3.2. During recording**

The recording contained four blocks: monosyllabic and disyllabic words in each of Nanning Mandarin and Nanning Cantonese. Monosyllabic words always went before disyllabic words; Each block consisted of two repetitions of the wordlist. Language order was counterbalanced across bilingual participants, and word order was randomised within each block. Preceding each formal trial, there was a practice trial containing five words outside the final corpus, which did not go through the

data analysis. This is especially helpful for the NM and NC bilingual speakers to switch between language varieties.

Target words were presented one by one on the screen; each monosyllabic word was presented within a disyllabic context for easier identification and disambiguation of multiple readings. Participants were asked to produce the words as naturally as possible; they were allowed to repeat a token if needed for correction. There were rest phases between each repetition of each block.

In total, 7486 monosyllabic NM words and 6905 monosyllabic NC words were collected.

## **2.2. Data processing and analysis**

### **2.2.1. Acoustic processing**

#### **2.2.1.1. Segmentation and alignment**

Recordings of Nanning Mandarin and Nanning Cantonese were automatically segmented at the syllable level using Praat (Boersma, 2002) scripts. The TextGrids with the silent and sounding intervals were created using *To TextGrid (silences)*... in Praat, with a silence threshold of  $-25$  dB. The Chinese character (*Hànzì*), IPA transcription, and tone category of each word were also aligned, along with the validation results (see § 2.2.1.2). We then used Montreal Forced Aligner (MFA) v2.0.0rc7 (McAuliffe et al., 2017) for phone-level segmentation. The acoustic model we used was the Mandarin MFA acoustic model v2.0.0 from the MFA official site; the lexicons were a dictionary constructed from the wordlist, using the corresponding Mandarin IPA transcriptions and tone values in the official MFA format from Mandarin (China) MFA dictionary v2.0.0.

We used the Mandarin acoustic model and dictionary for both Nanning Mandarin and Nanning Cantonese for the following reasons. (1) For Nanning Mandarin, the phonological transcriptions are fairly similar to Standard Mandarin since Nanning Mandarin is a close variety of Standard Mandarin. In addition, the segments of Nanning Mandarin are not the focus of this thesis. (2) For Nanning

Cantonese, there were no usable pre-trained Cantonese models at that time, and our data was not sufficient to train models. Since the NM and NC tokens shared the same or similar IPA transcriptions, it is reasonable and acceptable to use a Mandarin model for both varieties. The vowels unique to Nanning Cantonese were manually relabelled in TextGrids.

The output TextGrids of each step went through necessary manual inspections and corrections.

### **2.2.1.2. Validation of the NM accent**

Nanning Mandarin is an inherently unstable interlanguage variety that is highly susceptible to influence from Standard Mandarin. This is particularly due to both constant exposure to Standard Mandarin through media and the function of Standard Mandarin as a medium of instruction in formal settings. Given the formal register of the reading task, it was noted that some speakers can be induced to style-shift, i.e., neutralising their NM accent or even in favour of a more standard, SM-like production. Therefore, a robust validation process was essential to distinguish and select authentic, natural NM tokens. The process involved two components: (1) a primary auditory judgement by native speakers, supplemented by (2) a secondary check of speaker eligibility based on participants' demographics.

The native speakers' auditory judgement was initially performed by the author. During the recording, the author made preliminary judgements, flagging tokens that appeared overly SM-like and lacked characteristic NM features. To formalise and more objectively verify these intuitions, a panel of six NM native speakers (4 female; age 26–54, median age = 35.5) was recruited to perform a series of validation tasks.

Firstly, NM speakers were divided into five groups depending on a combination of speaker eligibility status and NM accent authenticity. The five designated groups were:

<b>Group</b>	<b>Speaker eligibility status</b>	<b>NM accent authenticity</b>
<b>(a)</b>	<b>Fully eligible</b>	<b>Authentic</b>
<b>(b)</b>	<b>Partially eligible</b>	<b>SM-like or very careful speech</b>
<b>(c)</b>	Fully eligible	Not authentic or ambiguous
<b>(d)</b>	Partially eligible	Not authentic or ambiguous
<b>(e)</b>	Partially eligible	Authentic

**Table 2.2.1:** Grouping of NM speakers. Group (a) and (b) (in bold) served as the reference points for rating.

For clarity in this section, ‘speaker eligibility’ is defined as ‘fully eligible’ if all eligibility requirements from § 2.1.2 were met, and ‘partially eligible’ if only non-essential, ‘preferred’ criteria were violated. ‘Non-authentic’ NM productions refer to speech that neutralises the NM features, while ‘ambiguous’ productions are those that fall intermediate between Nanning Mandarin and Standard Mandarin, not clearly identifiable as either.

To create the stimuli for the validation task, three to five random tokens (monosyllabic and disyllabic) from each speaker were extracted and concatenated, creating a short audio file for each of the five speaker groups. The six raters were asked to evaluate the five recordings and rate the authenticity of the NM accent for Groups (c), (d) and (e), which included 21 NM speakers, on a five-point Likert scale from 1 to 5. To anchor the raters’ judgements, they were also provided with the recordings from Groups (a) and (b) as reference points: Group (a) served as the high end of the scale (rated 4–5), while Group (b) as the low end (rated 1–2). Following the rating task, the raters were also asked about their perceptual threshold, i.e., whether a rating of 3 was acceptable as authentic Nanning Mandarin. The five-point scale was defined for the raters as follows, with 5 indicating the most authentic NM accent:

<b>Rating</b>	<b>Degree of NM authenticity</b>
<b>1</b>	Standard Mandarin / not Nanning Mandarin
<b>2</b>	SM-like or careful reading speech with little NM accent
<b>3</b>	Ambiguous or intermediate accent
<b>4</b>	<b>Mild or slightly unnatural NM accent</b>
<b>5</b>	<b>Strong and authentic NM accent</b>

**Table 2.2.2:** Detailed rating criteria. The rating of 4–5 (in bold) indicates an acceptable, authentic NM accent.

As a supplementary measure, a quantitative assessment of speaker eligibility was also conducted for the ‘partially eligible’ speakers. For this, the Weighted Constraints model (Pater, 2009) was adapted to calculate a penalty score. This score reflected the degree of a speaker’s violation of the ‘preferred’ (i.e., non-essential) criteria in the eligibility requirements (see § 2.1.2), providing a quantitative basis for the exclusion process.

The rating results provided a clear distinction between the groups, consistent with the author’s intuition. All six raters unanimously rated Group (e) as 5; in contrast, Groups (c) and (d) were rated 1–3. Regarding the perceptual threshold for authenticity, four of the six raters indicated that a rating of 3 was not acceptable as representative Nanning Mandarin; one rater considered it acceptable, while the sixth did not provide an answer.

Following the raters’ evaluation, a two-step data filtering process was implemented. First, at the speaker level, NM speakers rated as 4–5 were retained in the acoustic analysis. Second, at the token level, all tokens from the retained speakers underwent a final auditory validation by the author. In this phase, the author’s goal was to consistently apply the collective standard of NM authenticity established by the raters in the previous rating results. Any tokens that failed to meet this standard were flagged as unusable for the token-level exclusion from the final dataset (see § 2.2.3.3). After a

combination of validation tasks, 20 NM speakers were excluded.

## **2.2.2. Standard Mandarin data**

For Standard Mandarin, ten unique syllables were used to create a wordlist of 40 monosyllabic words, with each syllable realised in one of the four SM lexical tone categories. 21 fluent speakers of Standard Mandarin (13 females and eight males aged 18–53) were recruited from six Mandarin-dominant cities of China. Though these cities have their own regional dialects, the participants reported an average score of 4.14 on a 1–5 scale (5 as native-like fluency) on their fluency in Standard Mandarin.

Recordings were collected remotely via participant-controlled smartphones. Participants were instructed to install a specific recording app and record themselves while reading the words presented on the Gorilla Experiment Builder (Anwyl-Irvine et al., 2018). Words were presented trial by trial and fully randomised. Recordings were also automatically segmented and aligned using Praat (Boersma, 2002) and the MFA (McAuliffe et al., 2017) at the syllable and phone levels. Each audio used for analysis was also manually checked for its representativeness of Standard Mandarin. For more details, see Zhao (2024).

## **2.2.3. Acoustic analysis**

### **2.2.3.1. F0 contour extraction**

For all three language varieties, f0 values were automatically extracted from 11 equidistant points in the sonorant portion of each syllable, using raw autocorrelation *Sound: To Pitch (raw ac)* in Praat.

For Nanning Mandarin and Nanning Cantonese, a pitch range of 40–400 Hz was used, which generated the least number of undefined values; other parameters remained default. F0 values were extracted from sound clips of each token to, which was expected to reduce the undefined values. Each sound clip was wrapped by 0.2 seconds on both ends, to maximally retain the information at

boundaries. For Standard Mandarin, a pitch range of 75–500 Hz was used for females and 75–250 Hz for males. F0 values were extracted from each of the entire audio files. (See Zhao (2024) for details of Standard Mandarin data processing.)

### **2.2.3.2. Cleaning data**

Before any further steps of acoustic analysis, data were first cleaned in R.

Outliers, defined as data points exceeding  $\pm 2$  standard deviations from the mean, were identified in R and converted to NA values. This procedure resulted in an additional 4,162 missing data points, taking 4.17% of the total number.

Tokens were excluded if they met any of the following criteria: with production errors (e.g., mispronunciations) or pauses; with more than five (50%) unusable f0 values; or showing a boundary mismatch of over 0.15s between the start or end boundaries of the MFA intervals and the silent/sounding intervals. A total of 809 tokens were therefore excluded.

The first f0 point was excluded as its value was highly affected by the preceding consonant and was frequently undefined, much more than the other 10 points.

Speakers with at least 72 (50%) usable tokens were retained for acoustic analysis (two NM speakers were excluded).

### **2.2.3.3. Final datasets**

The final dataset contained 3390 tokens from 30 NM speakers, 5822 tokens from 48 NC speakers, and 793 tokens from 21 SM speakers.

For Nanning Mandarin and Nanning Cantonese, 30 NM speakers (female = 14; aged 26–66, median age = 43) and 48 NC speakers (female = 24; aged 26–67, median age = 51) were retained for further analysis, among which 4 were NM monolingual speakers. Speakers aged 20–29 and 30–39 were

grouped into one age group 20–39, because the sample size for each of the two younger decades individually was insufficient for robust analysis. Furthermore, this grouping enables a more meaningful contrast between ‘younger’ speakers and the ‘mid-to-older’ age groups (40–69) in the previous literature. This approach eventually resulted in a more balanced distribution of participants across the age groups.

For Standard Mandarin, 21 SM speakers (female = 13; aged 18–59, median age = 25) were retained for further analysis, were retained for further analysis.

The detailed age and gender distributions for all selected participants can be found in Table 2.2.3 below.

Nanning Mandarin speakers

Age \ Gender	20-39		40-49	50-59	60-69	Total
	20-29	30-39				
Female	4	2	3	2	3	14
Male	4	0	6	3	3	16
SUM	8	2	9	5	6	30

Nanning Cantonese speakers

Age \ Gender	20-39		40-49	50-59	60-69	Total
	20-29	30-39				
Female	1	2	5	8	8	24
Male	4	1	7	4	8	24
SUM	5	3	12	12	16	48

### Standard Mandarin speakers

Age \ Gender	20-39		40-49	50-59	60-69	Total
	18-29	30-39				
Female	11	1	0	0	1	13
Male	6	1	0	0	1	8
SUM	17	2	0	0	2	21

**Table 2.2.3:** Summary of the age and gender information of the selected speakers.

Due to time constraints and low task difficulty, validation was not conducted for Nanning Cantonese, even among younger speakers.

#### 2.2.3.4. F0 standardisation

The raw f0 values in Hertz (Hz) were standardised by speaker-specific f0 ranges for each language variety. The speaker-specific min and max f0 were summarised from the extracted f0 points. The standardised f0 values were then transformed into the tone values in real-valued continuous Chao tone numbers ranging from 0 to 5, using the Log T-value formula (F. Shi, 1986):

$$LT = \frac{\log_{10} \text{mean} - \log_{10} \text{min}}{\log_{10} \text{max} - \log_{10} \text{min}} \times 5$$

We compared several methods to standardise the raw f0 before finally deciding on the Log T-values. First, we tried **semitones** standardised by speaker-specific low f0, adapted from Yuan and Liberman (2014). While they originally defined the low f0 using the 5th percentiles of a speaker's f0 values, we adapted this baseline to speaker-specific mean – 1.5 standard deviations to reduce NA values. However, this adjustment in turn introduced negative semitone values, which are physically uninterpretable. Moreover, this method rescales the f0 only by the pitch floor, so it is not as comprehensive as the T-values. Second, we tried **T-values** standardised by the raw f0 averaged by

speaker and tone, which is then converted to logarithm Chao tone values using the Log T-value formula (F. Shi, 1986). This method rescales the raw f0 by speaker-specific pitch ranges and obtains tone values ranging from 1 to 5, is more comprehensive and easier to interpret. Third, we tried **Z-values** standardised by the raw f0 averaged by speaker and tone (Zhu, 2003). The result of this method is mathematically equivalent to those of T-values (Bei, 2020), but it is harder to interpret as it has negative values and is not rescaled from 1 to 5.

Therefore, we decided to use Log T-values for f0 standardisation. Reference to Chao tone numbers in this thesis thus reflects the Log T-values.

The Chao tone numbers were assigned based on the five equal intervals of the continuous Log T-values ranging from 0 to 5 (therefore 6 lines), rather than assigned to the closest integers. This is based on the division in Shi (1991), which is an improvement of Chao original division, where the five numbers were decided based on the five levels (therefore 4 intervals), rather than the five intervals. The new division greatly reduces the data loss due to data points falling outside the lines of integers.

#### **2.2.4. Contour modelling**

To statistically model and compare the tonal contours, Generalised Additive Mixed Models (GAMMs) were chosen for this analysis as they are a powerful and flexible method for modelling non-linear data, such as f0 contours. Their use in phonetic research is increasingly common because they can effectively model full curves directly from the data. Crucially, they can also account for the hierarchical dependencies and handle the within-contour autocorrelation (Sonderegger & Sóskuthy, 2025).

An alternative approach, polynomial regression, was also trialled. However, this method proved problematic. Models using orthogonal polynomials (which have independent coefficients) failed to adequately capture the true shape of the f0 contours. Conversely, models with potentially correlated regression coefficients produced more plausible shapes. GAMMs overcome some limitations of

polynomial regressions. Unlike polynomial regression, where the user must manually specify the basic functions, GAMMs estimate the required degree of smoothness from the data itself, which helps to avoid oversmoothing and overfitting. Furthermore, as Sonderegger and Sóskuthy (2025) noted, polynomial regression “does not straightforwardly handle nonindependence and autocorrelation within trajectories”.

The tone contours were compared within the tone groups that share similar contour shapes. We used *mgcv* package (Wood, 2011) in R to implement GAMMs and *itsadug* package (van Rij et al., 2015) in R for plotting. Following the instructions of Sóskuthy (2017) and Wieling (2018), we additively built up the model that can fit our data best. Due to time constraints, this process was only conducted on the level tone group.

#### **2.2.4.1. Description of the final model**

We fitted a model to predict the f0 contour in Log T-value Chao tone numbers, with the following elements:

- 1) Parametric terms
  - i. **tone** (treatment coded: NM Tone 1, NC Tone 1, NC Tone 5, SM Tone 1)
  - ii. **preceding consonant** (sum coded: sonorant, voiceless unaspirated, voiceless aspirated, fricative)
- 2) Smooth terms
  - i. **s(timepoint, bs = "cr", k = 10)**: a reference smooth over time point (indicating the reference level, i.e., when tone = NM Tone 1 and preceding consonant = average),
  - ii. **s(timepoint, by = tone.ordered, bs = "cr", k = 10)**: a difference smooth over time point by each tone (indicating the additive difference from each of the NC or SM tone to NM Tone 1), and
  - iii. **s(timepoint, by = preceding\_consonant, bs = "cr", k = 10)**: a difference smooth over time point by each preceding consonant (indicating the additive

difference from each preceding consonant to the average smooth)

3) Random slopes

i. `s(char, timepoint, bs="re")`: one over character (word) by each of the 10 time points, and

ii. `s(spkr, tone, bs="re")`: one over speakers by each tone

4) Random smooths over time point

i. `s(timepoint, char, bs = "fs", m = 1, k = 5)`: one by character (word),

ii. `s(timepoint, spkr, bs = "fs", m = 1, k = 5)`: one by speaker, and

iii. `s(timepoint, spkr, by = tone, bs = "fs", m = 1, k = 5)`: one by speaker and tone

5) The model also included an AR1 error model.

Other minor predictors (such as age group, gender and trial order, etc.) and interactions were excluded for the current stage, as the model was already complex and cost high computational resources. The process of model testing and justification will be described below in detail.

#### 2.2.4.2. (1) Basic model: f0 trajectory ~ tone

**Difference smooth.** Since the NM Tones are the anchor for cross-linguistic comparisons, we decided to fit a reference smooth for NM Tone 1 and a difference smooth for the differences between NM Tone 1 and each of the NC and SM Tones. By doing so, the model coefficients are easier to interpret than fitting separate smooths to each of the four tones, whereas the model fit remains the same. This is similar to the treatment coding in linear mixed-effects models.

**Model settings.** For the smooth terms, we used 10 knots ( $k=10$ ), cubic regression splines (`bs="cr"`), and fast restricted maximum likelihood estimation (`method="fREML"`).  $k=10$  is the default number of knots in *mgcv* package. `fREML` is the default estimation method for the `bam()` function in *mgcv* package. It runs much faster than other methods, such as maximum likelihood estimation (`method="ML"`), and thus is more computationally efficient.

### 2.2.4.3. (2) Add random structures

To account for hierarchical dependencies due to speakers and items, we tested several options and combinations of random structures.

**Random intercepts and slopes.** We first added a set of random intercepts and slopes. The full set included:

- Random intercepts:
  - `s(char, bs="re")`: one over character (word), and
  - `s(spkr, bs="re")`: one over speaker
- Random slopes:
  - `s(char, timepoint, bs="re")`: one over the combination of character (word) and time point, and
  - `s(spkr, tone, bs="re")`: one over the combination of speaker and tone

First, we added the random intercepts only. Model summary showed that the smooths for both random intercepts were significant ( $p_{char} < 0.001$ ;  $p_{spkr} < 0.001$ ), indicating the necessity to include random intercepts. Second, we added two additional random slopes. Model summary showed that the smooths for all four random effects were significant ( $p_{char} < 0.001$ ;  $p_{spkr} < 0.001$ ;  $p_{char,timepoint} < 0.001$ ;  $p_{spkr,tone} < 0.01$ ), indicating the necessity to include random slopes. By model comparison using `compareML()`, we confirmed that the model containing the full set of random effects is significantly better than those containing random intercepts only or containing random effects over character (word) only.

**Random smooths.** We further added random smooths to the model. The full set of random smooths is as follows:

- Random smooths over time point
  - `s(timepoint, char, ...)`: one by character (word),

- `s(timepoint, spkr, ...)`: one by speaker, and
- `s(timepoint, spkr, by = tone, ...)`: one by speaker and tone

First, we fitted the model with random smooths but without any random intercepts or slopes. To figure out whether the model was improved, the Akaike Information Criterion (AIC; Akaike, 1974) was used: the lower AIC value indicates the better model. Results showed that the model with random smooths only was better-performed than the previous model containing only random intercepts and slopes. The random smooths originally applied  $k=10$ . To reduce the computational inefficiency, we also tested a model using  $k=5$  in the random smooths. Model comparison showed that using  $k=5$  slightly improved the model. Statistical assessment of the  $k$  value was conducted on the final model.

Second, we added back the aforementioned random slopes to the model. Now, the model contained two random slopes and three random smooths. Random intercepts were not included, as random smooths already include the shift of intercept (Wieling, 2018). The model was further improved, supported by the lower AIC value and the significant result of model comparison.

Regarding whether to include random slopes when random smooths already exist in the model, neither Sós-kuthy (2017) nor Wieling (2018) included a random slope. According to Wieling (2018), this is because the random smooth by speaker and tone already contains the between-item variations. However, when comparing models, we found that including random slopes actually significantly improved the model. Therefore, we compared a set of models where combinations of random slopes and random smooths were manipulated, as shown in Table 2.2.4. AIC values showed that the model including the full set of random slopes and random smooths (i.e., model2) was the best-behaved model. This was further supported by model comparison.

Random effects		model1	model2	model3	model4	model5	model6
Random slope	s(char, timepoint, bs="re")						
	s(spkr, tone, bs="re")						
Random smooth	s(timepoint, char, bs = "fs", m = 1, k = 5)						
	s(timepoint, spkr, bs = "fs", m = 1, k = 5)						
	s(timepoint, spkr, by = tone.ordered, bs = "fs", m = 1, k = 5)						
AIC value		31328.61	31327.38	38526.38	31666.45	31328.61	32753.21

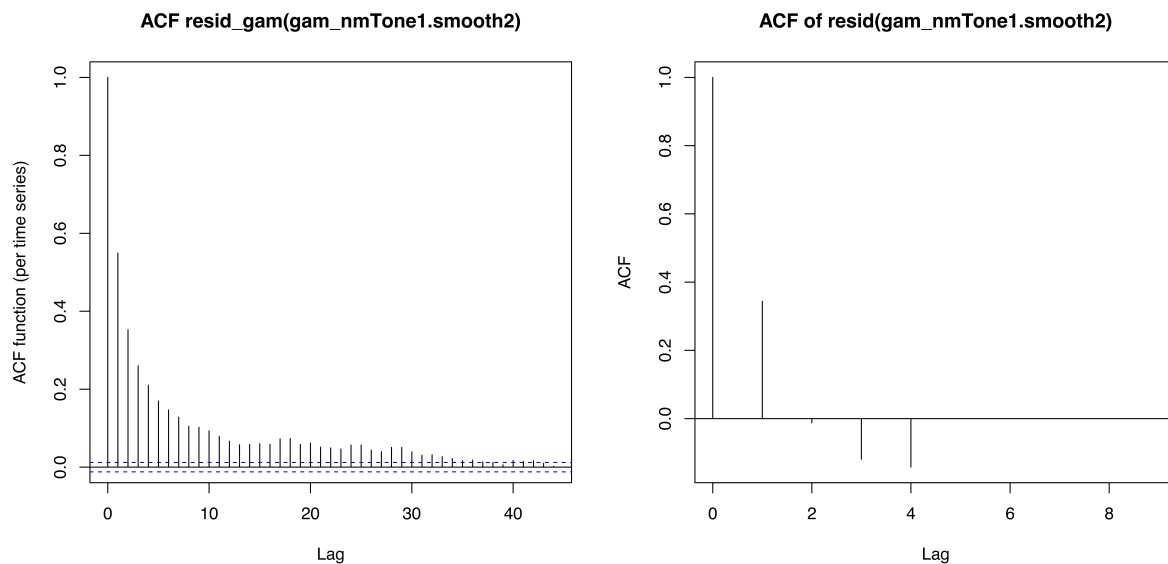
**Table 2.2.4:** Models containing different combinations of focus effects and their corresponding AIC values. Checkmarks indicate the contained random effects.

#### 2.2.4.4. (3) Add an AR1 error model

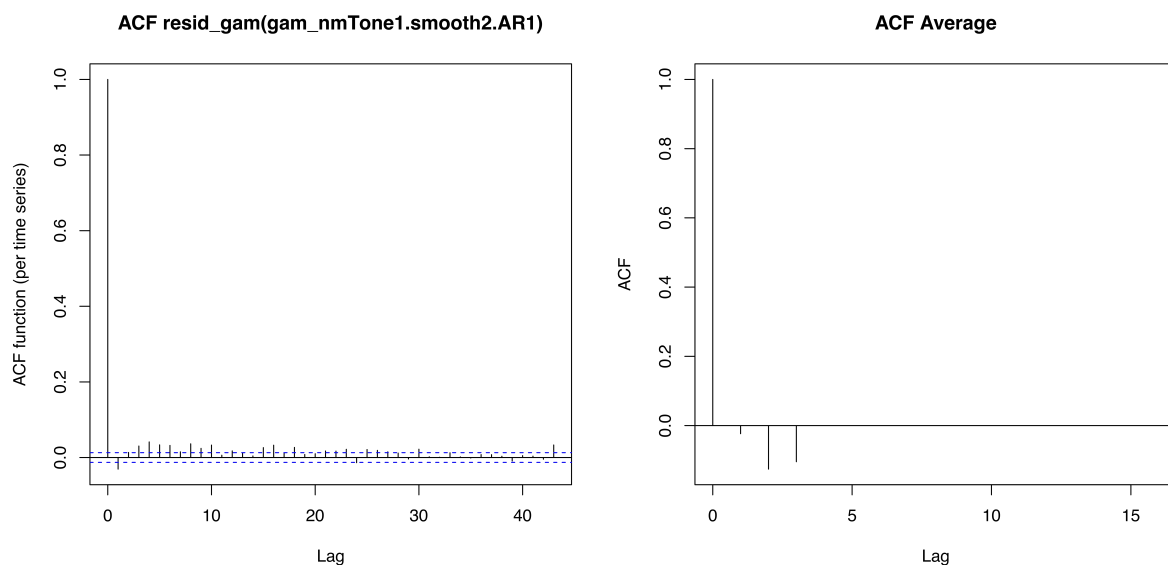
To examine the autocorrelation in the time series, we plotted the residual autocorrelation graph (i.e., ACF plots) for the so-far best model, model2, using `acf_resid()` in *itsadug* package. Figure 2.2.1 shows the residual autocorrelation overall and by trajectory. In each ACF plot, the height of the second line indicates the amount of autocorrelation at lag 1. In the left plot, the autocorrelation at lag 1 is 0.550, which means that each additional time point only yields about 43% additional information. The right plot shows the autocorrelation by trajectory, i.e., after excluding the autocorrelation between the last f0 point of the previous trajectory and the first f0 point of the next trajectory. Both ACF plots show that there is still an amount of autocorrelation remaining in the current model.

To further reduce residual autocorrelation, we also integrated an AR1 (auto-regression) error model. The AR1 error model took the *rho* value from the so-far best model ( $\rho = 0.550$ ). After adding the AR1 model, we plotted the ACF plots again, as shown in Figure 2.2.2. Both ACF plots show that the autocorrelation has been removed almost completely. According to model comparison, even after adding an AR1 model, the model containing the full set of random slopes and random smooths still

behaved significantly better than the previous models that contained no random structures or contained random smooths only.



**Figure 2.2.1:** Autocorrelation graph for the so-far best model, overall (left) and by trajectory (right). The height of the second line indicates the amount of autocorrelation at lag 1.



**Figure 2.2.2:** Autocorrelation graph for the so-far best model after adding an AR1 model ( $\rho = 0.550$ ), overall (left) and by trajectory (right). The height of the second line indicates the amount of autocorrelation at lag 1.

#### 2.2.4.5. (4) Add another fixed effect: f0 trajectory ~ tone + preceding consonant

To account for the potential microprosodic effect, we included another predictor of preceding consonant as the fixed effect in the model. To achieve this, we added a parametric parameter and a difference smooth over time point for the preceding consonant, because we assumed the preceding consonant would alter the shape of f0 contour.

Compared to the average smooth, significant differences were observed for the smooths of voiceless unaspirated, voiceless aspirated and fricative consonants ( $p_{vl.unasp} < 0.001$ ;  $p_{vl.asp} < 0.001$ ;  $p_{fricative} < 0.001$ ); the smooth for sonorant did not yield a significant difference from the average smooth ( $p_{sonorant} = 0.12$ ). This indicates the necessity of including the fix effect of the preceding consonant. Result of model comparison further showed that the model was significantly improved with the additional fixed effect. Along with the process of building the increasingly complex model, the confidence intervals of the estimated smooths became wider, and the significance patterns of difference smooths changed considerably. This also implies that the model was improved, as it was not overconfident anymore.

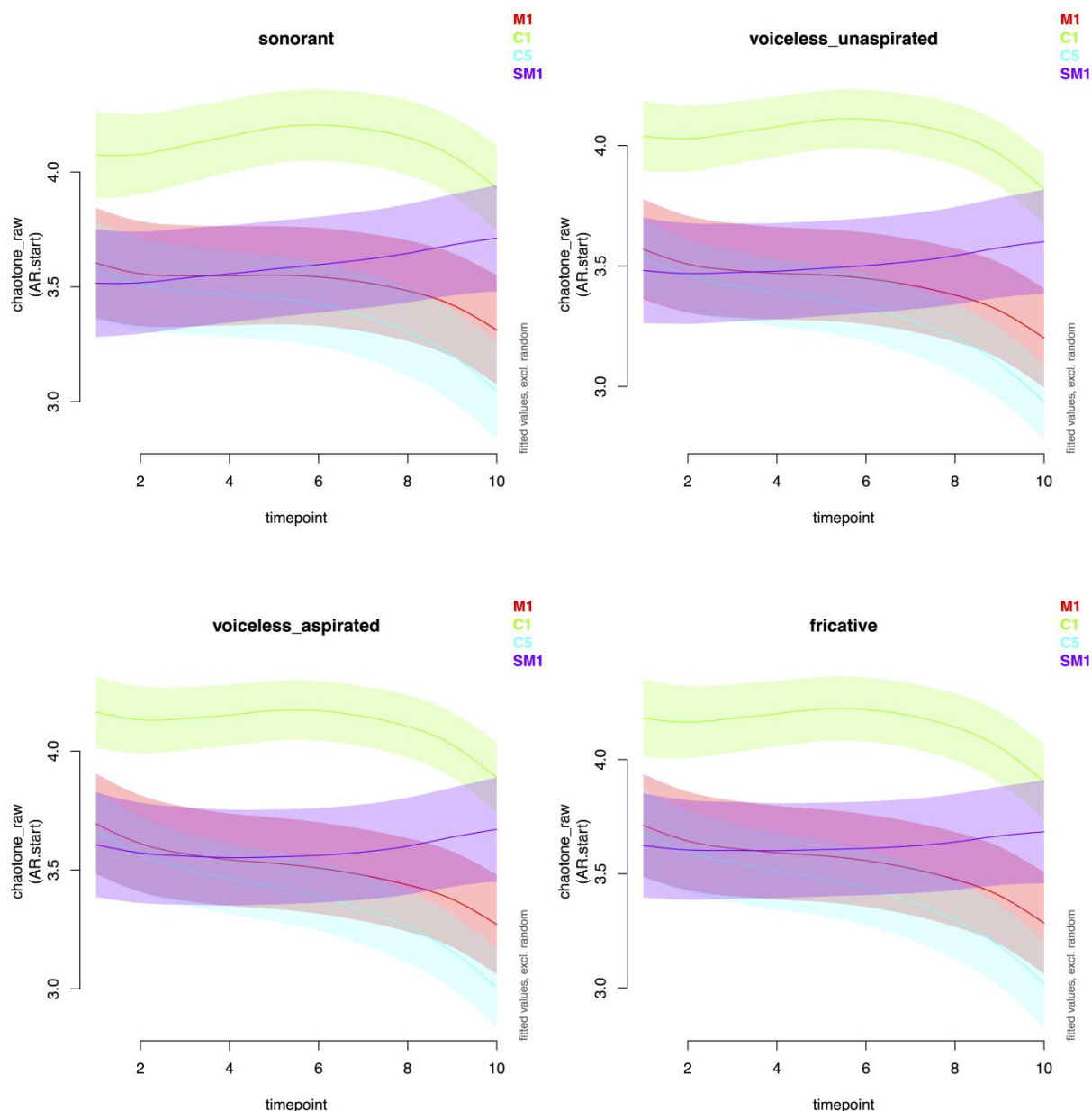
This resulted in the final version of the model described in § 2.2.4.1 at the beginning of this chapter.

#### 2.2.4.6. (5) Remaining discussions

***k* values.** We conducted a statistical assessment by running `gam.check()`. Results showed that all the smooths with *k* values had *k*-indices larger than 1, *p*-values far larger than 0.1 (either 0.99 or 1.00), and *edf* values much smaller than their corresponding *k'* values. This means that the *k* values we used were sufficiently large.

**Other predictors.** For now, we decided not to include other minor predictors, such as age groups, gender, and trial order. These variables are not the main research question of the current study, though it would be worthwhile to investigate. Due to the limitations of time and computational cost, they were not included at the current state.

**Interactions in main effects.** We did not further investigate the interactions between tone and preceding consonant. Our current model already showed significant main effects of the preceding consonant. In addition, this model was already complex and took a long computation time, even when using `fREML`. Therefore, we decided not to include these interactions in this model. This is also consistent with our statistical models for the intonation study.



**Figure 2.2.3:** Estimated  $f_0$  contours for the four level tones with four types of syllable-onset consonants.

To visually inspect the potential interaction between tone and preceding consonant, we plotted the smooths for each of the four consonant types, as in Figure 2.2.3. The smooth plots show that the preceding consonant only affected the onset of the f0 contour, as predicted for microprosodic effects, whereas the overall tonal pattern was not affected. Therefore, we considered the interaction between tone and preceding consonant as a secondary research question saving for future research.

**Interactions in random effects.** Given that the effect of tone category may vary across speakers, we also tried modifying the random smooths for speaker-specific variation (i.e., with the speaker variable) to examine the by-speaker tone effect. We created a combined grouping variable for speaker and tone using `interaction()` in *mgcv* package, and then replaced the original speaker variable to the interaction of speaker and tone. The resulting random smooths are as follows:

- `s(timepoint, SpkrTone.ord, bs = "fs", m = 1, k = 5)`
- `s(timepoint, SpkrTone.ord, by = tone.ordered, bs = "fs", m = 1, k = 5)`

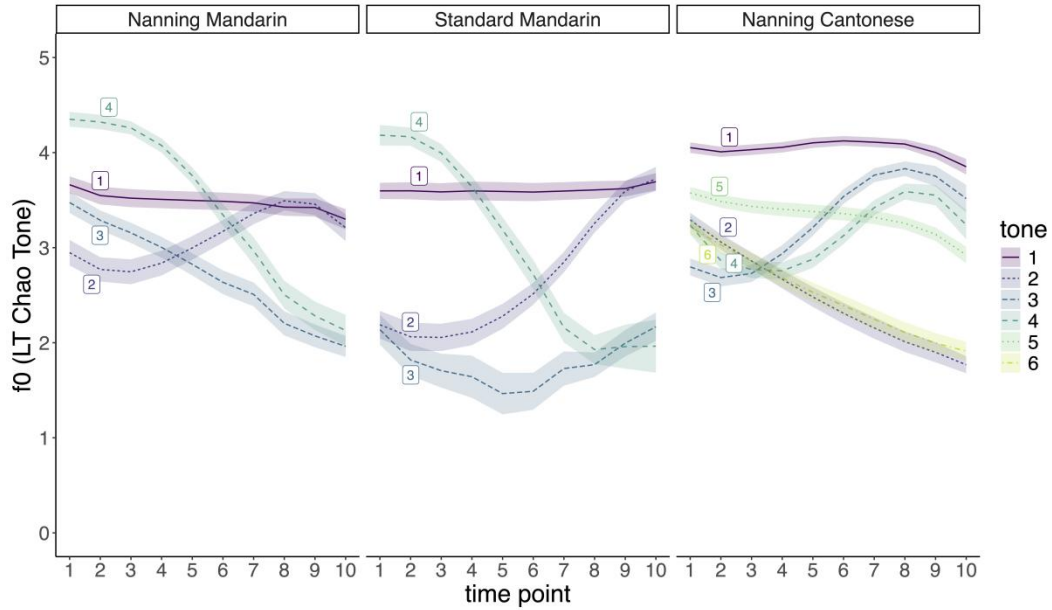
According to model comparison, including speaker–tone interaction did not significantly improve the model, and even more, raised the AIC value. Therefore, we would stick to the model without interaction in random smooths.

## 2.3. Results and discussions

### 2.3.1. Phonetic tone inventories

#### 2.3.1.1. Tone systems

Figure 2.3.1 displays the f0 contours in Chao tone numbers for each tone and language variety. Table 2.3.1 shows Chao tone numbers for each tone and language variety; the tone values were decided by the key f0 points, mostly the trajectories, in the ten-point f0 contours. The full set of f0 values in LT Chao tone numbers is presented in Appendix A.2.1.



**Figure 2.3.1:** F0 contours in Chao tone numbers of Nanning Mandarin, Standard Mandarin, and Nanning Cantonese for each phonological tone. Each ribbon represents 1 standard error from the mean.

Tone	Nanning Mandarin	Standard Mandarin	Tone	Nanning Cantonese
1	44	44	1	54
2	34	34	2	42
3	42	323	3	34
			4	34
4	53	52	5	43
			6	42

**Table 2.3.1:** Chao tone numbers for each phonological tone in Nanning Mandarin, Standard Mandarin, and Nanning Cantonese. Tones are aligned according to their historical correspondence.

Both Mandarin varieties have four phonological tones, whereas Cantonese has six phonological

tones, with a likely tone merger between Tone 2 and Tone 6. Although Nanning Mandarin inherited the ton categories of Standard Mandarin, when we zoom out to the use of the  $f_0$  space, Nanning Mandarin presents great similarity to Nanning Cantonese. The NM tone system shows an overall narrower  $f_0$  span than that of Standard Mandarin, mainly due to a higher low limit: SM Tone 3 reaches a tone height at about 1.5 LT values, while NM Tone 3 falls to only a height of 2 LT, similar to Nanning Cantonese. On the other hand, the upper limit of the NM tones, i.e., the onset of NM Tone 4, is the highest across the three varieties.

Furthermore, Nanning Mandarin appears to contrast with Standard Mandarin and shares greater resemblance to Nanning Cantonese. Unlike Standard Mandarin where a High–Low contrast is utilised, Nanning Mandarin displays a High–Mid–Low contrast, most saliently at the onset. Specifically, at the onset, NM Tone 4 begins in the high range, NM Tone 1 and Tone 3 begin in the middle range, and NM Tone 2 begins in the low-to-mid range; while at the offset, NM Tone 1 and Tone 2 end in the mid range, and NM Tone 3 and Tone 4 end in the low range. This three-level tonal contrast, however, resembles that of Nanning Cantonese, which is clearer at the offset.

Nevertheless, it is unclear why Nanning Mandarin and Nanning Cantonese, with one more level of contrasts than Standard Mandarin, occupied narrower  $f_0$  spans. This could be due to tonal factors: for instance, Nanning Mandarin and Nanning Cantonese might rely more heavily on height contrast than on contour contrast. If so, this perhaps also explains why two pairs of (near) tone merger occurred in Nanning Cantonese, although investigations in cue weighting require further perception study. It may also be related to temporal factors, such as shorter syllabic duration or faster speech rate, which can restrict the magnitude of contour change. This also requires future investigations to support.

In terms of the tonal inventories, the tonal descriptions of the present study will be compared to those in the previous literature. Tables below display the comparison of tone systems for Standard Mandarin (Table 2.3.2), Nanning Cantonese (Table 2.3.3), and Nanning Mandarin (Table 2.3.5). For Standard Mandarin, as previous literature and textbooks have reached consensus on the tonal description, only one reference was selected here for representation. Previous descriptions of the NC

and NM tones were repeated from Table 1.2.5 and Table 1.2.6 with some adjustments.

Overall, the tonal transcriptions of all three varieties are in line with the previous transcriptions.

For **Standard Mandarin** (Table 2.3.2), the patterns of tonal contour generally matched the previous descriptions, containing a high-level Tone 1 [44], a rising Tone 2 [34], a low-dipping Tone 3 [323], and a falling Tone 4 [52]. (For more detailed discussions, refer to the original work (Zhao, 2024).)

Source	Tone 1	Tone 2	Tone 3	Tone 4
W.-S. Lee & Zee (2003)	55	35	214	51
Tone pattern	high-level	mid-to high rising	low (dipping)	high-to-low falling
Present study	44	34	323	52
Tone pattern	high level	mid rising	low (dipping)	high-to-low falling

**Table 2.3.2:** Comparison of the SM tone systems from different sources.

For **Nanning Cantonese** (Table 2.3.3), the tonal patterns were also largely consistent with the previous auditory and acoustic transcriptions. Our findings identified a high-level Tone 1 [54], a mid-falling Tone 2 [42], a mid-rising Tone 3 [34] and Tone 4 [34], a mid-level Tone 5 [43], and another mid-falling Tone 6 [42].

Source	Tone 1	Tone 2	Tone 3	Tone 4	Tone 5	Tone 6
Xie (1994)	55	21	35	13	33	22
Hong (1989)						
H. Yang (1997)						
GX Doc (1998)						
NN Doc (1998)						
S. Li (2002)	55	21	35	24	33	22
Y. Lin & Qin (2008)						
H. Chen & Lin (2009)						
Teng (2018)						
Tang (2020)						
Xie (2007)	55	21	35	23	33	22
Shi S. (2007)	55	31	35	34	44	32
S. Wang (2015)	55	21	35	24	33	31
T. Luo (2015)	55	31	35	24	33	21
Tone pattern	high level	low falling	mid rising	low or mid rising	mid level	low level; low or mid falling
Present study	<b>54</b>	<b>42</b>	<b>34</b>	34	43	<b>42</b>
Tone pattern	high level	mid falling	mid rising	mid rising	mid level	mid falling

**Table 2.3.3:** Comparison of the NC tone systems from different sources. Bold indicates the tone values different from all the previous transcriptions.

First of all, the tonal descriptions of the level tones and rising tones in our findings were generally close to the previous findings. Both of the level tones (Tone 1 and Tone 5) had declined contours, different from the previous descriptions, where the tone contours remained at the same height. For the rising tones, our findings identified the same transcription of [34] for both Tone 3 and Tone 4, although the contour (in Figure 2.3.1) showed that Tone 3 had an earlier and slightly lower f<sub>0</sub> valley with a sharper rise to a higher peak, compared to Tone 4. The higher f<sub>0</sub> peak of Tone 3 than Tone 4

is in fact in line with previous transcriptions; the lower  $f_0$  valley of Tone 3, however, is not. Nevertheless, the discrepancy in  $f_0$  valley between the two rising tones was subtle, which may indicate an ongoing near tone merger between Tone 3 and Tone 4. It is also possible that the onset of Tone 4 is perceived as lower than that of Tone 3, as unanimously documented by the auditory transcriptions; perception studies are required to further validate this account.

Larger differences from previous transcriptions occurred in the falling tones, Tone 2 and Tone 6. For Tone 2, although our findings also identified a falling pattern, the tonal height was higher than that in previous descriptions, especially at the onset. For Tone 6, previous literature has reported discrepant patterns, with a trend of merger to Tone 2 especially by young speakers. We also observed this tone merger, consistent with findings in previous auditory (Hong, 1989; Y. Lin & Qin, 2008; GX Doc, 1998) and acoustic (T. Luo, 2015; S. Wang, 2015; Xian & Yang, 2022) studies. This merger is presented in the tone value (both [42]) and more clearly in the tone contour, which will be statistically compared by contour modelling in § 2.3.2.2. Similar to Tone 2, Tone 6 showed a higher  $f_0$  height than previous transcriptions, especially at the onset. Overall, the differences between our study and previous research are mainly in tonal height; specifically, the lowest tone registers (Tones 2, 4 and 6) are higher in our descriptions. In previous studies, the transcription of low tone registers was relatively inconsistent, either in the mid or low pitch range. This suggests that these lowest tone registers may inherently be less stable and be more variable, which is actually demonstrated in the tone contours from each individual speaker in Figure 2.3.3.

Table 2.3.4 summarises the tonal contrast of the NC tones in the previous study and our study. Before the T6–T2 merger, Tone 6 was a low-level tone, and the tonal contrast consisted of high, mid, and low level, mid and low rising, and low falling. After the T6–T2 merger, the low-level tone shifted to a mid–low fall, creating an even distribution of two tones for each contour type. Possibly due to this restructuring, a wide tonal space was no longer necessary, leading to the general upward shift of tones originally located in the mid to low range. As a result, the NC tone system no longer reaches a level as low as 1 LT. Consequently, the three varieties generally occupy a tonal space from 2 to 5 LT, though they differ in exact height.

Source	Tone contrast		
<b>Previous findings (before T6 merger)</b>	H level (T1)		
	M level (T5)	M rise (T3)	
	L level (T6)	L–M rise (T4)	L–M fall (T2)
<b>Previous findings (after T6 merger)</b>	H level (T1)		
	M level (T5)	M rise (T3)	
		L–M rise (T4)	L–M fall×2 (T2, T6)
<b>Our findings</b>	H level (T1)		
	M level (T5)	M rise×2 (T3, T4)	M fall×2 (T2, T6)

**Table 2.3.4:** Summary of tonal contrasts in Nanning Cantonese from different sources.

However, this mechanism fails to explain why tone merger occurs: why does raising the pitch of the lowest tone registers result in two pairs of (near) tone mergers, instead of other pitch adjustments to maintain tonal contrasts, such as raising the mid-rise and mid-fall tones? Tone merger may therefore be driven by additional factors, which is a field still remaining unclear (see Xian and Yang (2022) for a brief discussion).

For **Nanning Mandarin** (Table 2.3.5), at a glance, the NM tone–contour mappings reflect those of Standard Mandarin, but with a noticeable lack of dipping in NM Tone 3. Our tonal descriptions are generally consistent with previous studies. As the prior studies from small sample sizes display high variation, our NM transcriptions report identical tone values as some of the previous transcriptions. Our findings identified a mid-level Tone 1 [44], a mid-rising Tone 2 [34], a mid-falling Tone 3 [42], and a high-falling Tone 4 [53].

Source	Tone 1	Tone 2	Tone 3	Tone 4
<b>Y. Lin (2008)</b>	33/44	24 to 34/35	21	53/55
<b>Fan (2020)</b>	55 (sometimes 33/44)	35 (sometimes 24)	21 (sometimes 31/41/42; occasionally 213)	51
<b>Y. Liang (2005)</b>	44	24/35	21	53/54
<b>T. Li (2012)</b>	between 33 and 55	Narrower pitch range and lower f0 level than SM Tone 2	21	53/54
<b>H. Yang (2003)</b>	33/44	Mostly 12/13	falling without dipping	Mostly 31/42
<b>Flaming (2020)</b>				Falling, flatter with a narrower range than SM Tone 4
<b>Tone pattern</b>	mid or high level	mid or low rising	low or mid falling (occasionally dipping)	high level; high-to-mid, high-to-low, or mid falling
<b>Present study</b>	44	34	42	53
<b>Tone pattern</b>	mid level	mid rising	mid falling	high-to-mid falling

**Table 2.3.5:** Comparison of the NM tone systems from different sources.

For Tone 1, previous auditory studies reported that NM Tone 1 is lower than SM Tone 1 (Fan, 2020; T. Li, 2012; Y. Liang, 2005; Y. Lin, 2008). However, in terms of absolute pitch (LT Chao tone values), NM Tone 1 is not lower than SM Tone 1, as demonstrated in Figure 2.3.1 and Table 2.3.1. Nevertheless, within the tonal space, NM Tone 1 is indeed positioned relatively lower than SM Tone 1 when referenced against the highest point. Furthermore, it exhibits a declined contour, resulting in a final pitch considerably lower than SM Tone 1. These findings may account for the lower NM Tone 1 perceived in previous auditory transcriptions. This declined contour of NM Tone 1 is similar to NC Tone 5, although the slope of the decline is gentler; conversely, SM Tone 1 exhibits a slightly

rising contour.

For Tone 3, previous descriptions reported a low-falling tone, whereas we observed a mid-falling tone, where the contour and height are both considerably similar to the NC falling tones. The reason for the upward shift in tone height may parallel the explanation given previously for the rise of the lowest tone registers in Nanning Cantonese. Though NM Tone 3 did not dip like SM Tone 3, it did exhibit a similar initial fall. In Standard Mandarin, the realisation of Tone 3 is highly contextual, and critically only dips in phrase-final position (F. Shi & Ran, 2011; Yip, 2002). In Nanning Mandarin, it may be that only the core portion of Tone 3 was realised without the boundary rise. The mid-falling realisation of NM Tone 3 may also relate to the merger of Tone 6 (traditionally low-level) with Tone 2 (mid-falling). This tone merger resulted in a lack of a low-level tone in the NC inventory. Given that SM Tone 3 is believed to be underlyingly low-level (F. Shi & Ran, 2011; Zhou, 2017), the lack of a low-level template in Nanning Cantonese potentially accounts for why NM Tone 3 was not realised even as the non-phrase-final SM Tone 3 [21]. The falling NC Tone 6 could have potentially served as the initial template for NM Tone 3, giving rise to a mid-falling instead of low-falling tone [42].

It is worth noting that during the recording, we also observed instances of the dipping contour mentioned in Fan (2020). However, due to the formal reading setting, the researcher (a native NM speaker) judged that many participants switched to a more formal register while reading the items. Therefore, these tokens were excluded following auditory validation (discussed in detail in § 2.2.1.2). To establish whether NM speakers produce a dipping Tone 3 in naturalistic NM, a more ideal approach would be to embed it in utterances (e.g., the sentence-final position in the target sentences used in our intonation study).

For the dynamic tones (Tone 2 and Tone 4), we also observed a narrower pitch range and a flatter contour than the corresponding SM tones, consistent with previous studies. For Tone 2, the  $f_0$  range in our transcription is narrower than that in most of the previous studies. Previous auditory studies found a lower onset and offset, resulting in an overall lower  $f_0$  space than SM Tone 2. However, we

only observed a lower offset but a higher onset. Additionally, Li (2012) suggested that the initial low portion of NM Tone 2 is longer than in SM Tone 2, resulting in a curlier contour. In our data, however, this characteristic is reversed, appearing in SM Tone 2 instead. NM Tone 2 displays an S-shaped contour, more similar to the NC rising tones; whereas SM Tone 2 exhibits a longer and deeper dipping portion followed by a sharper rise, without an arch shape or a final fall at its f0 peak.

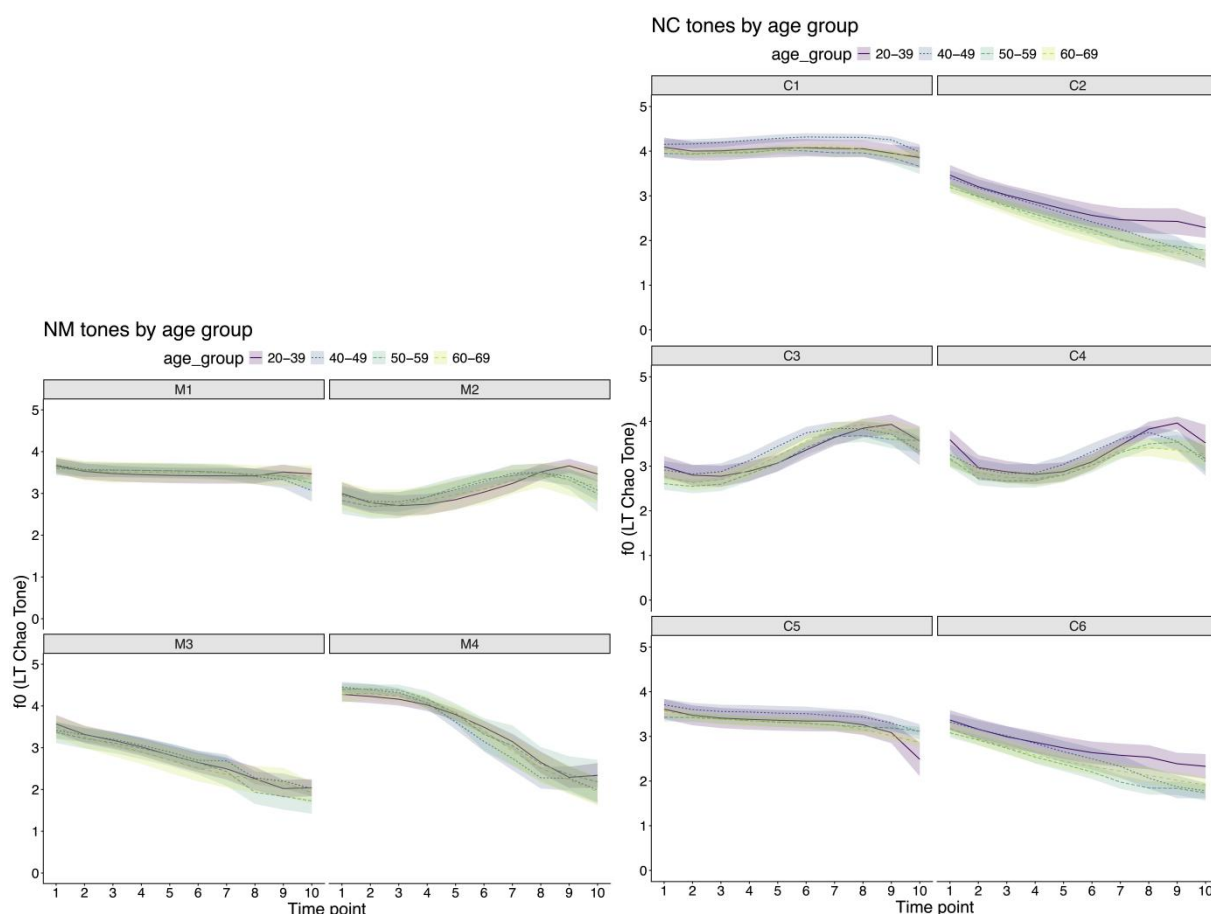
Tone 4 is the most controversial tone in previous transcriptions, although most sources agreed that the onset begins at the highest level [5]. Compared with the differences between NM and SM Tone 2, NM Tone 4 is not substantially narrower or flatter than SM Tone 4. We observed a higher onset and offset in NM Tone 4 than SM Tone 4; the higher offset is consistent with Liang (2005) and Lin (2008), whereas the higher onset, in our view, may be noteworthy. The absence of the Tone 3 pre-pausal rise resulted in two falling tones, NM Tone 3 and Tone 4. The dispersed f0 onsets and different slopes of the two falling tones could reflect a need for tonal contrast. The dispersed f0 onsets also correspond to the mid and high levels of the three-level tone system that is common to Nanning Cantonese, but distinct from Standard Mandarin.

The preceding discussion of the NM tones has inevitably involved comparisons with the SM and NC tones. Our findings show that the phonetic patterns of the NM tones are systematically different from the SM tones, although characteristics from both the target and source languages were observed. To enable a finer and more reliable comparison of the NM contour with the NC and SM contours, two methods will be employed in the subsequent sections: phonetic distance (§ 2.3.2.1) and contour modelling (§ 2.3.2.2 and § 2.3.2.3).

### **2.3.1.2. Age difference**

For Nanning Mandarin and Nanning Cantonese, f0 contours were also plotted by age group to investigate potential age-based variation. This analysis was motivated by different factors for each variety. Nanning Mandarin, as an unstable interlanguage, is experiencing ongoing change, particularly resulting from increasing contact with Standard Mandarin. For Nanning Cantonese, this

aligns with established distinctions in the literature, which often differentiate between an “old” and a “new” variety of the dialect (e.g., Lin & Qin, 2008).

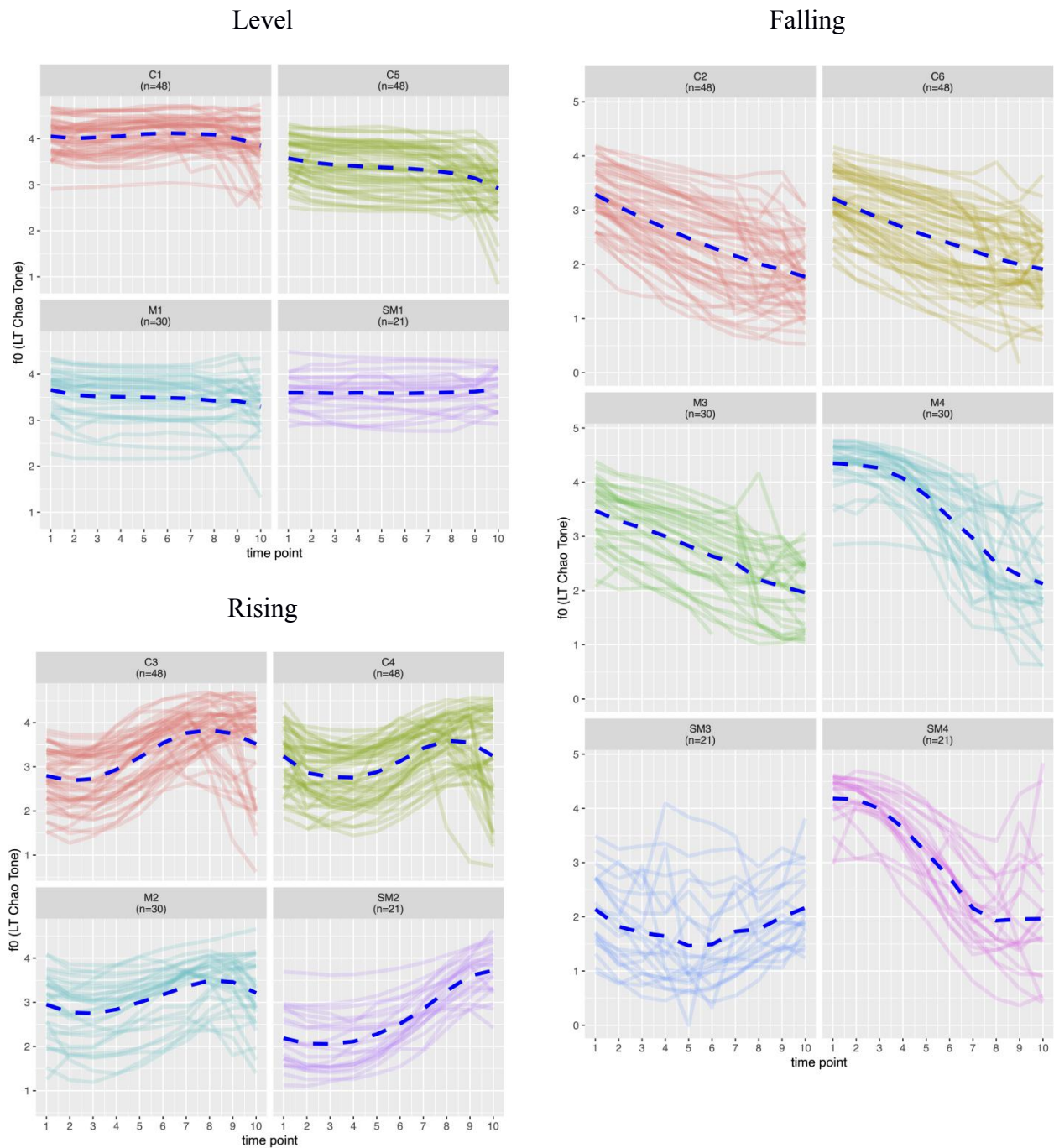


**Figure 2.3.2:** F0 contours in Chao tone numbers of Nanning Mandarin and Nanning Cantonese by each of the four age groups for each phonological tone. Each ribbon represents 1 standard error from the mean.

In Nanning Mandarin, we did not observe a salient age difference, except for in NM Tone 2, where the youngest speakers (aged 20–39) slightly stood out. In contrast, in Nanning Cantonese, the two younger age groups (20–39 and 40–49) stood out in several tones. Specifically, the youngest age group (20–39) presented a higher final pitch in the falling and rising tones, and a sudden falling tail in the mid-level Tone 5. Age group 40–49, on the other hand, showed the overall highest f0. More reliable analysis of the age effect requires statistical evidence.

### 2.3.1.3. Individual variation

To illustrate individual variation, individual-based  $f_0$  contours were plotted for each tone and grouped by similar contour shapes (i.e., level, rising, falling).



**Figure 2.3.3:** F0 contours in Chao tone numbers of Nanning Mandarin and Nanning Cantonese, plotted for individual speakers across phonological tones.

In the level and rising tones, a final fall in the contours was often observed in Nanning Mandarin and Nanning Cantonese, while absent in Standard Mandarin. The final portion of the contours displayed more variations across language varieties. When establishing a variation hierarchy within each language variety, both Mandarin varieties showed the least variation in the high-falling Tone 4 and the most variation in Tone 3. SM Tone 3, in particular, showed great individual variations along the entire contour, likely due to the co-occurrence with creaky voice. On the other hand, Nanning Cantonese showed the least variation in the level tones, and the most variation in the falling tones. While these patterns suggest structured individual variability within each variety, these observations remain descriptive and require further statistical modelling for confirmation.

### **2.3.2. Phonetic similarity of similar-contoured tones**

To investigate whether individual NM tones share greater similarity with Standard Mandarin or Nanning Cantonese, we examined the similarity of each NM f<sub>0</sub> contour to similarly-shaped NC and SM f<sub>0</sub> contours. Specifically, we kept Nanning Mandarin as the anchor and grouped the tones by phonetic contour shape (level, rising, low falling, and high falling):

- Level: Compare NM Tone 1 relates to NC Tone 1 and 5, SM Tone 1
- Rise: Compare NM Tone 2 relates to NC Tones 3 and 4, SM Tone 2
- Low fall: Compare NM Tone 3 relates to NC Tone 2 and 6, SM Tone 3 and 4
- High fall: Compare NM Tone 4 relates to NC Tone 2 and 6, SM Tone 3 and 4

The overall similarity of tone contours was investigated by root-mean-squared deviation (RMSD), and the time-varying similarity was investigated using GAMMs.

#### **2.3.2.1. Overall phonetic distance**

Root-mean-squared deviation (RMSD) between the averaged f<sub>0</sub> contours was calculated from all ten timepoints in Chao tone numbers. Lower RMSD values correspond to a higher similarity between

contours. As shown in Table 2.3.6, NM Tones 1 and 4 were most similar to SM f0 contours, though for NM Tone 1, this similarity was closely followed by that of NC Tone 5. On the other hand, NM Tones 2 and 3 were most similar to NC f0 contours.

To further investigate differences in the exact contour independent of f0 height and any potentially remaining individual variation, the RMSD values were also calculated after z-scoring the raw f0 in hertz by speaker and tone. The results still indicated that NM tones mostly shared the greatest similarity to NC tones. The only exceptional case is NM Tone 4, which was still most similar to SM Tone 4, though the other three falling tones became substantially more similar to NM Tone 4.

<b>NM Tone</b>	<b>Tone Comparisons</b>				
<b>NM1 (Level)</b>		<b>NC1</b>	<b>NC5</b>	<b>SM1</b>	
	raw	0.56	0.18	<u>0.17</u>	
	z-scored	0.27	<u>0.20</u>	0.23	
<b>NM2 (Rising)</b>		<b>NC3</b>	<b>NC4</b>	<b>SM2</b>	
	raw	0.26	<u>0.12</u>	0.60	
	z-scored	<u>0.15</u>	0.17	0.36	
<b>NM3 (Mid-fall)</b>		<b>NC2</b>	<b>NC6</b>	<b>SM3</b>	<b>SM4</b>
	raw	0.27	<u>0.24</u>	1.09	0.52
	z-scored	<u>0.09</u>	<u>0.09</u>	0.92	0.21
<b>NM4 (High-fall)</b>		<b>NC2</b>	<b>NC6</b>	<b>SM3</b>	<b>SM4</b>
	raw	1.03	1.00	1.86	<u>0.03</u>
	z-scored	0.25	0.25	1.01	<u>0.19</u>

**Table 2.3.6:** RMSD values on the Chao tone numbers listed first in each cell, and on the z-scored values, second. Each row presents a comparison of an NM tone to tones with a similar contour in Nanning Cantonese and Standard Mandarin. Underlining indicates the smallest RMSE numbers, that is, the highest similarity to NM contours.

### 2.3.2.2. Contour similarity with f0 height differences

Building on the initial RMSD comparison of averaged contours, we conducted a more dynamic, fine-grained, and statistically robust analysis using GAMMs. The tone contours were compared within the tone groups that share similar contour shapes. We used *mgcv* package (Wood, 2011) in R to implement GAMMs and *itsadug* package (van Rij et al., 2015) in R for plotting.

As per the final model for the level tone group, we built up the models for all four tone groups. One slight difference is the *rho* values: in the testing phrase, the *rho* values were from the model without the AR1 model and the predictor of preceding consonant. Now, we would use an updated *rho* value from the model without the AR1 model, but with the preceding consonant. This says, the final model and the model that the *rho* value was from only differed in the presence of the AR1 model. This would slightly change the *rho* value, but it would still be a good estimate of the amount of autocorrelation. All models underwent `gam.check()` to ensure the *k* values were sufficiently large.

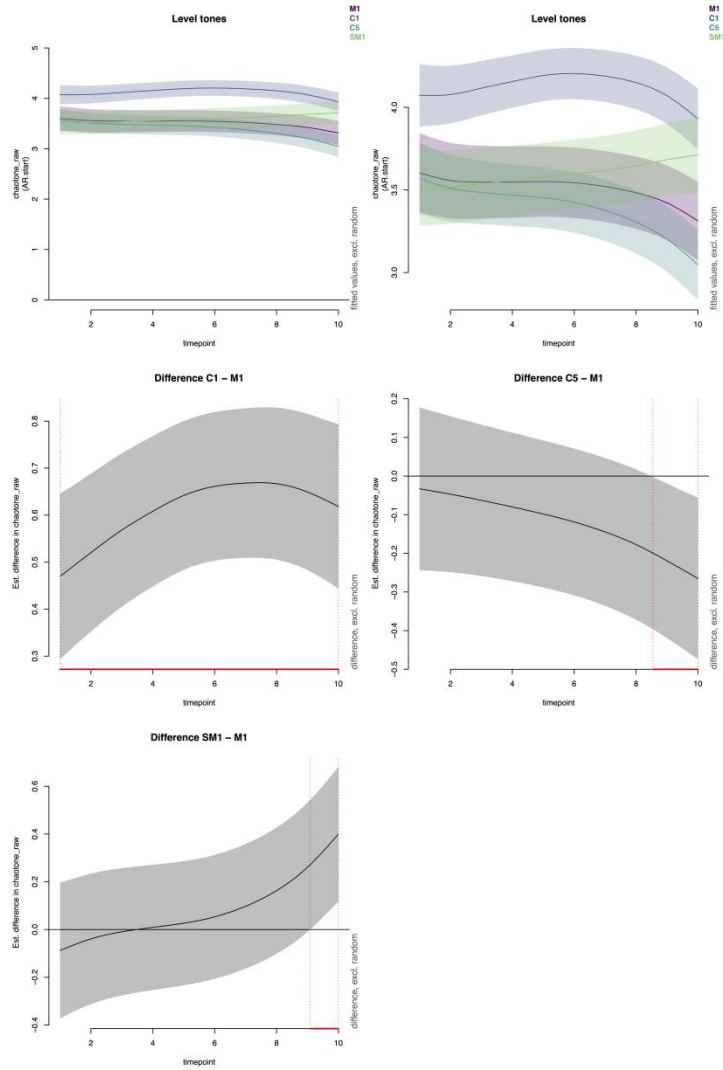
Following the tutorial of Sóskuthy (2017), we used a visual method for significant testing, i.e., plotting the difference smooths, which can demonstrate the exact nature of the difference between two contours. The model predictions of each tone, grouped by the similar contour, are also provided for graphical illustration. Figures Figure 2.3.4–Figure 2.3.7 show the predicted contours and the contour difference smooths for each of the four contour groups: level, rising, mid-falling, and high-falling. Within each figure, the two plots (in the top row) are the predicted contours of each member within the contour group. The preceding consonant was set to sonorant, because the *itsadug* package cannot plot on the average f0 smooths, and the sonorant onset is theoretically the condition without microprosodic effects on the tone contour. For the y-axes of the predicted contours, the left plot took the scale of Chao tone numbers (0–5), while the right plot took a scale automatically generated from the overall f0 space taken by each tone group, which can maximise the contour difference in display. The x-axis took the scale of the ten time points. The rest of the plots (in the middle and bottom rows) show the difference of pairwise comparisons within the contour group, i.e.,

comparing NC and SM tones against the NM tone baseline within the similar-contour group.

### **NM Tone 1: level**

NM and NC level tones had a slightly falling trend, while the SM level tone had a slightly rising trend. NC Tone 1 significantly differed from NM Tone 1 throughout the entire contour, mainly due to the higher f<sub>0</sub> height. This makes it necessary to further compare the f<sub>0</sub> contours after centring the f<sub>0</sub> height. NC Tone 1 was closest to NM Tone 1 at the onset, followed by the offset; the distance between f<sub>0</sub> points 6 to 8 reached the maximum. NC Tone 5 and SM Tone 1 diverged from NM Tone 1 only at the end of the contour: NC Tone 5 from about f<sub>0</sub> point 8, and SM Tone 1 from about f<sub>0</sub> point 9. NC Tone 5 showed a slightly greater declination, leading to a falling tail significantly lower than NM Tone 1; while SM Tone 1 had a rising tail, resulting in the divergence towards the end of the contour: NC Tone 5 was significantly lower than NM Tone 1, whereas SM Tone 1 was significantly higher than NM Tone 1.

Overall, NM Tone 1 showed more similarity to SM Tone 1, as indicated by the shortest significant region.



**Figure 2.3.4:** Overall f0 contours and difference curves for level tones. Shaded bands represent pointwise 95%-confidence intervals.

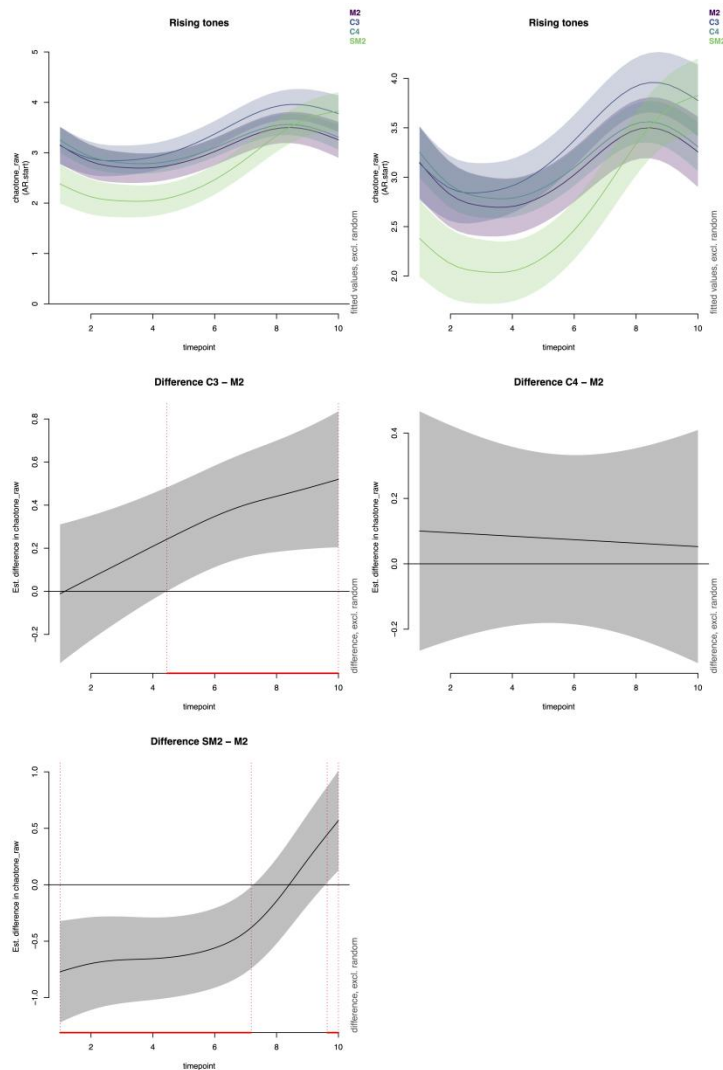
Top row: overall f0 contours (excluding random effects) for each tone by timepoint (1–10). Tones are ordered from dark to light: NM T1 (purple), NC T1 (blue), NC T5 (cyan), SM T1 (green).

Middle and bottom rows: difference curves of NC and SM tones relative to NM T1 (baseline). Difference between NM T1 and: NC T1 (significant at points 1–10); NC T5 (significant at points 8.5–10); SM T1 (significant at points 9.1–10).

## **NM Tone 2: rising**

The NM and NC rising tones presented an S-shape, whereas the SM rising tone had a different rising contour, with a longer and deeper dip, a sharper rise towards the end, and no final falling. SM Tone 2 had a significantly lower contour than NM Tone 2 until about point 7, followed by the intersection, and then the divergence at the end, resulting in a wider  $f_0$  range than NM Tone 2. The non-significant portion is mainly due to the intersection. On the other hand, the NC rising tones almost overlapped with the NM tone, especially at the onset. Specifically, NC Tone 3 showed the overlap at the onset and the divergence after point 4, where the contour was rising to a higher  $f_0$ . NC Tone 4, however, overlapped with the NM rising tone all along the contour, though with a very wide confidence interval in the difference smooth.

Overall, NM Tone 2 showed more similarity to NC Tone 4, as indicated by the non-significant difference smooth.



**Figure 2.3.5:** Overall f0 contours and difference curves for rising tones. Shaded bands represent pointwise 95%-confidence intervals.

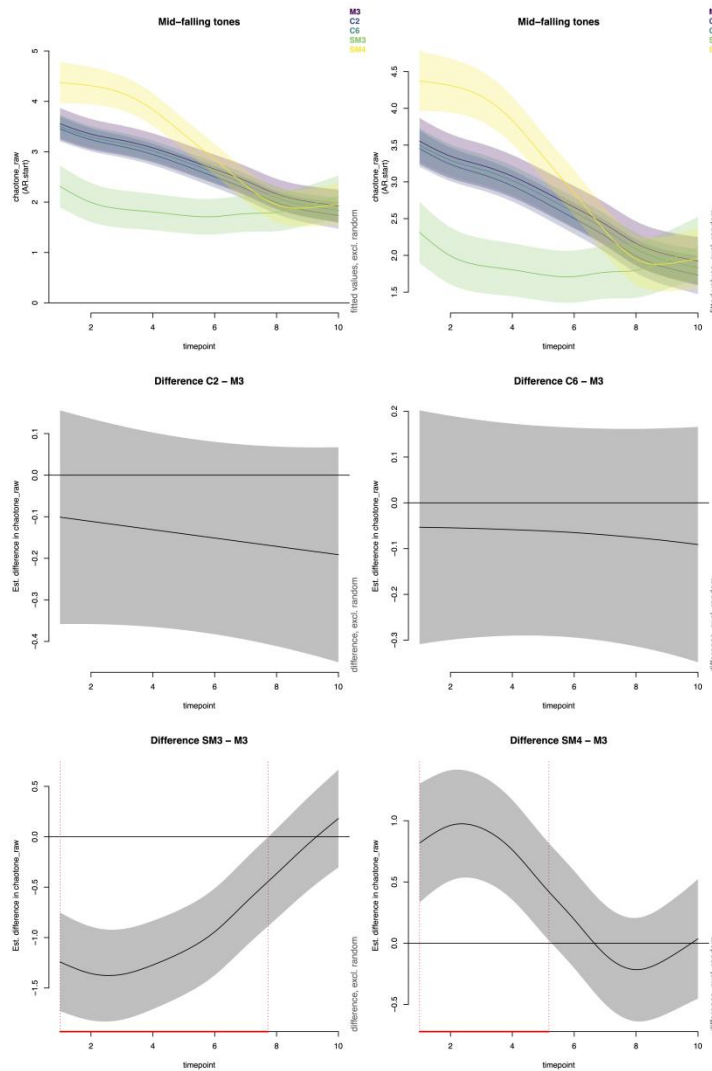
Top row: overall f0 contours (excluding random effects) for each tone by timepoint (1–10). Tones are ordered from dark to light: NM T2 (purple), NC T3 (blue), NC T4 (cyan), SM T2 (green).

Middle and bottom rows: difference curves of NC and SM tones relative to NM T2 (baseline). Difference between NM T2 and: NC T3 (significant at points 4.5–10); NC T4 (not significant); SM T2 (significant at points 1–7.2 and points 9.6–10).

### **NM Tone 3: mid-falling**

The NC and NM falling tones—all are mid-falling—showed an almost linear falling contour. In contrast, the SM falling tones had distinct contour shapes. SM Tone 3 was a low-dipping tone with slight wiggleness in the contour; the onset and offset of the contour were at about the same height. SM Tone 4 was a high-falling tone with a curved contour shape. Both NC mid-falling tones are not significantly different from the NM tone and showed very wide confidence intervals. Model prediction showed that the NC and NM tones almost overlapped along the entire contour, although a very slight divergence occurred towards the offset. Both SM tones got to overlap with the NC and NM tones towards the offset: SM Tone 3 from before point 8, and SM Tone 4 from about point 5. The overlapping portions lead to an insignificant difference from NM Tone 3.

Overall, NM Tone 3 showed more similarity to NC Tone 2 and 6, as indicated by the non-significant difference smooths.



**Figure 2.3.6:** Overall f0 contours and difference curves for mid-falling tones. Shaded bands represent pointwise 95%-confidence intervals.

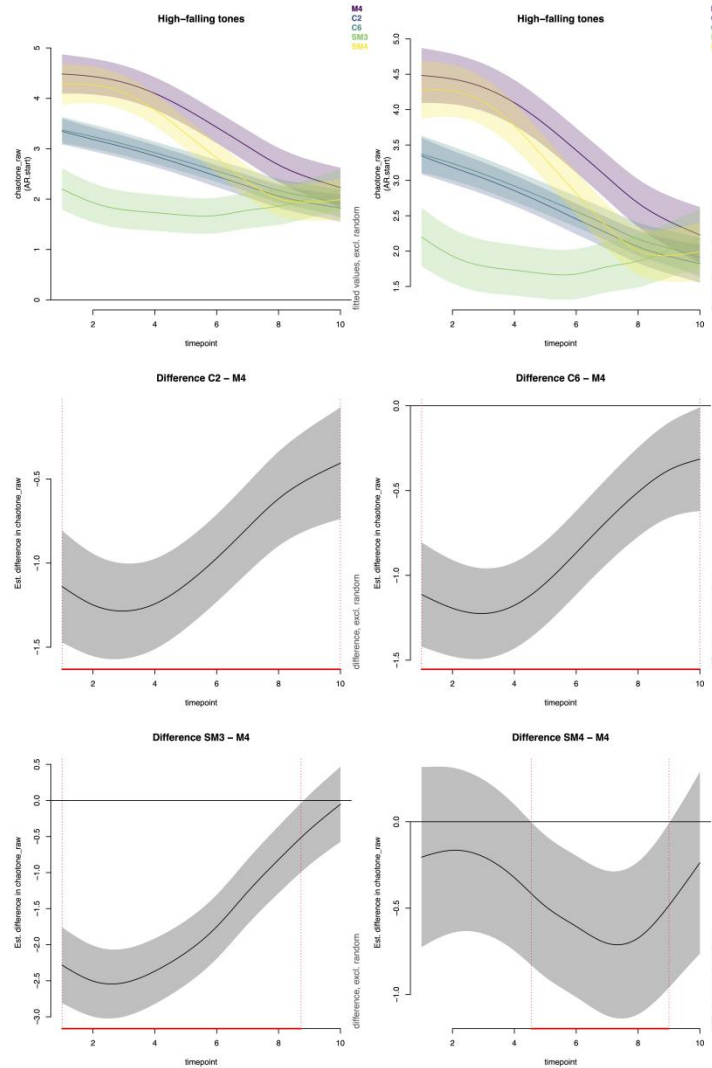
Top row: overall f0 contours (excluding random effects) for each tone by timepoint (1–10). Tones are ordered from dark to light: NM T3 (purple), NC T2 (blue), NC T6 (cyan), SM T3 (green), and SM T4 (yellow).

Middle and bottom rows: difference curves of NC and SM tones relative to NM T3 (baseline). Difference between NM T3 and: NC T2 (not significant); NC T6 (not significant); SM T3 (significant at points 1–7.7); SM T4 (significant at points 1–5.2).

#### **NM Tone 4: high-falling**

Similar to NM Tone 3, both SM tones were approaching NM Tone 4 towards the end, both at about point 9. Additionally, the initial portion of SM Tone 4 (before point 5) was not significantly different either. Previous visual observations identified a higher onset and offset for NM Tone 4 compared to SM Tone 4; however, this difference was not significant.

Overall, NM Tone 4 showed more similarity to SM Tone 4, as indicated by the shortest significant region.

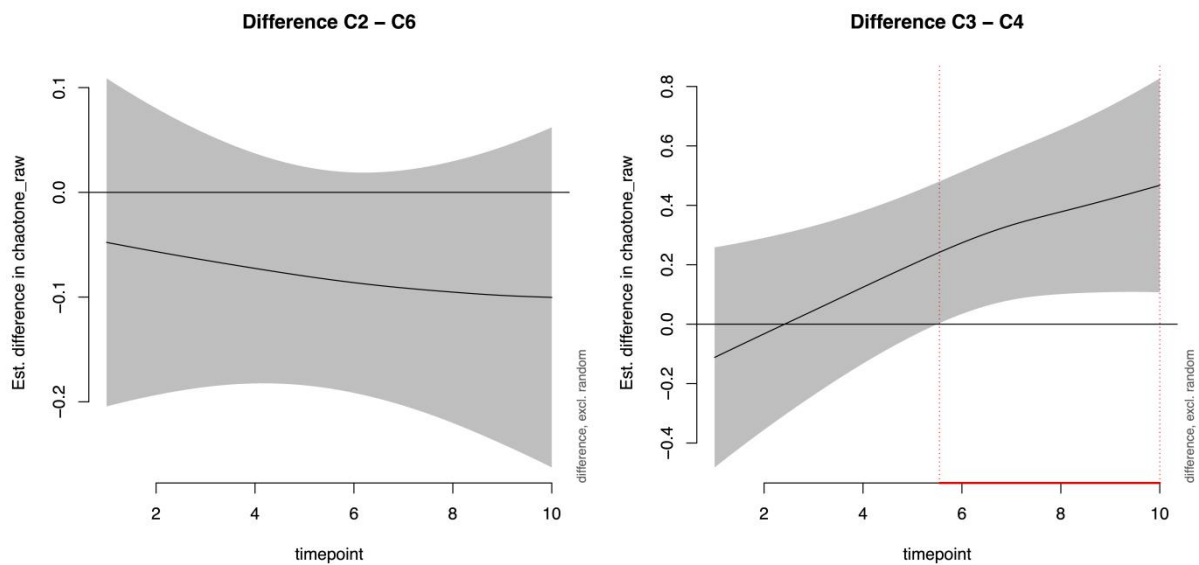


**Figure 2.3.7:** Overall f0 contours and difference curves for high-falling tones. Shaded bands represent pointwise 95%-confidence intervals.

Top row: overall f0 contours (excluding random effects) for each tone by timepoint (1–10). Tones are ordered from dark to light: NM T4 (purple), NC T2 (blue), NC T6 (cyan), SM T3 (green), and SM T4 (yellow).

Middle and bottom rows: difference curves of NC and SM tones relative to NM T4 (baseline). Difference between NM T4 and: NC T2 (significant at points 1–10); NC T6 (significant at points 1–10); SM T3 (significant at points 1–8.7); SM T4 (significant at points 4.5–9).

## Potential NC tone mergers



**Figure 2.3.8:** Difference curves between the two pairs of NC tone (near-) mergers. Left: difference between NC T2 and NC T6 (falling tones); difference not significant. Right: difference between NC T3 and NC T4 (rising tones); difference significant at points 5.5–10. Shaded bands represent pointwise 95%-confidence intervals.

We also ran post-hoc analysis on two pairs of NC tones with likely merger, i.e., NC Tones 2 and 6, and NC Tones 3 and 4. NC Tones 2 and 6 were not significantly different from each other, indicating a full tone merger in production. However, NC Tones 3 and 4 were different in contour after point 6. This is reflected in the estimated contours, where NC Tone 3 started at about the same height as NC Tone 4, but reached a higher  $f_0$  level at the end due to a sharper rise. Previous visual observations identified an earlier and slightly lower  $f_0$  valley for NC Tone 3 compared to NC Tone 4; however, this difference was not significant.

The merger of NC Tone 6 with Tone 2 is consistent with previous auditory and acoustic studies (Hong, 1989; Y. Lin & Qin, 2008; T. Luo, 2015; S. Wang, 2015; Xian & Yang, 2022; NN Doc, 1998); some of them only observed this merger in younger speakers' speech. In terms of the rising Tone 3 and Tone 4, a potential near-tone merger was identified. The majority of previous studies

transcribed NC Tone 3 and Tone 4 with different onsets and offsets, whereas our study identified non-significantly different onsets. This, however, is not in line with the findings in Xian and Yang (2022), where the Tone 3–4 pair had a higher degree of merger than the Tone 2–6 pair.

The divergence between the present GAMM results and the DTW-based findings reported by Xian and Yang (2022) can be attributed to methodological differences. DTW quantifies the overall shape difference by allowing temporal warping and summing up pointwise distances into a single distance measure (Senin, 2008). Therefore, it may be sensitive to local differences, which can increase the overall distance even if the contours are largely similar throughout most of their duration. In contrast, GAMMs assess whether two tonal contours differ systematically across time while taking speaker and token variability into account. Therefore, we assume it is possible that a pair of tones shows a comparatively larger DTW distance yet exhibits no consistent time-wide significant difference in the GAMM. From this perspective, GAMMs analysis may be able to reveal patterns of merger that a single-value distance measure like DTW cannot capture.

### **2.3.2.3. Contour similarity with f0 heights standardised**

Similar to the comparison of overall phonetic distance, we further examine contour similarity independent of the relative f0 height differences and any potentially residual individual variation. This need is particularly evident for level tones: visual inspection of Figure 2.3.4 suggests that NC Tone 1 had a contour shape similar to NM Tone 1; however, statistical analysis shows that the two differed significantly across the entire contour, due to the overall higher f0 height of NC Tone 1. This raises the question of whether two contours would share greater similarity if the f0 height difference were removed.

To address this, we centred the f0 values by *z*-scoring LT Chao tone values by speaker and tone. The same set of models was fitted on the *z*-scored data; random structures were maximally consistent across the tone groups, except for the models of falling tones, where the last random smooth `s(timepoint, spkr, by = tone)` were removed due to computational convergence issues. We

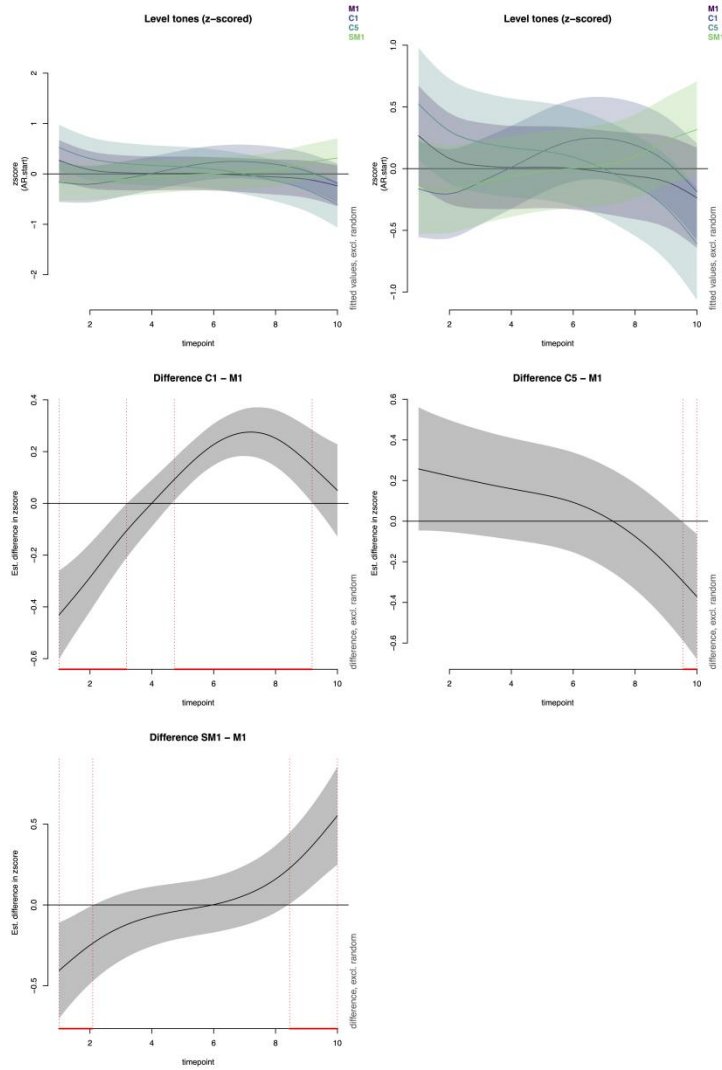
also examined the  $k$  values for the  $z$ -scored models using `gam.check()`. Although the results for rising tones showed some significant  $p$ -values, and the results for rising and high-falling tones showed some  $k$ -indices slightly lower than 1, the overall results indicated that the  $k$  values were appropriate for all four tone groups.

Figures Figure 2.3.8–Figure 2.3.12 show the predicted contours and the contour difference smooths after  $z$ -score standardisation, for each of the four contour groups: level, rising, mid-falling, and high-falling. Within each figure, the two plots (in the top row) are the predicted contours of each member within the contour group. The preceding consonant was set to sonorant, because the *itsadug* package cannot plot on the average  $f_0$  smooths, and the sonorant onset is theoretically the condition without microprosodic effects on the tone contour. For the  $y$ -axes of the predicted contours, the left plot took the same range of Chao tone numbers (0–5) but centred at zero, while the right plot took a scale automatically generated from the overall  $f_0$  space taken by each tone group, which can maximise the contour difference in display. The  $x$ -axis took the scale of the ten time points. The rest of the plots (in the middle and bottom rows) show the difference of pairwise comparisons within the contour group after  $z$ -score standardisation, i.e., comparing NC and SM tones against the NM tone baseline within the similar-contour group.

### **NM Tone 1: level**

After standardisation, all four level tones became largely overlapped. This is especially for NC Tone 1, which was located in a tone height clearly higher than the other level tones. Moreover, for the NC tones, the significant regions in the difference smooths were reduced. NC Tone 1 was no longer different from NM Tone 1 throughout the entire contour; the contour was still mostly different, nevertheless, except at (near-)intersections at about points 4 and 10. For NC Tone 5, the divergence point was delayed to near point 10. However, for SM Tone 1, the difference region was expanded and also occurred at the beginning. The onset of the contour became significantly different before about point 2, and the offset had an early divergence from after point 8.

Overall, after centring the relative tone height, NM Tone 1 was the most similar to NC Tone 5, as indicated by the shortest significant region.



**Figure 2.3.9:** Overall f0 contours and difference curves for level tones after *z-score* standardisation. Shaded bands represent pointwise 95%-confidence intervals.

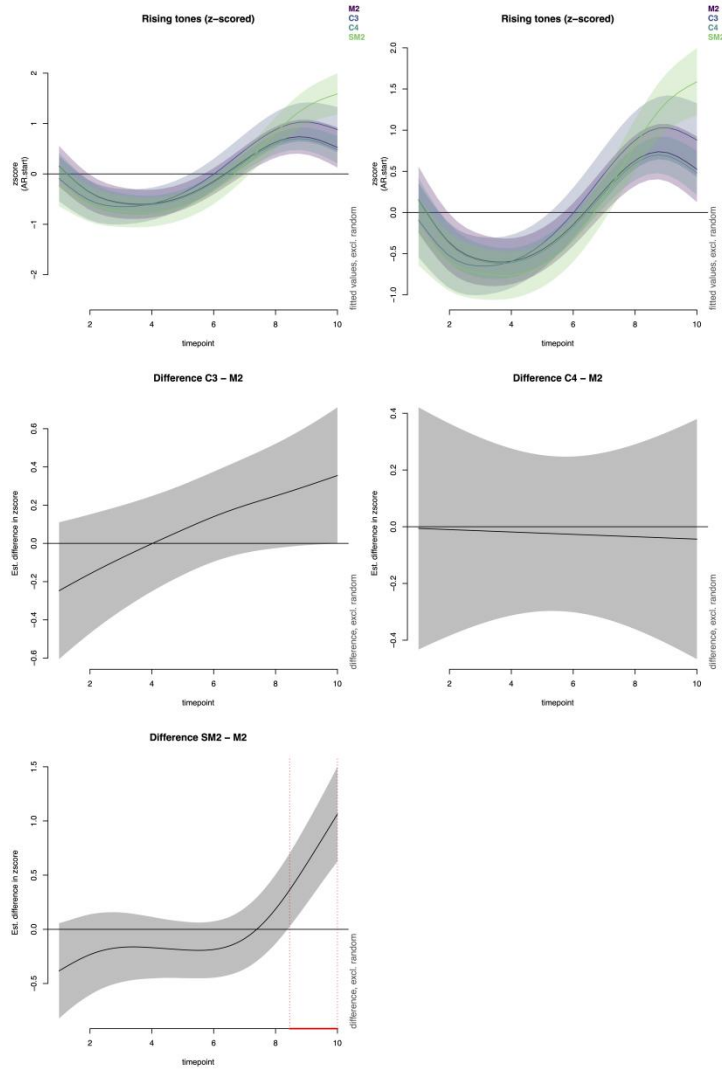
Top row: overall f0 contours (excluding random effects) for each tone by timepoint (1–10). Tones are ordered from dark to light: NM T1 (purple), NC T1 (blue), NC T5 (cyan), SM T1 (green).

Middle and bottom rows: difference curves of NC and SM tones relative to NM T1 (baseline). Difference between NM T1 and: NC T1 (significant at points 1–3.2 and 4.7–9.2); NC T5 (significant at points 9.5–10); SM T1 (significant at points 1–2.1 and 8.5–10).

## **NM Tone 2: rising**

Similar to the level tones, the extent of overlap between the rising tones was also increased after centring. Furthermore, all three pairwise comparisons showed reduced regions of significant difference. For Nanning Cantonese, NC Tone 3 became not significantly different along the entire contour, similar to NC Tone 4, where the insignificant difference remained. Consequently, both NC rising tones were not significantly different from NM Tone 2. For the SM tone, the difference region also greatly reduced after point 8. This reveals a key distinction between f0 height and contour shape: although the absolute f0 height of SM Tone 2 and NM Tone 2 was most similar at the end of the contour, this final portion was precisely where their contour shapes were most different.

Overall, after centring the relative tone height, NM Tone 1 was most similar to NC Tone 3 and 4, as indicated by the non-significant difference smooths. Nevertheless, NM Tone 2 was also mostly similar to SM Tone 2.



**Figure 2.3.10:** Overall f0 contours and difference curves for rising tones after *z-score* standardisation. Shaded bands represent pointwise 95%-confidence intervals.

Top row: overall f0 contours (excluding random effects) for each tone by timepoint (1–10). Tones are ordered from dark to light: NM T2 (purple), NC T3 (blue), NC T4 (cyan), SM T2 (green).

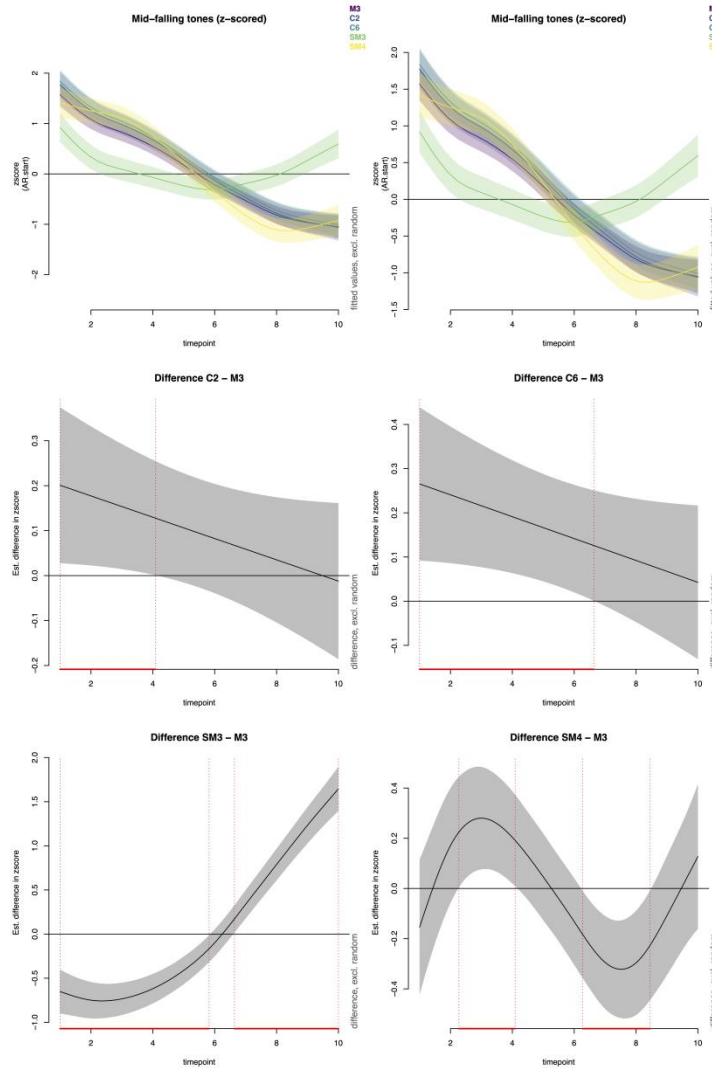
Middle and bottom rows: difference curves of NC and SM tones relative to NM T2 (baseline). Difference between NM T2 and: NC T3 (not significant); NC T4 (not significant); SM T2 (significant at points 8.5–10).

### **NM Tone 3: mid-falling**

Centring  $f_0$  heights altered the tonal patterns of the falling tones, both for mid-falling and high-falling tones. Instead of approximating each other towards the end, the tone contours now had several intersections with each other, particularly in the middle of the contours. Moreover, after  $z$ -scoring, the NC falling tones, which were originally insignificantly different from NM Tone 3, had significant divergences in the first halves of the contours: until about point 4 for NC Tone 2, while until about point 7 for NC Tone 6. The linear patterns of the three tones were weakened and became slightly concave.

For Standard Mandarin, the significance patterns of the contour differences were also greatly altered. Generally, the differences were significant only except when the two contours were close, intersected, or overlapped, since the two SM tones had more curvatures. Specifically, SM Tone 3 was significantly different from NM Tone 3 except when at the intersection (around point 6). SM Tone 4 was significantly different from NM Tone 3 at the two peaks of curvature, where it diverged the most from the SM tone.

Overall, after centring the relative tone height, NM Tone 3 was the most similar to NC Tone 2, as indicated by the shortest significant region.



**Figure 2.3.11:** Overall f0 contours and difference curves for mid-falling tones after *z-score* standardisation. Shaded bands represent pointwise 95%-confidence intervals.

Top row: overall f0 contours (excluding random effects) for each tone by timepoint (1–10). Tones are ordered from dark to light: NM T3 (purple), NC T2 (blue), NC T6 (cyan), SM T3 (green), and SM T4 (yellow).

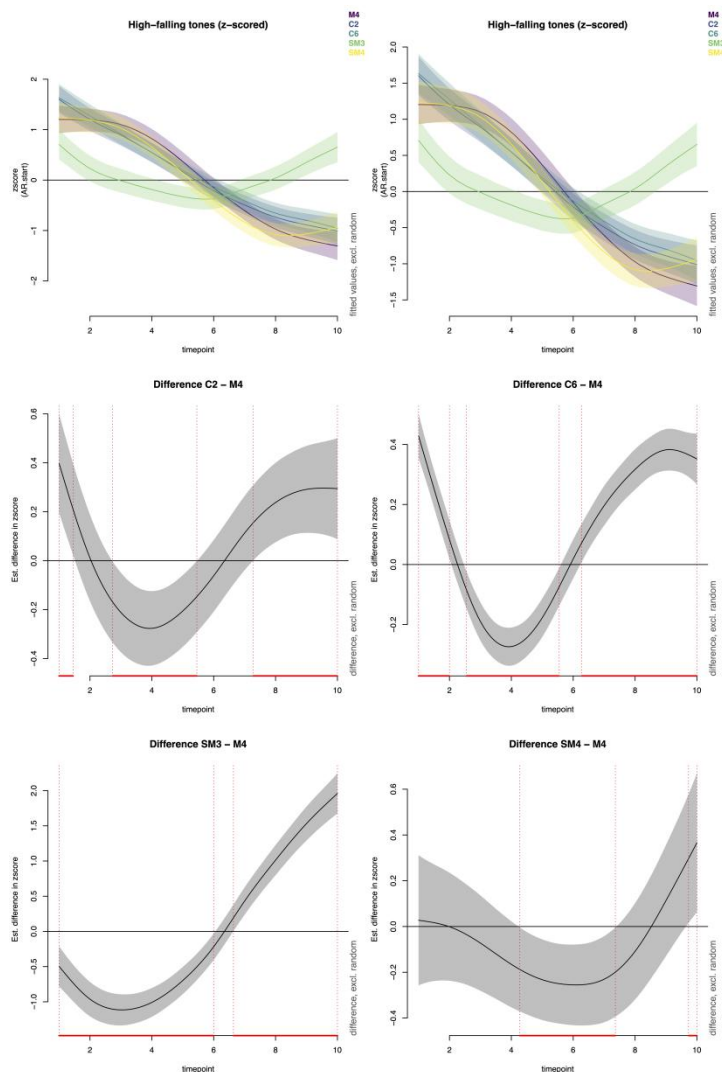
Middle and bottom rows: difference curves of NC and SM tones relative to NM T3 (baseline). Difference between NM T3 and: NC T2 (significant at points 1–4.1); NC T6 (significant at points 1–6.6); SM T3 (significant at points 1–5.8 and 6.6–10); SM T4 (significant at points 2.3–4.1 and 6.3–8.5).

### **NM Tone 4: high-falling**

For NC tones, centring f0 heights created intersections between the NC tones (low-falling, near-linear) and the NM tone (high-falling, convex-concave). (NM Tone 4 had a curved shape that was slightly flatter than SM Tone 4, which was distinct from the more linear NC falling tones.) Consequently, the difference smooths, originally with significant differences along the entire contours, now had insignificant regions only at intersections at about points 2 and 6. The two SM tones, however, showed different changes due to their distinct f0 contours and heights. For the dipping SM Tone 3, the insignificantly different region was changed, from the offset (i.e., the portion closest to SM Tone 4) to the intersection at about point 6.

On the other hand, SM Tone 4 became more closely overlapped with NM Tone 4, both of which were high-falling. Originally, SM Tone 4 had a f0 height slightly lower than NM Tone 4. Additionally, its onset and offset were close to NM Tone 4, but it had a sharper fall and greater curvature magnitude especially at the end, where a short rising tail was presented. After centring, the significant divergence in the middle was reduced (from about after point 4 to 7), where the sharper fall made SM Tone 4 gradually lower than NM Tone 4. Significant divergence also occurred at the offset (point 10) due to the rising tail.

Overall, after centring the relative tone height, NM Tone 4 was the most similar to SM Tone 4, as indicated by the shortest significant region.



**Figure 2.3.12:** Overall  $f_0$  contours and difference curves for high-falling tones after  $z$ -score standardisation. Shaded bands represent pointwise 95%-confidence intervals.

Top row: overall  $f_0$  contours (excluding random effects) for each tone by timepoint (1–10). Tones are ordered from dark to light: NM T4 (purple), NC T2 (blue), NC T6 (cyan), SM T3 (green), and SM T4 (yellow).

Middle and bottom rows: difference curves of NC and SM tones relative to NM T4 (baseline). Difference between NM T4 and: NC T2 (significant at points 1–1.5, 2.7–5.5, and 7.3–10); NC T6 (significant at points 1–2, 2.5–5.5, and 6.3–10); SM T3 (significant at points 1–6 and 6.6–10); SM T4 (significant at points 4.3–7.4 and 9.7–10).

### 2.3.2.4. Interim summary

The table below summarises the NC and SM tones that the four NM tones are most similar to, when the relative tone heights were retained or centred. Specifically:

- NM Tone 1 inherited the f<sub>0</sub> height of the lower NC and SM tones, while the contour solely was more similar to NC Tone 5.
- NM Tone 2 was not significantly different from NC Tone 4; the contour solely was also very similar to NC Tone 3, followed by SM Tone 2.
- NM Tone 3 was not significantly different from NC Tone 2 and 6, while the contour solely was more similar to NC Tone 2.
- NM Tone 4 was always most similar to SM Tone 4, retaining or removing the relative f<sub>0</sub> height.

	<b>Relative f<sub>0</sub> height retained</b>	<b>Relative f<sub>0</sub> height removed</b>
<b>NM Tone 1</b> <b>(Level)</b>	SM Tone 1, *closely followed by NC Tone 5	NC Tone 5
<b>NM Tone 2</b> <b>(Rising)</b>	NC Tone 4 (not sig)	NC Tones 3 and 4 (not sig) *SM Tone 2 largely similar
<b>NM Tone 3</b> <b>(Mid-falling)</b>	NC Tones 2 and 6 (not sig)	NC Tone 2
<b>NM Tone 4</b> <b>(High-falling)</b>	SM Tone 4	SM Tone 4

**Table 2.3.7:** Summary of the contour comparisons between the NM tones and the corresponding NC and SM tones.

The result of contour comparisons is generally consistent with the overall phonetic distance of tone contours in RMSE summarised in Table 2.3.6.

### **2.3.3. Formation of Nanning Mandarin citation tones**

#### **2.3.3.1. Source language as main template**

Based on the GAMMs results, our analysis proceeds from the hypothesis that the phonetic realisations of NM tones can be described in terms of template alignment. Specifically, NM tones tend to align more closely with tones in the source language, Nanning Cantonese, which serves as the main source of templates. The target language, Standard Mandarin, provides an additional source for templates when no suitable NC counterpart is available.

In Nanning Mandarin, the cases of the level Tone 1 and rising Tone 2 are simpler, as the corresponding NC and SM tones are largely similar in contour. In these cases, NM tones directly correspond to the NC contours. The falling Tones 3 and 4, however, present a more complex scenario that demonstrates the tendency of template alignment. The key question for NM Tone 3 is: if its prototypical target of approximation is the dipping SM Tone 3, and such a contour is absent in Nanning Cantonese as a readily available template, why does NM Tone 3 not simply replicate the contour of SM Tone 3? One possible reason may be related to the underlying target of tonal approximation. As discussed previously, the phonological essence of SM Tone 3 is not its dipping contour but its low register, while the rising tail is highly contextual. Therefore, the true target of approximation may actually be a low tone without contour inflexion. This ruled out the inflected SM Tone 3 as the phonetic template, and the most suitable candidate thus became the mid-falling tones in Nanning Cantonese.

This outcome, in turn, constrains the template alignment for NM Tone 4. As mentioned in earlier discussions, NM Tone 4, as another falling tone, has to be phonetically distinct from NM Tone 3 to maintain tonal contrast. Because the two NC falling tones are acoustically merged and have both served as the template for SM Tone 3, they cannot anymore provide an additional template. Consequently, NM Tone 4 instead corresponds to the falling contour in the target language, i.e., the high-falling SM Tone 4.

This hypothesis is to some extent in line with the previous view that the NM tone inventory is largely a product of L1 transfer or shift-induced interference (Chen Z., 2007). Here, we propose a more explicit formulation of the general tendency of template alignment: NM tones tend to align with patterns in the source language in most cases; only when a suitable template is absent in the source language does the realisation correspond to a template from the target language.

### **2.3.3.2. Hybrid features under joint influence**

However, this tendency of template alignment is not a simple replication of contour. As demonstrated by the GAMMs results, the differences between the NM tones and their NC or SM templates are not always insignificant throughout the contour. Specific cases can illustrate this adaptation. NM Tone 1 had a tail different from both its potential templates (NC Tone 5 and SM Tone 1). It exhibits a flatter contour at an intermediate  $f_0$  height, suggesting a hybrid realisation resulting from the combined influence of both templates. In terms of NM Tone 4, while its pattern is most similar to its SM Tone 4 template, its phonetic realisation is considerably different, particularly in the flatter contour shape. This flatter contour of NM Tone 4 is likely an influence from the more linear (non-curved) contour of the NC falling tones.

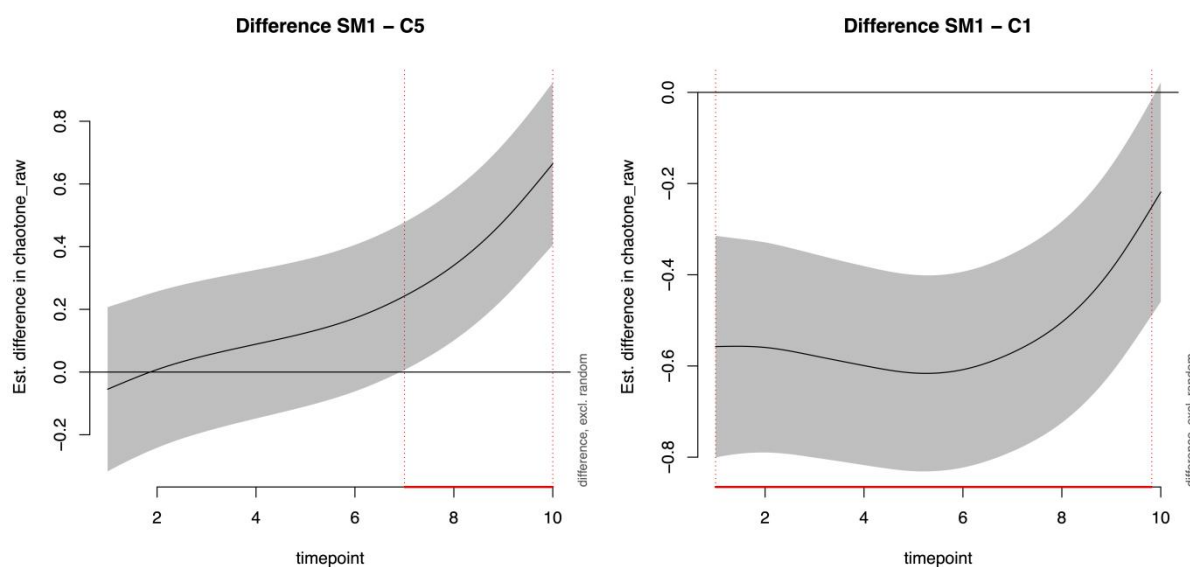
In conclusion, NM tones have developed unique features beyond the template alignment, which are likely hybrid or intermediate forms under the joint influence of both Nanning Cantonese and Standard Mandarin.

### **2.3.3.3. Target language as acoustic reference**

Having established that NM tones tend to favour the NC counterparts as the main source of templates, with Standard Mandarin serving as an additional source, a further question arises: when a specific contour type corresponds to multiple NC tones, what mechanism determines which specific tone is aligned with? For instance, within the level tones, why does NM Tone 1 align more closely with NC Tone 5 rather than NC Tone 1?

We propose a further hypothesis: the SM tone may serve as the acoustic reference for tone formation. Specifically, among multiple NC templates, an NM tone corresponds to the NC tone that is acoustically closest to its SM reference.

This hypothesis effectively explains the formation of NM Tones 1 and 3. For NM Tone 1, the acoustic reference, SM Tone 1, is a level tone with an  $f_0$  height between 3 and 4 LT. In the NC inventory, NC Tone 5 is the closest match to NM Tone 1 (at a similar  $f_0$  height), whereas NC Tone 1 is significantly higher ( $f_0$  height at about 4 LT). This observation is confirmed in the difference smooths where the relative  $f_0$  height was retained: as in Figure 2.3.13, NC Tone 5 was not significantly different from SM Tone 1 in over half of the normalised time. Therefore, NC Tone 5, rather than NC Tone 1, served as the template for NM Tone 1.

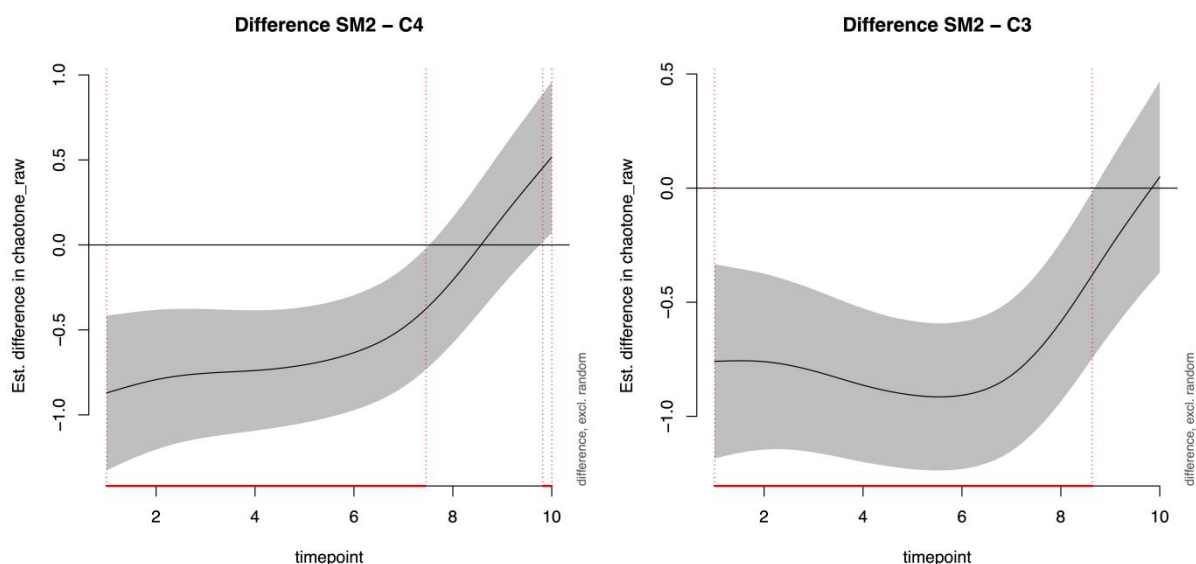


**Figure 2.3.13:** Difference curves between SM T1 and the corresponding NC level tones. Difference between SM T1 and: NC T1 (significant at points 1–9.8); NC T5 (significant at points 7–10). Shaded bands represent pointwise 95%-confidence intervals.

Regarding NM Tone 3, the acoustic reference, SM Tone 3, is underlyingly a low tone register without contour inflexion. NC Tones 2 and 6 represent the lowest pitch register and the only falling

tones in the NC tone inventory. Given that these two tones have undergone a full acoustic merger, they can arguably be treated as one contour. Therefore, NM Tone 3 has no other option within the NC low register than NC Tone 2 or Tone 6.

The case of NM Tone 2, however, presents a more complex challenge. Taking SM Tone 2 as the acoustic reference, a visual inspection of the model predictions with relative f0 height retained (as in Figure 2.3.5) suggests a trade-off: NC Tone 3 appears closer to SM Tone 2 in terms of contour slope (exhibiting a similarly sharper rise), whereas NC Tone 4 appears closer in terms of absolute f0 height (located at a similarly lower height).



**Figure 2.3.14:** Difference curves between SM T2 and the corresponding NC rising tones. Difference between SM T2 and: NC T3 (significant at points 1–8.6); NC T4 (significant at points 1–7.5 and 9.8–10). Shaded bands represent pointwise 95%-confidence intervals.

The difference smooths in Figure 2.3.14 reveal that both NC Tone 3 and Tone 4 exhibit similarity to SM Tone 2 only towards the contour offset. However, the insignificant regions suggest that NC Tone 4 was only slightly more similar to SM Tone 2, with a marginal difference of approximately 1 f0 point. In addition, within the significantly different regions, the magnitude of divergence between

NC Tone 4 and SM Tone 2 is smaller than that of NC Tone 3, even though this was not large enough to alter the statistical significance. It is possible that this subtle acoustic difference is sufficiently perceptible to speakers, serving as the decisive factor in their favour of NC Tone 4 as the template.

## 2.4. Summary

This chapter reports on the  $f_0$  contours in three varieties: Nanning Mandarin, Nanning Cantonese, and Standard Mandarin. Speech data were collected through controlled, isolated production of monosyllables from native speakers. In general, the tonal descriptions of the three varieties show similarity to the data in previous literature, but with some minor differences in tone values and contours. The tone–contour mapping in Nanning Mandarin resembled that in Standard Mandarin. However, from the global  $f_0$  space and tonal contrast to the fine-grained contour shapes, Nanning Mandarin shares greater similarities to Nanning Cantonese. Furthermore, the NM tone system has developed unique characteristics that are distinct from both the source and target languages, owing to the changes caused by the integration of influences from two linguistic sources.

Overall, the lexical tone system of Nanning Mandarin resembled Nanning Cantonese in the phonetic space; specifically, it spanned over a narrower  $f_0$  range and implemented a three-level tone system with High–Mid–Low contrasts. Moreover, individual NM tones demonstrated features similar to and unique from both the source and target languages. In terms of individual contours, the NM tones were predominantly most similar to those of Nanning Cantonese, based on results of RMSE and GAMMs in both Chao’ tone numbers and  $z$ -scored hertz. This suggests a clear influence of Nanning Cantonese on Nanning Mandarin. Nevertheless, NM tones clearly inherited the SM tone categories, and individual NM tones still approximated those of Standard Mandarin in  $f_0$  contours. Particularly, NM Tone 4 is most similar to SM Tone 4. However, a noticeable lack of dipping was observed in NM Tone 3. This can potentially be accounted for by the low-level underlying target and contextual final rise associated with SM Tone 3, and crucially, the constraint by the lack of a low-level template in the NC tones.

Results of the three language varieties show that Nanning Mandarin, as a variety of Standard Mandarin, has presented a hybrid tone system with a clear and large influence from Nanning Cantonese in its approximation to Standard Mandarin. To examine whether these similarities and differences are significant and affect pitch perception, contour comparisons were conducted using GAMMs.

The GAMMs results suggest a general tendency of tone formation for Nanning Mandarin, in which NM tones tend to align with NC tones when close acoustic correspondences are available, while correspond to SM tones in cases where such correspondences are lacking. This tendency can be characterised by the following observations:

1. The SM counterpart, which represents the target of tonal approximation, provides the acoustic reference;
2. The NC tone that is acoustically closest to the SM reference serves as the main template;
3. When there is no suitable candidate in Nanning Cantonese, the SM counterpart itself serves as the template.

To establish a more thorough understanding of the lexical tones and intonation, future studies should also investigate the tones in contexts, such as disyllabic tones and the  $f_0$  realisations of SM neutral tone morphemes in Nanning Mandarin. Regarding NC tone mergers, further research needs to include a predictor of age group for statistical analysis to examine the age variation reported in previous studies, more perception experiments are also necessary to examine the tone merger from the listeners' side. In addition, whether the similarities and differences between Nanning Mandarin and its source and target languages affect pitch perception still remains to be tested.

## Chapter 3: Production of intonational focus

This chapter presents the methodology and findings of the intonation study. The chapter begins with an introduction of methods, from building up the corpus (§3.1) to processing and analysing the data (§ 3.2). The main emphasis is to elicit maximally natural production of non-standard language varieties in a lab setting, and to deal with large-corpus data. Then § 3.3 describes the detailed findings in focus marking, in terms of main focus effect, tone-specific focus effect, and cross-linguistic comparisons of Nanning Mandarin and its source or target languages. Finally, § 3.4 summarises the main findings and our implications of focus marking and its interaction with tonal variation, with specific focus on cross-linguistic variations and language contact.

This chapter addresses Research Questions (RQ) 2 and 3, reformulated more concisely as follows:

RQ 2: intonational focus

- (2a) What are the primary f<sub>0</sub> range modulation used to mark focus in Nanning Mandarin and Nanning Cantonese?
- (2b) In what ways are the f<sub>0</sub> range modulation in NM focus marking similar to or different from those of its source and target languages?

RQ 3: tone–intonation interaction

- (3a) How does focus marking modulate the f<sub>0</sub> range of different tone categories?
- (3b) In what ways are these tone–intonation interaction patterns in Nanning Mandarin similar to or different from those of its source and target languages?

The experiment was designed first to answer RQ 2, by eliciting the same target sentences under different meaning conditions. Meanwhile, lexical sets varying in tone were also designed to investigate the interaction between tonal and intonational realisations, thereby addressing RQ 3.

### 3.1. Corpus build-up

This section introduces the careful considerations for collecting data that is as natural as possible to build up a relatively large corpus. The design of materials is described in § 3.1.1, and the design of experiments in § 3.1.2. The background of participants is introduced in § 3.1.3. The detailed procedure of the production experiment is introduced in § 3.1.4.

Nanning Mandarin is a colloquial, non-standard local variety of Mandarin, while Nanning Cantonese is the local Cantonese dialect. When designing the materials and experiment, the central goal was to activate a colloquial language environment to elicit the casual registers within the formal lab setting. This was particularly crucial for eliciting authentic NM accent and preventing speakers from shifting to a more standard, SM-like production.

To achieve this, several strategies were employed to elicit speech that was maximally vernacular and natural:

- (1) Naturalistic materials: Using common words, sentences and contexts that are natural and familiar to native speakers.
- (2) Conversational prompts: Using pre-recorded audio from a native interlocutor as the primary prompt to simulate conversations.
- (3) Relaxed atmosphere: Fostering a comfortable environment to reduce formality. In most of the cases, at least one mutual acquaintance of the author and participant was present during the recording. Participants were also allowed to chat informally before and between tasks to help them relax.

The production experiment adopted a question-and-answer paradigm, which is the conventional approach for investigating focus intonation. Each question–answer pairs of eliciting questions and target responds; by using this paradigm, participants’ focus was narrowed to the target items, making the tasks simpler than, for example, eliciting targets by providing contexts (, by reducing the cognitive load in processing complicated contexts). Though this paradigm might take slightly longer time than the methods present targets with contexts, it is relatively well-established and yields more

natural elicitation.

### **3.1.1. Materials**

When designing the materials, the first step is to determine what meaning condition is under investigation of intonation, which is described in § 3.1.1.1. The next step, also the key task, is to design semi-spontaneous question–answer pairs, which contain the same target sentences across different meaning conditions and different tones across different sentences. The structure and tonal variations of designed sentences is introduced in § 3.1.1.2, and then § 3.1.1.3 introduces word selection in terms of cross-linguistic naturalness and consistence. Finally, § 3.1.1.4 summarises the factors included in the intonation study. The full material for the production experiment is provided in Appendix B.1.

#### **3.1.1.1. Meaning conditions**

##### **Focus conditions**

We investigated four meaning conditions in terms of focus conditions and information structure: broad focus declarative with all new information (Broad Focus; BF), initial narrow focus declarative with given information (Initial Focus; IF), final narrow focus declarative with given information (Final Focus; FF), and broad focus interrogative with all new information (Broad Question; BQ). Table 3.1.1 below illustrates the elicitations and targets under the four meaning conditions.

We did not design a contrast between new and given information: broad focus and broad question indicates all-new context, while narrow focus indicates given context except for the focal item. The factors this study looks at are already too many; considering the number of materials and time constraints for experiments, we will not discuss the effect of information structure here, although we are aware that it may have an effect on the  $f_0$  realisations of intonation patterns.

As Broad Question was not investigated in the present study, only the three focus conditions (Broad

Focus, Initial Focus, and Final Focus) will be described in details.

Meaning condition	Elicitation	Response with target
<b>Broad Focus, All New (BF)</b>	<i>What did he just say to you?</i>	He said <u><b>A-Ying wanted to cook soup.</b></u>
<b>Initial Narrow Focus with Given (IF)</b>	<i>Who did he just say that wanted to cook soup?</i>	<u><b>A-Ying</b> wanted to cook soup.</u>
<b>Final Narrow Focus with Given (FF)</b>	<i>What did he just say that A-Ying wanted to cook?</i>	<u><b>A-Ying wanted to cook soup.</b></u>
<b>Broad Question, All New (BQ)</b>	<i>Guess what he just said to me.</i>	Did he say <u><b>A-Ying wanted to cook soup?</b></u>

**Table 3.1.1:** Four meaning conditions with elicitations and targets. Underlining indicates target sentences; bold indicates focused items.

### Focus types

The selected focus type is new-information focus elicited by *wh*- questions. This is the conventional paradigm for focus study when focus position effect is considered, widely used in the majority of research on varieties of China, including Standard Mandarin (Y. Xu, 1999; Y. Xu et al., 2012), Lanyin Mandarin (C. Shen & Xu, 2016), Hong Kong Cantonese (W. L. Wu & Xu, 2010), Nanchang Gan (B. Wang et al., 2011), and non-Sinitic varieties such as Yi (B. Wang et al., 2011). In the studies involved variations in focus position, only Gu and Lee (2007) used yes-no questions to elicit new-information focus in Hong Kong Cantonese.

A notable limitation was the difficulty in eliciting natural-sounding verb focus (i.e., middle focus, MF) in SVO sentences. Using *wh*- questions to elicit verb focus condition often sounds unnatural in the language varieties investigated. An alternative method was therefore piloted, i.e., using yes-no questions to elicit corrective focus on the subject, verb, or object, as verbs were more natural to be elicited by yes-no questions. However, this approach was ultimately abandoned. The three pilot participants reported that the corrective task was unintuitive, out of context, and thus difficult to understand. Consequently, their elicited productions featured weak or even no emphasis on the target word, even when responding to prompts with clear narrow focus. This focus pattern contrasted

clearly with the stronger emphasis observed in their spontaneous conversations. Such a qualitative observation differs from findings in Standard Mandarin. For instance, Chen and Braun (2006) reported that corrective focus was marked more prominently than new-information focus in terms of  $f_0$ . This discrepancy may reflect differences between language varieties and in experimental design, which would require further systematic investigation.

The weak emphasis in the corrective task may be attributed to: (1) Context familiarity: Natural production of corrective focus may require more familiar contexts. Since the entire target sentence was constructed and thus unfamiliar to the participants as a whole, the focused item may not clearly stand out. This possibly made it harder for speakers to carry strong emphasis when correcting an unfamiliar item, especially in a lab setting. (2) Pragmatic sufficiency: Prosodic prominence may not be obligatory to signal correction; simply uttering the full sentence with the corrected word is sufficient to convey the right information. (3) Material design: In our designed materials for corrective focus, a reporting clause “He said...” was required before each target sentence when responding to the yes-no question<sup>15</sup>. With the reporting clause, participants would tend to produce broad focus rather than narrow focus.

In contrast, pilot participants confirmed that responding to *wh*- questions felt more natural and straightforward, where their responses contained perceptibly clearer emphasis. Given that verb focus per se was not central to the study’s main goals, as subject and object focus are sufficient for investigating variation in focus position, the verb focus condition was therefore excluded from the final experiment.

We acknowledge that new-information focus and corrective focus can have different  $f_0$  realisations in Standard Mandarin (Y. Chen, 2022), although the main difference lies in the  $f_0$  contrast between

---

<sup>15</sup> The obligation of the reporting clause “He said” can be illustrated by the example below of a set of five meaning conditions:

- |  |   |
|--|---|
| BF: – Just now, what did he called to say?                           | – He said <b>A-Ying wanted to cook soup</b> .                 |
| IF: – Just now, did he called to say A-Long wanted to cook soup?     | – No, he said A- <b>Ying</b> wanted to cook soup.             |
| MF: – Just now, did he called to say A-Ying wanted to buy soup?      | – No, he said A-Ying wanted to <b>cook</b> soup.              |
| FF: – Just now, did he called to say A-Ying wanted to cook porridge? | – No, he said A-Ying wanted to cook <b>soup</b> .             |
| BQ: – Just now, he called in?  | – What did he say, saying <b>A-Ying wanted to cook soup</b> ? |

the focal and non-focal regions, rather than on the focal item itself (Y. Chen & Gussenhoven, 2008; Qin P., 2021). However, there is also evidence shows that new-information focus and alternative focus had similar realisations and thus perceptually hard to distinguish (Y. Yang, 2022).

To obtain speech that was as natural as possible under lab settings, we prioritised the method that participants found more intuitive. The final experiment therefore used *wh*- questions exclusively to elicit new-information focus on subjects and objects.

### **3.1.1.2. Sentence lists**

#### **Targets and tone variations**

Each target sentence has specifically selected subject (S), verb (V) and object (O) for the variation in focus position. The target sentences are constituted by five syllables in a simple SVO structure of <name prefix + name + modal verb + verb + noun>. Such sentence structure can maximally avoid cross-linguistic variations in sentence structure and functional words.

For each tone category, three target sentences were designed with distinct sets of subject, verb and object; within each sentence, the lexical tone was the same across the individual subject, verb and object. The Mandarin and Cantonese tone categories are systematically mapped, according to the same diachronic correspondence as that of our tone study, which is further illustrated with the target sentences in Table 3.1.2 below. During the recording, no participant reported noticing that the individual S, V, and O of each utterance bore the same tone; we therefore assume that the designed sentences were not perceived as unnatural in terms of their tone distribution.

In Mandarin, two adjacent Tone 3 syllables will result in tone sandhi in the first syllable, in the case of this study, the verb in the verb-object phrase. We decided not to insert any classifier or functional word between the verb and object due to cross-linguistic lexical variation. Specifically, these lexical items are often different across language varieties; moreover, including these words further reduce the number of usable VO phrases. The Tone 3 sandhi was presented across each focus condition, so

the comparison did not involve an additional variable of tone sandhi. In our data, non-sandhi Tone 3 could be analysed from the subject words, and sandhi Tone 3 could be analysed by comparing to the disyllabic words collected in our tone study in the future research.

Table 3.1.2 below illustrates the structure of the five-syllable target sentences across each tone in Nanning Mandarin, Standard Mandarin, and Nanning Cantonese. Tones are aligned according to their historical correspondence mentioned in § 1.2.2; numbers in brackets refer to Chao tone numerals from § 2.3.1. Phonological transcriptions are separated by semicolons for Mandarin and Cantonese; for Nanning Mandarin, segmental transcriptions use SM transcriptions. English glosses are in italics; for prefix and names, English glosses were adopted from *pīn yīn* reflecting SM pronunciations.

Tone			Target sentence				
NM	SM	NC	Name prefix	Subject (Proper name)	Modal verb	Verb	Object
1 (44)	1 (44)	1 (54)	阿 a;a A	英 jīŋ;jēŋ <i>Ying</i>	要 jau;jiu <i>want to;</i> <i>be going to,</i> <i>will</i>	煲 pau;pu <i>cook</i>	汤 tʰaŋ;tʰoŋ <i>soup</i>
2 (34)	2 (34)	2 (42)		莹 jīŋ;jēŋ <i>Ying</i>		还 xuan;wen <i>return</i>	钱 tɕʰien;tɕʰin <i>money</i>
3 (42)	3 (323)	3 (34)		影 jīŋ;jēŋ <i>Ying</i>		洗 ei;ɬei <i>wash</i>	碗 wan;wun <i>bowl</i>
		4 (34)		敏 min;men <i>Min</i>			
4 (53)	4 (52)	5 (43)		翠 tsʰuei;tsʰui <i>Cui</i>		寄 tɕei;ki <i>send</i>	信 ein;ɬen <i>letter</i>
		6 (42)		曼 man;man <i>Man</i>		练 lien;lin <i>practice</i>	字 tsɿ;tsi <i>writing</i>

**Table 3.1.2:** Examples of five-syllable target sentences.

Finally, 12 lexical sets (4 tones × 3 lexical sets) were designed for Nanning Mandarin and Standard Mandarin, and 18 lexical sets (6 tones × 3 lexical sets) were designed for Nanning Cantonese. Due to time constraints, Mandarin speakers only produced three sentences each for

Tone 3 and 4, rather than covering the full set of NC sentences as in the tone study.

### Reporting clauses

Under the Broad Focus and Broad Question condition, the five-syllable target sentences and the elicitations were preceded by a reporting clause “He said”, as a requirement for responses in broad focus. This also makes the sentence more similar to natural conversations. For Broad Question, some speakers tended to omit “He”, the information given in the elicitation, due to coherence; for the sake of consistence, we used “He said” for all the reporting clauses. In the responses under Initial Focus and Final Focus, the reporting clause was not presented, as the target alone as response is natural. Moreover, the reporting clause tends to elicit a broad focus, which will weaken the required narrow focus in production.

Table 3.1.3 below presents the question–answer pairs for one target sentence across each meaning condition in the three language varieties. Target sentences are underlined; focal components are in bold. Segmental transcriptions for Nanning Mandarin use SM transcriptions. Particles in Standard Mandarin are realised in neutral tones, while in Cantonese they are usually transcribed in full tones but highly variable. Therefore, tone values for particles were not transcribed here.

Meaning condition	Language	Elicitation	Response with target
Broad Focus	NM	kaŋ <sup>44</sup> kaŋ <sup>44</sup> tʰa <sup>44</sup> kən <sup>44</sup> ni <sup>42</sup> ʂu <sup>44</sup> ʂən <sup>42</sup> mɿ <sup>34</sup> a 刚刚他跟你说什么啊?	tʰa <sup>44</sup> ʂu <sup>44</sup> <u>a<sup>44</sup> jɿŋ<sup>44</sup> ja<sup>53</sup> pa<sup>54</sup> tʰaŋ<sup>44</sup></u> 他说阿英要煲汤。
	NC	ŋam <sup>54</sup> ŋam <sup>54</sup> kʰy <sup>34</sup> kən <sup>54</sup> ni <sup>34</sup> kəŋ <sup>34</sup> mət <sup>5</sup> jɛ <sup>34</sup> a 啱啱佢跟你讲乜嘢啊?	kʰy <sup>34</sup> kəŋ <sup>34</sup> <u>a<sup>43</sup> jɛŋ<sup>54</sup> jiu<sup>43</sup> pu<sup>54</sup> tʰəŋ<sup>54</sup></u> 佢讲阿英要煲汤。
	SM	kaŋ <sup>44</sup> kaŋ <sup>44</sup> tʰa <sup>44</sup> kən <sup>44</sup> ni <sup>323</sup> ʂu <sup>44</sup> ʂən <sup>34</sup> mɿ lɿ 刚刚他跟你说什么了?	tʰa <sup>44</sup> ʂu <sup>44</sup> <u>a<sup>44</sup> jɿŋ<sup>44</sup> ja<sup>52</sup> pa<sup>54</sup> tʰaŋ<sup>44</sup></u> 他说阿英要煲汤。
	Eng.	<i>What did he just say to you?</i>	<i>He said <u>A-Ying wanted to cook soup.</u></i>

Meaning condition	Language	Elicitation	Response with target
Initial Focus	NM	kaŋ44 kaŋ44 tʰa44 ʂuɔ44 ʂei34 jaʊ53 paʊ44 tʰaŋ44 刚刚他说谁要煲汤?	<u>a44 jin44 jaʊ53 paʊ44 tʰaŋ44</u> 阿英要煲汤。
	NC	ŋam54 ŋam54 kʰy34 kəŋ34 pin54 kɔ43 jiu43 pu54 tʰəŋ54 啱啱佢讲边个要煲汤?	<u>a43 jen54 jiu43 pu54 tʰəŋ54</u> 阿英要煲汤。
	SM	kaŋ44 kaŋ44 tʰa44 ʂuɔ44 ʂei34 jaʊ52 paʊ44 tʰaŋ44 刚刚他说谁要煲汤?	<u>a44 jin44 jaʊ52 paʊ44 tʰaŋ44</u> 阿英要煲汤。
	Eng.	<i>Who did he just say that wanted to cook soup?</i>	<u>A-Ying wanted to cook soup.</u>
Final Focus	NM	kaŋ44 kaŋ44 tʰa44 ʂuɔ44 a44 jin44 jaʊ53 paʊ44 ʂən42 mɿ34 刚刚他说阿英要煲什么?	<u>a44 jin44 jaʊ53 paʊ44 tʰaŋ44</u> 阿英要煲汤。
	NC	ŋam54 ŋam54 kʰy34 kəŋ34 a43 jen54 jiu43 pu54 met5 je34 啱啱佢讲阿英要煲乜嘢?	<u>a43 jen54 jiu43 pu54 tʰəŋ54</u> 阿英要煲汤。
	SM	kaŋ44 kaŋ44 tʰa44 ʂuɔ44 a44 jin44 jaʊ52 paʊ44 ʂən34 mɿ 刚刚他说阿英要煲什么?	<u>a44 jin44 jaʊ52 paʊ44 tʰaŋ44</u> 阿英要煲汤。
	Eng.	<i>What did he just say that A-Ying wanted to cook?</i>	<u>A-Ying wanted to cook soup.</u>
Broad Question	NM	ni42 tsʰai44 tʰa44 kaŋ44 kaŋ44 kən44 wɔ42 ʂuɔ44 ʂən42 mɿ34 你猜他刚刚跟我说什么?	tʰa44 ʂuɔ44 <u>a44 jin44 jaʊ53 paʊ44 tʰaŋ44</u> a 他说阿英要煲汤啊?
	NC	ni34 tsʰai54 kʰy34 ŋam54 ŋam54 kən54 ŋɔ34 kəŋ34 met5 je34 你猜佢啱啱跟我讲乜嘢?	kʰy34 kəŋ34 <u>a43 jen54 jiu43 pu54 tʰəŋ54</u> a 佢讲阿英要煲汤啊?
	SM	ni323 tsʰai44 tʰa44 kaŋ44 kaŋ44 kən44 wɔ323 ʂuɔ44 ʂən34 mɿ 你猜他刚刚跟我说什么?	tʰa44 ʂuɔ44 <u>a44 jin44 jaʊ52 paʊ44 tʰaŋ44</u> ma 他说阿英要煲汤吗?
	Eng.	<i>Guess what he just said to me.</i>	<i>Did he say <u>A-Ying wanted to cook soup?</u></i>

**Table 3.1.3:** Example of question–answer pairs across meaning conditions.

### 3.1.1.3. Word selection

Word selection of the intonation study followed similar criteria as in our previous tone study, i.e., **high frequency** and **similar phonetic context** cross-linguistically, with preferences on certain syllable structures and non-heteronyms (See § 2.1.1.2 for details). Additional considerations for the intonation study are described as follows.

#### Word frequency

The five-syllable target sentences were constituted by lexical equivalents across language varieties, identical in the written form. The subjects (person names) are all female or gender-neutral names that are frequently used, validated by two native or fluent (native-like) NM and/or NC speakers apart from the author. Although the Nanning varieties were prioritised over Standard Mandarin, materials of all three language varieties underwent validation of naturalness.

Two object words in the lexical sets were not ideal in their corresponding verb-object (VO) phrases, namely 面 ‘noodle’ in 卖面 ‘sell noodles’ and 鸟 ‘bird’ in 养鸟 ‘raise birds’. These items were the result of the constraints on word selection. In Nanning Cantonese, certain tone categories (such as NC Tone 6) contain very few common words. When taking into consideration the requirement of cross-linguistic shared cognates and the other phonetic criteria, the pool of eligible words became extremely small. During piloting and formal recording, native speakers noted the issues with these items. For 卖面, according to pilot participants (six native NC speakers and one Northern Mandarin speaker), the word 面 was considered less natural as a free morpheme after 卖 ‘sell’, as it is usually part of a compound word (e.g., 面条 ‘noodles’). (面, as a general term for wheat-based products, will have a reference that is too generic<sup>16</sup> and thus ambiguous.) Regardless, in fact, no

---

<sup>16</sup> NC native speakers reported that 面 ‘wheat-based products; noodles; powder/flour...’ is not used on its own in Nanning Cantonese, but often in disyllabic words such as 面包 ‘bread’, 面条 ‘noodles’, and 面粉 ‘flour’. This may be related to the fact that wheat-based products were traditionally less common in southern China. Since the monosyllabic use of 面 is accepted in Standard Mandarin, this monosyllabic use in Nanning Cantonese may sound somewhat SM-like. Similarly, the Northern Mandarin speaker reported that 面 is not usually used on its own after 卖 ‘sell’. Since 面 is a general term for wheat-based products, and 卖 ‘sell’

participants in the main experiment commented on its unnaturalness. For 养鸟, during the recording, speakers reported that 雀 is the more common colloquial term for ‘bird’ in Nanning Cantonese, while 鸟 is more typical of Nanning Mandarin or written-style Nanning Cantonese.

Despite these minor issues, the items were retained after careful consideration. The final selection was based on choosing the most acceptable options from a very limited set, guided by native speaker intuition. Crucially, the items are grammatical, intelligible, commonly used, lexically equivalent, and very few in the lexical sets. Any potential minor unnaturalness was not affecting the pronunciation and likely mitigated by the familiarisation phase before recording. The choice of 养鸟, for instance, illustrates the prioritisation in this study. We decided to use 鸟, which is more natural in Nanning Mandarin but less natural in Nanning Cantonese, for the following reasons: 1) No better options were available without creating other issues (e.g., reusing the V or O items to create a more natural VO phrase, or using different lexical sets for Mandarin and Cantonese). 2) Increasing repetitions of a natural VO phrase would have made the repetitive task even more monotonous. 3) Nanning Mandarin is the primary focus and anchor of the present study, rather than Nanning Cantonese.

### **Phonetic context**

**Syllable structure.** In previous studies, different syllabic structures were preferred for different purpose. For example, for easier segmentation, Man (1999) chose syllables with obstruent onset and monophthong (/si/ and /se/); while for obtaining continuous pitch traces, Xu (1999) chose sonorant syllables for the entire sentence. Unlike our previous tone study, the requirement for a cross-linguistically similar phonetic contexts was less strict in the present study, though it was preferred. This was a necessary compromise because we prioritised the use of cross-linguistic lexical equivalents, which severely limited the available word pool. As a result, more cross-linguistic differences were allowed and more syllable structures were accepted, including V, CV, VN, and

---

can take a wide range of arguments, the exact referent of 面 after 卖 would be ambiguous. It would sound more natural after verbs such as 煮 ‘cook’, where the verb restricts the interpretation of 面 to ‘noodles’.

CVN. Examples in Nanning Mandarin, Nanning Cantonese and Standard Mandarin include 买 ‘buy’ /mai<sup>41</sup>/–/mai<sup>34</sup>/–/mai<sup>214</sup>/; 牛 ‘ox’ /niou<sup>31</sup>/–/ŋeu<sup>34</sup>/–/niou<sup>35</sup>/; 问 ‘ask’ /wən<sup>52</sup>/–/mən<sup>31</sup>/–/wən<sup>53</sup>/; 洗 ‘wash’ /ɕi<sup>41</sup>/–/ɬei<sup>35</sup>/–/ɕi<sup>212</sup>/. All-sonorant syllables were preferred whenever possible, which unintendedly resulted in certain NC target sentences being articulatorily challenging due to sequences of successive glides (/j/). These sentences are 阿友要领米 ‘A-You will claim the rice’ and 阿美要养鸟 ‘A-Mei will keep a bird’.

**NC alternative pronunciations.** During the main experiment, some NC words were noted to have systematic alternative readings varying in the segments, not in the tones. This variation is termed 文白异读 (literally, ‘variation in the literary and colloquial pronunciations’), which has been introduced in § 1.2. Such words appeared in the responding sentences are listed below; underlined words only appeared in the reporting clauses. We use the symbol ~ to record the variants; the former one(s) is more common in Nanning Cantonese:

Words varying in vowel: /ɛ~a/ in 关, 还, 炒, 弹; /ɐi~i/ in 美; /i~ei/ in 寄; /ɛ~i/ in 骑; /ɔ~œ/ in 朵; /i~y/ in 佢; /u~ou/ in 路; /ɛ~e~ia/ in 听

Words varying in consonant: /t~s/ in 洗, 信

Participants were allowed to produce the variant they preferred and no need to follow the pre-recorded elicitation, although we forgot to explicitly mention this during the instructions.

(Selected NC items did not have multiple readings in tone. However, in the main experiment, three NC speakers tended to produce 凤 /fong<sup>22</sup>/ in an alternative reading of /fong<sup>35</sup>/, due to self-claimed influence by Guangdong Cantonese. However, when in a disyllabic context 凤凰 ‘phoenix’, they pronounced 凤 as /fong<sup>22</sup>/. Therefore, we believed 凤 had a single pronunciation of /fong<sup>22</sup>/, those participants were thus asked to produce /fong<sup>22</sup>/ in the experiment.)

### **Minor inconsistency in elicitations and reporting clauses**

To maximise the naturalness of the dialogue, a few unavoidable lexical variations across Mandarin and Cantonese were permitted in the elicitation questions and reporting clauses, not affecting the target sentences.

<b>Mandarin</b>	<b>Cantonese</b>	<b>English translation</b>	<b>Distribution</b>
刚刚	啱啱	‘just now’	elicitation
他	佢	‘third-person pronoun’	elicitation; reporting clause in target
说	讲	‘say’	elicitation; reporting clause in target
什么	乜嘢	‘what’	elicitation

**Table 3.1.4:** Lexical items inconsistent across the three varieties.

#### 3.1.1.4. Interim summary of materials

In summary, although all four meaning conditions were presented to participants, only the three focus conditions (i.e., Broad Focus, Initial Focus, Final Focus) were subjected to the subsequent analysis; Broad Question condition is reserved for future study. Regarding tones, all the non-checked tones of the three language varieties were under investigation. The table below summarises the factors involved in the present thesis.

<b>Factor</b>	<b>Value</b>
focus size + new/given information	broad focus, all new; narrow focus, given
Focus position	initial (subject); final (object)
Tone of SVO	NM Tone 1–4; NC Tone 1–6; SM Tone 1–4

**Table 3.1.5:** Factors under investigation of intonation.

#### 3.1.2. Experiment design

It was not feasible to create a unique, independent context for each question–answer pair, as this would be difficult to control alongside the many linguistic variables already under investigation (e.g.,

tone, focus). The experiment therefore employed a consistent, structurally repetitive task. This structural homogeneity could potentially lead to participant fatigue or a stereotyped, list intonation over the course of the experiment. To mitigate these potential effects, several experimental designs were implemented, including pseudo-randomising the experiment trials (§ 3.1.2.1), designing filler trials (§ 3.1.2.2), and using both visual prompts (§ 3.1.2.3) and pre-recorded audio prompts (§ 3.1.2.4) for elicitation. Finally, the validation and pilot phase of the experiment is sketched in § 3.1.2.5. In addition, the trial number of each token was also recorded in case it is necessary to statistically model any potential order effects.

### **3.1.2.1. Pseudo-randomisation of trial order**

Pseudo-randomisation has been applied: the trials are randomised but with the constraint that same meaning condition never appears consecutively, and same lexical set never appears consecutively. To prevent the participants from falling in to a rhythm or building strong expectations, we took the following steps to create psuedo-randomised sentence lists per language variety per participant:

- 1) Randomise the order of the 12 Mandarin and 18 Cantonese lexical sets per participant, and repeat that order 4 times.
- 2) Manually repeat the order of the 4 meaning conditions for 4 times for Mandarin and 2 times for Cantonese, obtaining two ordered lists. This is to make sure that the meaning conditions to be in fixed order, so that each lexical set will be unrepeatedly assigned to each of the four meaning conditions (when lowest common multiples are involved here).
- 3) Repeat the ordered list of meaning condition 3 times for Mandarin and 9 times for Cantonese, until 48 Mandarin and 72 Cantonese question–answer pairs are obtained. The obtained pairs will be unrepeated, and adjacent meaning conditions or lexical sets will be avoided as possible.
- 4) Repeat the 48 Mandarin and 72 Cantonese question–answer pairs twice. This creates the full stimuli lists with 96 Mandarin and 144 Cantonese formal trials.

- 5) Re-randomise every 4 question–answer pairs once the overall order has been fixed. This breaks the fixed order of the meaning conditions, so that the order within every four trials is shuffled once while the overall order is remained. Nevertheless, the trials may become (slightly) less dispersed, as this cannot be achieved with re-randomisation at the same time. But we have already achieved a good amount of randomisation in order, and we will have an extra step to insert fillers next.

By doing so, the meaning conditions and lexical sets were each randomised once and maximally dispersed.

### **3.1.2.2. Filler trials**

As mentioned above, each target sentences will be produced four times within one repetition per participant, as responses to the four meaning conditions. The repeated nature of the experiment may cause unnatural intonation due to various reasons. For example, speakers’ intonation patterns may tend to be fixed as the elicitations and targets are in similar structures, which bores the speakers through this large amount of similar trials. To reduce the effect of repetition and distract the speakers from the repeated utterance, we inserted a list of filler trials as a tool of “resetting”, which is provided in Appendix B.4.2. We truncated each trial list into blocks by the four meaning conditions, which has less elements than the lexical sets. Filler (trials) were inserted after each block, so that no meaning conditions or lexical sets appear repeatedly within one block. By doing so, the trials were presented in a pattern of “4 stimuli + 1 filler”, which might be noticed by the participants. However, if we re-randomise within each block with 5 trials, we will risk having the same lexical set or meaning condition in adjacent trials. Between the two, we chose to retain the “4 stimuli + 1 filler” pattern.

The fillers are a set of elicitation-response pairs that have different meaning conditions and sentence structures from the formal trials, but can still fit in the question–answer paradigm. By doing so, the fillers are expected to “reset” the speaker’s production “memory” to an extent, but not standing out from the trials in form and keep the entire experiment in a consistent format. The intonation types of

the fillers are listing sentences (including counting) and alternative questions, elicited by simple imperative sentences as instructions. Listing sentences were designed to vary in speech rate<sup>17</sup>, and alternative questions were designed to switch the order of the two alternative elements. Tone variations were not delicately considered for the filler trials, but the target sentences were designed to cover various tones in both Mandarin and Cantonese.

Similar to the formal trials, the fillers sentences are also identical across Mandarin and Cantonese, with few exceptions in the elicitations/instructions and the disjunctive conjunctions in the target alternative questions, not affecting the listing and alternative elements in the targets (see Table 3.1.6). In Nanning Cantonese, several disjunctive conjunctions can be used for alternative questions. According to ten native NC speakers and the existing literature and corpus (e.g., Y. Lin & Qin, 2008), only the most universally used one by all these sources were chosen for the fillers.

NM	NC	SM	English translation	Distribution
说	讲	说	‘say’	filler instruction
的	嘅	的	(structural particle)	filler instruction
这	阿	这	‘this’	filler instruction
个	只	个	(quantifier)	filler instruction
还是	重係	还是	‘or’ (disjunctive conjunctions)	filler

**Table 3.1.6:** Cross-linguistically inconsistent lexical items in filler trials.

In total, 8 filler trials were designed for Mandarin (6 listing sentences, with 3 each for normal and fast speech rate; 2 alternative questions, with 1 each for the two item orders), and 12 filler trials were designed for Cantonese (8 listing sentences, with 4 each for normal and fast speech rate; 4 alternative questions, with 2 each for the two item orders).

<sup>17</sup> Segmental omission is a salient feature of continuous speech in Nanning Mandarin. Also, fast speech as a challenge or game is expected to make the experiment more engaging.

The number of filler trials need to be the same as the number of formal blocks, so 24 (96 stimuli trials ÷ 4 meaning conditions) and 36 (144 stimuli trials ÷ 4 meaning conditions) filler trials are required for Mandarin and Cantonese respectively. For each participant and each language variety, the fillers were repeated three times and randomised within each repetition, creating 24 and 36 filler utterances for Mandarin and Cantonese respectively.

We used R scripts to conduct the pseudo-randomisation and generate different trial lists for each participant.

### **3.1.2.3. Visual prompts: screen display**

On screen, each question–answer pair was presented simultaneously, with the question appearing above the response. The entire response sentence (i.e., reporting clause + target sentence) was displayed in a larger, bold font, but crucially, no typographic cues (such as highlighting, underlining, or in italics) were used to mark the intended focus item. This was a deliberate choice to elicit natural focus productions based on the conversational context with different pragmatic meaning conditions, rather than testing a participant’s ability to follow explicit prosodic instructions which only reflects our expectation. Filler trials were presented in the same format to ensure consistency.

All instructions and stimuli were presented in standard Simplified Chinese characters, including characters used only in Written Vernacular Cantonese. For certain Cantonese words with multiple written forms (e.g., 重係 ‘or’), the characters were chosen based on the wide recognition among native NC speakers<sup>18</sup> and in previous literature (Y. Lin & Qin, 2008).

To clearly denoting the interlocutors, the elicitation and response was preceded by one of two icons respectively, representing the different speakers. A third telephone icon, used in the pilot phase to indicate the call recipient and create a context of phone call, was removed from the final experiment

---

<sup>18</sup> During experiment design, NC speakers mostly preferred 重係 over 仲係. However, during the main experiment, some participants reported that 仲係 was easier for them to recognise the word, potentially because 重 alone has multiple readings.

as participants reported that it was confusing. Despite this simplification, some participants still commented that the task felt as natural as participating in a series of phone calls.

Our experiment employed minimal contextual framing, such as images or narrative text, in addition to the question–answer pairs. This design was chosen for several reasons: the practical difficulty of creating hundreds of unambiguous images, the need to avoid participant fatigue from excessive reading, and the goal of minimising production variability across the three language varieties. While a simple question–answer paradigm may not elicit the most spontaneous speech, it is a well-established method in focus research. Given that this is first attempt of systematic studying NM and NC intonation, this controlled approach was deemed sufficient for establishing a foundational description of their prosodic profiles.

#### **3.1.2.4. Audio prompts: pre-recorded elicitations**

##### **Nanning Mandarin and Nanning Cantonese**

Presentation in written Chinese, even with vernacular lexical items, carries the risk of inadvertently triggering some extent of a more formal register of Mandarin, i.e., a stronger SM accent (than the desired colloquial Nanning Mandarin). This is particularly when the production is collected under a formal lab setting. This can be mitigated by using speeches of a native NM and NC bilingual speaker as the primary prompt, making the texts only supplement.

Questions and focus happen in real language with other human beings, so a more natural way of elicitations for the lab setting is using role-play dialogues. Since we were not able to have participants work in pairs, to simulate role-play dialogues, we instead used a pre-recorded interlocutor. Our expectation was that the participants are encouraged to quickly settle into using their indigenous dialect by hearing it. We understand that a Nanning native confederate should be a better interlocutor that can better elicit more natural intonation in vernacular, but the use of pre-recorded audio can minimise the variability of prompts.

The interlocutor is a native NM and NC bilingual speaker (female, age = 46, year of resident in Nanning = 46). The eliciting sentences were recorded using Awesome Voice Recorder X v 2.1.0 (Newkline.Co,Ltd.) on an iPhone 12 smartphone in a recording studio of Nanning Radio and Television Station; except for seven filler instructions (for NM filler 2, 4, 6, and NC filler 2, 4, 6, 8), which were recorded using the same application on an iPhone 13 Pro Max in the interlocutor's bedroom with a quiet environment. This is because this set of instructions were improved following one pilot participant's suggestion. Due to recording and time restrictions, they were not re-recorded in the studio. The audio was recorded in the same setting as the experiments of this study: WAV format, mono channel, with a sampling rate of 22,050 Hz, a bit rate of 320 kbps, and a high encode quality. When producing the elicitations for Initial Focus and Final Focus, the interlocutor was required to produce with emphasis on “who” and “what”, to implicitly guide the participants to produce the corresponding focus.

The eliciting recordings were automatically sliced and scaled to an intensity of 70 dB using Praat (Boersma, 2002), except for the three re-recorded NM filler instructions. They were scaled to an intensity of 68 dB, because they were still a bit louder at 70 dB compared to the rest, reported by one pilot participant and confirmed by the researcher. A set of Praat scripts was developed to facilitate the process. The pre-recorded audio was validated by two native or fluent speakers of Nanning Mandarin and/or Nanning Cantonese, to make sure it was maximally natural.

61 eliciting sentences for Nanning Mandarin (3 lexical sets  $\times$  4 tones  $\times$  4 meaning types + 8 fillers + 5 practice) and 89 eliciting sentences for Nanning Cantonese (3 lexical sets  $\times$  6 tones  $\times$  4 meaning types + 12 fillers + 5 practice) were used in the experiment. The instructions for fillers were also pre-recorded to reduce the inconsistency between filler and stimuli trials.

### **Standard Mandarin**

For SM speakers, originally, we decided to use the same set of NM elicitations, because one pilot speaker and Speaker 01 confirmed that they would not reply to Nanning Mandarin with an NM

accent. According to them, they had been resident in Nanning for 31 and 53 years respectively, and had been used to replying to Nanning Mandarin in Standard Mandarin in their daily conversations, because this is their daily language environment. The researcher had also confirmed that these speakers did not acquired an NM accent in the previous tone study. This is also due to the consideration of time limits.

However, Speaker 02 reported that they were slightly affected by the NM elicitations under the Broad Question condition every now and then, if without reminding themselves. This was probably due to the different question particles used in Nanning Mandarin and Standard Mandarin. Therefore, a different set of elicitations was recorded exclusively for the SM speakers from Speaker 03, to maximally remove the potential effect of the NM accent. The SM elicitations was recorded at a quiet bedroom using Awesome Voice Recorder X v2.1.0 (Newkline.Co,Ltd.) on an iPhone 12 smartphone, with all the other recording settings and processing procedure remains the same. The interlocutor was the author due to the lack of Northern Mandarin speakers in Nanning. Although the author was not a native speaker of Northern Mandarin, she had passed Putonghua (SM) proficiency test with a level of 1-B (equivalent to a grade of 93.1%). Eliciting sentences were produced with an intonation that is maximally similar to natural speech, and with no NM accent, if there was any southern China accent.

The first ten trials, covering all four meaning conditions and two filler conditions, were validated by five listeners. They are two female and one male Northern Mandarin speakers, one Guangxi Cantonese and Mandarin bilingual speaker, and one Wu dialect and Mandarin bilingual speaker. All three Northern Mandarin speakers were born in Henan Province. The male listener had been resident in Nanning for 53 years, and one female listener had been resident in Nanning for 31 years; however, their Northern Mandarin (accent) had been retained. The other female listener had never resident in Southern China (for more than a year). The Cantonese and Mandarin bilingual speaker had been resident in Nanning for 38 years, and the Wu and Mandarin bilingual speaker had been only resident in Zhejiang Province (Wu Chinese region).

All of them reported that the elicitations did not have (noticeable) NM accent and sounded like natural speech. More importantly, although they could tell from the speech that the interlocutor was not from Northern China, the Mandarin itself is standard enough.

### **Participants' feedback**

The pre-recorded elicitation proved highly effective, supported by participant feedback during the experiment. The design offered several distinct advantages. First, it helped maintain consistent acoustic quality in the recordings. The pre-recorded stimuli, with their stable vocal intensity and pitch, served as an auditory anchor that mitigated the common tendency for speakers' productions to decrease in volume and pitch due to fatigue or being nervous. Second, the semi-spontaneous question–answer paradigm enhanced participant engagement and simplified task logistics. This interactive format prevented the cognitive monotony of repetitive tasks. Moreover, the conversational turn-taking allowed participants to time their button presses more intuitively within the rhythm of the dialogue, reducing the cognitive load of coordinating simultaneous articulation and manual action. Finally, the naturalness of the stimuli facilitated more spontaneous responses. The authentic NM and NC accents and natural focus intonation of the interlocutor(s) prompted participants to reply in a manner that was less stilted and more closely resembled natural speech. The SM elicitation was also reported to be natural and without NM accent by Speaker 03.

#### **3.1.2.5. Validation and pilot study**

Prior to the pilot experiment, all text materials underwent a validation process with a panel of native speakers of Nanning Mandarin, Nanning Cantonese, and Standard Mandarin. The validation confirmed the (linguistic) naturalness and pragmatic coherence of the design. The language used in the elicitation questions, target answers, and on-screen instructions was confirmed to be natural and appropriate for all three language varieties, with no language contamination; in particular, the validators agreed that their natural NM accent were prompted. The semantic and pragmatic link between each question and its answer was verified to be logical and natural.

(The experiment was piloted in person on four participants during the design.) The pilot participants may not take Nanning Mandarin and Nanning Cantonese as their L1, but they were all born in Guangxi, had been long resident in Nanning, and were at least very familiar with the NM accent. In particular, the final version of the full experiment was piloted on one NM speaker (female, age = 29, born in Guangxi, year of resident in Nanning = 19) and one Cantonese-Mandarin bilingual speaker (female, age = 66, born in Guangxi, year of resident in Nanning = 38) who had native-like NM intuition. Productions of both pilot speakers sounded very natural.

The two pilot speakers confirmed that all sentences were brief and sounded natural in NM and NC daily speech. They reported that once they were familiar with the sentence list, they were generally able to produce the five-syllable sentences without internal pauses or a flat, ‘report’ intonation. In addition, the speakers did not develop a monotonous or fatigued intonation during the task. Instead, they felt their productions became smoother and more natural as they progressed, attributing this positive outcome to the conversation-like pre-recorded elicitations. These findings from the pilot study were subsequently confirmed during the main experiment.

The experiment was not piloted on SM speakers due to limited speakers in Nanning.

### **3.1.3. Participants**

When recruiting the participants, acquaintances of the researcher were preferred for the convenience sampling.

#### **3.1.3.1. Nanning Mandarin and Nanning Cantonese speakers**

For Nanning Mandarin and Nanning Cantonese, 54 native speakers were recruited. Among them, 49 were bilingual speakers, which were preferred for convenience; 5 were NM monolingual speakers. 23 bilingual speakers and 2 NM monolingual speakers also participated in our previous tone study. We aimed at collecting data from 30 usable speakers to reach a (relatively) large sample size. To ensure the sample size is large enough, we also conducted calculation based on the rule of thumb of

Central Limit Theorem and power analysis. Our goal was to obtain the power ( $1 - \beta$ ) of 0.80 (conventional) to detect a strong effect size of 0.30 (strong suggested by Cohen,  $R^2$  for linear regressions) at the standard 0.05 alpha error probability (conventional). To conduct a power analysis, we ran the Omnibus F Test ( $R^2 > 0$  in Linear Regression) using the function of *pwrss* package (Bulus & Jentschke, 2025) in R: `pwrss.f.reg(R2 = 0.30, k = 3, power = 0.80, alpha = 0.05)`, with three factors (tone, focus condition, and focus position), and the resultant sample size ( $n$ ) = 30. Therefore, 30 independent data points, i.e., 30 participants were needed.

The eligibility of participation was determined based on a demographic interview concerning language background, parallel with the tone study. With a few exceptions, participants met the following criteria:

- They had been residing in Nanning before the age of 12;
- They had acquired Nanning Mandarin or Nanning Cantonese before the age of 12 (NM: 5 exceptions; NC: 4 exceptions);
- They did not have close family members or friends who spoke Northern Mandarin dialects;
- They had not received formal training in linguistics.

Additional criteria were preferred:

- They had not stayed outside of Nanning for more than one year within the past ten years (15 exceptions, among which only 1 had stayed in Northern Mandarin-speaking region);
- They had not received any explicit training or test in Standard Mandarin (22 exceptions).

Participants also needed to be proficient in written Standard Simplified Chinese, and have no auditory or speaking impairment.

Similar to the tone study, our recruitment criteria for the intonation study were also inclusive in terms of the language backgrounds beyond Nanning Mandarin and Nanning Cantonese and the diverse social backgrounds (see § 2.1.2 for details). We additionally documented speakers' most

confident language variety and other language capacity: all speakers were most confident in Nanning Mandarin and/or Nanning Cantonese, and some speakers could additionally speak various Cantonese dialects, Zhuang (languages), Pinghua, and Southern Mandarin.

Participants were roughly balanced for gender (28 females and 26 males) and evenly distributed from young to old for age (20 for age 20–39, 15 for age 40–49, 9 for age 50–59, and 10 for age 60–69), with some leniency given the convenience sampling. Speakers aged 20 to 68 (median age = 45); among them, speakers aged 40 to 60 formed the majority, for the similar reason as in our previous tone study (see § 2.1.2 for details). Participants aged 20–39 were recruited in higher numbers than those in other age groups. This over-recruitment was unavoidable to secure a sufficient number of usable young NC speakers. Many younger candidates, while able to produce isolated words suitable for the previous tone study, lacked the conversational proficiency required for the present intonation study. This recruitment challenge reflects the current sociolinguistic state of Nanning Cantonese: the younger generation is now largely unable to speak it fluently. Due to the strong SM promotion, Standard Mandarin is mandatory in key domains such as schools and the workplace. This in turn gradually shifts people’s language attitudes, which tends to subordinate Nanning Cantonese to the more prestigious Standard Mandarin. Consequently, Nanning Cantonese furthermore lose its language environment at home and faces endangerment, particularly among younger speakers, most of which now acquire Nanning Mandarin as their L1.

Table 3.1.7 displays the detailed age and gender distribution of the recruited participants.

Age \ Gender	20–39		40–49	50–59	60–69	Total
	20–29	30–39				
Female	9	4	5	3	7	28
Male	5	2	10	6	3	26
SUM	14	6	15	9	10	54

**Table 3.1.7:** Age and gender distribution of the recruited participants of Nanning Mandarin and Nanning Cantonese.

### **3.1.3.2. Standard Mandarin speakers**

For Standard Mandarin, ten fluent speakers of Standard Mandarin were recruited also from Nanning. All of the SM speakers participated in our previous tone study. We only collected a small amount of SM speakers due to the time and place limit, and also because SM intonation is very well-investigated with numerous previous studies.

Although participants originated from various Northern Mandarin areas across China with their own regional dialects, all of them were native speakers of Northern Mandarin dialects, and eight of them also identified SM as their L1. Participants left their Northern Mandarin speaking hometown at an average age of 21.3. In their self-evaluation, all speakers confirmed Standard Mandarin as one of their most confident varieties, and they further reported an average score of 7.85 on a 1–10 scale (10 as native-like fluency) on their SM fluency.

Although all participants were residing in Nanning, most of them reported no (discernible) influence from the NM accent on their speech; three participants noted only a very slight influence. Individuals with formal broadcast training were deliberately excluded, as the prosodic and articulatory patterns of broadcast speech differ markedly from spontaneous, conversational speech. (Except for one speaker who could additionally speak Southwestern Mandarin, none of the speakers could speak other Chinese languages.)

Due to the very limited number of Northern Mandarin speakers in resident in Nanning, we did not balance the gender and age of speakers. Eventually, we recorded 3 female and 7 male speakers aged 25–61 (median age = 46.5).

Age Gender	20–39	40–49	50–59	60–69	Total
	Female	2	0	1	0
Male	2	1	3	1	7
SUM	4	1	4	1	10

**Table 3.1.8:** Age and gender distribution of the recruited participants of Standard Mandarin.

### 3.1.4. Recording procedure

#### 3.1.4.1. Before recording

##### Recording settings

Recordings were conducted in person locally in Nanning. The entire process took place by a desk in a quiet room or booth in a location convenient for the participant. The speaker sat in front of the laptop, and the researcher sat on the right of the speaker. The smartphone was mounted on a tripod at approximately mouth height, positioned 25 to 40 cm from and slightly to the right side of the speaker’s mouth, to prevent recording air bursts. Mutual acquaintances of the author and the participant and/or a native Guangxi Cantonese speaker were also present for most of the participants (except for 12 participants), to relax the participants and better communicate with the NC speakers.

The instructions and question–answer pairs were presented in Simplified Chinese on a Macbook Pro (Retina, 13-inch, Early 2015 for the first 38 NC and NC speakers and all 10 SM speakers; Apple M3 for the last 16 NM and NC speakers) via PsychoPy v.2021.2.3 (Peirce et al., 2019). The production was recorded using Awesome Voice Recorder X v2.1.0 (by Newkline.Co,Ltd.) on an iPhone 12 smartphone. The device used for our intonation study is different from that for the tone study (i.e., iPhone 7), but the effect should be minor, because the primary measurement of this study, i.e.,  $f_0$ , is very robust. The audio was recorded in WAV format, mono channel, with a sampling rate of 22,050

Hz, a bit rate of 320 kbps, and a high encode quality.

### **Demographic interview**

Before the study, participants received a recorded demographic interview about their language and social backgrounds, except for seven speakers who received the interview after the recording due to the limited time window of quiet environment. The demographic information includes the occasions and frequency of language use, SM and other language proficiency, educational backgrounds, occupations, age and gender. For Nanning Mandarin and Nanning Cantonese, we also recorded the age and approaches of NM and NC acquisition; for Standard Mandarin, we also recorded the age they left Northern Mandarin region, year of residence in Nanning, as well as self-evaluation of SM fluency and NM influence.

We also collected information about their experience in music training in terms of the type (instrumental/vocal, melody/rhythm) and length. According to Jansen (2023), musical abilities was found to have a significant correlation with speech prosody perception. This correlation is stronger for music perception than music training, and stronger for prosodic pitch than prosodic time perception. We have included a short question regarding music training, but we were not able to collect data of participants' music perception.

### **Instructions**

In the instructions, participants were made sure to understand the conversational context of the lab speech, the procedure of the PsychoPy experiment, and the task requirements. They also had an overview of the stimuli and filler lists, to ensure that they were able to recognise the texts and produce the sentences smoothly. This is particularly for the few characters used uniquely in Cantonese, which commonly do not appear in written texts. Participants were also reminded of the tone of the proper names 莹 /jiŋ/ and 影 /jiŋ/, of which the tones sounded like switched in Mandarin and Cantonese and so thus easy to be confused and produced mistakenly. The participants could raise relevant questions or provide feedback during the introductions, the practice phases and

after the experiment.

### **3.1.4.2. During recording**

The recording of each language variety was preceded with a practice phase with five trials. The five practice trials were not included in the formal trial list, covering all four meaning conditions plus one listing filler, and randomised in order. This is to confirm the speakers with the required language environment, especially for the bilingual speakers who needed to switch the second language variety.

Language order was counterbalanced across NM and NC bilingual participants; trials were repeated twice and pre-randomised in order, using the aforementioned pseudo-randomisation method. Question–answer pairs were presented one by one on the screen; the participant controlled the progression by pressing the right button on the keyboard. Participants pressed the space button to play the pre-recorded elicitation, produced their response, and pressed the right button to move on.

Participants were allowed to repeat a token if needed for correction, simply by pressing the space button to replay the sound and response. The author would not validate a speaker’s productions regarding whether the intonation realisation is “correct” or “proper”, as we look at the prosody people actually produce. But if the author, as a native NM speaker and can understand Nanning Cantonese, believe that the speaker did not produce the utterance according to the elicitation, i.e., did not meet the task requirement, the author would ask the speaker to produce it again.

After every 30 trials, there was a 60-second short rest (thus, three rests for Mandarin recording and five rests for Cantonese recording), serving as psychological milestones. This was suggested by one pilot participant to rest their eyes and brains from the repeated nature of the recording, and was also useful for the participants to provide feedback and apply necessary adjustment. Participants were also allowed to pause any time when they feel needed, as they can control the experiment progress by key press. They can also skip or end the rest early by key press. Additionally, bilingual speakers had a rest between the two language varieties, and the duration was unrestricted.

The pre-recorded audio was played by the speaker of the laptop, which was confirmed not affecting the speech naturalness by the pilot speakers. The reasons are as follows. 1) The researcher attended all the experiments in-person to monitor the sound volume. 2) Presumably, using a speaker should make the question–answer more conversation-like than using a headphone or earphone, which may make the participants less nervous. 3) Using a speaker is more convenient for the researcher to monitor the recording than using a headphone or earphone. 4) Using a Bluetooth speaker will have better sound quality but also with frequent electric noise. The sound quality of MacBook Pro M3 is close to real human, which is way better than MacBook Pro 2015. We believe that this did not affect the experiment quality, as most of the participants reported that the form of experiment was natural; with a less real sound quality, it sounded like telephone call.

Due to the restriction of the recording venue (i.e., not in a professional recording studio), instances of recorded tokens were found to have background noises, among them the most common one is the cooling pads of MacBook Pro 2015. In the pilots, we had confirmed that this degree of noise did not affect the pitch trace of the speech in Praat. During the practice phase, we made sure that the speaker’s speech was naturally loud without holding their voice, and the intensity of noise was no higher than 1 unit on the volume bar on AVR. The measuring unit of intensity in AVR is unclear, but this can serve as a rough reference at the recording.

In total, 2976 NM, 4464 NC and 960 SM sentences were collected.

## **3.2. Data processing and analysis**

The following sections will describe the detailed pipeline of processing and analysing the production data. The procedure of data processing includes annotating the data (§ 3.2.1), validating the accents and naturalness (§ 3.2.2), and then selecting the data subjected for analysis (§ 3.2.3). For acoustic analysis, the data underwent f0 extraction (§ 3.2.4) and f0 standardisation (§ 3.2.5).

### **3.2.1. Annotation and segmentation**

We used Praat (Boersma, 2002) and Python scripts for word-level segmentation. We first ran Python scripts to tokenise the utterances using two segmenters, i.e., *pkuseg* package (R. Luo et al., 2022) for Mandarin and *pycantonese* package (J. L. Lee et al., 2022) for Cantonese. The TextGrids with the silent and sounding intervals were then created using *To TextGrid (silences)*... in Praat. For the silence threshold, we used -40 dB for most speakers, -45 dB for some speakers (7 NM, 5 NC, 7 SM), and -35 dB for only few speakers (5 NM and 4 NC), to obtain the most accurate TextGrids and reduce the manual workload.

The TextGrids were then attached with the tokenised sentences, as well as the information of validation result (see § 3.2.2), meaning condition, lexical set, tone category, and type (elicitation / target response). The output TextGrids of each step was manually checked. We did not use automatic speech recognition, such as Whisper, for segmentation, because both the time stamps and transcriptions it generated were far from accurate.

We then used Montreal Forced Aligner (MFA) v3.2.0 (McAuliffe et al., 2017) for phone-level segmentation. We used the up-to-date acoustic models and lexicons of the VoxCommunis Corpus (Ahn & Chodroff, 2022), which is derived from the Mozilla Common Voice Corpus (Ardila et al., 2020). For Mandarin, the model zh-CN\_vxc\_acoustic17.zip was trained using the Common Voice Corpus 17.0 (CV 17) of Chinese (China) (zh-hk) (updated on 2024/03/20, 234 validated hours), and the lexicon was adapted from zh-CN\_vxc\_lexicon17.txt; for Cantonese, the model yue\_epi\_acoustic19.zip was trained using the Common Voice Corpus 19.0 (CV 19) of Cantonese (yue) (updated on 2024/09/18, 209 validated hours), and the lexicon was adapted from yue\_epi\_lexicon19.txt.

For Mandarin, the VoxCommunis model generally behaved better than the MFA default model, as its phones were trained uses a lexicon without the attached tones. For Cantonese, we did not use the HKCantonese\_models (So et al., 2025), as they were trained from the Common Voice Corpus of Hong Kong Chinese (zh-HK) in addition to Cantonese (yue) corpus. The Hong Kong Chinese corpus was less recommended as the Cantonese were unnatural due to the use of Mandarin syntax

and lexicon. (The text prompts in Common Voice’s *yue* subset are generally in Written Vernacular Cantonese, while Common Voice’s *zh-hk* subset contained a mixture of Standard Written Chinese and Cantonese (Lau et al., 2025).) We should also note that the sound system of Nanning Cantonese (Yong-Xun subgroup) is quite different from those of Guangzhou and Hong Kong Cantonese (Guangfu subgroup), so neither MFA models would be highly accurate for Nanning Cantonese. This is not a major issue in the present study as the focus of this study is not on segments. We did not use the other forced aligners, such as Charsil aligner, as MFA was trained from larger datasets.

Initial inspection of the automatic alignment revealed inaccuracies. In some instances, MFA incorrectly modified sentence boundaries, and phone boundaries were also sometimes misaligned. These errors are likely attributable to the connected and coarticulated nature of the semi-spontaneous speech (Nanning Mandarin in particular) collected in the experiment. Creating a script for automatic correction was not feasible, as the error patterns were inconsistent across conditions (e.g., speakers and items). Therefore, we manually corrected the most significantly misaligned boundaries for all TextGrids (to ensure data reliability); the rest of the automatic alignment was considered reliable.

### **3.2.2. Validation of accents and naturalness**

The exclusion of speakers and tokens were based on the combination of the linguistic and demographic profiles and the accent validation by native speakers.

#### **Nanning Cantonese young speakers**

As mentioned in § 3.1.3, due to the strong promotion Standard Mandarin, the younger generation in Nanning commonly take Nanning Mandarin as their native language, resulting in greatly reduced NC speakers. Therefore, Nanning Cantonese effectively became their heritage language, or even worse, the younger speakers only have lower NC proficiency or competence. During the data collection, the author noticed that younger speakers, especially those aged below 35, could hardly speak fluent Nanning Cantonese and mostly did not choose Nanning Cantonese as their confident language variety, even though they claimed to have the capability to speak Nanning Cantonese. This is mainly

because they had very low exposure to Nanning Cantonese during growth and seldom had the language environment to use Nanning Cantonese.

Therefore, validation of NC proficiency is necessary for the 16 NC speakers aged 20–39, to exclude those who could not speak fluent Nanning Cantonese. The NC proficiency accent were validated from two aspects: (1) Cantonese proficiency: is speaker's Cantonese production correct, accurate and fluent? (2) NC accent: is the Cantonese production in an authentic NC accent, rather than other Cantonese accents, especially Guangdong or Hong Kong accent?

The recordings were audited by three listeners including the author.

- 1) The author is a native NM speaker, can mostly understand Nanning Cantonese and speak basic Nanning Cantonese.
- 2) The second listener was born in Guangxi, takes Yong-Xun Cantonese as L1, and had been resident in Nanning for 38 years. She had native-like NM intuition and thus was able to identify the unauthentic or less proficient Guangxi Cantonese and the nuance in accent under the influence from of Guangdong Cantonese or Mandarin. (The second listener participated in the pilot study.)
- 3) The third listener was born and has been always resident in Nanning (year of resident = 46), taking Nanning Mandarin and Nanning Cantonese as her L1. She was able to confirm the most authentic NC accent. (The third listener was also the interlocutor of the pre-recorded elicitations.)

All three listeners were required to mark the recording of each speaker on a scale 1–5 (5 as authentic Nanning Cantonese, 0 as cannot speak Nanning Cantonese at all), where intervals of 0.5 were allowed. A score of 4–5 would be considered as acceptable Nanning Cantonese. The second and third listeners were not provided with the ratings of their previous raters, to avoid any potential interference in judgement. The procedure was as follows:

1. The recordings went through an initial validation by the author during the recording. Any instances that were incorrect, inaccurate, or not smooth, were marked by the author. An evaluation was generated based on the demographic information and the detailed notes of productions. Based on this evaluation, speakers that would undergo further validation were selected.
2. The second listener also attended most of the in-person recordings and had first impressions of the speakers and their productions. After the recording, she went through all the recordings, marking all the instances that are not accurate, smooth, or with accents outside Guangxi.
3. According to the consensus of both the author and the second listener, speakers with a score of 5 were accepted, and those with a score 1–3 were excluded. These two sets of speakers then served as the top and bottom references for the third speaker to rate, with additional brief comments on the reasons of low-rated speakers. The rest of the speakers with score between 3 and 5 ( $3 < x < 5$ ) were assigned to the third listener for validation.
4. The third listener was asked to listen to the recordings of each of those speakers. She was allowed to select and skip when listening, no need to go through the entire recording of each speaker; she ended up to listen to the recordings of the first repetition of each speaker. Notes taken by the second listener were provided as supplementary material for the third listener to learn about the general patterns, especially for the portions she skipped listening to. All recordings were anonymised, as the third speaker did not attend any of the experiment of the validated speakers. Notes of the second listener were agreed by the third listener and additional notes were added.

Finally, 9 young NC speakers with an average score of 4–5 were retained for the further analysis.

### **Nanning Mandarin and Standard Mandarin speakers**

Similar validation procedure was also applied to Nanning Mandarin and Standard Mandarin.

For Nanning Mandarin, 8 speakers were selected by the author to be rated by the third listener on a scale 1–5. These speakers were initially evaluated by the author to potentially have a weak NM accent due to the influence of other accents, or to be less natural in speech. After the validation, 4 speakers with an average score of 4-5 were retained for further analysis. Unlike the reading task in our tone study, the question-answer paradigm using pre-recorded elicitations largely improved the naturalness of the sentence productions.

For Standard Mandarin, one speaker was initially evaluated by the author to have less natural speech. This speaker was validated by a Northern Mandarin speaker and finally retained. All 10 speakers were then rated by two Northern Mandarin speakers on a scale 1 to 5. Both listeners had also involved in the validation of the pre-recorded elicitation; for details of their language backgrounds, see § 3.1.2.5.

The auditors were asked to rate the SM speaker by their pronunciation and naturalness. Pronunciation was the primary criteria: speakers were expected to speak representative Northern Mandarin, with minimal accent of regional dialects, but no need to apply strict requirement as that to broadcast hosts. Speakers' self-rating scores were provided for the listeners as rating references, but the listeners were not provided with each other's rating. After the validation, all 10 SM speakers got an average score of 3.8–4.5, among which 9 speakers scored 4–4.5. Therefore, all 10 SM speakers were retained for the further analysis.

### **3.2.3. Data selection**

Due to the limited time and the large amount of data, for Nanning Mandarin and Cantonese, only bilingual speakers with fully usable Nanning Mandarin and Nanning Cantonese were retained for further analysis. Speakers who were NM monolingual, had either Nanning Mandarin or Nanning Cantonese excluded after validation, or have partial usability (e.g., less clear speech, less natural speech, or consistent problems in producing particular meaning conditions), were temporarily excluded from the analysis for this thesis.

After the validation and selection, 31 NM and NC bilingual speakers (female = 15; aged 27–67, median age = 48) and 10 SM speakers (female = 3; aged 25–61, median age = 46.5) were retained for further analysis. Speakers aged 20–29 and 30–39 were grouped into one age group 20–39 for the similar reason in § 2.1. Participants’ gender remained balanced and age group remained evenly distributed. The detailed age and gender distributions for all selected participants could be found in Table 3.2.1 below.

From the selected speakers, 2976 target utterances in Nanning Mandarin, 4464 target utterances in Nanning Cantonese, and 960 utterances in Standard Mandarin were remained for further analysis.

Nanning Mandarin and Nanning Cantonese bilingual speakers

Age \ Gender	20-39		40-49	50-59	60-69	Total
	20-29	30-39				
Female	2	3	2	3	5	15
Male	2	1	6	4	3	16
SUM	4	4	8	7	8	31

Standard Mandarin speakers

Age \ Gender	20-39		40-49	50-59	60-69	Total
	20-29	30-39				
Female	1	1	0	1	0	3
Male	0	2	1	3	1	7
SUM	1	3	1	4	1	10

**Table 3.2.1:** Summary of the age and gender information of the selected speakers.

### 3.2.4. Extracting f0 min, max, and range

As reviewed in § 1.3, previous studies on focus intonation mostly examined f0 range and/or mean f0

contour. To establish a fundamental investigation of the focus effect on f0, we started with f0 range, a simpler but representative acoustic measure of f0 modulation. Although f0 range may be less sensitive for level tones, it is included as a baseline acoustic measure, and the results in f0 range, minimum (min) and maximum (max) may further indicate where contour analysis is needed under specific focus conditions or tone categories.

### 3.2.4.1. F0 extraction

F0 min and max were automatically extracted in Praat using an adapted two-pass method. The f0 measurements were first extracted from the sonorant portion of each syllable in the five-syllable target sentences. From these values, only the f0 values from the subject and object words were manually corrected, and then were re-extracted using the same method and settings.

**Two-pass method.** To reduce the undefined values, we improved the extraction methods. We used the two-pass method adapted from Hirst (2011) to automatically generate speaker-specific pitch ranges for f0 extraction. In the first pass of extraction, pitch objects were created for each sound file (i.e., each speaker in each language variety) using filtered autocorrelation and the Praat default pitch range (50–800 Hz). From these extracted values, the 25% and 75% quantiles ( $Q1$  and  $Q3$ ) of each speaker were summarised. The second pass then used raw autocorrelation with a pitch floor of  $Q1*0.6$  and a pitch top<sup>19</sup> of  $Q3*2.5$  to create pitch objects for each sound file. Unless mentioned, other parameters remained default. From these pitch objects, f0 min and max were extracted from the sonorant portion of each syllable in the five-syllable targets.

During the manual f0 correction (see § 3.2.4.2), a re-analysis procedure was implemented for TextGrids with numerous pitch measurement errors (e.g., halving/doubling, breaks, or erroneous tracks). These files went through the extraction again with an adjusted pitch floor, but only if this new setting was observed to yield a cleaner pitch track. Eventually, this adjustment was necessary

---

<sup>19</sup> When using filtered autocorrelation, the term “top” and “ceiling” are distinguished, as the pitch ceiling in filtered autocorrelation is skewed. See “[Pitch Analysis by Filtered Autocorrelation](#)” for details.

for only six speakers: the pitch floor of  $Q1*0.6$  was changed for two NC speakers (05, 19) to  $Q1*0.75$ ; two NC speakers (23, 24) to  $Q1*0.55$ ; and one SM speaker (02) and one NM speaker (24) to  $Q1*0.5$ .

**Algorithm.** We used filtered autocorrelation *Sound: To Pitch (filtered ac)* in the first pass because it is the new algorithm recommended by Praat manual for intonation analysis since 2023. However, we had to use raw autocorrelation *Sound: To Pitch (raw ac)* at the second pass, because filtered autocorrelation would generate much more undefined values.

**Pitch range.** In the first pass, we used the default pitch range because it generated the least number of undefined values across all three language varieties, among all the settings we had tested. At the second pass, we used a lower pitch floor and a higher pitch top than those in Hirst (2011), because it is more suitable to capture low tones and the high final rising in Broad Question.

**Quantiles.**  $Q1$  and  $Q3$  were extracted from the entire sound file, which also contained the pre-recorded elicitations, the carriers and the fillers, instead of from the target sentences only. This was to maintain consistency and was expected to only have minor effects.

To obtain more accurate results, different methods and settings were tested and compared using a subset of data. The detailed process of trial and error will be described below.

### 1) Testing method, algorithm, and pitch range

*Data used:* Speaker Batch 1, including four NM and NC bilingual speakers (female = 2) and 2 SM speakers (female = 1).

For pitch ranges, we tested the one-pass and two-pass methods using different settings. For Praat algorithms, we tested the traditional raw autocorrelation (raw AC) and the new filtered autocorrelation (filtered AC); the latter is recommended for intonation analysis by Praat manual.

**Pitch range.** Our initial approach to pitch extraction was the conventional one, which we term the

“one-pass method”. This involved a trial-and-error process using raw AC, to find a single pitch range that yielded the fewest undefined values across the dataset. In testing, however, we observed that the optimal pitch range differed across language varieties, and should also differ by speaker, given the intrinsic pitch variation. While manually determining a specific range for each speaker is too time-consuming to be practical for our large corpus.

To improve automation and replicability, we therefore tested the two-pass method proposed by Hirst (2011). In the first pass, a wide, fixed pitch range was used for each speaker, i.e., 50–700 Hz (Hirst, 2011) or 60–700 Hz (Hirst, 2022). From the extracted values, the 25% and 75% quantiles ( $Q1$  and  $Q3$ ) of the speaker’s  $f_0$  distribution were calculated. In the second pass, a pitch floor of  $Q1*0.75$  and a pitch top of  $Q3*1.5$  or  $Q3*2.5$  (Hirst, 2011, 2022) were used to automatically generate a speaker-specific pitch range.

**Algorithm.** Both algorithms were tested for the one-pass method. For the two-pass method, Hirst (2011)’s original two-pass method used raw AC in Praat, as this pre-dated the newer filtered AC (proposed in 2023). To compare the performance of these two algorithms within the two-pass framework, we tested all four possible combinations: *raw+raw*, *raw+filtered*, *filtered+raw*, and *filtered+filtered*. (In this notation, the first and second terms refer to the algorithm applied in the first and second pass of the extraction, respectively.)

**Extraction results.** We inspected the number of undefined values in the extracted output. TextGrids with and without boundary correction were first compared. In general, corrected TextGrids yielded slightly less undefined values than uncorrected TextGrids (as expected).

Regarding the performance of the methods, both one-pass and two-pass methods yielded the fewest undefined values. In the one-pass method, filtered AC produced slightly more undefined values than raw AC. For the two-pass method, however, any combination using filtered AC in the second pass (i.e., *raw+filtered* and *filtered+filtered*) caused a substantial increase in undefined values. A potential reason is that filtered AC requires a higher pitch top than raw AC, according to the Praat

manual.

The optimal algorithm combination for the two-pass method also differed by language variety. For Nanning Mandarin and Nanning Cantonese, the *raw+raw* combination produced the fewest undefined values; while for Standard Mandarin, it was the *filtered+raw* combination. This divergence may be attributed to the composition of participants: the NM and NC data were collected from the same set of bilingual speakers, whereas the SM data was sourced from a separate, distinct group of speakers. For the corrected TextGrids, only three methods could consistently generate a lower number of undefined values for all three language varieties: the one-pass method using raw AC (code: *yr1-1*) or filtered AC (code: *yf1-1*), and the two pass method using *filtered+raw* combination with the Praat default pitch range in the first pass (code: *yf2r-3*).

We recognised that a lower count of undefined values does not necessarily equate to higher extraction accuracy. Therefore, to assess the qualitative accuracy of the three methods (i.e., *yr1-1*, *yf1-1*, and *yf2r-3*), we manually inspected the uncorrected TextGrids, examining the boundary alignment, pitch traces and corresponding word of the problematic tokens. (Due to time restrictions, we did not inspect the corrected TextGrids.) This inspection revealed that the one-pass method using raw AC (*yr1-1*) generated the most accurate results overall, in terms of the mapping between the undefined values and their corresponding boundaries and pitch traces, although it did not always produce the fewest undefined values among the three methods.

Regarding the corresponding words, we found that the undefined values occurred most often in the low-starting tones (i.e., rising, low-falling, and low-dipping); this is particularly (common) for SM speakers and male speakers. Moreover, in Nanning Mandarin and Nanning Cantonese, undefined values mostly appeared on the name prefix (often short and creaky). In Standard Mandarin, the problematic items are generally the sentence-final object words under Initial Focus condition, likely due to substantial post-focus compression (PFC). While this aligns with our predictions, the greatly weakened production resulted in many undefined values and subsequent data loss.

Based on this comprehensive evaluation, a final decision was made. Despite the high accuracy of the one-pass method, its reliance on a time-consuming, trial-and-error process made it impractical for the full dataset. Instead, we selected the two-pass method with the *filtered+raw* combination (*yf2r-3*), as it is more systematic and replicable among the three methods. Although TextGrid inspection showed its output to be slightly less accurate in some cases, this method yielded one of the lowest counts of undefined values for all three language varieties. Moreover, it employed the more advanced filtered AC algorithm, and automated the process of generating speaker-specific pitch ranges.

## 2) Testing pitch range (further) and areas concerning extraction

*Data used:* Speaker Batches 1 and 2 (SM speakers excluded), including 12 NM and NC bilingual speakers (female = 7).

For this subset of 12 speakers, we used the selected two-pass method from the previous testing (*filtered+raw*, code: *yf2r-3*) for extraction. During the manual f0 correction after extraction (see § 3.2.4.2), we observed that the pitch traces of some speakers seemed unreasonably problematic, such as pitch halving, doubling or breaks when the audio was fine (e.g., no creaks or noise). This is particularly for the low tones. To check the reliability of the pitch measurements extracted using method *yf2r-3*, we inspected the entire TextGrid from each of these speakers.

**F0 floor.** For the TextGrids that extensively showed problematic pitch traces, we tested different pitch floors by manipulating the coefficients multiplied with  $Q1$ . Results showed that the pitch traces could be greatly improved by using a lower pitch floor. We assumed that this is potentially because the tonal languages investigated in our study can reach a much lower pitch floor

After comparing the extraction output, a lower pitch floor of  $Q1*0.6$  was selected as default, as it yielded better results for most speakers. However, during the manual f0 correction, we noted that a few speakers were better extracted with a different coefficient (other than 0.6). These exceptions did not follow a predictable pattern, so we were not able to set one criterion to automatically decide the speaker-specific coefficient. Therefore, a case-by-case protocol was adopted: (during manual f0

correction,) if a TextGrid was observed to contain significant pitch (measurement) errors, the coefficient was manually tested to find an optimised setting; the extraction was then re-run with the adjusted pitch floor. This approach acknowledges that no single pitch floor is perfect, as a pitch floor that is too low or too high can cause pitch halving or doubling respectively (Hirst, 2011).

**F0 top.** In TextGrid inspection, we observed that f0 values frequently exceeded the initial pitch top of  $Q3*1.5$ . This was particularly true for Nanning Mandarin and Nanning Cantonese, where the question -final particle 啊 /a/ often involved a sharp pitch rise; for Standard Mandarin, the overall f0 ranges could reach a high pitch top. Therefore, a higher pitch ceiling of  $Q3*2.5$  was adopted for the extraction. This adjustment is methodologically sound: according to Praat manual, the pitch top setting does not affect pitch measurement (e.g., analysis window), and previous research found no systematic errors from setting the pitch top too high (Hirst, 2011).

**Area for calculating  $Q1$  and  $Q3$ .** Currently,  $Q1$  and  $Q3$  were calculated from the entire duration of each speaker's recording. We acknowledge that an ideal approach would be to calculate the quantiles exclusively from the sonorant portions of each speaker's productions, including the reporting clauses and fillers. However, our simpler method was used to maintain consistency with data processed during the method testing phases and to adhere to the project's time constraints.

This methodological compromise is considered acceptable for two main reasons. First, quantile measurements are relatively stable and are not expected to vary dramatically based on the calculation area. Second, these quantiles served only as an intermediate step to estimate speaker-specific pitch ranges for the f0 extraction; they were not used as a direct measurement in the final acoustic analysis.

**Area for f0 extraction.** The f0 standardisation procedure in this study required speaker-specific f0 ranges. Currently, these f0 ranges were summarised from the f0 min and max that had already been corrected and re-extracted from the subject and object words of the five-syllable target sentences. (The full procedure for the subsequent manual correction and re-extraction of f0 values is detailed in

§ 3.2.4.2.) Theoretically, a speaker's true pitch range is best represented by the full extent of their productions. This raised a key methodological question: was our subset of data sufficient for a reliable calculation, or should a more inclusive set of segments (i.e., also containing fillers and reporting clauses) be incorporated?

Fillers, however, were immediately excluded. This was a practical necessity, as they had been segmented by word rather than by syllable, which disabled the isolation of the sonorant portions required for accurate f0 measurement. (Since fillers were not investigated in this study, they were thus safely excluded.)

To determine the effects of including the reporting clauses, we conducted a direct comparison. F0 min and max values were extracted under two conditions: 1) from the target sentences only (conducted in the previous steps), and 2) from the entire response, including the reporting clause. Resulting f0 and time measurements were identical for both conditions, according to the comparison in R. This finding is logical, as the S, V, and O were already designed to cover the full tonal inventory of each language variety and were therefore sufficient to cover the speakers' full pitch ranges. Therefore, the reporting clauses were also excluded from the final f0 extraction. This focused approach simplifies the procedure and maintains methodological consistency without sacrificing the quality of the pitch range data.

### **3.2.4.2. Manual f0 correction and re-extraction**

In TextGrid inspection of the initially extracted subset, we observed erroneous f0 min and max resulting from e.g., f0 measurement errors, noise, creaky voice, and microprosodic effects. To determine if these errors systematically impacted the results, we conducted a test on this subset of data. For each speaker, the first half of the data was obtained from the automatic extraction, while the second half was manually corrected. We then included a fixed effect for "experiment half" in our preliminary linear mixed-effects models; results showed significant interactions with tone and focus in several measures. This confirmed that the extraction errors were not random and necessitated a

full manual correction.

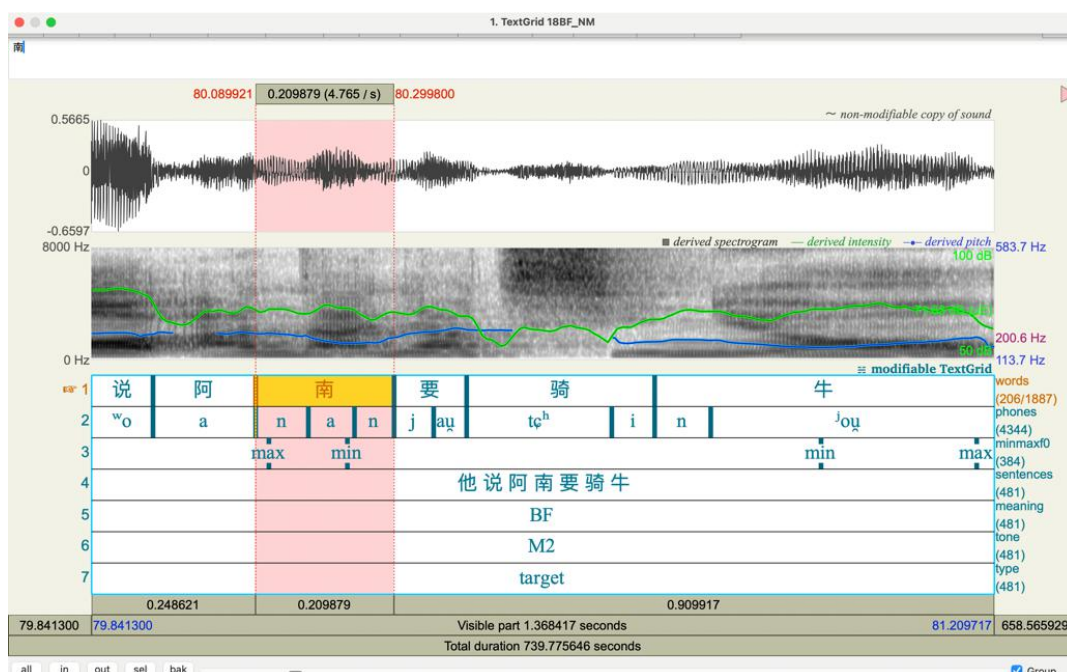
Given the statistical impact of these errors, we decided to manually check the extracted f0 values for all TextGrids. For the current f0 range analysis, this manual correction and the subsequent re-extraction were focused on the subject and object words only. This is to establish a baseline analysis on the primary carriers of tone and focus, which are crucial for the foundational stage of this investigation. A future analysis of f0 contour will investigate the entire target sentences, where all syllables (including the question particle) need to undergo the steps in this section, making a syllable-by-syllable f0 range analysis at this stage redundant. For the residual pitch track errors, no further manual calculation of f0 values was performed, as this was not feasible for this large corpus.

**Controlling for microprosodic effects.** As mentioned in § 3.1.1.3, the selected syllables included various onset types, i.e., sonorant, voiceless unaspirated, voiceless aspirated, and fricative. This potentially have microprosodic effects on the f0 contour at the start of each syllable. By definition, we assumed the microprosodic effects would be “micro” and thus minimal. Furthermore, because the same lexical items were used across all meaning conditions, any such effects were expected to be consistent and therefore not confound the main comparisons.

Despite this, we still controlled for their influence. First, erroneous f0 min and max caused by clearly identifiable microprosodic effects were manually corrected, guided by both visual inspection of pitch traces and auditory perception. A predictor for preceding consonant was included in the subsequent statistical models to account for any remaining systematic effects (see § 3.3 for details).

**Perceptually rising/falling tones.** Our initial plan was to manually correct the f0 min and max for perceptually dynamic tones (rising and falling), where the extracted f0 min and max within a syllable’s boundaries did not match its perceived contour. (Low-falling tones were perceptually low-level to the author’s ear, so it was not involved in this step.) For example, if a phonologically rising tone was measured with a falling contour, we intended to manually reverse the order of the f0 min and max points.

These mismatches are a common result of coarticulation, particularly the carryover effects, in connected speech. For instance, as shown in Figure 3.2.1, the syllable 南 /nan<sup>34</sup>/ is a rising tone. However, due to carryover from the preceding high-level tone (阿 /a<sup>44</sup>/) and the delay of the rise into the following syllable (要 /jau<sup>52</sup>/), its f0 contour within the segmental boundaries of /nan/ is actually falling. Consequently, the f0 max of /nan/ was actually extracted from part of the high-level tone.



**Figure 3.2.1:** Example of carryover effect in the subject with NM Tone 2.

However, when attempting this manual correction, it became clear that the process was both highly subjective and methodologically unsound. It is difficult to determine the “true” f0 min or max when a tone’s contour is truncated by, or spills across, segmental boundaries. Taking the above example, as the rise of /nan/ is mostly in the following syllable /jau/, if we put the f0 max to the right boundary of /nan/ to match with its perceptual rising contour, the modified f0 max would be more of an artificial measurement. We concluded that this tone-segment misalignment, as a fundamental feature of connected speech, should be more carefully dealt with rather than simply corrected as an error. Furthermore, as this is a production-only study, relying on researcher’s impressionistic judgements without a formal perception experiment would be unreliable.

Therefore, to avoid introducing subjective errors, we abandoned the plan for manual correcting for the perceptual dynamic contours. The f0 range analysis presented here is based on the automatically extracted f0 values aligned with the segmental boundaries, as an inevitable compromise. A more sophisticated analysis that takes into consideration the nature of tone in connected speech will be applied in future investigations of f0 contour, where the f0 range analysis could be updated accordingly.

**F0 re-extraction.** Following manual f0 correction, the (corrected) f0 min and max were re-extracted applying the same method as in the pre-correction extraction, i.e., the *filtered+raw* two-pass method with the same pitch settings.

### 3.2.5. Cleaning and standardisation

#### Cleaning data

Before any further steps of acoustic analysis, data were firstly cleaned in R.

Tokens (syllables) were excluded if they met any of the following criteria: with production errors (e.g., mispronunciations) or pauses; or with f0 min larger than f0 max due to wrong measurement of Praat. A total of 219 tokens were therefore excluded.

Outliers, defined as data points exceeding  $\pm 2$  standard deviations from the mean, were identified in R: if either f0 min or f0 max of a token is outlier, the corresponding token was removed. A total of 1995 tokens were additionally excluded.

A small number of remaining pitch-doubling or halving errors were retained in the final dataset, because such errors could only be resolved based on f0 contour extraction. This was considered acceptable, because initial manual inspection found such errors relatively infrequent in proportion to the large corpus, and the majority had already been removed during preceding data cleaning steps.

The final dataset contained 2877 NM and 4289 NC utterances from 31 NM and NC bilingual

speakers, as well as 944 utterances from 10 SM speakers.

### **F0 standardisation**

The raw f0 min and max values (in Hz) were standardised by speaker-specific f0 ranges for each language variety. The speaker-specific f0 ranges were summarised from the corrected and re-extracted f0 ranges in the subject and object positions only, as these were considered the most reliable.

The standardised f0 values were then transformed into the tone values in real-valued continuous Chao tone numbers ranging from 0 to 5, using the Log T-value formula (F. Shi, 1986):

$$LT = \frac{\log_{10} mean - \log_{10} min}{\log_{10} max - \log_{10} min} \times 5$$

LT Chao tone f0 ranges were then calculated from the standardised f0 values.

### **3.3. Results of f0 min, max, and range**

Related to the research questions proposed for this chapter, the following analyses aim to investigate (1) the f0 range modulations across focus conditions in each language variety (§ 3.3.1), (2) the effect of focus across different tone categories (§ 3.3.2), and (3) whether the focus-marking pattern in Nanning Mandarin aligns more closely with Nanning Cantonese or Standard Mandarin (§ 3.3.3).

Regarding sentence-level focus effects, we predicted Standard Mandarin to exhibit the typical focus-marking pattern reported in previous studies, namely on-focus expansion and out-of-focus compression. Particularly, we expected post-focus compression (PFC) to be present and pre-focus compression to be weaker. In contrast, Nanning Cantonese was expected to show limited f0 modulation and no local focus contrast; accordingly, PFC was not expected. For Nanning Mandarin, considering the long-term contact with Nanning Cantonese and its historically predominant status in the city, it is reasonable to expect a strong NC influence on Nanning Mandarin. We therefore expect

that the focus-related f0 range patterns in Nanning Mandarin are overall more similar to those in Nanning Cantonese. For example, local f0 contrast between on-focus and out-of-focus items may be absent, including post-focus compression (PFC), which resembles the case of Taiwan Mandarin speakers (Y. Xu et al., 2012).

The specific modulations of f0 min and max, however, may be tone-specific. Previous studies have shown considerable tonal variation in focus realisation, particularly in Cantonese. We therefore expected focus effects to vary across tone categories. In general, dynamic tones were expected to show greater f0 modulation, in particular, greater f0 range expansion under focus, which was commonly found in both Standard Mandarin and standard Cantonese (Y. Chen & Gussenhoven, 2008; Gu & Lee, 2009; W. L. Wu & Xu, 2010); level tones, in contrast, may show more limited f0 range modulation.

To examine the questions (1) and (2), a mixed-effects linear regression analysis tested the effects of Initial and Final Focus against Broad Focus in each language variety, as well as the interaction between each focus condition and each tone. The model was run on the LT Chao tone f0 max, min, and range in the subject and object nouns with fixed effects of **focus condition** (broad focus, initial focus, final focus), **tone** (Mandarin: 1, 2, 3, 4; Cantonese: 1, 2, 3, 4, 5, 6), **preceding consonant** (sonorant, voiceless unaspirated, fricative, voiceless aspirated), **age group** (20–39, 40–49, 50–59, 60–69), the full **interactions** between focus condition and tone, and a **random intercept** for speaker and target sentence. Additional **by-speaker slopes** did not converge. For Standard Mandarin, the random intercept for target sentence was removed as it did not converge due to the smaller sample size.

**Focus condition** was treatment-coded with Broad Focus as the control condition (Initial Focus +1 vs. Broad Focus 0; Final Focus +1 vs. Broad Focus 0). **Tone** was sum-coded with Mandarin Tone 4 and Cantonese Tone 6 as the held-out tone in the first iteration and Tone 1 as the held-out tone in the second iteration. **Preceding consonant** was sum-coded with fricative as the held-out consonant in the first iteration and sonorant as the held-out consonant in the second iteration. **Age group** was

sum-coded with age 60–69 as the held-out tone in the first iteration and age 20–39 as the held-out tone in the second iteration. This model rotation was implemented for interpretation of all potential focus–tone interactions, as well as main effects of preceding consonants and age group. (Baseline of comparison: broad focus, average consonant, average tone, average age.) The model results will be described in § 3.3.1 for the main focus effect and in § 3.3.2.1 for the tone-specific focus effect. The full set of models and their results are provided in Appendix B.2.1. For better interpreting the tone-specific focus effect, an additional post-hoc analysis is described in § 3.3.2.2, with (codes and) results provided in Appendix B.2.2.

To examine the question (3), an additional analysis tested the cross-linguistic differences in the overall focus effect. The model was conducted that directly compared the narrow focus effects among the three language varieties, including an **interaction** between the between-language difference and between-focus-condition difference (language varieties: Nanning Mandarin, Nanning Cantonese, Standard Mandarin; focus condition: broad focus, initial focus, final focus), and a **random intercept** for speaker and target sentence. Additional **by-speaker slopes** did not converge.

**Language** was simple-coded with Nanning Mandarin as the control condition (Nanning Cantonese  $2/3$  vs. Nanning Mandarin  $-1/3$ ; Standard Mandarin vs.  $2/3$  vs. Nanning Mandarin  $-1/3$ ). **Focus condition** was simple-coded with Broad Focus as the control condition (Initial Focus  $2/3$  vs. Broad Focus  $-1/3$ ; Final Focus vs.  $2/3$  vs. Broad Focus  $-1/3$ ). In this way, the main effects of language and focus condition would indicate the distance to the control condition, and the full interactions between language and focus condition would indicate the language variation addition to the narrow focus effect, or vice versa, the narrow focus effect addition to the language variation. The model results of cross-linguistic comparisons will be described in § 3.3.3. The full set of models and their results are provided in Appendix B.2.3.

All models were implemented with the *lme4* package with the *lmerTest* package for estimation of *p* values (Bates et al., 2015; Kuznetsova et al., 2017). Plots for raw acoustics were generated by

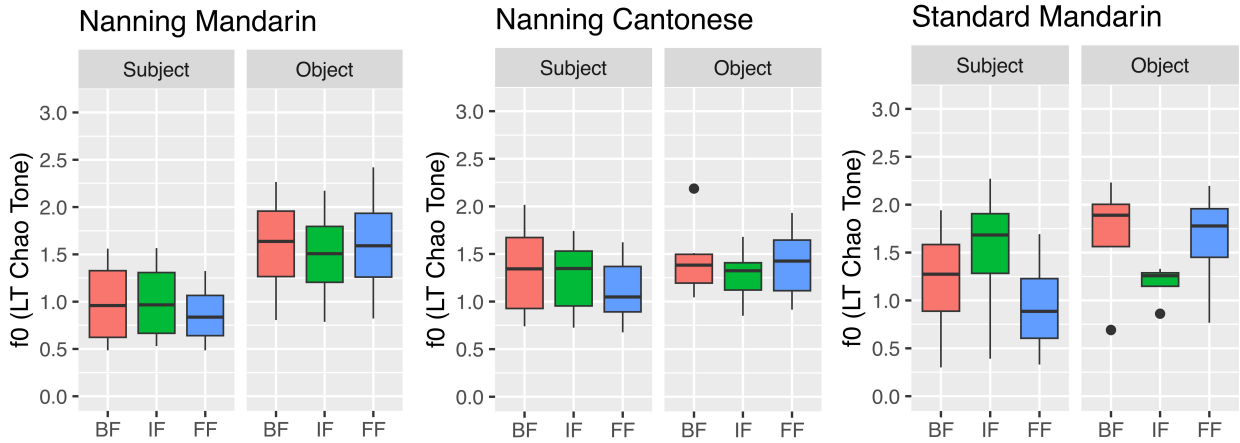
the *ggplot2* package (Wickham, 2016). Post-hoc analyses and its regarding plots for model-estimated marginal means (EMMs) were generated by the *emmeans* package (Lenth & Piaskowski, 2025).

### 3.3.1. Focus effects

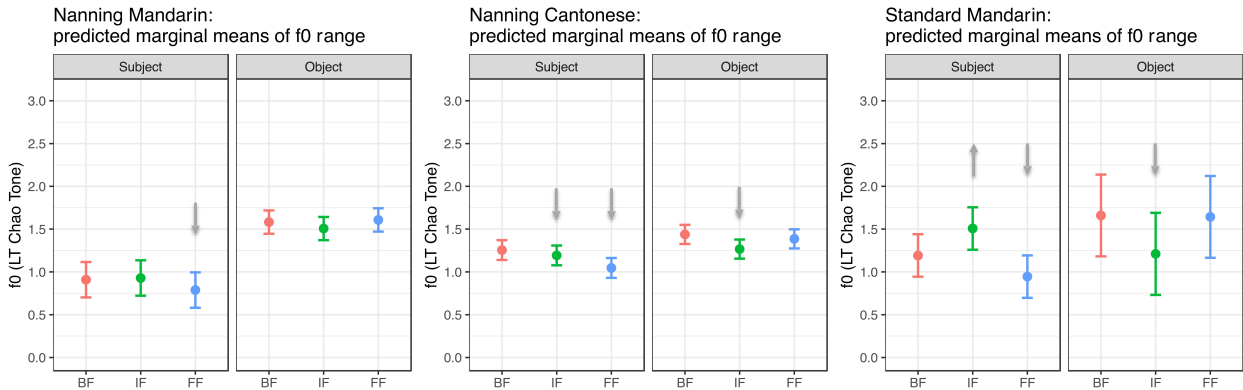
The first aspect under investigation of this chapter is the main focus effect, i.e., whether narrow focus conditions systematically modulate the f0 range (compared to Broad Focus). Our results show that different focus conditions yielded different effects, and the focus-marking patterns are different across language varieties. The figures below displays the raw acoustics (Figure 3.3.1) and model-estimated results (Figure 3.3.2) for the average f0 range in log-transformed LT Chao tone numbers across each focus condition in the subject and object positions, averaged across tones. In the latter figure, arrows indicate the direction of significant f0 modulations compared with Broad Focus: ↑ for expansion, ↓ for compression.

Estimated marginal means are model predictions averaged over the other covariates in the model (i.e., preceding consonant and age group). It is crucial to note that, due to the significant interaction between focus and tone detected in the model (see § 3.3.2), the main effect of focus does not fully capture the relationship; *emmeans* also issued a note advising that interpreting the main effect in isolation may be misleading. Therefore, Figure 3.3.2 should be interpreted as a generalised summary of the focus effect, averaged across all tones. The nature of this interaction is therefore explored next.

The model predictions actually display similar patterns to the f0 range in LT Chao tone. At a glance, Nanning Mandarin showed very limited f0 range modulation. The compression in the subject under Final Focus is a common modulation across the three varieties; in contrast, cross-linguistically various patterns were shown in the subject f0 range under Initial Focus. More detailed descriptions will be presented below, and the full set of models and their results are provided in Appendix B.2.1.



**Figure 3.3.1:** Average f0 range across each focus condition.



**Figure 3.3.2:** Model-estimated average f0 range across each focus condition. Arrows indicate significant expansion (↑) or compression (↓) compared to Broad Focus.

### 3.3.1.1. Nanning Mandarin

**Initial subject focus.** Initial Focus significantly modulated the f0 max, while no main effects of f0 range and min were observed. Compared to Broad Focus, the subject position under Initial Focus showed a significantly higher f0 max (max:  $\beta_{IF} = 0.06$ ,  $p < 0.01$ ), while the object position showed a significantly lower f0 max (max:  $\beta_{IF} = -0.06$ ,  $p < 0.05$ ).

**Final object focus.** Final focus significantly modulated the f0 min and range within the subject

position: a significantly higher subject f0 min and a narrower subject f0 range were observed (min:  $\beta_{FF} = 0.15$ ,  $p < 0.001$ ; range:  $\beta_{FF} = -0.12$ ,  $p < 0.001$ ), suggesting a compression effect. However, within the object position, no main effects of final focus reached significance for f0 max, min, or range.

**Preceding consonants.** Within the subject position, onset consonants significantly modulated the f0 min. Specifically, a sonorant onset and a voiceless unaspirated onset significantly lowered the f0 min (min:  $\beta_{sonorant} = -0.17$ ,  $p < 0.05$ ;  $\beta_{vcl.unasp} = -0.19$ ,  $p < 0.05$ ), while a voiceless aspirated onset significantly raised the f0 min (min:  $\beta_{vcl.asp} = 0.53$ ,  $p < 0.01$ ) (, compared to average).

**Age group.** No main effects of age group were observed on the subject position, whereas on the object position, the oldest age group (Age Group 60–69) showed an overall wider f0 range and higher f0 max than average (min:  $\beta_{60-69} = 0.20$ ,  $p < 0.05$ ; max:  $\beta_{60-69} = 0.22$ ,  $p < 0.05$ ).

### 3.3.1.2. Nanning Cantonese

**Initial subject focus.** Initial Focus compressed the f0 range in both the subject and object. The subject f0 range was significantly compressed mainly by raising the f0 min (range:  $\beta_{IF} = -0.06$ ,  $p < 0.01$ ; min:  $\beta_{IF} = 0.04$ ,  $p < 0.05$ ). The object f0 range was also significantly compressed, with a lower f0 max and a higher f0 min (range:  $\beta_{IF} = -0.17$ ,  $p < 0.001$ ; min:  $\beta_{IF} = 0.05$ ,  $p < 0.05$ ; max:  $\beta_{IF} = -0.12$ ,  $p < 0.001$ ).

**Final object focus.** Final Focus also modulated the f0 range, min and max. Similar to Initial Focus, Final Focus also significantly compressed the subject f0 range by raising the f0 min (range:  $\beta_{FF} = -0.21$ ,  $p < 0.001$ ; min:  $\beta_{FF} = 0.23$ ,  $p < 0.001$ ). On the object position, both f0 min and max were significantly raised without significant modulation on f0 range (min:  $\beta_{FF} = 0.11$ ,  $p < 0.001$ ; max:  $\beta_{FF} = 0.05$ ,  $p < 0.05$ ; range:  $\beta_{FF} = -0.05$ ,  $p = 0.06$ ).

**Preceding consonants.** Onset consonants significantly modulated the f0 range and max, mainly on the subject position. Within the subject, sonorant and voiceless aspirated onsets expanded the f0

range (range:  $\beta_{sonorant} = 0.16, p < 0.05$ ;  $\beta_{vcl.asp} = 0.41, p < 0.05$ ); additionally, voiceless aspirated onsets raised the f0 max (max:  $\beta_{vcl.asp} = -0.39, p < 0.01$ ). On the other hand, fricative and voiceless unaspirated onsets compressed the f0 range (range:  $\beta_{fricative} = -0.25, p < 0.05$ ;  $\beta_{vcl.unasp} = -0.33, p < 0.01$ ); additionally, voiceless unaspirated onsets lowered the f0 max (max:  $\beta_{vcl.unasp} = -0.33, p < 0.01$ ). Regarding the object position, only voiceless aspirated onsets showed a significant effect of f0 range expansion (range:  $\beta_{vcl.asp} = 0.18, p < 0.05$ ).

**Age group.** Significant effects of age groups were observed on f0 max and range. Age Group 60–69 showed a raised f0 max on both subject and object position (S max:  $\beta_{60-69} = 0.14, p < 0.05$ ; O max:  $\beta_{60-69} = 0.14, p < 0.05$ ), and an expanded f0 range on the object position (range:  $\beta_{60-69} = 0.15, p < 0.05$ ). In contrast, Age Group 20–39 showed a lower f0 max on the subject (max:  $\beta_{20-39} = -0.12, p < 0.05$ ), and Age Group 40–49 showed a lower subject f0 max on the object (max:  $\beta_{40-49} = -0.15, p < 0.05$ ).

### 3.3.1.3. Standard Mandarin

**Initial subject focus.** Initial Focus significantly modulated the f0 range, min and max on both the subject and object positions. Compared to Broad Focus, the subject position under Initial Focus showed an expanded f0 range, mainly by a higher f0 max and lower f0 min (range:  $\beta_{IF} = 0.32, p < 0.001$ ; max:  $\beta_{IF} = 0.19, p < 0.001$ ; min:  $\beta_{IF} = -0.12, p < 0.01$ ). Within the object position, the f0 range was compressed, with a lowered f0 min and max (range:  $\beta_{IF} = -0.45, p < 0.001$ ; max:  $\beta_{IF} = -1.34, p < 0.001$ ; min:  $\beta_{IF} = -\text{range}, p < 0.001$ ).

**Final object focus.** Final Focus significantly modulated the f0 range, min and max on the subject position. Compared to Broad Focus, the subject position under Final Focus showed a compressed f0 range, mainly by a lower f0 max and a higher f0 min (range:  $\beta_{FF} = -0.25, p < 0.001$ ; max:  $\beta_{FF} = -0.15, p < 0.001$ ; min:  $\beta_{FF} = 0.10, p < 0.05$ ). No significant modulation was observed within the object position.

**Preceding consonants.** Extensive modulations were observed on the f0 min, max and range, varying

by onset consonants. Sonorant onsets showed an expanded f0 range mainly by a higher f0 max, on both the subject and object (S range:  $\beta_{sonorant} = 0.20, p < 0.001$ ; S max:  $\beta_{sonorant} = 0.13, p < 0.001$ ; O range:  $\beta_{sonorant} = 0.21, p < 0.05$ ; O max:  $\beta_{sonorant} = 0.34, p < 0.001$ ). Voiceless unaspirated onsets showed a compressed f0 range mainly by a lower f0 max, on both the subject and object (S range:  $\beta_{vcl.unasp} = -0.36, p < 0.001$ ; S max:  $\beta_{vcl.unasp} = -0.40, p < 0.001$ ; O range:  $\beta_{vcl.unasp} = -0.22, p < 0.01$ ; O max:  $\beta_{vcl.unasp} = -0.20, p < 0.001$ ). Voiceless aspirated onsets showed an expanded f0 range with higher f0 max and min on the subject position (range:  $\beta_{vcl.asp} = -0.22, p < 0.01$ ; max:  $\beta_{vcl.asp} = 0.44, p < 0.001$ ; min:  $\beta_{vcl.asp} = 0.22, p < 0.001$ ), and a lowered f0 max on the object position (max:  $\beta_{vcl.asp} = -0.13, p < 0.01$ ). Fricative onsets only showed a lower f0 max on the subject (max:  $\beta_{fricative} = -0.17, p < 0.001$ ).

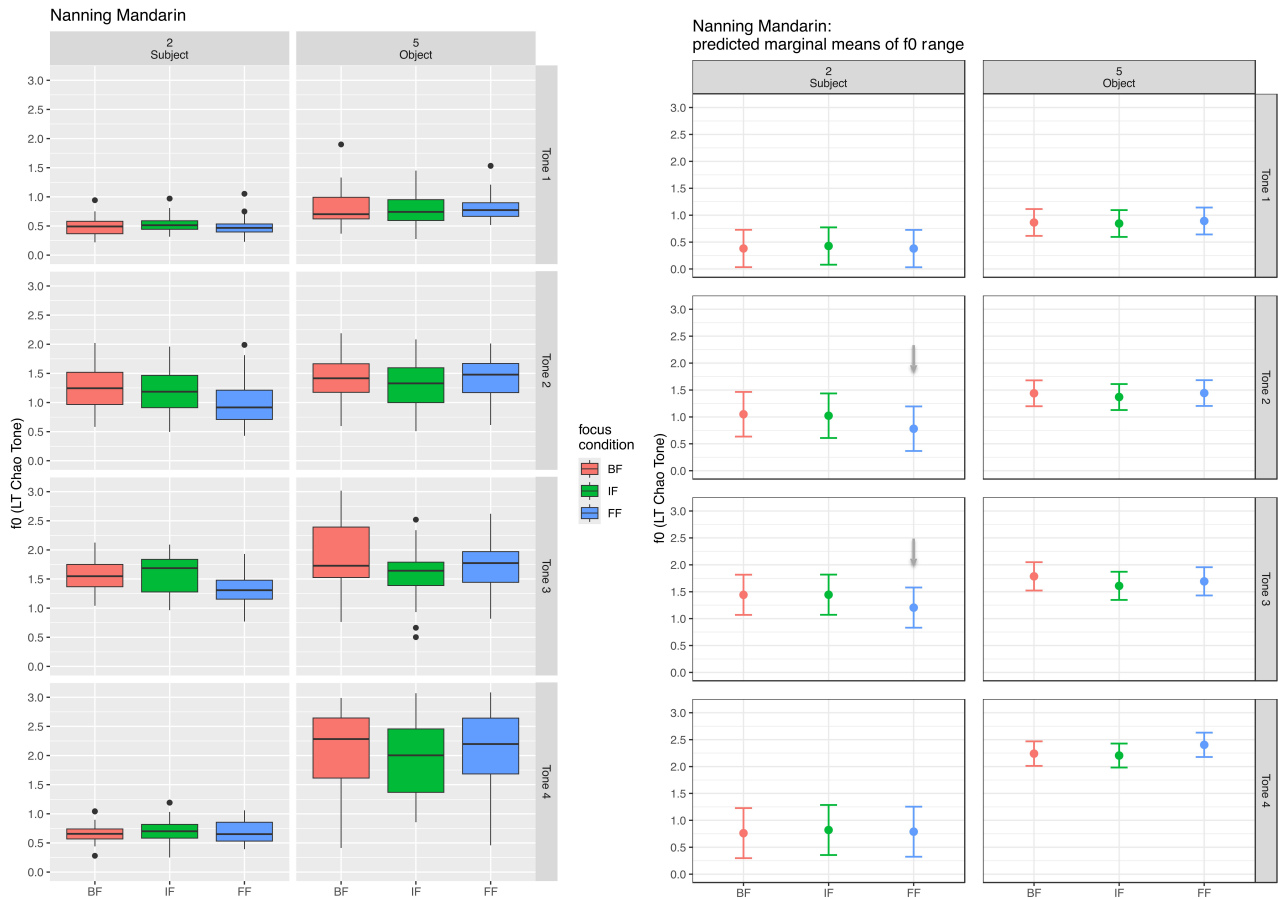
**Age group.** No significant effects of age groups were observed in Standard Mandarin.

### 3.3.2. Tone effects

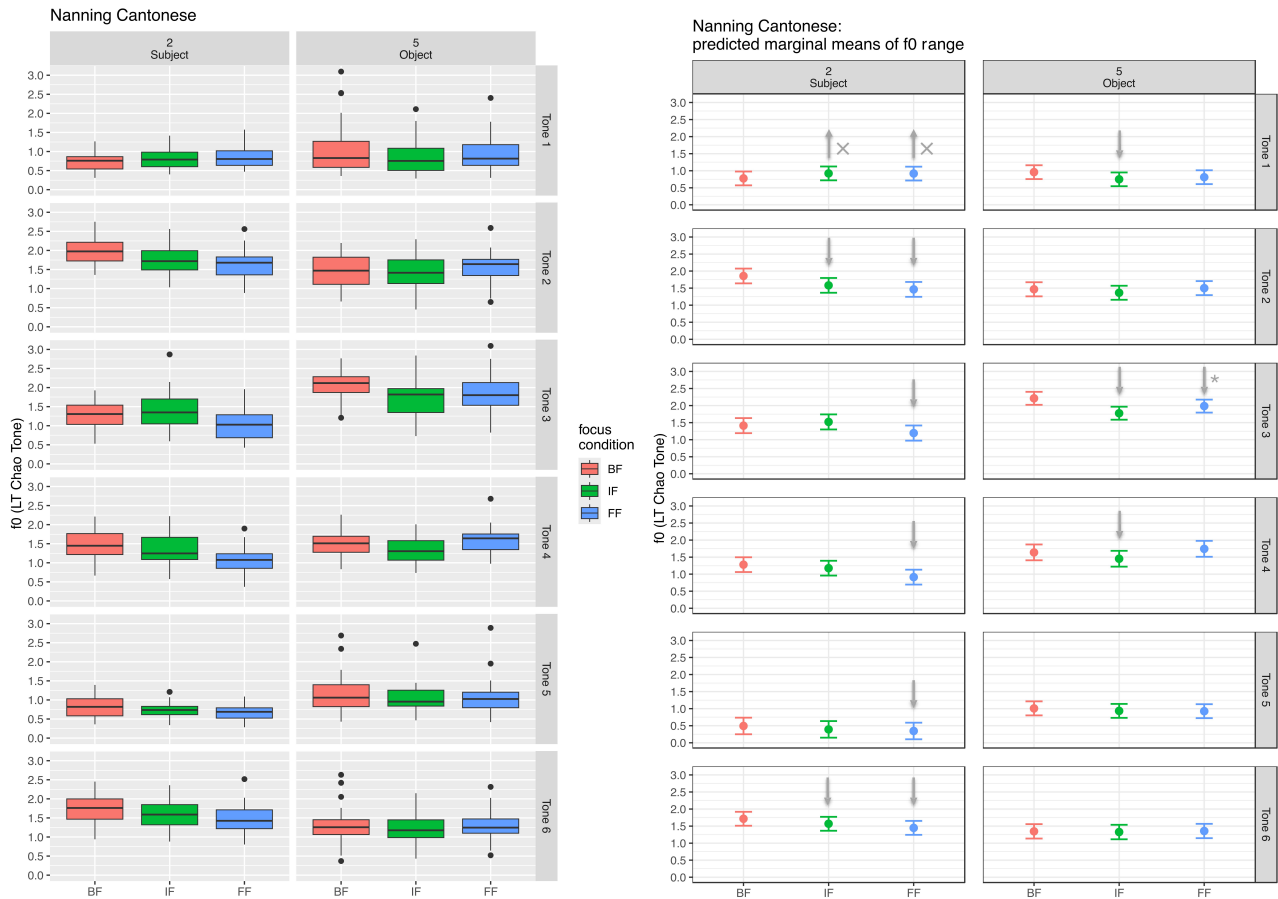
After examining the main focus effect in each language variety, we further investigate the interaction between focus condition and tone category. More specifically, we expect the effect of focus to differ across tone categories. Figures Figure 3.3.3–Figure 3.3.5 display the results for the average Chao tone f0 range in the subject and object, across each focus condition and phonological tone in Nanning Mandarin (Figure 3.3.3), Nanning Cantonese (Figure 3.3.4) and Standard Mandarin (Figure 3.3.5). Within each figure, the left plot displays the raw acoustics, with boxplots showing the means, 25th and 75th percentiles, maximum and minimum values, and some outliers. The right plot is the estimated marginal means predicted from the models; points represent the marginal means averaged over all levels of tone and other covariates, and error bars represent 95% confidence intervals. Arrows indicate the direction of significant f0 modulations compared with Broad Focus:  $\uparrow$  for expansion,  $\downarrow$  for compression. Crosses ( $\times$ ) and asterisks ( $*$ ) indicate the direction of significant f0 modulations compared with the main focus effects:  $\times$  indicates the opposite direction,  $*$  indicates the absence of main effects; tonal effects without specific notification were in the same direction as the main effects.

In general, the model predictions display similar patterns to the raw f0 range in LT Chao tone. At a glance, in Nanning Mandarin, the f0 range realisations under narrow focus conditions do not seem to substantially differentiate from Broad Focus. In contrast, Standard Mandarin clearly shows f0 range modulation with great magnitude.

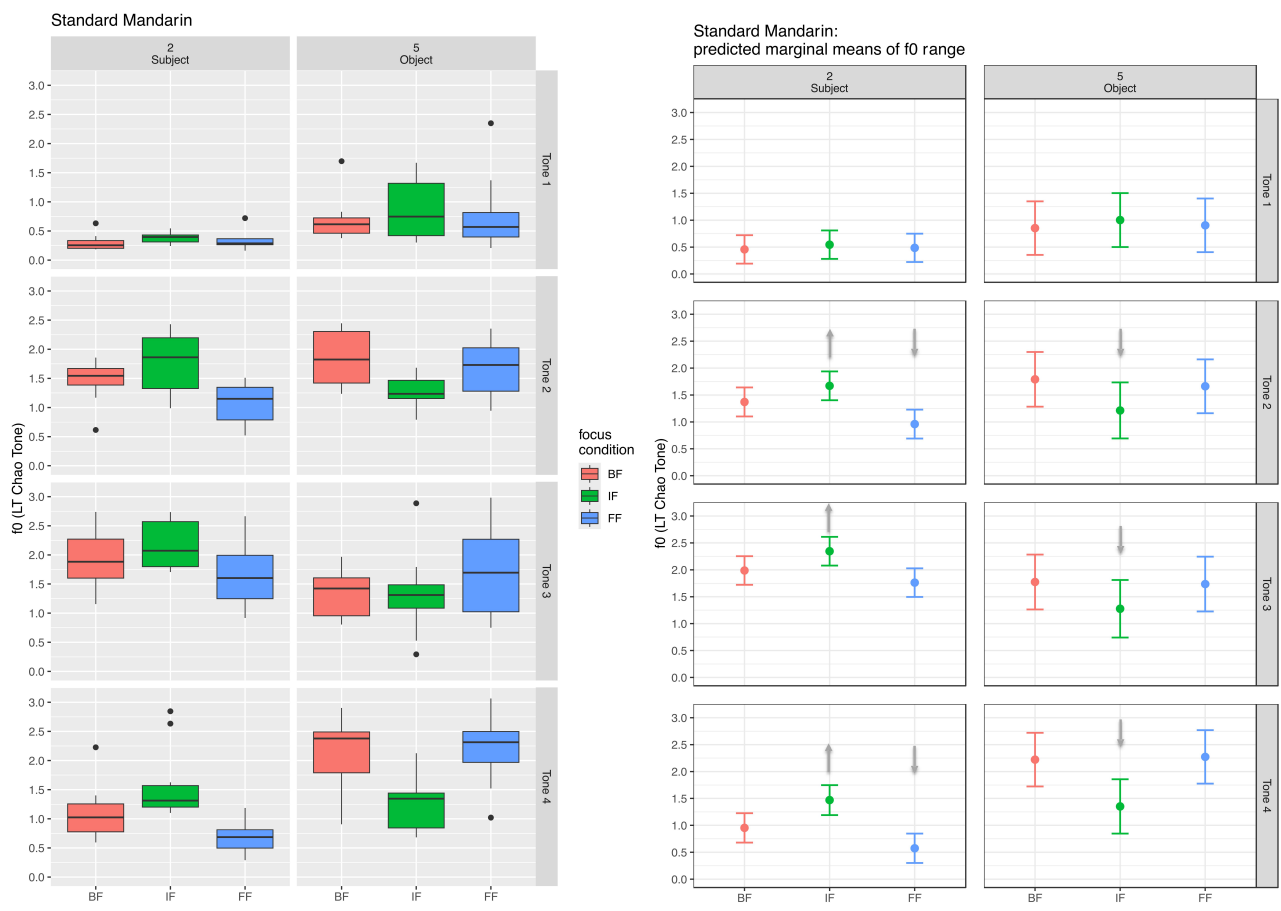
In this section, we will first present the results of the interaction between focus and tone in § 3.3.2.1. This interaction tests whether the focus effect (i.e., the difference between Initial/Final Focus and Broad Focus) varies across different tones. As the fixed effect coefficients contrast tone-specific effects against the average focus effect across all tones, interpreting the absolute effect size for any single tone becomes less direct. To provide a clearer and more intuitive understanding of the specific effects, we then conducted a post-hoc analysis using the *emmeans* package in § 3.3.2.2. This analysis illustrates the simple effect of focus by directly comparing the narrow and broad focus realisations within each specific tone. The models and results described in § 3.3.2.1 are provided in Appendix B.2.1, and the (codes) and results of post-hoc analysis are provided in Appendix B.2.2.



**Figure 3.3.3:** Average  $f_0$  range (in LT Chao tone) in Nanning Mandarin, across each focus condition and tone. Left plot is the raw acoustics; right plot is the model estimation.



**Figure 3.3.4:** Average  $f_0$  range (in LT Chao tone) in Nanning Cantonese, across each focus condition and tone. Left plot is the raw acoustics; right plot is the model estimation.



**Figure 3.3.5:** Average f0 range (in LT Chao tone) in Standard Mandarin, across each focus condition and tone. Left plot is the raw acoustics; right plot is the model estimation.

### 3.3.2.1. Differential focus effects: tone × meaning interactions

#### Nanning Mandarin

Though general tone differences were observed for these measures, interactions between tone and focus condition were observed in the f0 range and min only in the Final Focus condition and mainly within the subject position.

**Final object focus.** Within the subject position, NM Tone 2 and Tone 3 yielded a significantly narrower f0 range (range:  $\beta_{FF:Tone2} = -0.15, p < 0.001$ ;  $\beta_{FF:Tone3} = -0.12, p < 0.001$ ), enhancing the compression effect. In contrast, NM Tones 1 and 4 yielded a significantly wider f0 range than the broad focus condition (range:  $\beta_{FF:Tone1} = 0.12, p < 0.001$ ;  $\beta_{FF:Tone4} = 0.15, p < 0.001$ ), effectively cancelling out the general compression effect. Relatedly, NM Tone 2 and Tone 3 had a significantly higher f0 min (min:  $\beta_{FF:Tone2} = 0.17, p < 0.001$ ;  $\beta_{FF:Tone3} = 0.11, p < 0.01$ ), enhancing the raising effect. In contrast, NM Tones 1 and Tone 4 had a lower f0 min (min:  $\beta_{FF:Tone1} = -0.10, p < 0.01$ ;  $\beta_{FF:Tone4} = -0.19, p < 0.001$ ), effectively cancelling out the raising effect. Within the object position, NM Tone 4 had a significantly lower f0 min (min:  $\beta_{FF:Tone4} = -0.15, p < 0.05$ ), despite the absence of significant modulation in f0 min.

#### Nanning Cantonese

Interactions between tone and focus condition were extensively observed in f0 range, min and max. In both narrow focus conditions, tone yielded significant effects on f0 range, min and max on the subject, while it only modulated the f0 range and max on the object, not on the f0 min.

**Initial subject focus.** Within the subject position, NC Tone 2 yielded a significantly narrower f0 range (range:  $\beta_{IF:Tone2} = -0.21, p < 0.001$ ), enhancing the compression effect. In contrast, NC Tone 1 and Tone 3 had a significantly wider f0 range (range:  $\beta_{IF:Tone1} = 0.21, p < 0.001$ ;  $\beta_{IF:Tone3} = 0.17, p < 0.001$ ), overriding the general compression effect. Relatedly, NC Tone 2 yielded a higher f0 min and a lower f0 max (min:  $\beta_{IF:Tone2} = 0.10, p < 0.05$ ; max:  $\beta_{IF:Tone2} = -0.12, p < 0.01$ ); similarly, NC

Tone 6 also had a higher f0 min (min:  $\beta_{IF:Tone6} = 0.10, p < 0.05$ ). (These two tones thereby enhanced the raising effect of f0 min.) On the other hand, NC Tone 1 had a lower f0 min (min:  $\beta_{IF:Tone1} = -0.21, p < 0.001$ ), overriding the raising effect; NC Tone 3 had a higher f0 max (max:  $\beta_{IF:Tone3} = 0.19, p < 0.001$ ).

In the object position, NC Tone 3 yielded a significantly narrower f0 range with a lower f0 max (range:  $\beta_{IF:Tone3} = -0.27, p < 0.001$ ; max:  $\beta_{IF:Tone3} = -0.26, p < 0.001$ ), enhancing the compression effect. In contrast, NC Tone 6 yielded a significantly wider f0 range (range:  $\beta_{IF:Tone6} = 0.15, p < 0.05$ ), cancelling out the compression effect; nevertheless, it yielded a lower f0 max (max:  $\beta_{IF:Tone6} = -0.20, p < 0.001$ ). On the other hand, NC Tone 5 yielded a significantly higher f0 max (max:  $\beta_{IF:Tone5} = 0.11, p < 0.05$ ), cancelling out the lowering effect.

**Final object focus.** Within the subject position, NC Tone 1 yielded a significantly wider f0 range (range:  $\beta_{FF:Tone1} = 0.35, p < 0.001$ ), effectively overriding the compression effect. In contrast, NC Tone 2 and Tone 4 yielded a significantly narrower f0 range (range:  $\beta_{FF:Tone2} = -0.18, p < 0.001$ ;  $\beta_{FF:Tone4} = -0.16, p < 0.001$ ), enhancing the compression effect. Relatedly, NC Tone 1 had a lower f0 min; NC Tone 2 had a lower f0 max; NC Tone 4 had a higher f0 min (min:  $\beta_{FF:Tone1} = -0.31, p < 0.001$ ; max:  $\beta_{FF:Tone2} = -0.11, p < 0.05$ ; min:  $\beta_{FF:Tone4} = 0.11, p < 0.01$ ), enhancing the raising effect. Apart from these, NC Tone 3 and Tone 6 yielded a significantly higher f0 min (min:  $\beta_{FF:Tone3} = 0.13, p < 0.01$ ;  $\beta_{FF:Tone6} = 0.09, p < 0.05$ ), enhancing the raising effect. NC Tone 3 also had a higher f0 max (max:  $\beta_{FF:Tone3} = 0.12, p < 0.01$ ), while on the other hand, NC Tone 5 had a lower f0 min (min:  $\beta_{FF:Tone5} = -0.09, p < 0.05$ ).

Within the object position, NC Tone 3 yielded a significantly narrower f0 range, while NC Tone 4 yielded a significantly wider f0 range (range:  $\beta_{FF:Tone3} = -0.18, p < 0.01$ ;  $\beta_{FF:Tone4} = 0.16, p < 0.01$ ). In addition, NC Tone 1 and Tone 3 had a lower f0 max (max:  $\beta_{FF:Tone1} = -0.11, p < 0.05$ ;  $\beta_{FF:Tone3} = -0.12, p < 0.05$ ), effectively overriding the raising effect.

## Standard Mandarin

Interactions between tone and focus condition were observed in f0 range, min and max, mainly under Initial Focus; significant interactions under Final Focus were limited within SM Tone 1 .

**Initial subject focus.** Within the subject position, SM Tone 1 yielded a significantly narrower f0 range (range:  $\beta_{IF:Tone1} = -0.23, p < 0.01$ ), weakening the general expansion effect. In contrast, SM Tone 4 yielded a significantly wider f0 range (range:  $\beta_{IF:Tone4} = 0.20, p < 0.05$ ), enhancing the expansion effect. Relatedly, SM Tone 1 had significantly higher f0 min and max (min:  $\beta_{IF:Tone1} = 0.36, p < 0.001$ ; max:  $\beta_{IF:Tone1} = 0.13, p < 0.05$ ), effectively overriding the lowering effect of f0 min, but enhancing the raising effect of f0 max. In addition, SM Tone 3 yielded a significantly lower f0 min (min:  $\beta_{IF:Tone3} = -0.16, p < 0.05$ ), enhancing the lowering effect; SM Tone 2 yielded a significantly lower f0 max (max:  $\beta_{IF:Tone2} = -0.13, p < 0.05$ ), cancelling out the raising effect.

Within the object position, SM Tone 1 yielded a significantly wider f0 range (range:  $\beta_{IF:Tone1} = 0.60, p < 0.001$ ), effectively overriding the general compression effect. In contrast, SM Tone 4 yielded a significantly narrower f0 range (range:  $\beta_{IF:Tone4} = -0.42, p < 0.01$ ), enhancing the compression effect. Relatedly, SM Tone 1 had a lower f0 min (min:  $\beta_{IF:Tone1} = -0.52, p < 0.01$ ), while SM Tone 4 had a lower f0 max (max:  $\beta_{IF:Tone4} = -0.68, p < 0.001$ ), both enhancing the lowering effect. In addition, SM Tone 2 showed higher f0 min and max (min:  $\beta_{IF:Tone2} = 0.52, p < 0.001$ ; max:  $\beta_{IF:Tone2} = 0.39, p < 0.001$ ), and SM Tone 3 also showed a higher f0 max (max:  $\beta_{IF:Tone3} = 0.22, p < 0.05$ ), weakening the general lowering effects on f0 min and max.

**Final object focus.** Within the subject position, SM Tone 1 showed a significantly wider f0 range and lower f0 min (range:  $\beta_{FF:Tone1} = 0.28, p < 0.01$ ; min:  $\beta_{FF:Tone1} = -0.22, p < 0.01$ ), which cancelled out the compression effect of f0 range and weakened the raising effect of f0 min. No significant interactions were observed in the object position.

### 3.3.2.2. Post-hoc comparisons: EMMs results

For easier interpretation and more direct understanding of the tone-specific focus effects, this section will examine the simple effect of focus by directly comparing the narrow and broad focus

realisations within each specific tone. To assess the magnitude and direction of the focus effect, all contrast estimates ( $\Delta$ ) presented below are calculated as the difference between the Broad Focus condition and the respective narrow focus condition (i.e., BF–IF or BF–FF). A positive estimate ( $\Delta > 0$ ) indicates a compression effect in the f0 range or a lowering effect in the f0 min and max under narrow focus.

### **Nanning Mandarin**

Tonal variations on the focus effects were found to be limited in Nanning Mandarin. Final focus only modulated the f0 range and min of certain tones within the subject position, while Initial Focus did not yield any tone-specific f0 modulation.

**Final object focus.** Within the subject position, NM Tone 2 and Tone 3 both yielded a significantly narrower f0 range (range:  $\Delta_{BF-FF:Tone2} = 0.27, p < 0.001$ ;  $\beta_{BF-FF:Tone3} = 0.24, p < 0.001$ ), mainly related to the significantly higher f0 min (min:  $\Delta_{BF-FF:Tone2} = -0.33, p < 0.001$ ;  $\Delta_{BF-FF:Tone3} = -0.26, p < 0.001$ ). However, no f0 modulation reached significant in f0 max; NM Tone 1 and Tone 4 did not yield any significant f0 modulation. This suggests that the overall compression observed across all tones was mainly driven by the significant focus effects found in NM Tone 2 and Tone 3.

### **Nanning Cantonese**

In Nanning Cantonese, tone-specific focus effects were extensively observed in f0 range, min and max, mainly within the subject position.

**Initial subject focus.** Within the subject position, NC Tone 2 and Tone 6 both showed significant f0 range compression with the raising of f0 min (range:  $\Delta_{BF-IF:Tone2} = 0.27, p < 0.001$ ;  $\Delta_{BF-IF:Tone6} = 0.15, p < 0.01$ ; min:  $\Delta_{BF-IF:Tone2} = -0.14, p < 0.01$ ;  $\Delta_{BF-IF:Tone6} = -0.14, p < 0.01$ ). In addition, the f0 range compression of NC Tone 2 was also contributed by a lowered f0 max (max:  $\Delta_{BF-IF:Tone2} = 0.14, p < 0.05$ ). In contrast, NC Tone 1 had a significantly wider f0 range, mainly due to a lowered f0 min (range:  $\Delta_{BF-IF:Tone1} = -0.15, p < 0.05$ ; min:  $\Delta_{BF-IF:Tone1} = 0.17, p < 0.01$ ); both f0 adjustments

are in the opposite direction to the main focus effects. NC Tone 3 showed a significantly higher max (max:  $\Delta_{BF-IF:Tone3} = -0.17, p < 0.01$ ), but did not yield any significant adjustments in f0 range.

In the object position, the tone-specific f0 modulation displayed a more consistent pattern, compared to the subject position. NC Tones 1, 3 and 4 all yielded compression effect in f0 range, with the lowering of f0 max (range:  $\Delta_{BF-IF:Tone1} = 0.21, p < 0.01$ ;  $\Delta_{BF-IF:Tone3} = 0.44, p < 0.001$ ;  $\Delta_{BF-IF:Tone4} = 0.19, p < 0.05$ ; max:  $\Delta_{BF-IF:Tone1} = 0.21, p < 0.001$ ;  $\Delta_{BF-IF:Tone3} = 0.37, p < 0.001$ ;  $\Delta_{BF-IF:Tone4} = 0.15, p < 0.05$ ). NC Tone 2 and Tone 6 did not show any significant f0 adjustments, despite their focus effects in the subject position.

**Final object focus.** Final Focus yielded a pre-focus compression effect on most of the tones, but the object-related effects were limited. Within the subject position, all NC tones except for Tone 1 compressed the f0 range and raised the f0 min (range:  $\Delta_{BF-FF:Tone2} = 0.39, p < 0.01$ ;  $\Delta_{BF-FF:Tone3} = 0.22, p < 0.001$ ;  $\Delta_{BF-FF:Tone4} = 0.37, p < 0.001$ ;  $\Delta_{BF-FF:Tone5} = 0.15, p < 0.01$ ;  $\Delta_{BF-FF:Tone6} = 0.27, p < 0.001$ ; min:  $\Delta_{BF-FF:Tone2} = -0.30, p < 0.001$ ;  $\Delta_{BF-FF:Tone3} = -0.36, p < 0.001$ ;  $\Delta_{BF-FF:Tone4} = -0.34, p < 0.001$ ;  $\Delta_{BF-FF:Tone5} = -0.14, p < 0.01$ ;  $\Delta_{BF-FF:Tone6} = -0.32, p < 0.001$ ). In addition, NC Tone 3 also had a raised f0 max (max:  $\Delta_{BF-FF:Tone3} = -0.14, p < 0.001$ ). In contrast, NC Tone 1 yielded a significantly wider f0 range (range:  $\Delta_{BF-FF:Tone1} = -0.14, p < 0.05$ ), which, again, is in the opposite direction to the main compression effect.

Within the object position, only NC Tone 3 showed a narrower f0 range, mainly caused by a higher f0 min (range:  $\Delta_{BF-FF:Tone3} = 0.23, p < 0.01$ ; min:  $\Delta_{BF-FF:Tone3} = -0.16, p < 0.05$ ). In addition, NC Tone 4 had a higher f0 max, and NC Tone 5 had a higher f0 min (max:  $\Delta_{BF-FF:Tone4} = -0.14, p < 0.05$ ; min:  $\Delta_{BF-FF:Tone5} = -0.20, p < 0.01$ ); neither of these led to a significant change in f0 range.

## Standard Mandarin

Similar to Nanning Cantonese, Standard Mandarin was also found to have extensive tonal variation in f0 range, min and max, mostly within the subject position.

**Initial subject focus.** Within the subject position, all four tones were observed to yield significant f0 adjustments. Except for SM Tone 1, all the other tones expanded the on-focus f0 range (range:  $\Delta_{BF-IF:Tone2} = -0.30, p < 0.01$ ;  $\Delta_{BF-IF:Tone3} = -0.36, p < 0.001$ ;  $\Delta_{BF-IF:Tone4} = -0.52, p < 0.001$ ). Nevertheless, the motivations of expansion are various: SM Tones 2 and 3 lowered the f0 min, while SM Tone 4 raised the f0 max (min:  $\Delta_{BF-IF:Tone2} = 0.23, p < 0.05$ ;  $\Delta_{BF-IF:Tone3} = 0.28, p < 0.01$ ; max:  $\Delta_{BF-IF:Tone4} = -0.31, p < 0.001$ ). In addition, SM Tone 1 simultaneously raised the f0 min and max, without adjusting the f0 range (min:  $\Delta_{BF-IF:Tone1} = -0.25, p < 0.05$ ; max:  $\Delta_{BF-IF:Tone1} = -0.32, p < 0.001$ ).

Within the object position, all tones except for SM Tone 1 compressed the post-focus f0 range (range:  $\Delta_{BF-IF:Tone2} = 0.58, p < 0.01$ ;  $\Delta_{BF-IF:Tone3} = 0.50, p < 0.05$ ;  $\Delta_{BF-IF:Tone4} = 0.87, p < 0.001$ ). In terms of f0 min and max, all tones except for SM Tone 2 lowered the f0 min and max simultaneously, while SM Tone 2 only lowered the f0 max (min:  $\Delta_{BF-IF:Tone1} = 1.42, p < 0.001$ ;  $\Delta_{BF-IF:Tone3} = 0.63, p < 0.01$ ;  $\Delta_{BF-IF:Tone4} = 1.15, p < 0.001$ ; max:  $\Delta_{BF-IF:Tone1} = 1.27, p < 0.001$ ;  $\Delta_{BF-IF:Tone2} = 0.96, p < 0.001$ ;  $\Delta_{BF-IF:Tone3} = 1.13, p < 0.001$ ;  $\Delta_{BF-IF:Tone4} = 2.02, p < 0.001$ ).

**Final object focus.** Within the subject position, SM Tone 2 and Tone 4 both compressed the pre-focus f0 range by lowering the f0 max (range:  $\Delta_{BF-FF:Tone2} = 0.41, p < 0.001$ ;  $\Delta_{BF-FF:Tone4} = 0.38, p < 0.001$ ; max:  $\Delta_{BF-FF:Tone2} = 0.17, p < 0.05$ ;  $\Delta_{BF-FF:Tone4} = 0.24, p < 0.01$ ). Additionally, NM Tone 2 raised the f0 min (min:  $\Delta_{BF-FF:Tone2} = -0.24, p < 0.05$ ). However, no significant f0 adjustments were observed in SM Tone 1 and Tone 3, nor in the object position.

### 3.3.3. Cross-linguistic comparisons of main focus effect

The final analysis in this chapter addressed the question of whether the focus-marking pattern in Nanning Mandarin is more similar to that in Nanning Cantonese or in Standard Mandarin. The tables below show the significant interactions between language and focus condition, which indicate the cross-linguistic differences under narrow focus compared to broad focus. Cross-linguistic comparisons were conducted on six measures, i.e., f0 range, min and max on the subject and object

positions. Within each cell, significant levels are indicated by asterisks, and  $\beta$  coefficients represent the numerical distance between the source/target languages and Nanning Mandarin. Numerically closer distance to Nanning Mandarin is indicated by highlighted cells. The full set of models and their results are provided in Appendix B.2.3.

#### Nanning Cantonese vs. Nanning Mandarin

	S range	S max	S min	O range	O max	O min
<b>Initial Focus</b>	**	*		*		
<b>nlang1:nmeaning1</b>	-0.10	-0.08	0.01	-0.11	-0.09	0.02
<b>Final Focus</b>	**		*			
<b>nlang1:nmeaning2</b>	-0.10	-0.02	0.09	-0.07	0.01	0.08

#### Standard Mandarin vs. Nanning Mandarin

	S range	S max	S min	O range	O max	O min
<b>Initial Focus</b>	***	*	**	***	***	***
<b>nmeaning1:nlang2</b>	0.27	0.11	-0.16	-0.34	-1.34	-1.00
<b>Final Focus</b>	*	***				
<b>nmeaning2:nlang2</b>	-0.13	-0.19	-0.06	-0.04	-0.11	-0.08

**Table 3.3.1:** Differences of the main focus effect between Nanning Mandarin and its source (top) and target (bottom) languages, compared to Broad Focus. Significance levels:  $p < 0.001$ , \*\*\*,  $p < 0.01$ , \*\*,  $p < 0.05$ , \*.

#### 3.3.3.1. Nanning Cantonese vs. Nanning Mandarin

Differences between Nanning Cantonese and Nanning Mandarin were primarily within the subject position under both Initial and Final Focus; differences in object-related measures were limited (in f0

range under Initial Focus).

**Initial subject focus.** Significant differences were observed in f0 range and max. Within the subject position, Nanning Cantonese showed a narrower f0 range and lower f0 max than Nanning Mandarin, compared to the Broad Focus condition (range:  $\beta_{NC-NM:IF-BF} = -0.10$ ,  $p < 0.01$ ; max:  $\beta_{NC-NM:IF-BF} = -0.08$ ,  $p < 0.05$ ). This is essentially because Nanning Cantonese had f0 range compression without modulation on the f0 max, whereas Nanning Mandarin had f0 max raising without modulation on the f0 range (as mentioned in § 3.3.1; same for the following). Among object-related measures, differences were only observed in f0 range: although both language varieties compressed the f0 range under Initial Focus, Nanning Cantonese narrowed the f0 range even further (range:  $\beta_{NC-NM:IF-BF} = -0.11$ ,  $p < 0.05$ ).

**Final object focus.** Significant differences were observed in the subject f0 range and min, while object-related measures did not show any significant differences. Within the subject position, although both language varieties showed compressed f0 range and raised f0 min under Final Focus, Nanning Cantonese yielded an even narrower f0 range and higher f0 min than Nanning Mandarin (range:  $\beta_{NC-NM:FF-BF} = -0.10$ ,  $p < 0.01$ ; max:  $\beta_{NC-NM:FF-BF} = 0.09$ ,  $p < 0.05$ ).

### 3.3.3.2. Standard Mandarin vs. Nanning Mandarin

Differences between Nanning Mandarin and Standard Mandarin extensively exist in the Initial Focus condition; differences under Final Focus were limited within the subject position.

**Initial subject focus.** Significant differences were observed in all six measures, i.e., f0 range, min, and max on both the subject and object positions. Within the subject position, Standard Mandarin showed an expanded f0 range with a higher f0 max and a lower f0 min than Nanning Mandarin, compared to Broad Focus (range:  $\beta_{SM-NM:IF-BF} = 0.27$ ,  $p < 0.001$ ; max:  $\beta_{SM-NM:IF-BF} = 0.11$ ,  $p < 0.05$ ; min:  $\beta_{SM-NM:IF-BF} = -0.16$ ,  $p < 0.01$ ). This is mainly due to the f0 range expansion and f0 min lowering in Standard Mandarin under Initial Focus, which were absent in Nanning Mandarin. Nevertheless, although f0 max raising was observed in both language varieties, the magnitude in

Standard Mandarin was larger. Within the object position, Standard Mandarin showed an even narrower f0 range with lower f0 min and max than Nanning Mandarin, compared to Broad Focus (range:  $\beta_{SM-NM:IF-BF} = -0.34$ ,  $p < 0.001$ ; min:  $\beta_{SM-NM:IF-BF} = -1.00$ ,  $p < 0.001$ ; max:  $\beta_{SM-NM:IF-BF} = -1.34$ ,  $p < 0.001$ ). This is related to the earlier results, where under Initial Focus, Standard Mandarin showed f0 min lowering while Nanning Mandarin showed no modulation of f0 min. Nevertheless, although both language varieties had f0 range compression and f0 max lowering, the magnitude was larger in Standard Mandarin.

**Final object focus.** Significant differences were observed in subject f0 range and max, while object-related measures showed no significant differences. Within the subject position, Standard Mandarin yielded an even narrower f0 range and lower f0 max than Nanning Mandarin (range:  $\beta_{SM-NM:FF-BF} = -0.13$ ,  $p < 0.05$ ; max:  $\beta_{SM-NM:FF-BF} = -0.19$ ,  $p < 0.001$ ). Standard Mandarin further compressed the f0 range than Nanning Mandarin, although both language varieties showed f0 range compression under Final Focus; additionally, Standard Mandarin lowered the f0 max when Nanning Mandarin showed no significant modulation.

### 3.3.3.3. Numerical distance

To further compare numerical distances between Nanning Mandarin and its source/target languages, the absolute value of the  $\beta$  coefficient in each of the measures was also compared, regardless of significance. In general, Nanning Cantonese was numerically closer to Nanning Mandarin, with respect to the tested f0 measures. In only two measures, i.e., the subject f0 min and object f0 range under Final Focus, Standard Mandarin was numerically closer to Nanning Mandarin.

## 3.4. Discussions

This chapter examined focus realisations by manipulating f0 range, min and max in Nanning Mandarin, Nanning Cantonese and Standard Mandarin. Simple SVO target sentences were elicited by *wh*- questions, with focus positioned on the initial subject or the final object. The main effects of

Initial and Final focus were examined through pairwise comparisons with the Broad Focus condition within each variety. An additional cross-linguistic comparison assessed the similarities and differences in focus effects between Nanning Mandarin and its source and target languages. The tone-specific focus effects were also examined by comparing Initial and Final Focus to Broad Focus within each individual tone.

### 3.4.1. Focus effects

Table 3.4.1 summarises the significant main effects of focus in f0 range, min and max, which were presented in the results section. Arrows indicate the direction of significant f0 modulations compared with Broad Focus: upward arrow (↑) for expanded f0 range or higher f0 min and max, downward arrow (↓) for compressed f0 range or lower f0 min and max.

Language	Initial Focus		Final Focus	
	Subject [focused]	Object	Subject	Object [focus]
<b>Nanning Mandarin</b>	max↑	max↓	range↓ min↑	
<b>Nanning Cantonese</b>	range↓ min↑	range↓ max↓ min↑	range↓ min↑	max↑ min↑
<b>Standard Mandarin</b>	range↑ max↑ min↓	range↓ max↓ min↓	range↓ max↓ min↑	

**Table 3.4.1:** Summary of the main effects of the narrow focus in f0 range, min and max. ↑ indicates expanded f0 range or higher f0 min/max; ↓ indicates compressed f0 range or lower f0 min/max.

#### 3.4.1.1. Focus marking in the three varieties

At a glance, the main focus effects presented clear cross-linguistic variation. Among the three

varieties, Nanning Mandarin showed the most limited focus effects, especially in f0 range. In contrast, Nanning Cantonese and Standard Mandarin both manipulated f0 range, min and max more extensively, except at the final focus position (i.e., the object under Final Focus), although the two varieties employed divergent f0 patterns.

In Nanning Mandarin, only the pre-focus position showed a compression effect in f0 range by raising the f0 min; otherwise, no f0 range manipulation was observed. Nevertheless, Initial Focus yielded on-focus raising and post-focus lowering in f0 max, suggesting a potential local contrast in mean f0 contour similar to Standard Mandarin.

On the other hand, Nanning Cantonese always compressed the f0 range under both Initial and Final Focus, except for the final focus position; post-focus compression was thus absent. Within the final focus position, f0 min and max were both higher without adjusting the f0 range; this was absent in the two Mandarin varieties and suggests a possible on-focus raising of the mean f0. In addition, f0 min raising was found at all conditions, i.e., at the subject and object positions under both Initial and Final Focus, while f0 max modulation was only found in the objects.

The focus marking in Standard Mandarin generally aligns with the previous findings: results show clear on-focus expansion and out-of-focus compression in f0 range, all of which involved the manipulation of f0 min and max. The post-focus compression in f0 range, particularly, involved the lowering of both f0 min and max, suggesting the additional suppression in the mean f0 contour. The pre-focus f0 range compression displayed a smaller magnitude than the post-focus compression. Similar to Nanning Mandarin, the final focus position did not show any f0 modulation.

### **3.4.1.2. Cross-linguistic comparison**

If we compare the f0 modulation by focus condition across varieties, Initial Focus produced more cross-linguistically differentiated patterns, while Final Focus showed greater overall similarity. Under Initial Focus, Nanning Mandarin did not manipulate the f0 range, Nanning Cantonese consistently compressed the f0 range, while Standard Mandarin contrasted the on-focus and

out-of-focus f0 range. Furthermore, Nanning Mandarin adjusted only the f0 max, Nanning Cantonese relied more on adjusting the f0 min, and Standard Mandarin adjusted both f0 values. In contrast, Final Focus only yielded pre-focus f0 range compression without modulating the on-focus f0 range in all three varieties. Although the specific adjustments in f0 min and max varied across varieties, all three varieties raised the on-focus f0 min.

To account for the convergence under Final Focus, one possible explanation comes from the similar sentence-final f0 realisation under Broad and Final Focus in Standard Mandarin. Previous studies have noted this similarity, interpreting it either as final prominence of Broad Focus (Jin, 1996; F. Shi et al., 2009), less efficient encoding of Final Focus (F. Liu & Xu, 2005; Y. Xu, 2019), or a combination of both (Shi F. & Xia, 2022). These studies, however, offer different explanations. Jin (1996) relates the similarity to sentence-final prominence conditioned by semantic and syntactic structure under both Broad and Final Focus. Explanations within the less efficient encoding account also differ. Liu and Xu (2005) propose a functional competition in the final position between marking focus and distinguishing sentence types. Xu (2019) instead attributes this reduced efficiency to the lack of PFC, a critical focus signal, in Final Focus. Shi and Xia (2022), meanwhile, explained it in terms of physiological constraints, namely reduced airflow towards utterance-final positions.

However, these non-physiological accounts may not straightforwardly extend to Cantonese, where prominence is not reliably associated with f0 range expansion, question and statement intonation are signalled differently from Standard Mandarin (Y. Chen, 2022), and Initial Focus remains distinct from Broad Focus despite the lack of PFC. As for the sentence-initial differences between Broad and Final Focus, previous interpretations mainly analyse the narrower initial f0 range under Final Focus as out-of-focus compression. In contrast, Jin (1996) interprets Broad Focus itself as involving initial prominence. Cross-linguistic evidence is therefore needed to develop a more generalisable account of both sentence-final similarity and sentence-initial differences.

Despite the lack of a unified explanation, these studies consistently reported that Final Focus showed similar final f0 range but narrower initial f0 range relative to Broad Focus. One possible account is

that the sentence-final similarity between Broad and Final Focus constrains the remaining modifiable f<sub>0</sub> behaviour in the out-of-focus region under Final Focus. This account, nevertheless, may not fully explain the lack of f<sub>0</sub> range modulation in NM Initial Focus, suggesting potential contribution from other acoustic correlates or prosodic mechanisms.

Beyond the broad cross-linguistic similarities described above, taking Nanning Mandarin as the anchor for pairwise comparison reveals additional patterns of similarity with its source and target languages, even within the Initial Focus condition that exhibited more language-specific realisations. Under Initial Focus, Nanning Mandarin did not expand the on-focus f<sub>0</sub> range or compress the post-focus f<sub>0</sub> range; this absence of local f<sub>0</sub> range contrast resembles Nanning Cantonese. Particularly, the absence of PFC in the Nanning varieties resembles that of Mandarin by bilingual Min-Mandarin speakers ([Y. Chen, 2014](#); [Y. Chen et al., 2015](#); [Y. Xu et al., 2012](#)). At the same time, the presence of a local contrast in f<sub>0</sub> max is, to some extent, similar to Standard Mandarin. Moreover, both Mandarin varieties showed no modulation in f<sub>0</sub> range, min, or max at the final focus position.

However, the statistical comparisons in § 3.3.3 show that focus realisation in Nanning Mandarin differs significantly from that in both Nanning Cantonese and Standard Mandarin in many measures, despite its overall greater similarity to Nanning Cantonese. This indicates features of focus marking specific to Nanning Mandarin. Crucially, even in the cases where cross-linguistic similarities were observed (as discussed above), these similarities were only surface-level: the corresponding effects in Nanning Mandarin were significantly weaker. For example, although similarities were observed across all three varieties (e.g., pre-focus f<sub>0</sub> range compression) and across pairs (e.g., f<sub>0</sub> max contrast shared by Nanning Mandarin and Standard Mandarin, f<sub>0</sub> min raising shared by Nanning Mandarin and Nanning Cantonese), Nanning Mandarin realised these shared effects with significantly smaller magnitudes. Taken together, while Nanning Mandarin used the same f<sub>0</sub> measures as Nanning Cantonese and Standard Mandarin, it differs in both measure selection and effect magnitude. These findings support the view that Nanning Mandarin, while shaped by the influence from both its source and target languages, employs a distinct strategy in focus marking as an independent linguistic system.

Overall, the observed patterns suggest that Nanning Mandarin does not reliably employ  $f_0$  as a primary cue for focus marking. Cross-linguistically, the presence or absence of PFC appears to be close related to broader focus-marking strategies: Standard Mandarin employs local  $f_0$  contrast, realised through on-focus enhancement and post-focus compression, whereas Nanning Cantonese tends to maintain the global  $f_0$  pattern, resulting in the absence of PFC. These differences may be partly attributed to cue weighting, as discussed in Chapter 1: Standard Mandarin relies primarily on  $f_0$  to mark focus, whereas Cantonese makes greater use of duration and intensity, with  $f_0$  playing a less central role. In this respect, Nanning Mandarin shows greater similarity to Nanning Cantonese, with even more limited use of  $f_0$  in focus marking. To examine whether Nanning Mandarin exhibits a cue-weighting strategy similar to Nanning Cantonese, further investigation of other acoustic correlates is therefore needed.

In addition, it should be noted that the present analysis reflects overall patterns across speakers, without systematic modelling of speaker-generation or individual variation. While descriptive inspection of tonal data (Chapter 2) did not reveal clear age-related patterns, and individual variation appeared to be structured primarily by contour type, further quantitative analysis would be required to assess whether such variation plays a role in focus-marking patterns in Nanning Mandarin. Given that Nanning Mandarin is undergoing contact-induced change, and that the transferability of PFC has been linked to L2 experience and proficiency (as reviewed in Chapter 1), variation across age groups may provide useful insight into the development of intonation. Moreover, considering the unstable nature of Nanning Mandarin as a Local Mandarin variety, together with the multilingual environment and the diversity of speakers' language backgrounds, a relatively high degree of individual variation is to be expected. Future work incorporating speaker-level modelling and apparent-time analysis would be necessary to further clarify these patterns.

### **3.4.2. Tone effects**

This chapter examines the tone-specific focus realisation in two ways: one is examining the tonal variation compared to the averaged tone, the other is examining the simple effects of the narrow

focus effects within each specific tone. For a more direct and meaningful interpretation of the tone-specific focus effects, only the results of the second approach will be discussed here.

Tone	Initial Focus		Final Focus	
	Subject [focused]	Object	Subject	Object [focused]
NM T1				
NM T2			range↓ min↑	
NM T3			range↓ min↑	
NM T4				
NC T1	<u>range↑</u> <u>min↓</u>	range↓ max↓	<u>range↑</u>	
NC T2	range↓ min↑ max↓*		range↓ min↑	
NC T3	max↓*	range↓ max↓	range↓ min↑ max↑*	range↓* min↑
NC T4		range↓ max↓	range↓ min↑	max↓
NC T5			range↓ min↑	min↑
NC T6	range↓ min↑		range↓ min↑	
SM T1	<u>min↑</u> max↑	min↓ max↓		
SM T2	range↑ min↓	range↓ max↓	range↓ min↑ max↓	
SM T3	range↑ min↓	range↓ min↓ max↓		
SM T4	range↑ max↑	range↓ min↓ max↓	range↓ max↓	

**Table 3.4.2:** Summary of the tone-specific focus effects in f0 range, min and max. ↑ indicates expanded f0 range or higher f0 min/max; ↓ indicates compressed f0 range or lower f0 min/max.

Table 3.4.2 summarises the f0 modulation under the Initial and Final Focus conditions across each tone, which has been presented in § 3.3.2. Arrows indicate the direction of significant f0 modulations compared with Broad Focus: upward arrow (↑) for expanded f0 range or higher f0 min and max, downward arrow (↓) for compressed f0 range or lower f0 min and max. Underlining and asterisks indicate the direction of significant f0 modulations compared with the main focus effects: underlining for the opposite direction, asterisks indicate the absence of main effects; tonal effects

without specific notification were in the same direction as the main effects. The examination of tone-specific focus effects reveals further evidence of different patterns across the three language varieties. At a glance, the final focus position generally displayed minimal focus effects across all tones, consistent with the main effects. Among the three varieties, Nanning Mandarin displayed the most restricted tonal-specific focus effects, mirroring its overall main effect pattern; Nanning Cantonese showed the greatest tonal variation.

Nanning Mandarin (NM) tones yielded even more restricted focus effects than the main focus effects, which was only observed within the pre-focus region. Under Initial Focus, the significant main effect in  $f_0$  max disappeared at the tone level, as none of the individual tones showed a corresponding modulation. For Final Focus, the main effect of pre-focus compression was driven only by the mid-onset dynamic tones (NM Tone 2 [34] and Tone 3 [42]). This parallels Standard Mandarin to some extent, where pre-focus compression was observed in dynamic tones; however, NM Tone 4 (falling) was not involved in this pattern, despite its contour being most similar to SM Tone 4. A plausible explanation is the need to maintain contrast between the two NM falling tones: applying  $f_0$  range compression to NM Tone 4 [53] may reduce its distinctiveness relative to NM Tone 3 [42]. This issue does not arise in Standard Mandarin, which has only one falling tone.

In Nanning Cantonese (NC), the tonal contributions to the main effect of Initial Focus were dispersed across different tones; in contrast, almost every tone contributed to the main effect of Final Focus, except for NC Tone 1. Although the main effect of Initial Focus showed a symmetric compression effect of the  $f_0$  range in both subject and object positions, the tone-specific analysis revealed that compression was present in only one of the two positions. The only tone showing a symmetric effect on the  $f_0$  range was NC Tone 3 under Final Focus, yet its on-focus compression did not appear in the corresponding main effect.

NC Tone 1 was the most atypical among the six tones. It was the only tone that exhibited focus effects in the opposite direction to the main effects under both narrow focus conditions, though only within the subject position. Under Initial Focus, NC Tone 1 also showed an asymmetric effect that

enhanced the local  $f_0$  contrast, resembling the pattern observed in Standard Mandarin. As Tone 1 was the only Nanning Cantonese tone that yielded on-focus  $f_0$  range expansion, our findings reverse the previously reported pattern in standard Cantonese, where all tones except Tone 1 yielded on-focus expansion in some or all focus positions (Gu & Lee, 2009; W. L. Wu & Xu, 2010). Finally, it was rarely the case that a tone raised or lowered both  $f_0$  min and  $f_0$  max within the same position, suggesting that  $f_0$  mean shifts may not be a dominant cue.

In Standard Mandarin, all four tones contributed to the on-focus expansion effect under Initial Focus. Previous literature reported that the on-focus expansion was greatest on SM Tone 4, followed by Tone 2, while Tone 1 exhibited the weakest effect, attributed to different mechanisms (Y. Chen & Gussenhoven, 2008). Our current finding is largely consistent with this previous finding: SM Tone 4 showed the greatest on-focus expansion, followed by Tone 3 and Tone 2, while Tone 1 did not modulate the  $f_0$  range. However, the observed tone-dependent mechanisms differ from those reported by Chen and Gussenhoven (2008). Despite the absence of  $f_0$  range manipulation under Initial Focus, SM Tone 1 did show on-focus raising and post-focus lowering of both  $f_0$  min and  $f_0$  max. This pattern potentially suggests the raising and lowering of the mean  $f_0$  contour, indicating that  $f_0$  range alone may not fully capture focus-related modulation in level tones such as SM Tone 1. Under Final Focus, only the dynamic tones contributed to the main effect of pre-focus compression. One possible account may be that the level tones lacked the necessary space for range compression.

Building on the cross-varietal contour similarities and template selection process of lexical tones discussed in Chapter 2, the following analysis examines focus effects within groups of tones that share comparable contour shapes. Within the level tones, NM Tone 1 and its tonal templates (NC Tone 5 and SM Tone 1) commonly showed (overall) minimal  $f_0$  modulation. By comparison, NC Tone 1, which has a higher pitch register than all three, showed an effect that ran in the opposite direction from the main effect within the subject. This discrepancy may be related to the preceding tone. In the target sentences, the name prefix before the subject always carried the lower level tone (NM Tone 1, NC Tone 5 and SM Tone 1). Consequently, subjects carrying NM Tone 1, NC Tone 5 or SM Tone 1 were always preceded by the same tone, whereas subjects bearing NC Tone 1 were not.

Given that the preceding NC Tone 5 is lower than NC Tone 1, a carryover effect may have widened the f0 range of NC Tone 1 in the subject position. Notably, in both Nanning Cantonese and Standard Mandarin, only the highest level tone registers displayed focus effects in the opposite direction of the main effects, again in the subject position. It remains unclear whether high-register tones are more susceptible to such exceptions.

The rising tones in the three varieties all share mid-rising contours and showed broadly similar f0 modulation patterns, especially under Final Focus. Under Initial Focus, SM Tone 2 showed the most extensive f0 adjustment (in both subject and object positions), followed by NC Tone 2 (mainly in the object position), while NM Tone 2 showed no observable modulation. The falling tones displayed a similar cross-varietal hierarchy. Specifically, the mid-falling NM Tone 3 yielded the same focus effects as its template tones (NC Tones 2 and 6) under Final Focus but diverged under Initial Focus, where no f0 adjustment was observed. The high-falling NM Tone 4 diverged substantially from its template (SM Tone 4) in most positions, except in the final focus position, where an overall convergent pattern was found across varieties and tones.

Taken together, across the three language varieties, mid dynamic tones generally exhibited the most extensive f0 modulation, whereas high-level tones tended to show the least modulation or the most exceptional patterns. Nevertheless, the specific manifestation of these tendencies remained language-dependent. A detailed contour-based analysis is necessary to further examine the interaction between intonational focus and tone.

## **Chapter 4: Concluding remarks**

At the beginning of this thesis, research questions were proposed regarding lexical tone, intonational focus and their interaction in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin. Specifically:

**RQ 1:** How are the lexical tone systems of Nanning Mandarin, Nanning Cantonese, and Standard Mandarin acoustically realised and compared?

**RQ 2:** How are intonational focus in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin acoustically realised and compared?

**RQ 3:** How do the patterns of tone–intonation interaction in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin manifest acoustically, and how do they compare?

In the final section of this thesis, each of these research questions will be addressed with summary of findings, further discussions and concluding remarks.

### **4.1. Summary of the key findings**

#### **4.1.1. RQ 1: hybrid tone system with greater source influence in Nanning Mandarin**

**RQ 1:** How are the lexical tone systems of Nanning Mandarin, Nanning Cantonese, and Standard Mandarin acoustically realised and compared?

To address RQ 1, a production study was conducted to investigate the citation tones from monosyllables (Chapter 2). The tone study of this thesis comprises the acoustic-phonetic analyses of lexical tone systems in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin. This study is the first large-scale attempt to acoustically examine the tones of the two low-resourced Nanning

varieties; the results provided the contour shapes of the tone inventories and updated the Chao tone values. According to the acoustic measures and GAMMs results, Nanning Mandarin tones, although inheriting the mapping between tone category and contour from Standard Mandarin, showed greater similarities to Nanning Cantonese in f<sub>0</sub> space, tonal contrast, and individual tone contours. Similar to Nanning Cantonese, Nanning Mandarin tones occupied a narrower f<sub>0</sub> space than Standard Mandarin, and employed a High–Mid–Low tonal contrast, while Standard Mandarin presented a High–Mid two-level contrast. Moreover, cross-linguistic contour comparisons showed that individual Nanning Mandarin tones primarily resemble Nanning Cantonese tones, except for NM Tone 4, which was most similar to corresponding Standard Mandarin tones. A hypothesis about tone formation in Nanning Mandarin is therefore proposed: taking the corresponding Standard Mandarin tone as the target of approximation, Nanning Mandarin tones tend to align more closely with the contours in the source language (Nanning Cantonese) that was most similar to the target tone in the target language (Standard Mandarin), while Standard Mandarin only serves as an additional source of templates. We concluded that Nanning Mandarin possessed a hybrid tone system with greater influence from its source language.

#### **4.1.2. RQ 2 and 3: unique focus marking with restricted f<sub>0</sub> modulation in Nanning Mandarin**

**RQ 2:** How are intonational focus in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin acoustically realised and compared?

To address RQ 2, a production study was conducted to investigate the f<sub>0</sub> realisations of focus marking from simple SVO sentences (Chapter 3). The intonation study of this thesis examined the effects of Initial and Final Focus on f<sub>0</sub> range, min and max in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin. Acoustic measures and statistical modelling show divergent patterns under Initial Focus across the three varieties, while Final Focus yielded more cross-linguistically convergent patterns. Standard Mandarin extensively manipulated f<sub>0</sub> range, min and max, particularly by expanding the non-final on-focus f<sub>0</sub> range and compressing the out-of-focus

f0 range, which is in line with previous findings. Nanning Cantonese, though also made use of f0 in focus marking, consistently compressed the f0 range except in the final focus position. This resulted in the lack of local f0 contrast, including post-focus compression. More crucially, Nanning Mandarin displayed focus marking patterns featuring highly restricted f0 manipulation, which were distinct from its source and target languages; post-focus compression was also absent. The uniqueness of focus marking in Nanning Mandarin was further confirmed by pairwise comparisons: although greater similarities to Nanning Cantonese were observed across the acoustic measures, Nanning Mandarin were significantly different from both its source and target languages in many measures.

**RQ 3:** How do the patterns of tone–intonation interaction in Nanning Mandarin, Nanning Cantonese, and Standard Mandarin manifest acoustically, and how do they compare?

To address RQ 3, the intonation study also investigated the tone-dependent f0 adjustments in focus marking across the three varieties (Chapter 3). In general, the tone-specific focus effects mirrored the main focus effects, with more extensive f0 modulation in the mid dynamic tones and least in the high-level tones. However, specific patterns of tonal variations were largely language-dependent. Nanning Mandarin, again, presented minimal f0 modulation by individual tones; Nanning Cantonese displayed great tonal variation under Initial Focus and more consistent patterns under Final Focus; whereas Standard Mandarin tones in general evenly contributed to the main Initial Focus effects. Despite the broad and partial commonality observed, Nanning Mandarin tones did not seem to yield focus effects similar to their tonal templates.

In conclusion, focus realisations in f0 were found to be different across the three varieties, although the relatively convergent pattern under Final Focus may be explained by the cross-linguistically limited on-focus modulation. Regarding Nanning Mandarin, focus yielded particularly restricted effects, distinct from Nanning Cantonese and Standard Mandarin.

## **4.2. Further discussions: interpretation and theoretical implications**

### **4.2.1. Limitations of synchronic data**

The present analysis is based on synchronic acoustic data on both tone and intonation, which reveal the outcomes of language contact but do not directly determine the cognitive, perceptual, or diachronic mechanisms underlying the formation of Nanning Mandarin prosody.

Regarding the tone formation hypothesis, the current account is framed in terms of template alignment based on cross-variety acoustic correspondence. The notion of template alignment, however, should not be taken to imply a deliberate selection process. Rather, the observed patterns may partly reflect perceptual assimilation between unfamiliar and familiar tone categories. Under this interpretation, unfamiliar Standard Mandarin tone categories may be perceptually mapped onto the closest available Nanning Cantonese tone categories, giving rise to the pattern of template alignment observed in Nanning Mandarin. However, this perceptual account alone may not fully explain cases in which a Nanning Mandarin tone shows greater similarity to the corresponding Standard Mandarin tone, suggesting that additional factors, such as the need to maintain tonal contrast, may also play a role. More broadly, the formation of Nanning Mandarin necessarily involves diachronic development across generations. The mechanisms underlying this development therefore cannot be fully established from synchronic acoustic data alone.

Similar limitations of synchronic inference extend to intonation patterns, as intonational realisations in Nanning Mandarin appear even more complex. The current results on intonational focus do not display a straightforward pattern that can be readily attributed to a contact mechanism in the same way as lexical tones.

Nevertheless, further apparent-time analysis of the current data may provide insights into how language contact has shaped Nanning Mandarin prosody over time. For example, future analysis may examine whether older speakers produce tones that are more similar to Nanning Cantonese, while younger speakers show greater similarity to Standard Mandarin. To better understand the formation of prosodic patterns in Nanning Mandarin and other such intermediate varieties, further studies may

start with apparent-time comparisons across age groups, and extend to a wider range of Local Mandarin and intermediate varieties, perception and acquisition studies, as well as longitudinal data.

#### **4.2.2. Focus-marking strategy and prosodic variation in language contact**

As discussed in Chapter 3, clear cross-linguistic differences in focus marking were observed: Nanning Cantonese tends to maintain the global  $f_0$  pattern without reliable use of  $f_0$ , whereas Standard Mandarin employs  $f_0$  as the main acoustic correlate to create local focus contrasts. The mechanisms underlying these differences may involve multiple factors. One potential factor is cue weighting, as discussed in § 3.4.1.2, whereby Standard Mandarin relies primarily on  $f_0$ , while Cantonese makes greater use of duration and intensity. Another factor may lie in tone–intonation interaction. Previous studies on question intonation show that Cantonese allows greater modification of tonal contours to realise intonation, whereas Standard Mandarin tends to preserve tone identity, resulting in weaker intonational signalling (Y. Xu, 2019; J. Zhang et al., 2020). This suggests that the weighting between tonal and intonational contrasts in  $f_0$  realisation may vary not only across languages, but also across different prosodic functions within the same language. Extending this reasoning, in focus marking, Cantonese may instead prioritise the preservation of tonal contours, while Standard Mandarin allows greater modification of  $f_0$  to encode focus.

Turning to Nanning Mandarin, the results show that it does not reliably employ  $f_0$  to mark focus, in this respect resembling Nanning Cantonese. In particular, the absence or weakening of post-focus compression (PFC) in Nanning Mandarin is consistent with observations in language contact settings reviewed in Chapter 1, and is especially comparable to Taiwan Mandarin. This pattern lends further support to the view that PFC is difficult to transfer across languages (Y. Chen, 2015). Cross-linguistic comparison further shows that focus-making patterns in Nanning Mandarin share greater similarity with those of Nanning Cantonese, while still exhibiting limited similarities to Standard Mandarin. Within a contact continuum, this places Nanning Mandarin closer to Nanning Cantonese in terms of focus-marking strategy. Whether the underlying mechanisms are also similar, however, requires further investigation, including analyses of additional acoustic correlates, detailed

f0 contour examination, and perceptual evidence from both Nanning Mandarin and Nanning Cantonese.

Despite these surface-level similarities, focus-marking patterns in Nanning Mandarin remain statistically distinct from both its source and target languages. Specifically, this distinctiveness is reflected in its highly limited focus effects, in terms of both measure selection and effect magnitude. Compared even with Taiwan Mandarin, another Local Mandarin variety, Nanning Mandarin still displays notable differences, such as the absence of on-focus expansion. This suggests that contact-induced varieties require case-specific analysis.

To account for the limited focus effects in Nanning Mandarin, one possible explanation is that conflicting strategies from the source and target languages lead to a partial neutralisation of prosodic effects. However, unlike the relatively transparent template-alignment pattern observed in lexical tones, focus marking shows a less straightforward pattern of alignment. It is possible that the cross-linguistic differences already present at the tonal level further amplify the divergence at the intonational level, such that, despite the overall greater similarity to Nanning Cantonese, the pattern of alignment becomes less clear in intonation.

Furthermore, building on its hybrid tonal system and distinct focus-marking strategy, Nanning Mandarin also displays patterns of tone–intonation interaction that differ from both its source and target languages. Notably, the overall focus effects in Nanning Mandarin are already limited, and tone-specific focus effects are even more restricted. One possibility is that reduced focus marking constrains the extent to which tone-specific focus effects can be realised. This reduction at the intonational and tone–intonation interaction levels recalls the levelling mechanisms discussed in Chapter 1. However, unlike the convergence between varieties in contact proposed in classic accounts of dialect levelling (e.g., Trudgill, 1986), the reduction in Nanning Mandarin result in prosodic patterns that are distinct from both. Therefore, the underlying mechanisms behind the formation of Nanning Mandarin prosody remain to be further explored.

Taken together, these observations suggest that as more prosodic dimensions are involved, Nanning Mandarin exhibits increasingly distinct patterns as an independent linguistic system. While tonal patterns may still show relatively clear links to the source and target languages, intonational patterns display greater divergence. Given its hybrid tonal system and distinct focus-marking strategy, the mechanisms underlying tone–intonation interaction in Nanning Mandarin may also differ from those in either Nanning Cantonese or Standard Mandarin. These patterns do not necessarily imply that the mechanisms are entirely language-specific. Rather, they point to the need for a more comprehensive account that can accommodate cross-linguistic variation while still capturing shared mechanisms underlying tone–intonation interaction.

#### **4.2.3. Tone–intonation interaction: additive model?**

The empirical fact that tones with similar contours can exhibit distinct focus realisations provides evidence for the relative independence of tone and intonation. Even tones with not significantly different contours do not necessarily lead to similar focus realisations. This divergence is primarily due to two factors. Firstly, focus realisation is highly language-specific. For example, NM Tone 3 and its tonal templates (NC Tones 2 and 6) were not significantly different in contour, but yielded largely different  $f_0$  realisations. Secondly, even within a single variety, similar contours do not guarantee identical focus effects. For instance, NC Tones 2 and 6 are acoustically fully merged and thus share acoustically the same tonal context in the target sentences, yet their focus realisations are not entirely identical. Although general commonalities were observed among specific tone contour groups and suggest some contour constraints, whenever there is room for tonal variation, the focus effects yielded by individual tones will largely be language-specific, in aspects such as which tones are affected and through which acoustic measures.

Consequently, when viewed longitudinally, the surface  $f_0$  realisation of an utterance is the result of the combined action of at least three interacting sources: shared contour constraints, language-specific tone inventory features, and highly variable intonational adjustments (language- and tone-dependent). However, it remains unclear whether the manner of this interaction aligns with

the simple additive models previously proposed by Chao (1933). An additive model should predict consistent focus effects across tones within a language variety, but our data show clear tonal variation. Given the continuous nature of connected speech, a non-local model taking horizontal influences into consideration is required to definitively confirm how tone and intonation specifically contribute to the surface f<sub>0</sub> realisation. Therefore, a more fine-grained contour analysis is necessary with examination of tonal contexts and coarticulation.

### **4.3. Conclusion**

This thesis investigated the production of lexical tones and intonational focus of a non-standard intermediate Mandarin variant, namely Nanning Mandarin, compared to its source language, Nanning Cantonese, and target language, Standard Mandarin. The tone study suggested that Nanning Mandarin, although largely inheriting the mapping between tone category and contour, showed substantially greater similarities to Nanning Cantonese tones. We therefore proposed a hypothetical tone formation process to account for the emergence of Nanning Mandarin tones, where Nanning Cantonese tones were primarily selected as the tonal templates. The intonation study revealed cross-linguistically different patterns of focus realisation by f<sub>0</sub> range, min and max. Although Nanning Mandarin resembled Nanning Cantonese more than Standard Mandarin in focus marking patterns, distinct and restricted focus effects were observed in most places. These findings all suggest that Nanning Mandarin has developed unique prosodic features beyond the influence of its source and target languages. Tonal variation in focus realisation further support the relatively independent relationship between tone and intonation; the exact mechanism of their interaction requires more fine-grained examination of f<sub>0</sub> contour in continuous speech.

## Appendix A: Tone study

### A.1. Speech materials and experiments

Below are the wordlists of 72 monosyllabic words and 72 disyllabic words. The monosyllabic words cover all four Mandarin tone categories (M1–M4) and all six Cantonese tone categories (C1–C4), and the disyllabic words cover all 16 tone combinations in Mandarin and all 36 tone combinations in Cantonese. Mandarin and Cantonese tones are historically corresponded and numbered according to the systematic correspondence. The cross-linguistic tone mapping in this wordlist reflect the main corresponding relationship displayed in Table 2.1.1. IPA transcriptions for Mandarin were adapted from Lee and Zee’s (2003) Standard Mandarin transcriptions; for Nanning Cantonese, IPA transcriptions were adapted from Lin and Qin (2008).

#### A.1.1. Monosyllabic wordlists

##### Practice phase

Character	Nanning Mandarin and Standard Mandarin		Nanning Cantonese		English gloss	Disyllabic context
	IPA	Tone	IPA	Tone		
书	ʃu	M1	ʃy	C1	<i>book</i>	(读书)
江	teiaŋ	M1	kəŋ	C1	<i>river</i>	(江水)
记	tei	M4	ki	C5	<i>note</i>	(记得)
犯	fan	M4	fan	C6	<i>commit</i>	(犯错)
水	ʃuei	M3	ʃui	C3	<i>water</i>	(水果)

## Formal phase

Character	Nanning Mandarin and Standard Mandarin		Nanning Cantonese		English gloss	Disyllabic context
	IPA	Tone	IPA	Tone		
翻	fan	M1	fan	C1	<i>turn over</i>	(翻身)
夫	fu	M1	fu	C1	<i>man</i>	(夫人)
姑	ku	M1	ku	C1	<i>aunt (father's sister)</i>	(姑妈)
瓜	kua	M1	k <sup>w</sup> a	C1	<i>melon</i>	(西瓜)
夸	k <sup>h</sup> ua	M1	k <sup>wh</sup> a	C1	<i>boast</i>	(夸奖)
巴	pa	M1	pa	C1	<i>adhere</i>	(大巴)
班	pan	M1	pan	C1	<i>class</i>	(上班)
披	p <sup>h</sup> i	M1	p <sup>h</sup> i	C1	<i>drape</i>	(披风)
他	t <sup>h</sup> a	M1	t <sup>h</sup> a	C1	<i>he</i>	(其他)
单	tan	M1	tan	C1	<i>single</i>	(单独)
摊	t <sup>h</sup> an	M1	t <sup>h</sup> an	C1	<i>spread out</i>	(摆摊)
妈	ma	M1	ma	C1	<i>mother</i>	(妈妈)
烦	fan	M2	fan	C2	<i>annoyed</i>	(烦人)
扶	fu	M2	fu	C2	<i>support with hand</i>	(扶手)
爬	p <sup>h</sup> a	M2	p <sup>h</sup> a	C2	<i>climb</i>	(爬树)
排	p <sup>h</sup> ai	M2	p <sup>h</sup> ai	C2	<i>line up</i>	(排队)
婆	p <sup>h</sup> uo	M2	p <sup>h</sup> ɔ	C2	<i>old woman</i>	(阿婆)
皮	p <sup>h</sup> i	M2	p <sup>h</sup> i	C2	<i>skin</i>	(皮肤)
谈	t <sup>h</sup> an	M2	t <sup>h</sup> am	C2	<i>talk</i>	(谈话)
弹	tan	M2	t <sup>h</sup> an	C2	<i>bounce</i>	(弹琴)
图	t <sup>h</sup> u	M2	t <sup>h</sup> u	C2	<i>picture</i>	(图片)
同	t <sup>h</sup> uŋ	M2	t <sup>h</sup> ɔŋ	C2	<i>same</i>	(同学)
麻	ma	M2	ma	C2	<i>hemp</i>	(芝麻)
南	nan	M2	nam	C2	<i>south</i>	(南宁)
反	fan	M3	fan	C3	<i>opposite</i>	(反对)
府	fu	M3	fu	C3	<i>mansion</i>	(首府)
古	ku	M3	ku	C3	<i>ancient</i>	(古代)
摆	pai	M3	pai	C3	<i>place</i>	(摆放)
板	pan	M3	pan	C3	<i>board</i>	(黑板)
比	pi	M3	pi	C3	<i>compare</i>	(比赛)
补	pu	M3	pu	C3	<i>supplement</i>	(补课)
普	p <sup>h</sup> u	M3	p <sup>h</sup> u	C3	<i>ordinary</i>	(普通)
打	ta	M3	ta	C3	<i>hit</i>	(打球)
毯	t <sup>h</sup> an	M3	t <sup>h</sup> an	C3	<i>blanket</i>	(毛毯)
赌	tu	M3	tu	C3	<i>gamble</i>	(赌钱)

Character	Nanning Mandarin and Standard Mandarin		Nanning Cantonese		English gloss	Disyllabic context
	IPA	Tone	IPA	Tone		
土	tʰu	M3	tʰu	C3	<i>soil</i>	(土地)
野	ie	M3	je	C4	<i>wild</i>	(野菜)
以	i	M3	ji	C4	<i>by/with</i>	(以为)
勇	juŋ	M3	juŋ	C4	<i>brave</i>	(勇敢)
雨	y	M3	jy	C4	<i>rain</i>	(大雨)
马	ma	M3	ma	C4	<i>horse</i>	(马路)
买	mai	M3	mai	C4	<i>buy</i>	(买菜)
奶	nai	M3	nai	C4	<i>milk</i>	(奶茶)
你	ni	M3	ni	C4	<i>you</i>	(你好)
努	nu	M3	nu	C4	<i>exert</i>	(努力)
女	ny	M3	ny	C4	<i>female</i>	(女人)
有	jou	M3	jeu	C4	<i>have</i>	(所有)
母	mu	M3	mu	C4	<i>mother</i>	(父母)
富	fu	M4	fu	C5	<i>rich</i>	(富贵)
故	ku	M4	ku	C5	<i>reason</i>	(故事)
霸	pa	M4	pa	C5	<i>bossy</i>	(霸道)
怕	pʰa	M4	pʰa	C5	<i>fear</i>	(害怕)
拜	pai	M4	pai	C5	<i>worship</i>	(拜访)
臂	pi	M4	pi	C5	<i>arm</i>	(手臂)
屁	pʰi	M4	pʰi	C5	<i>fart</i>	(放屁)
布	pu	M4	pu	C5	<i>cloth</i>	(棉布)
带	tai	M4	tai	C5	<i>bring</i>	(皮带)
太	tʰai	M4	tʰai	C5	<i>too</i>	(太阳)
兔	tʰu	M4	tʰu	C5	<i>rabbit</i>	(白兔)
挂	kua	M4	kʷa	C5	<i>hang</i>	(挂历)
饭	fan	M4	fan	C6	<i>cooked rice</i>	(饭店)
父	fu	M4	fu	C6	<i>father</i>	(父母)
罢	pa	M4	pa	C6	<i>quit</i>	(罢工)
败	pai	M4	pai	C6	<i>defeat</i>	(失败)
办	pan	M4	pan	C6	<i>handle</i>	(办法)
避	pi	M4	pi	C6	<i>avoid</i>	(避开)
部	pu	M4	pu	C6	<i>department</i>	(部长)
蛋	tan	M4	tan	C6	<i>egg</i>	(鸡蛋)
地	ti	M4	ti	C6	<i>ground</i>	(地方)
渡	tu	M4	tu	C6	<i>ferry</i>	(渡河)
病	piŋ	M4	peŋ	C6	<i>illness</i>	(生病)
动	tuŋ	M4	tøŋ	C6	<i>move</i>	(动物)

## A.1.2. Disyllabic wordlists

### Practice phase

Character	Nanning Mandarin and Standard Mandarin		Nanning Cantonese		English gloss
	IPA	Tone	IPA	Tone	
天平	t <sup>h</sup> ien p <sup>h</sup> iŋ	M1+M2	t <sup>h</sup> in p <sup>h</sup> eŋ	C1+C2	balance scale
根本	kən pən	M1+M3	kən pun	C1+C3	foundation
头顶	t <sup>h</sup> ou tiŋ	M2+M3	t <sup>h</sup> eu teŋ	C2+C3	top of the head
红豆	xuŋ tou	M2+M4	h <sup>h</sup> ouŋ teu	C2+C6	red bean
海带	xai tai	M3+M4	h <sup>h</sup> oi tai	C3+C5	kelp

### Formal phase

Character	Nanning Mandarin and Standard Mandarin		Nanning Cantonese		English gloss
	IPA	Tone	IPA	Tone	
孤单	ku tan	M1+M1	ku tan	C1+C1	<i>lonely</i>
锅巴	kuo pa	M1+M1	k <sup>o</sup> pa	C1+C1	<i>rice crust</i>
姑婆	ku p <sup>h</sup> uo	M1+M2	ku p <sup>h</sup> ɔ	C1+C2	<i>grandaunt</i>
瓜皮	kua p <sup>h</sup> i	M1+M2	k <sup>wa</sup> p <sup>h</sup> i	C1+C2	<i>melon rind</i>
汤粉	t <sup>h</sup> aŋ fən	M1+M3	t <sup>h</sup> ɔŋ fən	C1+C3	<i>rice noodles in soup</i>
拖把	t <sup>h</sup> uo pa	M1+M3	t <sup>h</sup> ɔ pa	C1+C3	<i>mop</i>
巴马	pa ma	M1+M3	pa ma	C1+C4	<i>Bama (place name)</i>
斑马	pan ma	M1+M3	pan ma	C1+C4	<i>zebra</i>
公布	kuŋ pu	M1+M4	k <sup>o</sup> ŋ pu	C1+C5	<i>announce</i>
通过	t <sup>h</sup> uŋ kuo	M1+M4	t <sup>h</sup> ɔŋ k <sup>o</sup>	C1+C5	<i>pass through</i>
拖地	t <sup>h</sup> uo ti	M1+M4	t <sup>h</sup> ɔ ti	C1+C6	<i>mop the floor</i>
工地	kuŋ ti	M1+M4	k <sup>o</sup> ŋ ti	C1+C6	<i>construction site</i>
皮肤	p <sup>h</sup> i fu	M2+M1	p <sup>h</sup> i fu	C2+C1	<i>skin</i>
台风	t <sup>h</sup> ai fəŋ	M2+M1	t <sup>h</sup> ɔi fəŋ	C2+C1	<i>typhoon</i>
平房	p <sup>h</sup> iŋ faŋ	M2+M2	p <sup>h</sup> eŋ fəŋ	C2+C2	<i>single-storey house</i>
南宁	nan niŋ	M2+M2	nam neŋ	C2+C2	<i>Nanning (place name)</i>
苹果	p <sup>h</sup> iŋ kuo	M2+M3	p <sup>h</sup> eŋ k <sup>o</sup>	C2+C3	<i>apple</i>
糖果	t <sup>h</sup> aŋ kuo	M2+M3	t <sup>h</sup> ɔŋ k <sup>o</sup>	C2+C3	<i>candy</i>
男友	nan jou	M2+M3	nam j <sup>eu</sup>	C2+C4	<i>boyfriend</i>
朋友	p <sup>h</sup> əŋ jou	M2+M3	p <sup>h</sup> eŋ j <sup>eu</sup>	C2+C4	<i>friend</i>
皮带	p <sup>h</sup> i tai	M2+M4	p <sup>h</sup> i tai	C2+C5	<i>belt</i>

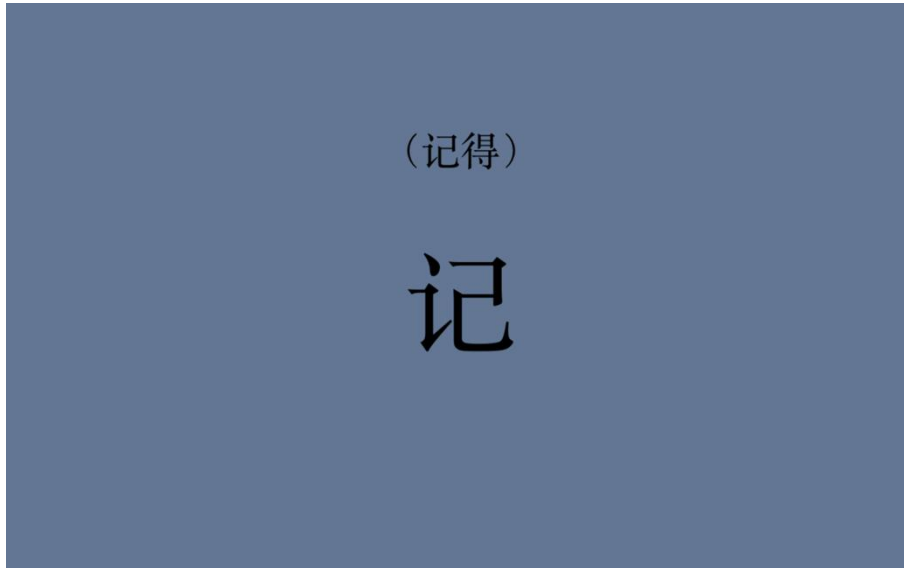
Character	Nanning Mandarin and Standard Mandarin		Nanning Cantonese		English gloss
	IPA	Tone	IPA	Tone	
*头痛 <sup>20</sup>	t <sup>h</sup> ou t <sup>h</sup> uŋ	M2+M4	t <sup>h</sup> eu t <sup>h</sup> oŋ	C2+C5	<i>headache</i>
*难过	nan kuo	M2+M4	nan kɔ	C2+C5	<i>sad</i>
皮蛋	p <sup>h</sup> i tan	M2+M4	p <sup>h</sup> i tan	C2+C6	<i>century egg</i>
平地	p <sup>h</sup> iŋ ti	M2+M4	p <sup>h</sup> eŋ ti	C2+C6	<i>flat ground</i>
摆摊	pai t <sup>h</sup> an	M3+M1	pai t <sup>h</sup> an	C3+C1	<i>set up a stall</i>
普通	p <sup>h</sup> u t <sup>h</sup> uŋ	M3+M1	p <sup>h</sup> u t <sup>h</sup> oŋ	C3+C1	<i>ordinary</i>
果皮	kuo p <sup>h</sup> i	M3+M2	kɔ p <sup>h</sup> i	C3+C2	<i>fruit peel</i>
打牌	ta p <sup>h</sup> ai	M3+M2	ta p <sup>h</sup> ai	C3+C2	<i>play cards</i>
古板	ku pan	M3+M3	ku pan	C3+C3	<i>old-fashioned</i>
打鼓	ta ku	M3+M3	ta ku	C3+C3	<i>beat a drum</i>
好友	xau jou	M3+M3	hu jɛu	C3+C4	<i>good friend</i>
宝马	pau ma	M3+M3	pu ma	C3+C4	<i>BMW</i>
反对	fan tuei	M3+M4	fan tui	C3+C5	<i>oppose</i>
广告	kuaŋ kau	M3+M4	kɔŋ ku	C3+C5	<i>advertisement</i>
打败	ta pai	M3+M4	ta pai	C3+C6	<i>defeat</i>
土地	t <sup>h</sup> u ti	M3+M4	t <sup>h</sup> u ti	C3+C6	<i>land</i>
马蜂	ma fɔŋ	M3+M1	ma fɔŋ	C4+C1	<i>wasp</i>
*买单	mai tan	M3+M1	mai tan	C4+C1	<i>pay the bill</i>
*女工	ny kuŋ	M3+M1	ny kɔŋ	C4+C1	<i>female worker</i>
买房	mai faŋ	M3+M2	mai fɔŋ	C4+C2	<i>buy a house</i>
奶瓶	nai p <sup>h</sup> iŋ	M3+M2	nai p <sup>h</sup> eŋ	C4+C2	<i>milk bottle</i>
奶粉	nai fən	M3+M3	nai fən	C4+C3	<i>milk powder</i>
米粉	mi fən	M3+M3	mɛi fən	C4+C3	<i>rice noodles</i>
老友	lau jou	M3+M3	lu jɛu	C4+C4	<i>old friend</i>
买米	mai mi	M3+M3	mai mɛi	C4+C4	<i>buy rice</i>
冷冻	lɔŋ tuŋ	M3+M4	laŋ tɔŋ	C4+C5	<i>freeze</i>
买票	mai p <sup>h</sup> iau	M3+M4	mai p <sup>h</sup> iu	C4+C5	<i>buy a ticket</i>
有病	jou piŋ	M3+M4	jɛu pɛŋ	C4+C6	<i>sick</i>
米饭	mi fan	M3+M4	mɛi fan	C4+C6	<i>cooked rice</i>
过冬	kuo tuŋ	M4+M1	kɔ tɔŋ	C5+C1	<i>get through winter</i>
对方	tuei faŋ	M4+M1	tui fɔŋ	C5+C1	<i>the other side</i>
过年	kuo nien	M4+M2	kɔ nin	C5+C2	<i>celebrate the New Year</i>

<sup>20</sup> Two pairs of words are marked with an asterisk. Within each tone combination, the upper word was used in the initial stage of data collection but was later replaced by the lower word due to frequent misreadings. Two participants produced the upper words (头痛 ‘headache’ and 买单 ‘pay the bill’), while the remaining participants produced the lower words (难过 ‘sad’ and 女工 ‘female worker’).

Character	Nanning Mandarin and Standard Mandarin		Nanning Cantonese		English gloss
	IPA	Tone	IPA	Tone	
拜年	pai nien	M4+M2	pai nin	C5+C2	<i>pay a New Year visit</i>
拜访	pai faŋ	M4+M3	pai fəŋ	C5+C3	<i>pay a visit</i>
对比	tuei pi	M4+M3	tui pi	C5+C3	<i>compare</i>
购买	kou mai	M4+M3	kəu mai	C5+C4	<i>purchase</i>
过敏	kuo min	M4+M3	kə mən	C5+C4	<i>allergy</i>
怕痛	p <sup>h</sup> a t <sup>h</sup> uŋ	M4+M4	p <sup>h</sup> a t <sup>h</sup> əŋ	C5+C5	<i>fear pain</i>
放屁	faŋ p <sup>h</sup> i	M4+M4	fəŋ p <sup>h</sup> i	C5+C5	<i>fart</i>
带队	tai tuei	M4+M4	tai tui	C5+C6	<i>lead a team</i>
对面	tuei mien	M4+M4	tui min	C5+C6	<i>opposite side</i>
地方	ti faŋ	M4+M1	ti fəŋ	C6+C1	<i>place; local</i>
办公	pan kuŋ	M4+M1	pan kəŋ	C6+C1	<i>office work</i>
地皮	ti p <sup>h</sup> i	M4+M2	ti p <sup>h</sup> i	C6+C2	<i>land plot</i>
地图	ti t <sup>h</sup> u	M4+M2	ti t <sup>h</sup> u	C6+C2	<i>map</i>
地板	ti pan	M4+M3	ti pan	C6+C3	<i>floor</i>
地毯	ti t <sup>h</sup> an	M4+M3	ti t <sup>h</sup> an	C6+C3	<i>carpet</i>
父母	fu mu	M4+M3	fu mu	C6+C4	<i>parents</i>
豆奶	tou nai	M4+M3	təu nai	C6+C4	<i>soy milk</i>
害怕	xai p <sup>h</sup> a	M4+M4	həi p <sup>h</sup> a	C6+C5	<i>be afraid</i>
后退	xou t <sup>h</sup> uei	M4+M4	həu t <sup>h</sup> ui	C6+C5	<i>retreat</i>
后代	xou tai	M4+M4	həu tɔi	C6+C6	<i>descendants</i>
部队	pu tuei	M4+M4	pu tui	C6+C6	<i>army</i>

### A.1.3. Screenshot of trial display

For monosyllabic words, the stimulus was in larger font size, located in the centre of the screen; the disyllabic context, in a smaller font size, was located above the stimulus.



For disyllabic words, the stimulus was in a large font size located in the centre of the screen.



## A.2. Acoustic results

### A.2.1. Raw f0 values in Log-T Chao tone numbers

The table below shows the average f0 values in LT Chao tone across each of the ten f0 points for each phonological tone in each language variety. Based on these f0 values, the tonal contours were plotted (Figure 1.2.1) and the tone values were summarised (Table 1.2.1).

F0 values were rounded up by three digits. Highlighted cells with underlined values are the representative or key points for determining the tone values, most of which are the contour trajectories.

Raw f0 values in Log-T Chao tone numbers

Language	Tone	f0_1	f0_2	f0_3	f0_4	f0_5	f0_6	f0_7	f0_8	f0_9	f0_10	n_spr
Nanning Mandarin	1 [44]	<u>3.66</u>	3.55	3.52	3.51	3.50	3.49	3.47	3.43	3.42	<u>3.30</u>	30
	2 [34]	2.95	2.77	<u>2.75</u>	2.84	3.00	3.17	3.36	<u>3.49</u>	3.46	3.21	30
	3 [42]	<u>3.47</u>	3.29	3.16	3.00	2.82	2.64	2.51	2.20	2.07	<u>1.96</u>	30
	4 [53]	<u>4.35</u>	4.32	4.26	4.07	3.76	3.34	2.97	2.51	2.28	<u>2.13</u>	30
Nanning Cantonese	1 [54]	4.05	<u>4.01</u>	4.03	4.06	4.10	<u>4.12</u>	4.11	4.09	4.00	<u>3.85</u>	48
	2 [42]	<u>3.29</u>	3.06	2.86	2.67	2.48	2.31	2.16	2.01	1.90	<u>1.77</u>	48
	3 [34]	2.80	<u>2.69</u>	2.73	2.94	3.21	3.54	3.76	<u>3.83</u>	3.75	3.52	48
	4 [34]	3.24	2.86	2.77	<u>2.76</u>	2.88	3.13	3.42	<u>3.59</u>	3.55	3.24	48
	5 [43]	<u>3.58</u>	3.49	3.43	3.40	3.38	3.36	3.32	3.26	3.14	<u>2.92</u>	48
	6 [42]	<u>3.22</u>	3.03	2.86	2.68	2.53	2.39	2.25	2.11	2.00	<u>1.91</u>	48
Standard Mandarin	1 [44]	3.60	3.60	<u>3.59</u>	3.60	3.59	<u>3.58</u>	3.60	3.61	3.62	<u>3.69</u>	21
	2 [34]	2.19	<u>2.06</u>	2.06	2.11	2.28	2.51	2.85	3.26	3.60	<u>3.72</u>	21
	3 [323]	<u>2.14</u>	1.82	1.71	1.64	<u>1.47</u>	1.49	1.73	1.77	2.00	<u>2.17</u>	21
	4 [52]	<u>4.18</u>	4.17	4.00	3.64	3.19	2.72	2.16	<u>1.93</u>	1.96	1.96	21

## Appendix B: Intonation study

### B.1. Speech materials and experiments

The table below presents the five-syllable target sentences with English translation, together with the tonal correspondence between Mandarin and Cantonese. To keep the intonation experiment at a manageable length, Mandarin Tone 3 and Tone 4 were each assigned only three sentences, rather than using the full set of corresponding Cantonese sentences (i.e., six sentences each for Mandarin Tone 3 and Tone 4). This resulted in 12 target sentences for Nanning Mandarin and Standard Mandarin, and 18 target sentences for Nanning Cantonese. Person names in English are translated using Standard Mandarin *pīn yīn*.

#### B.1.1. Target sentences

Tone		Target sentence					English translation
Mandarin	Cantonese	Prefix	S (person name)	Modal v.	V	O	
1	1	阿	英	要	煲	汤	<i>A-Ying will make soup</i>
1	1	阿	冰	要	关	窗	<i>A-Bing will close the window</i>
1	1	阿	芳	要	听	歌	<i>A-Fang will listen to songs</i>
2	2	阿	莹	要	还	钱	<i>A-Ying will pay back the money</i>
2	2	阿	玲	要	弹	琴	<i>A-Ling will play the piano</i>
2	2	阿	南	要	骑	牛	<i>A-Nan will ride a cow</i>
3	3	阿	影	要	洗	碗	<i>A-Ying will wash the dishes</i>
3	3	阿	果	要	炒	粉	<i>A-Guo will stir-fry rice noodles</i>
	3	阿	朵	要	剪	纸	<i>A-Duo will cut paper</i>
3	4	阿	敏	要	买	马	<i>A-Min will buy a horse</i>
	4	阿	友	要	领	米	<i>A-You will claim the rice</i>
	4	阿	美	要	养	鸟	<i>A-Mei will keep a bird</i>
4	5	阿	翠	要	寄	信	<i>A-Cui will post a letter</i>
4	5	阿	佩	要	种	菜	<i>A-Pei will plant vegetables</i>
	5	阿	燕	要	退	票	<i>A-Yan will return the ticket (for a refund)</i>
	6	阿	凤	要	问	路	<i>A-Feng will ask for directions</i>
4	6	阿	曼	要	练	字	<i>A-Man will practice calligraphy</i>
	6	阿	露	要	卖	面	<i>A-Lu will sell noodles</i>

## B.1.2. Filler trials

The table below presents the full sets instructions and target response for the filler trials, where the upper rows (shaded) indicate the instructions and the lower rows indicate the target responses. Underlined and bold texts indicate cross-linguistic lexical variations. As Cantonese has a more than Mandarin, the Cantonese experiment contained more trials and therefore required more filler trials; accordingly, four more filler sentences were designed for Nanning Cantonese.

Nanning Mandarin and Standard Mandarin	Nanning Cantonese	Type
从 1 数到 10: <i>Count from 1 to 10:</i>	从 1 数到 10: <i>Count from 1 to 10:</i>	
一 二 三 四 五 六 七 八 九 十 <i>One two three four five six seven eight nine ten</i>	一 二 三 四 五 六 七 八 九 十 <i>One two three four five six seven eight nine ten</i>	
用最 <u>快</u> 的速度从 1 数到 10: <i>Count from 1 to 10 the fastest you can:</i>	用最 <u>快</u> 嘅速度从 1 数到 10: <i>Count from 1 to 10 the fastest you can:</i>	
一 二 三 四 五 六 七 八 九 十 <i>One two three four five six seven eight nine ten</i>	一 二 三 四 五 六 七 八 九 十 <i>One two three four five six seven eight nine ten</i>	count- ing
从 10 倒数到 1: <i>Count from 10 to 1:</i>	从 10 倒数到 1: <i>Count from 10 to 1:</i>	
十 九 八 七 六 五 四 三 二 一 <i>Ten nine eight seven six five four three two one</i>	十 九 八 七 六 五 四 三 二 一 <i>Ten nine eight seven six five four three two one</i>	
用最 <u>快</u> 的速度从 10 倒数到 1: <i>Count from 10 to 1 the fastest you can:</i>	用最 <u>快</u> 嘅速度从 10 倒数到 1: <i>Count from 10 to 1 the fastest you can:</i>	
十 九 八 七 六 五 四 三 二 一 <i>Ten nine eight seven six five four three two one</i>	十 九 八 七 六 五 四 三 二 一 <i>Ten nine eight seven six five four three two one</i>	
<u>说</u> 下面 <u>这</u> 句话: <i>Say the sentence below:</i>	<u>讲</u> 下面 <u>阿</u> 句话: <i>Say the sentence below:</i>	
我等下去买苹果、香蕉、龙眼、菠萝。 <i>Later I'm going to buy apples, bananas, longans, pineapples.</i>	我等下去买苹果、香蕉、龙眼、菠萝。 <i>Later I'm going to buy apples, bananas, longans, pineapples.</i>	list into- nation
用最 <u>快</u> 的速度 <u>说</u> 下面 <u>这</u> 句话: <i>Say the sentence below the fastest you can:</i>	用最 <u>快</u> 嘅速度 <u>讲</u> 下面 <u>阿</u> 句话: <i>Say the sentence below the fastest you can:</i>	
我等下去买苹果、香蕉、龙眼、菠萝。 <i>Later I'm going to buy apples, bananas, longans, pineapples.</i>	我等下去买苹果、香蕉、龙眼、菠萝。 <i>Later I'm going to buy apples, bananas, longans, pineapples.</i>	

Nanning Mandarin and Standard Mandarin	Nanning Cantonese	Type
	<p>讲下面阿句话: Say the sentence below: 我去过广东、四川、海南、山西。 I have been to Guangdong, Sichuan, Hainan, Shaanxi.</p>	
	<p>用最快嘅速度讲下面阿句话: Say the sentence below the fastest you can: 我去过广东、四川、海南、山西。 I have been to Guangdong, Sichuan, Hainan, Shaanxi.</p>	
<p>问下面这个问题: Ask the question below: 你要老友粉还是螺蛳粉? Will you have Lao-you rice noodles or Luo-si rice noodles?</p>	<p>问下面阿只问题: Ask the question below: 你要老友粉重係螺蛳粉? Will you have Lao-you rice noodles or Luo-si rice noodles?</p>	
<p>问下面这个问题: Ask the question below: 你要老友粉还是螺蛳粉? Will you have Lao-you rice noodles or Luo-si rice noodles?</p>	<p>问下面阿只问题: Ask the question below: 你要螺蛳粉重係老友粉? Will you have Lao-you rice noodles or Luo-si rice noodles?</p>	alter- native question
	<p>问下面阿只问题: Ask the question below: 今晚开车去重係打车去? Tonight will we go there by car or by taxi?</p>	
	<p>问下面阿只问题: Ask the question below: 今晚开车去重係打车去? Tonight will we go there by car or by taxi?</p>	

### B.1.3. Full set of question-answer pairs

The tables below present the full practice trials (including one filler trial), and the formal trials without the filler trials.

#### Nanning Mandarin

##### Practice phase

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Broad Focus	4	刚刚他跟你说什么啊?	他说阿亮要办事。	-
Initial Focus	4	刚刚他说谁要送货?	阿泰要送货。	
Final Focus	1	刚刚他说阿丹要搬什么?	阿丹要搬书。	
Broad Question	1	你猜他刚刚跟我说什么?	他说阿东要开车啊?	
Filler	-	用最快的速度说下面这句话:	合同第二、第四、第五条有问题。	

##### Formal phase

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Broad Focus	1	刚刚他跟你说什么啊?	他说阿英要煲汤。	1
Initial Focus		刚刚他说谁要煲汤?	阿英要煲汤。	
Final Focus		刚刚他说阿英要煲什么?	阿英要煲汤。	
Broad Question		你猜他刚刚跟我说什么?	他说阿英要煲汤啊?	
Broad Focus	1	刚刚他跟你说什么啊?	他说阿冰要关窗。	2
Initial Focus		刚刚他说谁要关窗?	阿冰要关窗。	
Final Focus		刚刚他说阿冰要关什么?	阿冰要关窗。	
Broad Question		你猜他刚刚跟我说什么?	他说阿冰要关窗啊?	
Broad Focus	1	刚刚他跟你说什么啊?	他说阿芳要听歌。	3
Initial Focus		刚刚他说谁要听歌?	阿芳要听歌。	
Final Focus		刚刚他说阿芳要听什么?	阿芳要听歌。	
Broad Question		你猜他刚刚跟我说什么?	他说阿芳要听歌啊?	
Broad Focus	2	刚刚他跟你说什么啊?	他说阿莹要还钱。	4
Initial Focus		刚刚他说谁要还钱?	阿莹要还钱。	
Final Focus		刚刚他说阿莹要还什么?	阿莹要还钱。	
Broad Question		你猜他刚刚跟我说什么?	他说阿莹要还钱啊?	
Broad Focus	2	刚刚他跟你说什么啊?	他说阿玲要弹琴。	5
Initial Focus		刚刚他说谁要弹琴?	阿玲要弹琴。	
Final Focus		刚刚他说阿玲要弹什么?	阿玲要弹琴。	
Broad Question		你猜他刚刚跟我说什么?	他说阿玲要弹琴啊?	

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Broad Focus	2	刚刚他跟你说什么啊?	他说阿南要骑牛。	6
Initial Focus		刚刚他说谁要骑牛?	阿南要骑牛。	
Final Focus		刚刚他说阿南要骑什么?	阿南要骑牛。	
Broad Question		你猜他刚刚跟我说什么?	他说阿南要骑牛啊?	
Broad Focus	3	刚刚他跟你说什么啊?	他说阿影要洗碗。	7
Initial Focus		刚刚他说谁要洗碗?	阿影要洗碗。	
Final Focus		刚刚他说阿影要洗什么?	阿影要洗碗。	
Broad Question		你猜他刚刚跟我说什么?	他说阿影要洗碗啊?	
Broad Focus	3	刚刚他跟你说什么啊?	他说阿果要炒粉。	8
Initial Focus		刚刚他说谁要炒粉?	阿果要炒粉。	
Final Focus		刚刚他说阿果要炒什么?	阿果要炒粉。	
Broad Question		你猜他刚刚跟我说什么?	他说阿果要炒粉啊?	
Broad Focus	3	刚刚他跟你说什么啊?	他说阿敏要买马。	9
Initial Focus		刚刚他说谁要买马?	阿敏要买马。	
Final Focus		刚刚他说阿敏要买什么?	阿敏要买马。	
Broad Question		你猜他刚刚跟我说什么?	他说阿敏要买马啊?	
Broad Focus	4	刚刚他跟你说什么啊?	他说阿翠要寄信。	10
Initial Focus		刚刚他说谁要寄信?	阿翠要寄信。	
Final Focus		刚刚他说阿翠要寄什么?	阿翠要寄信。	
Broad Question		你猜他刚刚跟我说什么?	他说阿翠要寄信啊?	
Broad Focus	4	刚刚他跟你说什么啊?	他说阿佩要种菜。	11
Initial Focus		刚刚他说谁要种菜?	阿佩要种菜。	
Final Focus		刚刚他说阿佩要种什么?	阿佩要种菜。	
Broad Question		你猜他刚刚跟我说什么?	他说阿佩要种菜啊?	
Broad Focus	4	刚刚他跟你说什么啊?	他说阿曼要练字。	12
Initial Focus		刚刚他说谁要练字?	阿曼要练字。	
Final Focus		刚刚他说阿曼要练什么?	阿曼要练字。	
Broad Question		你猜他刚刚跟我说什么?	他说阿曼要练字啊?	

## Nanning Cantonese

### Practice phase

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Broad Focus	4	啱啱佢跟你讲乜嘢啊?	佢讲阿亮要办事。	-
Initial Focus	4	啱啱佢讲边个要送货?	阿泰要送货。	
Final Focus	6	啱啱佢讲阿丹要搬乜嘢?	阿丹要搬书。	
Broad Question	5	你猜佢刚刚跟我讲乜嘢?	佢讲阿东要开车啊?	
Filler	-	用最快嘅速度讲下面阿句话:	合同第二、第四、第五条有问题。	

Formal phase

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Broad Focus	1	啱啱佢跟你讲乜嘢啊?	佢讲阿英要煲汤。	1
Initial Focus		啱啱佢讲边个要煲汤?	阿英要煲汤。	
Final Focus		啱啱佢讲阿英要煲乜嘢?	阿英要煲汤。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿英要煲汤啊?	
Broad Focus	1	啱啱佢跟你讲乜嘢啊?	佢讲阿冰要关窗。	2
Initial Focus		啱啱佢讲边个要关窗?	阿冰要关窗。	
Final Focus		啱啱佢讲阿冰要关乜嘢?	阿冰要关窗。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿冰要关窗啊?	
Broad Focus	1	啱啱佢跟你讲乜嘢啊?	佢讲阿芳要听歌。	3
Initial Focus		啱啱佢讲边个要听歌?	阿芳要听歌。	
Final Focus		啱啱佢讲阿芳要听乜嘢?	阿芳要听歌。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿芳要听歌啊?	
Broad Focus	2	啱啱佢跟你讲乜嘢啊?	佢讲阿莹要还钱。	4
Initial Focus		啱啱佢讲边个要还钱?	阿莹要还钱。	
Final Focus		啱啱佢讲阿莹要还乜嘢?	阿莹要还钱。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿莹要还钱啊?	
Broad Focus	2	啱啱佢跟你讲乜嘢啊?	佢讲阿玲要弹琴。	5
Initial Focus		啱啱佢讲边个要弹琴?	阿玲要弹琴。	
Final Focus		啱啱佢讲阿玲要弹乜嘢?	阿玲要弹琴。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿玲要弹琴啊?	
Broad Focus	2	啱啱佢跟你讲乜嘢啊?	佢讲阿南要骑牛。	6
Initial Focus		啱啱佢讲边个要骑牛?	阿南要骑牛。	
Final Focus		啱啱佢讲阿南要骑乜嘢?	阿南要骑牛。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿南要骑牛啊?	
Broad Focus	3	啱啱佢跟你讲乜嘢啊?	佢讲阿影要洗碗。	7
Initial Focus		啱啱佢讲边个要洗碗?	阿影要洗碗。	
Final Focus		啱啱佢讲阿影要洗乜嘢?	阿影要洗碗。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿影要洗碗啊?	
Broad Focus	3	啱啱佢跟你讲乜嘢啊?	佢讲阿果要炒粉。	8
Initial Focus		啱啱佢讲边个要炒粉?	阿果要炒粉。	
Final Focus		啱啱佢讲阿果要炒乜嘢?	阿果要炒粉。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿果要炒粉啊?	
Broad Focus	3	啱啱佢跟你讲乜嘢啊?	佢讲阿朵要剪纸。	9
Initial Focus		啱啱佢讲边个要剪纸?	阿朵要剪纸。	
Final Focus		啱啱佢讲阿朵要剪乜嘢?	阿朵要剪纸。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿朵要剪纸啊?	
Broad Focus	4	啱啱佢跟你讲乜嘢啊?	佢讲阿敏要买马。	10

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Initial Focus		啱佢讲边个要买马?	阿敏要买马。	
Final Focus		啱佢讲阿敏要买乜嘢?	阿敏要买马。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿敏要买马啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿友要领米。	
Initial Focus	4	啱佢讲边个要领米?	阿友要领米。	11
Final Focus		啱佢讲阿友要领乜嘢?	阿友要领米。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿友要领米啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿美要养鸟。	
Initial Focus	4	啱佢讲边个要养鸟?	阿美要养鸟。	12
Final Focus		啱佢讲阿美要养乜嘢?	阿美要养鸟。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿美要养鸟啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿翠要寄信。	
Initial Focus	5	啱佢讲边个要寄信?	阿翠要寄信。	13
Final Focus		啱佢讲阿翠要寄乜嘢?	阿翠要寄信。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿翠要寄信啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿佩要种菜。	
Initial Focus	5	啱佢讲边个要种菜?	阿佩要种菜。	14
Final Focus		啱佢讲阿佩要种乜嘢?	阿佩要种菜。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿佩要种菜啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿燕要退票。	
Initial Focus	5	啱佢讲边个要退票?	阿燕要退票。	15
Final Focus		啱佢讲阿燕要退乜嘢?	阿燕要退票。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿燕要退票啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿凤要问路。	
Initial Focus	6	啱佢讲边个要问路?	阿凤要问路。	16
Final Focus		啱佢讲阿凤要问乜嘢?	阿凤要问路。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿凤要问路啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿曼要练字。	
Initial Focus	6	啱佢讲边个要练字?	阿曼要练字。	17
Final Focus		啱佢讲阿曼要练乜嘢?	阿曼要练字。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿曼要练字啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿露要卖面。	
Initial Focus	6	啱佢讲边个要卖面?	阿露要卖面。	18
Final Focus		啱佢讲阿露要卖乜嘢?	阿露要卖面。	
Broad Question		你猜佢啱啱跟我讲乜嘢?	佢讲阿露要卖面啊?	
Broad Focus		啱佢跟你讲乜嘢啊?	佢讲阿露要卖面。	

## Standard Mandarin

### Practice phase

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Broad Focus	4	刚刚他跟你说了什么了?	他说阿亮要办事。	-
Initial Focus	4	刚刚他说谁要送货?	阿泰要送货。	
Final Focus	1	刚刚他说阿丹要搬什么?	阿丹要搬书。	
Broad Question	1	你猜他刚刚跟我说什么?	他说阿东要开车吗?	
Filler	-	用最快的速度说下面这句话:	合同第二、第四、第五条有问题。	

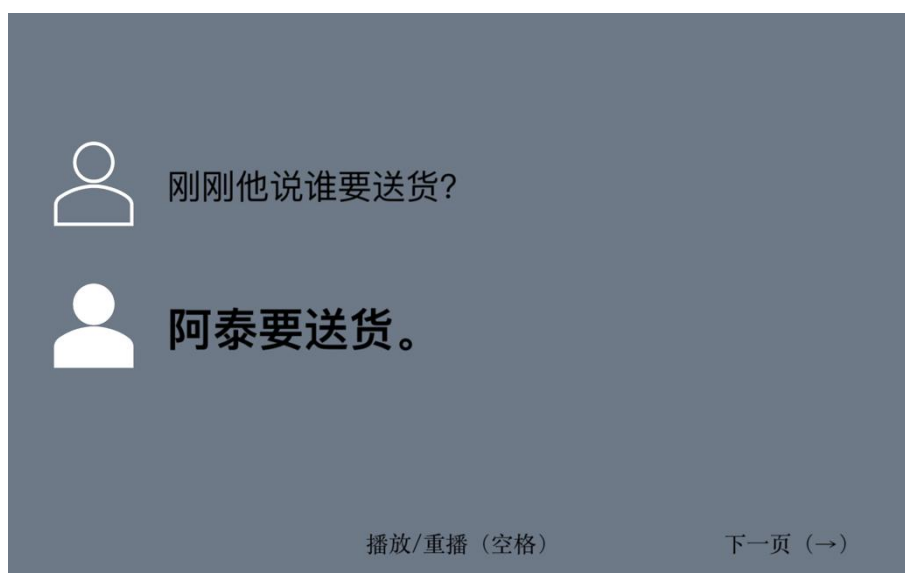
### Formal phase

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Broad Focus	1	刚刚他跟你说了什么了?	他说阿英要煲汤。	1
Initial Focus		刚刚他说谁要煲汤?	阿英要煲汤。	
Final Focus		刚刚他说阿英要煲什么?	阿英要煲汤。	
Broad Question		你猜他刚刚跟我说什么?	他说阿英要煲汤吗?	
Broad Focus	1	刚刚他跟你说了什么了?	他说阿冰要关窗。	2
Initial Focus		刚刚他说谁要关窗?	阿冰要关窗。	
Final Focus		刚刚他说阿冰要关什么?	阿冰要关窗。	
Broad Question		你猜他刚刚跟我说什么?	他说阿冰要关窗吗?	
Broad Focus	1	刚刚他跟你说了什么了?	他说阿芳要听歌。	3
Initial Focus		刚刚他说谁要听歌?	阿芳要听歌。	
Final Focus		刚刚他说阿芳要听什么?	阿芳要听歌。	
Broad Question		你猜他刚刚跟我说什么?	他说阿芳要听歌吗?	
Broad Focus	2	刚刚他跟你说了什么了?	他说阿莹要还钱。	4
Initial Focus		刚刚他说谁要还钱?	阿莹要还钱。	
Final Focus		刚刚他说阿莹要还什么?	阿莹要还钱。	
Broad Question		你猜他刚刚跟我说什么?	他说阿莹要还钱吗?	
Broad Focus	2	刚刚他跟你说了什么了?	他说阿玲要弹琴。	5
Initial Focus		刚刚他说谁要弹琴?	阿玲要弹琴。	
Final Focus		刚刚他说阿玲要弹什么?	阿玲要弹琴。	
Broad Question		你猜他刚刚跟我说什么?	他说阿玲要弹琴吗?	
Broad Focus	2	刚刚他跟你说了什么了?	他说阿南要骑牛。	6
Initial Focus		刚刚他说谁要骑牛?	阿南要骑牛。	
Final Focus		刚刚他说阿南要骑什么?	阿南要骑牛。	
Broad Question		你猜他刚刚跟我说什么?	他说阿南要骑牛吗?	
Broad Focus	3	刚刚他跟你说了什么了?	他说阿影要洗碗。	7

Meaning	Tone	Elicitation/Question	Target/Answer	Item
Initial Focus		刚刚他说谁要洗碗?	阿影要洗碗。	
Final Focus		刚刚他说阿影要洗什么?	阿影要洗碗。	
Broad Question		你猜他刚刚跟我说什么?	他说阿影要洗碗吗?	
Broad Focus	3	刚刚他跟你说了什么了?	他说阿果要炒粉。	8
Initial Focus		刚刚他说谁要炒粉?	阿果要炒粉。	
Final Focus		刚刚他说阿果要炒什么?	阿果要炒粉。	
Broad Question		你猜他刚刚跟我说什么?	他说阿果要炒粉吗?	
Broad Focus	3	刚刚他跟你说了什么了?	他说阿敏要买马。	9
Initial Focus		刚刚他说谁要买马?	阿敏要买马。	
Final Focus		刚刚他说阿敏要买什么?	阿敏要买马。	
Broad Question		你猜他刚刚跟我说什么?	他说阿敏要买马吗?	
Broad Focus	4	刚刚他跟你说了什么了?	他说阿翠要寄信。	10
Initial Focus		刚刚他说谁要寄信?	阿翠要寄信。	
Final Focus		刚刚他说阿翠要寄什么?	阿翠要寄信。	
Broad Question		你猜他刚刚跟我说什么?	他说阿翠要寄信吗?	
Broad Focus	4	刚刚他跟你说了什么了?	他说阿佩要种菜。	11
Initial Focus		刚刚他说谁要种菜?	阿佩要种菜。	
Final Focus		刚刚他说阿佩要种什么?	阿佩要种菜。	
Broad Question		你猜他刚刚跟我说什么?	他说阿佩要种菜吗?	
Broad Focus	4	刚刚他跟你说了什么了?	他说阿曼要练字。	12
Initial Focus		刚刚他说谁要练字?	阿曼要练字。	
Final Focus		刚刚他说阿曼要练什么?	阿曼要练字。	
Broad Question		你猜他刚刚跟我说什么?	他说阿曼要练字吗?	

### B.1.4. Screenshot of trial display

Two different icons indicate the pre-recorded interlocutor and the participant. Eliciting questions were presented in a smaller, non-bold font. Target sentences were presented in a larger, bold font. At the bottom are the instructions for pressing keyboard.



## B.2. Statistical models and results

### B.2.1. Analysis of focus effect (§ 3.3.1 and § 3.3.2.1)

For clearer presentation and to save space, each table of model results includes the results from both iterations of the rotated model. Significance levels throughout the appendix are indicated as follows:  $p < 0.001$ , \*\*\*,  $p < 0.01$ , \*\*,  $p < 0.05$ , \*.

## Nanning Mandarin

Subject f0 range

```
fit_subject_range_nm <- lmer(chaotone_range ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll2)
```

Model results for f0 range on the subject in Nanning Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	9.089e-01	8.613e-02	6.537e+00	10.552	2.41e-05	***
initial focus	1.988e-02	2.022e-02	2.058e+03	0.983	0.325696	
final focus	-1.202e-01	2.019e-02	2.058e+03	-5.953	3.09e-09	***
tone1	-5.275e-01	1.38e-01	5.214e+00	-3.823	0.011399	*
tone2	1.406e-01	1.192e-01	5.29e+00	1.180	0.288490	
tone3	5.335e-01	1.151e-01	5.325e+00	4.635	0.004826	**
tone4	-1.466e-01	1.667e-01	5.162e+00	-0.879	0.418493	
sonorant	2.002e-01	1.12e-01	5.007e+00	1.787	0.133827	
voiceless unaspirated	-4.851e-02	1.458e-01	5.005e+00	-0.333	0.752858	
voiceless aspirated	-3.013e-01	2.076e-01	5.016e+00	-1.451	0.206284	
fricative	1.496e-01	1.967e-01	5.002e+00	0.761	0.481121	
age 20-39	2.452e-03	5.029e-02	2.694e+01	0.049	0.961471	
age 40-49	3.196e-02	5.033e-02	2.701e+01	0.635	0.530769	
age 50-59	4.435e-03	5.262e-02	2.696e+01	0.084	0.933452	
age 60-69	-3.885e-02	5.032e-02	2.7e+01	-0.772	0.446841	
IF:tone1	2.63e-02	3.448e-02	2.058e+03	0.763	0.445723	
FF:tone1	1.19e-01	3.45e-02	2.058e+03	3.448	0.000576	***
IF:tone2	-4.705e-02	3.442e-02	2.058e+03	-1.367	0.171742	
FF:tone2	-1.49e-01	3.435e-02	2.058e+03	-4.339	1.5e-05	***
IF:tone3	-1.816e-02	3.465e-02	2.058e+03	-0.524	0.600161	
FF:tone3	-1.176e-01	3.47e-02	2.058e+03	-3.390	0.000713	***
IF:tone4	3.892e-02	3.647e-02	2.059e+03	1.067	0.286015	
FF:tone4	1.477e-01	3.633e-02	2.059e+03	4.065	4.99e-05	***

## Subject f0 min

```
fit_subject_min_nm <- lmer(chaotone_min ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll12)
```

Model results for f0 min on the subject in Nanning Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	2.697e+00	5.953e-02	2.979e+01	45.312	<2e-16	***
IF	3.869e-02	2.086e-02	2.058e+03	1.855	0.06374	.
FF	1.543e-01	2.084e-02	2.058e+03	7.408	1.86e-13	***
tone1	5.652e-01	5.648e-02	6.46e+00	10.007	3.61e-05	***
tone2	-4.16e-01	4.97e-02	7.067e+00	-8.369	6.47e-05	***
tone3	-7.53e-01	4.84e-02	7.354e+00	-15.557	6.79e-07	***
tone4	6.037e-01	6.733e-02	6.056e+00	8.967	0.000102	***
sonorant	-1.657e-01	4.324e-02	4.914e+00	-3.832	0.01264	*
voiceless unaspirated	-1.856e-01	5.627e-02	4.9e+00	-3.298	0.02219	*
voiceless aspirated	5.273e-01	8.04e-02	4.977e+00	6.558	0.00126	**
fricative	-1.761e-01	7.585e-02	4.88e+00	-2.321	0.069216	.
age 20-39	-1.573e-01	8.558e-02	2.701e+01	-1.838	0.07710	.
age 40-49	-5.757e-02	8.56e-02	2.704e+01	-0.673	0.50694	
age 50-59	6.482e-02	8.952e-02	2.702e+01	0.724	0.47525	
age 60-69	1.5e-01	8.559e-02	2.703e+01	1.753	0.090956	.
IF:tone1	9.883e-03	3.558e-02	2.058e+03	0.278	0.78120	
FF:tone1	-9.642e-02	3.56e-02	2.058e+03	-2.709	0.00681	**
IF:tone2	4.161e-02	3.551e-02	2.058e+03	1.172	0.24138	
FF:tone2	1.74e-01	3.543e-02	2.058e+03	4.909	9.87e-07	***
IF:tone3	-4.131e-04	3.575e-02	2.058e+03	-0.012	0.99078	
FF:tone3	1.093e-01	3.58e-02	2.058e+03	3.053	0.00229	**
IF:tone4	-5.108e-02	3.762e-02	2.059e+03	-1.358	0.174695	
FF:tone4	-1.868e-01	3.749e-02	2.058e+03	-4.984	6.73e-07	***

## Subject f0 max

```
fit_subject_max_nm <- lmer(chaotone_max ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll12)
```

Model results for f0 max on the subject in Nanning Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	3.606e+00	9.617e-02	9.263e+00	37.500	1.96e-11	***
initial focus	5.889e-02	2.083e-02	2.058e+03	2.827	0.00475	**
final focus	3.43e-02	2.081e-02	2.058e+03	1.648	0.09942	.
tone1	3.801e-02	1.405e-01	5.216e+00	0.271	0.79711	
tone2	-2.753e-01	1.213e-01	5.293e+00	-2.269	0.06958	.
tone3	-2.198e-01	1.172e-01	5.329e+00	-1.876	0.11596	
tone4	4.570e-01	1.697e-01	5.162e+00	2.693	0.04178	*
sonorant	3.432e-02	1.14e-01	5.003e+00	0.301	0.77544	
voiceless unaspirated	-2.344e-01	1.484e-01	5.001e+00	-1.580	0.17497	
voiceless aspirated	2.269e-01	2.113e-01	5.012e+00	1.074	0.33193	
fricative	-1.548e-01	8.485e-02	2.701e+01	-1.824	0.07921	.
age 20-39	-2.58e-02	8.487e-02	2.703e+01	-0.304	0.76349	
age 40-49	6.93e-02	8.876e-02	2.701e+01	0.781	0.44174	
age 50-59	3.571e-02	3.553e-02	2.058e+03	1.005	0.31500	
age 60-69	2.248e-02	3.555e-02	2.058e+03	0.632	0.52727	
IF:tone1	-5.853e-03	3.546e-02	2.058e+03	-0.165	0.86892	
FF:tone1	2.476e-02	3.539e-02	2.058e+03	0.700	0.48416	
IF:tone2	-1.848e-02	3.57e-02	2.058e+03	-0.518	0.60482	
FF:tone2	-8.274e-03	3.575e-02	2.058e+03	-0.231	0.81701	
IF:tone3	-1.138e-02	3.758e-02	2.058e+03	-0.303	0.76207	
FF:tone3	-3.896e-02	3.744e-02	2.058e+03	-1.041	0.29806	
IF:tone4	-2.676e-02	2.002e-01	4.998e+00	-0.134	0.89886	
FF:tone4	1.113e-01	8.487e-02	2.703e+01	1.311	0.20080	

## Object f0 range

```
fit_object_range_nm <- lmer(chaotone_range ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll15)
```

Model results for f0 range on the object in Nanning Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.582e+00	6.609e-02	2.396e+01	23.937	<2e-16	***
initial focus	-7.508e-02	3.912e-02	1.715e+03	-1.919	0.0551	.
final focus	2.579e-02	3.937e-02	1.715e+03	0.655	0.5125	
tone1	-7.18e-01	8.863e-02	7.085e+00	-8.102	7.87e-04	***
tone2	-1.434e-01	8.21e-02	8.276e+00	-1.747	0.1175	
tone3	2.036e-01	1.03e-01	7.041e+00	1.978	0.0883	.
tone4	6.579e-01	8.629e-02	8.957e+00	7.624	3.33e-05	***
sonorant	-4.007e-02	9.006e-02	5.027e+00	-0.445	0.6749	
voiceless unaspirated	-1.294e-01	9.288e-02	4.946e+00	-1.393	0.2230	
voiceless aspirated	-2.363e-02	7.166e-02	4.925e+00	-0.330	0.7551	
fricative	1.931e-01	8.744e-02	5.191e+00	2.208	0.0763	.
age 20-39	-1.254e-01	8.359e-02	2.719e+01	-1.500	0.1450	
age 40-49	-8.468e-02	8.332e-02	2.687e+01	-1.016	0.3185	
age 50-59	5.156e-03	8.734e-02	2.711e+01	0.059	0.9534	
age 60-69	2.049e-01	8.356e-02	2.719e+01	2.452	0.0209	*
IF:tone1	5.584e-02	6.356e-02	1.715e+03	0.879	0.3798	
FF:tone1	2.675e-03	6.389e-02	1.715e+03	0.042	0.9666	
IF:tone2	6.015e-03	6.882e-02	1.716e+03	0.087	0.9304	
FF:tone2	-2.118e-02	6.768e-02	1.715e+03	-0.313	0.7544	
IF:tone3	-1.018e-01	6.855e-02	1.716e+03	-1.485	0.1378	
FF:tone3	-1.19e-01	6.898e-02	1.716e+03	-1.725	0.0847	.
IF:tone4	3.994e-02	7.018e-02	1.716e+03	0.569	0.5693	
FF:tone4	1.375e-01	7.212e-02	1.716e+03	1.906	0.0568	.

## Object f0 min

```
fit_object_min_nm <- lmer(chaotone_min ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll15)
```

Model results for f0 min on the object in Nanning Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.491e+00	5.379e-02	3.054e+01	27.725	<2e-16	***
initial focus	1.463e-02	3.716e-02	1.714e+03	0.394	0.6938	
final focus	1.753e-02	3.739e-02	1.714e+03	0.469	0.6393	
tone1	6.983e-01	6.681e-02	8.806e+00	10.452	2.93e-06	***
tone2	-5.368e-01	6.323e-02	1.121e+01	-8.489	3.25e-06	***
tone3	-8.283e-01	7.754e-02	8.708e+00	-10.683	2.68e-06	***
tone4	6.668e-01	6.713e-02	1.262e+01	9.934	2.51e-07	***
sonorant	7.577e-02	6.405e-02	4.947e+00	1.183	0.2905	
voiceless unaspirated	3.96e-02	6.584e-02	4.798e+00	0.602	0.5748	
voiceless aspirated	1.945e-03	5.075e-02	4.771e+00	0.038	0.9710	
fricative	-1.173e-01	6.257e-02	5.244e+00	-1.875	0.1170	
age 20-39	7.353e-02	7.233e-02	2.691e+01	1.017	0.3184	
age 40-49	-1.149e-01	7.205e-02	2.654e+01	-1.595	0.1226	
age 50-59	2.268e-02	7.557e-02	2.682e+01	0.300	0.7664	
age 60-69	1.871e-02	7.231e-02	2.691e+01	0.259	0.7978	
IF:tone1	-6.821e-03	6.037e-02	1.714e+03	-0.113	0.9101	
FF:tone1	3.162e-02	6.069e-02	1.714e+03	0.521	0.6024	
IF:tone2	2.935e-02	6.536e-02	1.716e+03	0.449	0.6534	
FF:tone2	7.733e-03	6.429e-02	1.715e+03	0.120	0.9043	
IF:tone3	4.027e-02	6.511e-02	1.716e+03	0.619	0.5363	
FF:tone3	1.095e-01	6.552e-02	1.715e+03	1.671	0.0948	.
IF:tone4	-6.28e-02	6.666e-02	1.716e+03	-0.942	0.3462	
FF:tone4	-1.489e-01	6.849e-02	1.716e+03	-2.173	0.0299	*

## Object f0 max

```
fit_object_max_nm <- lmer(chaotone_max ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll15)
```

Model results for f0 max on the object in Nanning Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	3.07451	0.06081	32.98433	50.562	<2e-16	***
initial focus	-0.06086	0.02756	1714.75019	-2.208	0.0274	*
final focus	0.04236	0.02773	1714.43768	1.527	0.1269	
tone1	-0.02372	0.04047	14.20793	-0.586	0.5669	
tone2	-0.68045	0.03937	20.17501	-17.283	1.46e-13	***
tone3	-0.61908	0.04689	13.89721	-13.204	2.97e-09	***
tone4	1.32325	0.04230	23.80405	31.280	<2e-16	***
sonorant	0.03276	0.03540	5.52903	0.926	0.3933	
voiceless unaspirated	-0.08837	0.03615	5.20097	-2.445	0.0564	.
voiceless aspirated	-0.01944	0.02784	5.18265	-0.698	0.5152	
fricative	0.07505	0.03498	6.16006	2.145	0.0744	.
age 20-39	-0.05316	0.09836	27.12707	-0.540	0.5933	
age 40-49	-0.19858	0.09824	27.00002	-2.021	0.0533	.
age 50-59	0.02928	0.10284	27.08445	0.285	0.7780	
age 60-69	0.22246	0.09835	27.11494	2.262	0.0319	*
IF:tone1	0.05082	0.04477	1714.56843	1.135	0.2565	
FF:tone1	0.03556	0.04501	1714.58431	0.790	0.4296	
IF:tone2	0.03459	0.04849	1715.00355	0.713	0.4758	
FF:tone2	-0.01314	0.04769	1714.62081	-0.275	0.7830	
IF:tone3	-0.06332	0.04829	1715.87191	-1.311	0.1900	
FF:tone3	-0.00868	0.04860	1714.81590	-0.179	0.8583	
IF:tone4	-0.02209	0.04945	1715.12606	-0.447	0.6551	
FF:tone4	-0.01375	0.05081	1715.08458	-0.271	0.7867	

## Nanning Cantonese

Subject f0 range

```
fit_subject_range_nc <- lmer(chaotone_range ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll2)
```

Model results for f0 range on the subject in Nanning Cantonese

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.255e+00	5.905e-02	2.27e+01	21.254	<2e-16	***
initial focus	-6.167e-02	2.007e-02	3.048e+03	-3.073	0.002135	**
final focus	-2.082e-01	2.01e-02	3.047e+03	-10.360	<2e-16	***
tone1	-4.805e-01	8.855e-02	1.147e+01	-5.426	0.000180	***
tone2	6.013e-01	8.294e-02	1.089e+01	7.250	1.74e-05	***
tone3	1.574e-01	9.856e-02	1.045e+01	1.597	0.139949	
tone4	2.429e-02	8.291e-02	1.087e+01	0.293	0.775063	
tone5	-7.63e-01	1.231e-01	9.795e+00	-6.197	0.000111	***
tone6	4.605e-01	8.397e-02	1.095e+01	5.484	0.000194	***
sonorant	1.581e-01	6.384e-02	9.006e+00	2.477	0.035162	*
voiceless unaspirated	-3.258e-01	9.89e-02	9.222e+00	-3.294	0.009015	**
voiceless aspirated	4.14e-01	1.342e-01	8.973e+00	3.085	0.013088	*
fricative	-2.463e-01	9.886e-02	9.207e+00	-2.492	0.033781	*
age 20-39	-2.506e-02	6.312e-02	2.702e+01	-0.397	0.694484	
age 40-49	-7.629e-03	6.308e-02	2.695e+01	-0.121	0.904640	
age 50-59	5.018e-02	6.601e-02	2.699e+01	0.760	0.453774	
age 60-69	-1.749e-02	6.309e-02	2.695e+01	-0.277	0.783760	
IF:tone1	2.081e-01	5.006e-02	3.049e+03	4.157	3.32e-05	***
FF:tone1	3.503e-01	5.033e-02	3.048e+03	6.961	4.13e-12	***
IF:tone2	-2.127e-01	4.359e-02	3.047e+03	-4.879	1.12e-06	***
FF:tone2	-1.839e-01	4.375e-02	3.047e+03	-4.204	2.7e-05	***
IF:tone3	1.707e-01	4.369e-02	3.047e+03	3.907	9.56e-05	***
FF:tone3	-1.03e-02	4.385e-02	3.047e+03	-0.235	0.814301	
IF:tone4	-4.187e-02	4.368e-02	3.047e+03	-0.959	0.337868	
FF:tone4	-1.577e-01	4.375e-02	3.047e+03	-3.606	0.000316	***
IF:tone5	-3.861e-02	4.392e-02	3.047e+03	-0.879	0.379453	
FF:tone5	6.156e-02	4.394e-02	3.047e+03	1.401	0.161367	
IF:tone6	-8.56e-02	4.388e-02	3.047e+03	-1.951	0.051198	.
FF:tone6	-5.993e-02	4.371e-02	3.047e+03	-1.371	0.170475	

## Subject f0 min

```
fit_subject_min_nc <- lmer(chaotone_min ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll12)
```

Model results for f0 min on the subject in Nanning Cantonese

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	2.429e+00	8.08e-02	1.843e+01	30.065	<2e-16	***
initial focus	3.926e-02	1.864e-02	3.047e+03	2.106	0.035295	*
final focus	2.318e-01	1.867e-02	3.047e+03	12.413	<2e-16	***
tone1	1.204e+00	1.251e-01	9.924e+00	9.621	2.4e-06	***
tone2	-5.88e-01	1.181e-01	9.702e+00	-4.980	0.000606	***
tone3	-4.174e-01	1.411e-01	9.537e+00	-2.957	0.015083	*
tone4	-1.863e-01	1.181e-01	9.697e+00	-1.578	0.146564	
tone5	5.681e-01	1.78e-01	9.281e+00	3.192	0.010562	*
tone6	-5.802e-01	1.194e-01	9.726e+00	-4.858	0.000719	***
sonorant	-5.89e-02	9.344e-02	8.969e+00	-0.630	0.544186	
voiceless unaspirated	-5.812e-03	1.443e-01	9.057e+00	-0.040	0.968735	
voiceless aspirated	-2.144e-02	1.966e-01	8.956e+00	-0.109	0.915568	
fricative	8.615e-02	1.442e-01	9.051e+00	0.597	0.564937	
age 20-39	-9.547e-02	7.786e-02	2.701e+01	-1.226	0.230714	
age 40-49	-8.301e-02	7.783e-02	2.697e+01	-1.067	0.295652	
age 50-59	2.157e-02	8.143e-02	2.699e+01	0.265	0.793080	
age 60-69	1.569e-01	7.784e-02	2.697e+01	2.016	0.053875	.
IF:tone1	-2.111e-01	4.65e-02	3.048e+03	-4.540	5.85e-06	***
FF:tone1	-3.116e-01	4.676e-02	3.048e+03	-6.663	3.16e-11	***
IF:tone2	9.7e-02	4.049e-02	3.047e+03	2.396	0.016654	*
FF:tone2	7.158e-02	4.064e-02	3.047e+03	1.761	0.078291	.
IF:tone3	2.149e-02	4.058e-02	3.047e+03	0.529	0.596541	
FF:tone3	1.307e-01	4.074e-02	3.047e+03	3.208	0.001351	**
IF:tone4	-1.835e-02	4.058e-02	3.047e+03	-0.452	0.651077	
FF:tone4	1.127e-01	4.064e-02	3.047e+03	2.774	0.005574	**
IF:tone5	1.307e-02	4.08e-02	3.047e+03	0.320	0.748666	
FF:tone5	-8.847e-02	4.082e-02	3.047e+03	-2.167	0.030292	*
IF:tone6	9.79e-02	4.076e-02	3.047e+03	2.402	0.016379	*
FF:tone6	8.505e-02	4.061e-02	3.047e+03	2.094	0.036306	*

## Subject f0 max

```
fit_subject_max_nc <- lmer(chaotone_max ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll2)
```

Model results for f0 max on the subject in Nanning Cantonese

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	2.429e+00	8.08e-02	1.843e+01	30.065	<2e-16	***
initial focus	3.926e-02	1.864e-02	3.047e+03	2.106	0.035295	*
final focus	2.318e-01	1.867e-02	3.047e+03	12.413	<2e-16	***
tone1	1.204e+00	1.251e-01	9.924e+00	9.621	2.4e-06	***
tone2	-5.88e-01	1.181e-01	9.702e+00	-4.980	0.000606	***
tone3	-4.174e-01	1.411e-01	9.537e+00	-2.957	0.015083	*
tone4	-1.863e-01	1.181e-01	9.697e+00	-1.578	0.146564	
tone5	5.681e-01	1.78e-01	9.281e+00	3.192	0.010562	*
tone6	-0.11934	0.06620	12.31509	-1.803	0.09596	.
sonorant	-5.89e-02	9.344e-02	8.969e+00	-0.630	0.544186	
voiceless unaspirated	-5.812e-03	1.443e-01	9.057e+00	-0.040	0.968735	
voiceless aspirated	-2.144e-02	1.966e-01	8.956e+00	-0.109	0.915568	
fricative	8.615e-02	1.442e-01	9.051e+00	0.597	0.564937	
age 20-39	-9.547e-02	7.786e-02	2.701e+01	-1.226	0.230714	
age 40-49	-8.301e-02	7.783e-02	2.697e+01	-1.067	0.295652	
age 50-59	2.157e-02	8.143e-02	2.699e+01	0.265	0.793080	
age 60-69	1.569e-01	7.784e-02	2.697e+01	2.016	0.053875	.
IF:tone1	-2.111e-01	4.65e-02	3.048e+03	-4.540	5.85e-06	***
FF:tone1	-3.116e-01	4.676e-02	3.048e+03	-6.663	3.16e-11	***
IF:tone2	9.7e-02	4.049e-02	3.047e+03	2.396	0.016654	*
FF:tone2	7.158e-02	4.064e-02	3.047e+03	1.761	0.078291	.
IF:tone3	2.149e-02	4.058e-02	3.047e+03	0.529	0.596541	
FF:tone3	1.307e-01	4.074e-02	3.047e+03	3.208	0.001351	**
IF:tone4	-1.835e-02	4.058e-02	3.047e+03	-0.452	0.651077	
FF:tone4	1.127e-01	4.064e-02	3.047e+03	2.774	0.005574	**
IF:tone5	1.307e-02	4.08e-02	3.047e+03	0.320	0.748666	
FF:tone5	-8.847e-02	4.082e-02	3.047e+03	-2.167	0.030292	*
IF:tone6	9.79e-02	4.076e-02	3.047e+03	2.402	0.016379	*
FF:tone6	8.505e-02	4.061e-02	3.047e+03	2.094	0.036306	*

## Object f0 range

```
fit_object_range_nc <- lmer(chaotone_range ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll15)
```

Model results for f0 range on the object in Nanning Cantonese

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.438e+00	5.526e-02	3.571e+01	26.026	<2e-16	***
initial focus	-1.714e-01	2.745e-02	2.578e+03	-6.242	5.02e-10	***
final focus	-5.174e-02	2.733e-02	2.578e+03	-1.893	0.058485	.
tone1	-4.805e-01	8.137e-02	1.296e+01	-5.905	5.26e-05	***
tone2	2.904e-02	7.916e-02	1.527e+01	0.367	0.718769	
tone3	7.742e-01	7.935e-02	1.395e+01	9.756	1.31e-07	***
tone4	1.997e-01	8.889e-02	1.193e+01	2.247	0.044355	*
tone5	-4.306e-01	8.607e-02	1.232e+01	-5.004	0.000284	***
tone6	-9.182e-02	8.296e-02	1.436e+01	-1.107	0.286532	
sonorant	-1.167e-01	6.535e-02	8.994e+00	-1.786	0.107808	
voiceless unaspirated	-5.245e-02	6.51e-02	8.857e+00	-0.806	0.441497	
voiceless aspirated	1.762e-01	6.3e-02	8.716e+00	2.797	0.021449	*
fricative	-7.073e-03	7.808e-02	8.745e+00	-0.091	0.929856	
age 20-39	-8.862e-02	7.304e-02	2.747e+01	-1.213	0.235355	
age 40-49	-1.216e-01	7.258e-02	2.681e+01	-1.676	0.105451	
age 50-59	5.634e-04	7.612e-02	2.708e+01	0.007	0.994149	
age 60-69	2.097e-01	7.267e-02	2.694e+01	2.885	0.007606	**
IF:tone1	-3.828e-02	5.89e-02	2.578e+03	-0.650	0.515846	
FF:tone1	-9.644e-02	5.942e-02	2.578e+03	-1.623	0.104718	
IF:tone2	6.885e-02	6.47e-02	2.579e+03	1.064	0.287312	
FF:tone2	8.414e-02	6.441e-02	2.578e+03	1.306	0.191576	
IF:tone3	-2.654e-01	6.221e-02	2.578e+03	-4.267	2.05e-05	***
FF:tone3	-1.75e-01	6.193e-02	2.577e+03	-2.826	0.004743	**
IF:tone4	-1.435e-02	5.913e-02	2.579e+03	-0.243	0.808348	
FF:tone4	1.563e-01	5.961e-02	2.578e+03	2.622	0.008797	**
IF:tone5	9.841e-02	5.845e-02	2.578e+03	1.684	0.092357	.
FF:tone5	-3.038e-02	5.829e-02	2.578e+03	-0.521	0.602332	
IF:tone6	1.508e-01	6.457e-02	2.579e+03	2.336	0.019586	*
FF:tone6	6.142e-02	6.275e-02	2.579e+03	0.979	0.327824	

Object f0 min

```
fit_object_min_nc <- lmer(chaotone_min ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll15,
control=lmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)) )
```

Model results for f0 min on the object in Nanning Cantonese

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.433e+00	5.244e-02	3.743e+01	27.332	< 2e-16	***
initial focus	5.457e-02	2.532e-02	2.579e+03	2.155	0.03127	*
final focus	1.063e-01	2.521e-02	2.578e+03	4.216	2.57e-05	***
tone1	1.687e+00	7.455e-02	1.35e+01	22.625	3.98e-12	***
tone2	-8.764e-01	7.256e-02	1.594e+01	-12.078	1.95e-09	***
tone3	-2.823e-01	7.272e-02	1.454e+01	-3.882	0.00155	**
tone4	-3.013e-01	8.142e-02	1.241e+01	-3.700	0.00287	**
tone5	6.85e-01	7.884e-02	1.282e+01	8.689	9.9e-07	***
tone6	-9.118e-01	7.603e-02	1.498e+01	-11.992	4.42e-09	***
sonorant	1.038e-01	5.979e-02	9.319e+00	1.737	0.11531	
voiceless unaspirated	-3.344e-02	5.956e-02	9.174e+00	-0.561	0.58796	
voiceless aspirated	-5.107e-03	5.764e-02	9.026e+00	-0.089	0.93133	
fricative	-6.529e-02	7.143e-02	9.057e+00	-0.914	0.38438	
age 20-39	1.008e-01	7.091e-02	2.736e+01	1.422	0.16630	
age 40-49	-3.082e-02	7.051e-02	2.676e+01	-0.437	0.66552	
age 50-59	-8.656e-03	7.392e-02	2.7e+01	-0.117	0.90765	
age 60-69	-6.136e-02	7.058e-02	2.688e+01	-0.869	0.39234	
IF:tone1	-5.566e-02	5.434e-02	2.578e+03	-1.024	0.30574	
FF:tone1	-1.497e-02	5.481e-02	2.578e+03	-0.273	0.78483	
IF:tone2	-4.569e-03	5.968e-02	2.579e+03	-0.077	0.93898	
FF:tone2	-2.715e-02	5.942e-02	2.578e+03	-0.457	0.64780	
IF:tone3	1.019e-02	5.739e-02	2.578e+03	0.178	0.85908	
FF:tone3	5.62e-02	5.712e-02	2.578e+03	0.984	0.32533	
IF:tone4	-1.635e-02	5.455e-02	2.579e+03	-0.300	0.76438	
FF:tone4	-6.735e-02	5.499e-02	2.578e+03	-1.225	0.22073	
IF:tone5	1.445e-02	5.391e-02	2.578e+03	0.268	0.78869	
FF:tone5	9.484e-02	5.377e-02	2.578e+03	1.764	0.07788	.
IF:tone6	5.194e-02	5.956e-02	2.579e+03	0.872	0.38323	
FF:tone6	-4.157e-02	5.789e-02	2.579e+03	-0.718	0.47276	

## Object f0 max

```
fit_object_max_nc <- lmer(chaotone_max ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr) + (1 | lexical), f0raw_syll15)
```

### Model results for f0 max on the object in Nanning Cantonese

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	2.872e+00	6.182e-02	2.673e+01	46.455	<2e-16	***
initial focus	-1.169e-01	2.506e-02	2.578e+03	-4.663	3.27e-06	***
final focus	5.47e-02	2.495e-02	2.578e+03	2.192	0.028435	*
tone1	1.206e+00	1.033e-01	1.086e+01	11.671	1.74e-07	***
tone2	-8.47e-01	9.856e-02	1.185e+01	-8.594	1.96e-06	***
tone3	4.919e-01	9.985e-02	1.129e+01	4.926	0.000419	***
tone4	-1.013e-01	1.14e-01	1.041e+01	-0.889	0.394306	
tone5	2.543e-01	1.099e-01	1.058e+01	2.313	0.041956	*
tone6	-1.004e+00	1.040e-01	1.147e+01	-9.646	7.62e-07	***
sonorant	-1.329e-02	8.689e-02	9.069e+00	-0.153	0.881789	
voiceless unaspirated	-8.606e-02	8.673e-02	9.005e+00	-0.992	0.347008	
voiceless aspirated	1.721e-01	8.411e-02	8.938e+00	2.046	0.071233	.
fricative	-7.279e-02	1.042e-01	8.951e+00	-0.699	0.502535	
age 20-39	1.275e-02	7.149e-02	2.738e+01	0.178	0.859806	
age 40-49	-1.527e-01	7.11e-02	2.681e+01	-2.147	0.040981	*
age 50-59	-8.274e-03	7.453e-02	2.704e+01	-0.111	0.912423	
age 60-69	1.482e-01	7.117e-02	2.692e+01	2.082	0.046960	*
IF:tone1	-9.38e-02	5.377e-02	2.578e+03	-1.745	0.081162	.
FF:tone1	-1.117e-01	5.424e-02	2.578e+03	-2.059	0.039591	*
IF:tone2	6.302e-02	5.906e-02	2.578e+03	1.067	0.285991	
FF:tone2	5.678e-02	5.879e-02	2.578e+03	0.966	0.334231	
IF:tone3	-2.554e-01	5.678e-02	2.578e+03	-4.499	7.15e-06	***
FF:tone3	-1.185e-01	5.652e-02	2.577e+03	-2.096	0.036174	*
IF:tone4	-2.966e-02	5.398e-02	2.578e+03	-0.549	0.582755	
FF:tone4	8.93e-02	5.441e-02	2.578e+03	1.641	0.100869	
IF:tone5	1.127e-01	5.335e-02	2.578e+03	2.112	0.034805	*
FF:tone5	6.396e-02	5.321e-02	2.578e+03	1.202	0.229443	
IF:tone6	2.032e-01	5.894e-02	2.578e+03	3.448	0.000574	***
FF:tone6	2.011e-02	5.728e-02	2.578e+03	0.351	0.725530	

## Standard Mandarin

Subject f0 range

```
fit_subject_range_sm <- lmer(chaotone_range ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr), f0raw_syll2)
```

Model results for f0 range on the subject in Standard Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.19197	0.10622	7.42937	11.222	6.36e-06	***
initial focus	0.31538	0.04939	662.02873	6.385	3.23e-10	***
final focus	-0.24659	0.04854	662.01167	-5.080	4.92e-07	***
tone1	-0.73511	0.06783	662.03906	-10.838	<2e-16	***
tone2	0.17831	0.06326	662.04729	2.819	0.00497	**
tone3	0.79526	0.06226	662.03801	12.773	<2e-16	***
tone4.....	-0.23847	0.07441	662.01358	-3.205	0.00142	**
sonorant	0.19737	0.03848	662.05502	5.130	3.81e-07	***
voiceless unaspirated	-0.36301	0.05004	662.06915	-7.254	1.14e-12	***
voiceless aspirated	0.21693	0.07182	662.05105	3.021	0.00262	**
fricative	-0.07697	0.13499	5.98291	-0.570	0.58932	
age 20-39	-0.13827	0.20607	5.96746	-0.671	0.52732	
age 40-49	-0.19938	0.13502	5.98751	-1.477	0.19033	
age 50-59	-0.22790	0.08457	662.02013	-2.695	0.00722	**
age 60-69	0.27528	0.08384	662.00924	3.283	0.00108	**
IF:tone1	-0.01549	0.08434	662.00781	-0.184	0.85431	
FF:tone1	-0.16431	0.08385	662.02369	-1.960	0.05047	.
IF:tone2	0.04262	0.08506	662.01254	0.501	0.61651	
FF:tone2	0.02123	0.08456	662.01301	0.251	0.80182	
IF:tone3	0.20077	0.08817	662.02692	2.277	0.02309	*
FF:tone3	-0.13221	0.08406	662.01531	-1.573	0.11626	
IF:tone4	-0.05130	0.06667	661.99817	-0.769	0.44190	
FF:tone4	0.41462	0.20621	5.98356	2.011	0.09120	.

## Subject f0 min

```
fit_subject_min_sm <- lmer(chaotone_min ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr), f0raw_syll2)
```

### Model results for f0 min on the subject in Standard Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	2.52866	0.10731	7.16490	23.564	4.71e-08	***
initial focus	-0.12095	0.04558	662.02815	-2.654	0.008152	**
final focus	0.10118	0.04479	662.01418	2.259	0.024222	*
tone1	0.88499	0.06259	662.03660	14.140	<2e-16	***
tone2	-0.34235	0.05838	662.04334	-5.865	7.11e-09	***
tone3	-0.94300	0.05745	662.03574	-16.414	<2e-16	***
tone4.....	0.40036	0.06866	662.01574	5.831	8.62e-09	***
sonorant	-0.06910	0.03550	662.04967	-1.946	0.052026	.
voiceless unaspirated	-0.03872	0.04618	662.06125	-0.838	0.402112	
voiceless aspirated	0.22219	0.06627	662.04643	3.353	0.000845	***
fricative	-0.11437	0.06152	662.00313	-1.859	0.063451	.
age 20-39	0.20445	0.13767	5.99066	1.485	0.188144	
age 40-49	-0.13519	0.21018	5.97798	-0.643	0.543949	
age 50-59	0.26635	0.13769	5.99442	1.934	0.101273	
age 60-69	-0.33560	0.21030	5.99119	-1.596	0.161712	
IF:tone1	0.35602	0.07804	662.02111	4.562	6.04e-06	***
FF:tone1	-0.22135	0.07737	662.01219	-2.861	0.004355	**
IF:tone2	-0.11179	0.07782	662.01102	-1.436	0.151353	
FF:tone2	0.13871	0.07737	662.02402	1.793	0.073469	.
IF:tone3	-0.16072	0.07849	662.01488	-2.048	0.040982	*
FF:tone3	0.04705	0.07803	662.01527	0.603	0.546741	
IF:tone4	-0.08351	0.08136	662.02667	-1.027	0.305018	
FF:tone4	0.03560	0.07757	662.01715	0.459	0.646446	

## Subject f0 max

```
fit_subject_max_sm <- lmer(chaotone_max ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr), f0raw_syll2)
```

Model results for f0 max on the subject in Standard Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	3.72075	0.06495	8.27115	57.289	4.78e-12	***
initial focus	0.19433	0.03650	662.04936	5.325	1.39e-07	***
final focus	-0.14542	0.03587	662.02314	-4.054	5.63e-05	***
tone1	0.14958	0.05012	662.06520	2.985	0.002944	**
tone2	-0.16384	0.04674	662.07783	-3.505	0.000487	***
tone3	-0.14751	0.04600	662.06366	-3.207	0.001407	**
tone4.....	0.16178	0.05498	662.02605	2.943	0.003368	**
sonorant	0.12810	0.02843	662.08972	4.506	7.81e-06	***
voiceless unaspirated	-0.40158	0.03698	662.11139	-10.861	<2e-16	***
voiceless aspirated	0.43905	0.05306	662.08361	8.274	7.12e-16	***
fricative	0.12747	0.08034	5.97888	1.587	0.163853	
age 20-39	-0.27348	0.12260	5.95511	-2.231	0.067538	.
age 40-49	0.06699	0.08036	5.98599	0.834	0.436479	
age 50-59	0.12825	0.06249	662.03611	2.052	0.040530	*
age 60-69	0.05409	0.06195	662.01939	0.873	0.382888	
IF:tone1	-0.12733	0.06231	662.01718	-2.043	0.041419	*
FF:tone1	-0.02574	0.06195	662.04160	-0.416	0.677894	
IF:tone2	-0.11813	0.06285	662.02449	-1.880	0.060592	.
FF:tone2	0.06830	0.06248	662.02522	1.093	0.274685	
IF:tone3	0.11721	0.06514	662.04655	1.799	0.072423	.
FF:tone3	-0.09665	0.06211	662.02874	-1.556	0.120168	
IF:tone4	-0.16557	0.04926	662.00236	-3.361	0.000821	***
FF:tone4	0.07901	0.12273	5.97991	0.644	0.543602	

## Object f0 range

```
fit_object_range_sm <- lmer(chaotone_range ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr), f0raw_syll5)
```

Model results for f0 range on the object in Standard Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.65966	0.20112	6.79381	8.252	8.81e-05	***
initial focus	-0.44886	0.08949	527.46653	-5.016	7.22e-07	***
final focus	-0.01619	0.08218	527.11773	-0.197	0.84389	
tone1	-0.80732	0.10819	527.45749	-7.462	3.53e-13	***
tone2	0.13131	0.11199	527.15796	1.172	0.24154	
tone3	0.11420	0.12145	527.30557	0.940	0.34750	
tone4.....	0.56182	0.10961	527.13439	5.126	4.17e-07	***
sonorant	0.21363	0.08983	527.82564	2.378	0.01775	*
voiceless unaspirated	-0.21532	0.08222	527.06569	-2.619	0.00908	**
voiceless aspirated	-0.08634	0.06560	527.53337	-1.316	0.18867	
fricative	0.08803	0.08539	527.37738	1.031	0.30304	
age 20-39	-0.14319	0.26115	5.96124	-0.548	0.60339	
age 40-49	-0.18415	0.39917	5.97663	-0.461	0.66087	
age 50-59	0.07503	0.26143	5.98681	0.287	0.78378	
age 60-69	0.25231	0.39853	5.93849	0.633	0.55025	
IF:tone1	0.59879	0.14189	527.03654	4.220	2.88e-05	***
FF:tone1	0.06792	0.13562	527.09310	0.501	0.61670	
IF:tone2	-0.12955	0.16215	527.29215	-0.799	0.42468	
FF:tone2	-0.11274	0.14695	527.07240	-0.767	0.44328	
IF:tone3	-0.04822	0.16172	527.05151	-0.298	0.76568	
FF:tone3	-0.02246	0.14451	527.06109	-0.155	0.87657	
IF:tone4	-0.42102	0.15082	527.29957	-2.792	0.00544	**
FF:tone4	0.06728	0.14161	527.12517	0.475	0.63493	

## Object f0 min

```
fit_object_min_sm <- lmer(chaotone_min ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr), f0raw_syll5)
```

Model results for f0 min on the object in Standard Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.77350	0.12917	8.00170	13.730	7.62e-07	***
initial focus	-0.89433	0.08467	528.10680	-10.563	< 2e-16	***
final focus	-0.04626	0.07777	527.31450	-0.595	0.552200	
tone1	1.07121	0.10236	528.06342	10.465	< 2e-16	***
tone2	-0.68443	0.10598	527.41228	-6.458	2.41e-10	***
tone3	-0.70246	0.11492	527.74276	-6.112	1.91e-09	***
tone4.....	0.31569	0.10373	527.35710	3.043	0.002456	**
sonorant	0.12571	0.08497	528.85305	1.479	0.139631	
voiceless unaspirated	0.01393	0.07782	527.19932	0.179	0.857950	
voiceless aspirated	-0.04130	0.06207	528.24579	-0.665	0.506091	
fricative	-0.09835	0.08079	527.90671	-1.217	0.224050	
age 20-39	0.15957	0.16095	5.95542	0.991	0.360046	
age 40-49	-0.01476	0.24623	5.99251	-0.060	0.954136	
age 50-59	-0.01377	0.16135	6.01370	-0.085	0.934769	
age 60-69	-0.13103	0.24530	5.90270	-0.534	0.612731	
IF:tone1	-0.52488	0.13429	527.13095	-3.909	0.000105	***
FF:tone1	-0.08862	0.12835	527.25570	-0.690	0.490208	
IF:tone2	0.51823	0.15344	527.71278	3.377	0.000786	***
FF:tone2	-0.02068	0.13907	527.21366	-0.149	0.881825	
IF:tone3	0.26272	0.15305	527.16591	1.717	0.086649	.
FF:tone3	0.01920	0.13677	527.18812	0.140	0.888408	
IF:tone4	-0.25607	0.14272	527.72946	-1.794	0.073347	.
FF:tone4	0.09010	0.13402	527.33083	0.672	0.501673	

## Object f0 max

```
fit_object_max_sm <- lmer(chaotone_max ~ meaning*tone +
preceding_consonant + age_group + (1 | spkr), f0raw_syll5)
```

Model results for f0 max on the object in Standard Mandarin

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	3.433395	0.111753	6.933853	30.723	1.14e-08	***
initial focus	-1.344591	0.054045	527.539423	-24.879	<2e-16	***
final focus	-0.062607	0.049633	527.122753	-1.261	0.20772	
tone1	0.262426	0.065339	527.526807	4.016	6.77e-05	***
tone2	-0.551051	0.067639	527.171297	-8.147	2.72e-15	***
tone3	-0.588815	0.073349	527.347247	-8.028	6.51e-15	***
tone4.....	0.877439	0.066198	527.143008	13.255	<2e-16	***
sonorant	0.338830	0.054251	527.963014	6.246	8.69e-10	***
voiceless unaspirated	-0.200070	0.049660	527.060819	-4.029	6.43e-05	***
voiceless aspirated	-0.126367	0.039619	527.618258	-3.190	0.00151	**
fricative	-0.012393	0.051569	527.433098	-0.240	0.81017	
age 20-39	0.014947	0.144217	5.935965	0.104	0.92087	
age 40-49	-0.198833	0.220464	5.954361	-0.902	0.40212	
age 50-59	0.062605	0.144403	5.966273	0.434	0.67986	
age 60-69	0.121281	0.220042	5.908936	0.551	0.60172	
IF:tone1	0.073076	0.085698	527.025846	0.853	0.39420	
FF:tone1	-0.020483	0.081910	527.093121	-0.250	0.80263	
IF:tone2	0.387955	0.097931	527.331258	3.961	8.47e-05	***
FF:tone2	-0.134729	0.088752	527.068769	-1.518	0.12960	
IF:tone3	0.216326	0.097671	527.043799	2.215	0.02720	*
FF:tone3	-0.002161	0.087281	527.055279	-0.025	0.98026	
IF:tone4	-0.677356	0.091088	527.340101	-7.436	4.22e-13	***
FF:tone4	0.157373	0.085530	527.131596	1.840	0.06633	.

## B.2.2. Post-hoc pairwise comparisons of tone-specific focus effect (§ 3.3.2.2)

### Nanning Mandarin

F0 range

Post-hoc pairwise comparison: tonal variations in f0 range under focus in Nanning Mandarin

position	tone	contrast	estimate	SE	df	t.ratio	p.value
subject	1	BF - IF	-0.05	0.0395	2058	-1.169	0.4720
		BF - FF	0.00	0.0396	2058	0.032	0.9994
	2	BF - IF	0.03	0.0394	2058	0.690	0.7694
		BF - FF	0.27	0.0393	2058	6.853	<.0001 ***
	3	BF - IF	-0.00	0.0398	2058	-0.043	0.9990
		BF - FF	0.24	0.0399	2058	5.960	<.0001 ***
	4	BF - IF	-0.06	0.0429	2059	-1.369	0.3572
		BF - FF	-0.03	0.0427	2059	-0.643	0.7964
object	1	BF - IF	0.02	0.0708	1714	0.272	0.9602
		BF - FF	-0.03	0.0712	1715	-0.400	0.9156
	2	BF - IF	0.07	0.0800	1715	0.863	0.6635
		BF - FF	-0.00	0.0778	1715	-0.059	0.9981
	3	BF - IF	0.18	0.0795	1715	2.224	0.0675
		BF - FF	0.09	0.0801	1715	1.164	0.4751
	4	BF - IF	0.04	0.0824	1716	0.427	0.9046
		BF - FF	-0.16	0.0854	1716	-1.912	0.1356

F0 min

Post-hoc pairwise comparison: tonal variations in f0 min under focus in Nanning Mandarin

position	tone	contrast	estimate	SE	df	t.ratio	p.value
subject	1	BF - IF	-0.05	0.0408	2058	-1.192	0.4581
		BF - FF	-0.06	0.0408	2058	-1.419	0.3311
	2	BF - IF	-0.08	0.0406	2058	-1.976	0.1185
		BF - FF	-0.33	0.0405	2058	-8.099	<.0001 ***
	3	BF - IF	-0.04	0.0411	2058	-0.932	0.6198
		BF - FF	-0.26	0.0412	2058	-6.404	<.0001 ***
	4	BF - IF	0.01	0.0443	2059	0.280	0.9578
		BF - FF	0.03	0.0441	2058	0.737	0.7412
object	1	BF - IF	-0.01	0.0673	1715	-0.116	0.9926

		BF - FF	-0.05	0.0676	1715	-0.727	0.7475
	2	BF - IF	-0.04	0.0760	1716	-0.579	0.8314
		BF - FF	-0.03	0.0739	1715	-0.342	0.9376
	3	BF - IF	-0.05	0.0755	1716	-0.727	0.7476
		BF - FF	-0.13	0.0761	1716	-1.670	0.2169
	4	BF - IF	0.05	0.0782	1716	0.616	0.8115
		BF - FF	0.13	0.0811	1717	1.619	0.2378

F0 max

Post-hoc pairwise comparison: tonal variations in f0 max under focus in Nanning Mandarin

position	tone	contrast	estimate	SE	df	t.ratio	p.value
subject	1	BF - IF	-0.09	0.0407	2058	-2.324	0.0527
		BF - FF	-0.06	0.0408	2058	-1.393	0.3448
	2	BF - IF	-0.05	0.0406	2058	-1.307	0.3916
		BF - FF	-0.06	0.0405	2058	-1.459	0.3110
	3	BF - IF	-0.04	0.0410	2058	-0.986	0.5860
		BF - FF	-0.03	0.0411	2058	-0.633	0.8020
	4	BF - IF	-0.05	0.0442	2058	-1.074	0.5303
		BF - FF	0.00	0.0440	2058	0.106	0.9938
object	1	BF - IF	0.01	0.0499	1714	0.201	0.9779
		BF - FF	-0.08	0.0501	1714	-1.554	0.2661
	2	BF - IF	0.03	0.0564	1715	0.466	0.8871
		BF - FF	-0.03	0.0548	1714	-0.533	0.8550
	3	BF - IF	0.12	0.0560	1716	2.216	0.0687
		BF - FF	-0.03	0.0564	1715	-0.597	0.8217
	4	BF - IF	0.08	0.0580	1715	1.430	0.3258
		BF - FF	-0.03	0.0602	1715	-0.475	0.8829

## Nanning Cantonese

F0 range

Post-hoc pairwise comparison: tonal variations in f0 range under focus in Nanning Cantonese

position	tone	contrast	estimate	SE	df	t.ratio	p.value	
subject	1	BF - IF	-0.15	0.0562	Inf	-2.606	0.0248	*
		BF - FF	-0.14	0.0565	Inf	-2.515	0.0319	*
	2	BF - IF	0.27	0.0474	Inf	5.788	<.0001	***
		BF - FF	0.39	0.0476	Inf	8.241	<.0001	***
	3	BF - IF	-0.11	0.0475	Inf	-2.293	0.0567	
		BF - FF	0.22	0.0477	Inf	4.579	<.0001	***
	4	BF - IF	0.10	0.0475	Inf	2.179	0.0748	
		BF - FF	0.37	0.0476	Inf	7.690	<.0001	***
	5	BF - IF	0.10	0.0479	Inf	2.096	0.0907	
		BF - FF	0.15	0.0479	Inf	3.065	0.0062	**
	6	BF - IF	0.15	0.0478	Inf	3.081	0.0058	**
		BF - FF	0.27	0.0475	Inf	5.641	<.0001	***
object	1	BF - IF	0.21	0.0638	2578	3.285	0.0030	**
		BF - FF	0.15	0.0646	2578	2.293	0.0569	
	2	BF - IF	0.10	0.0718	2579	1.428	0.3265	
		BF - FF	-0.03	0.0715	2578	-0.453	0.8929	
	3	BF - IF	0.44	0.0684	2578	6.390	<.0001	***
		BF - FF	0.23	0.0681	2577	3.332	0.0025	**
	4	BF - IF	0.19	0.0642	2579	2.894	0.0107	*
		BF - FF	-0.10	0.0649	2578	-1.611	0.2413	
	5	BF - IF	0.07	0.0632	2578	1.155	0.4806	
		BF - FF	0.08	0.0630	2578	1.303	0.3938	
	6	BF - IF	0.02	0.0716	2580	0.287	0.9556	
		BF - FF	-0.01	0.0692	2579	-0.140	0.9893	

F0 min

Post-hoc pairwise comparison: tonal variations in f0 min under focus in Nanning Cantonese

position	tone	contrast	estimate	SE	df	t.ratio	p.value	
subject	1	BF - IF	0.17	0.0522	Inf	3.293	0.0028	**
		BF - FF	0.08	0.0525	Inf	1.520	0.2817	
	2	BF - IF	-0.14	0.0440	Inf	-3.095	0.0056	**
		BF - FF	-0.30	0.0442	Inf	-6.863	<.0001	***
	3	BF - IF	-0.06	0.0441	Inf	-1.376	0.3536	
		BF - FF	-0.36	0.0443	Inf	-8.176	<.0001	***
	4	BF - IF	-0.02	0.0441	Inf	-0.474	0.8838	
		BF - FF	-0.34	0.0442	Inf	-7.793	<.0001	***
	5	BF - IF	-0.05	0.0445	Inf	-1.177	0.4668	
		BF - FF	-0.14	0.0445	Inf	-3.224	0.0036	**
	6	BF - IF	-0.14	0.0444	Inf	-3.089	0.0057	**
		BF - FF	-0.32	0.0442	Inf	-7.174	<.0001	***
object	1	BF - IF	0.00	0.0589	2578	0.019	0.9998	
		BF - FF	-0.09	0.0596	2578	-1.533	0.2758	
	2	BF - IF	-0.05	0.0662	2579	-0.755	0.7305	
		BF - FF	-0.08	0.0659	2578	-1.201	0.4527	
	3	BF - IF	-0.06	0.0631	2578	-1.027	0.5599	
		BF - FF	-0.16	0.0628	2577	-2.588	0.0262	*
	4	BF - IF	-0.04	0.0592	2579	-0.646	0.7948	
		BF - FF	-0.04	0.0599	2578	-0.651	0.7920	
	5	BF - IF	-0.07	0.0583	2577	-1.184	0.4628	
		BF - FF	-0.20	0.0582	2578	-3.459	0.0016	**
	6	BF - IF	-0.11	0.0660	2580	-1.613	0.2403	
		BF - FF	-0.06	0.0638	2579	-1.014	0.5678	

F0 max

Post-hoc pairwise comparison: tonal variations in f0 max under focus in Nanning Cantonese

position	tone	contrast	estimate	SE	df	t.ratio	p.value
subject	1	BF - IF	0.02	0.0569	Inf	0.406	0.9130
		BF - FF	-0.06	0.0573	Inf	-1.100	0.5138
	2	BF - IF	0.14	0.0481	Inf	2.871	0.0114 *
		BF - FF	0.09	0.0482	Inf	1.839	0.1569
	3	BF - IF	-0.17	0.0482	Inf	-3.526	0.0012 **
		BF - FF	-0.14	0.0484	Inf	-2.978	0.0082 **
	4	BF - IF	0.08	0.0482	Inf	1.712	0.2006
		BF - FF	0.02	0.0483	Inf	0.443	0.8973
	5	BF - IF	0.05	0.0485	Inf	0.989	0.5836
		BF - FF	0.00	0.0485	Inf	0.067	0.9975
	6	BF - IF	0.01	0.0485	Inf	0.207	0.9766
		BF - FF	-0.05	0.0482	Inf	-1.011	0.5698
object	1	BF - IF	0.21	0.0582	2578	3.617	0.0009 ***
		BF - FF	0.06	0.0590	2578	0.966	0.5984
	2	BF - IF	0.05	0.0655	2579	0.822	0.6896
		BF - FF	-0.11	0.0652	2578	-1.709	0.2019
	3	BF - IF	0.37	0.0624	2578	5.967	<.0001 ***
		BF - FF	0.06	0.0621	2577	1.027	0.5601
	4	BF - IF	0.15	0.0586	2579	2.502	0.0332 *
		BF - FF	-0.14	0.0592	2578	-2.431	0.0401 *
	5	BF - IF	0.00	0.0577	2577	0.073	0.9971
		BF - FF	-0.12	0.0575	2578	-2.062	0.0980
	6	BF - IF	-0.09	0.0653	2579	-1.322	0.3831
		BF - FF	-0.07	0.0632	2579	-1.185	0.4624

## Standard Mandarin

F0 range

Post-hoc pairwise comparison: tonal variations in f0 range under focus in Standard Mandarin

position	tone	contrast	estimate	SE	df	t.ratio	p.value
subject	1	BF - IF	-0.09	0.0971	662	-0.901	0.6398
		BF - FF	-0.03	0.0967	662	-0.297	0.9526
	2	BF - IF	-0.30	0.0967	662	-3.102	0.0057 **
		BF - FF	0.41	0.0967	662	4.250	0.0001 ***
	3	BF - IF	-0.36	0.0979	662	-3.655	0.0008 ***
		BF - FF	0.23	0.0979	662	2.302	0.0563
	4	BF - IF	-0.52	0.1030	662	-4.997	<.0001 ***
		BF - FF	0.38	0.0971	662	3.903	0.0003 ***
object	1	BF - IF	-0.15	0.156	527	-0.959	0.6034
		BF - FF	-0.05	0.153	527	-0.339	0.9386
	2	BF - IF	0.58	0.192	528	3.011	0.0077 **
		BF - FF	0.13	0.172	527	0.748	0.7349
	3	BF - IF	0.50	0.191	527	2.599	0.0260 *
		BF - FF	0.04	0.168	527	0.230	0.9713
	4	BF - IF	0.87	0.172	527	5.072	<.0001 ***
		BF - FF	-0.05	0.163	527	-0.313	0.9475

F0 min

Post-hoc pairwise comparison: tonal variations in f0 min under focus in Standard Mandarin

position	tone	contrast	estimate	SE	df	t.ratio	p.value
subject	1	BF - IF	-0.24	0.0896	662	-2.624	0.0241 *
		BF - FF	0.12	0.0892	662	1.347	0.3696
	2	BF - IF	0.23	0.0892	662	2.609	0.0252 *
		BF - FF	-0.24	0.0892	662	-2.689	0.0201 *
	3	BF - IF	0.28	0.0904	662	3.117	0.0054 **
		BF - FF	-0.15	0.0903	662	-1.641	0.2294
	4	BF - IF	0.20	0.0953	662	2.145	0.0817
		BF - FF	-0.14	0.0896	662	-1.527	0.2788
object	1	BF - IF	1.42	0.148	527	9.587	<.0001 ***
		BF - FF	0.13	0.144	527	0.935	0.6187
	2	BF - IF	0.38	0.182	528	2.069	0.0974

		BF - FF	0.07	0.163	527	0.410	0.9114	
	3	BF - IF	0.63	0.181	527	3.489	0.0015	**
		BF - FF	0.03	0.159	527	0.170	0.9842	
	4	BF - IF	1.15	0.162	527	7.087	<.0001	***
		BF - FF	-0.04	0.155	528	-0.284	0.9566	

F0 max

Post-hoc pairwise comparison: tonal variations in f0 max under focus in Standard Mandarin

position	tone	contrast	estimate	SE	df	t.ratio	p.value	
subject	1	BF - IF	-0.32	0.0717	662	-4.497	<.0001	***
		BF - FF	0.09	0.0714	662	1.278	0.4078	
	2	BF - IF	-0.07	0.0714	662	-0.938	0.6165	
		BF - FF	0.17	0.0714	662	2.396	0.0444	*
	3	BF - IF	-0.08	0.0724	662	-1.053	0.5438	
		BF - FF	0.08	0.0723	662	1.066	0.5356	
	4	BF - IF	-0.31	0.0763	662	-4.082	0.0001	***
		BF - FF	0.24	0.0717	662	3.375	0.0022	**
object	1	BF - IF	1.27	0.0945	527	13.460	<.0001	***
		BF - FF	0.08	0.0921	527	0.902	0.6392	
	2	BF - IF	0.96	0.1160	528	8.244	<.0001	***
		BF - FF	0.20	0.1040	527	1.896	0.1409	
	3	BF - IF	1.13	0.1160	527	9.767	<.0001	***
		BF - FF	0.06	0.1020	527	0.637	0.7996	
	4	BF - IF	2.02	0.1040	527	19.518	<.0001	***
		BF - FF	-0.09	0.0986	527	-0.961	0.6020	

### B.2.3. Cross-linguistic comparisons (§ 3.3.3)

Subject f0 range

```
fit_subject_range.c <- lmer(chaotone_range ~ nlang1*nmeaning1 +
nlang1*nmeaning2 + nlang2*nmeaning1 + nlang2*nmeaning2 + (1 | spkr) +
(1 | lexical), f0raw_syll2)
```

Model results for cross-linguistic comparison of f0 range on the subject

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.11171	0.10333	19.27924	10.759	1.37e-09	***
nlang1	0.30514	0.04965	71.33448	6.146	4.09e-08	***
nmeaning1	0.08099	0.01926	5806.69367	4.205	2.65e-05	***
nmeaning2	-0.20104	0.01908	5806.56370	-10.535	<2e-16	***
nlang2	0.25519	0.07097	70.85495	3.596	0.000595	***
nlang1:nmeaning1	-0.09588	0.03421	5807.02800	-2.803	0.005079	**
nlang1:nmeaning2	-0.10205	0.03421	5806.84560	-2.983	0.002867	**
nmeaning1:nlang2	0.27499	0.05353	5806.64565	5.137	2.88e-07	***
nmeaning2:nlang2	-0.12704	0.05294	5806.55020	-2.399	0.016450	*

Subject f0 min

```
fit_subject_min.c <- lmer(chaotone_min ~ nlang1*nmeaning1 +
nlang1*nmeaning2 + nlang2*nmeaning1 + nlang2*nmeaning2 + (1 | spkr) +
(1 | lexical), f0raw_syll2)
```

Model results for cross-linguistic comparison of f0 min on the subject

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	2.61074	0.13982	19.72275	18.672	5.28e-14	***
nlang1	-0.23507	0.07163	70.13587	-3.282	0.00161	**
nmeaning1	-0.01144	0.02023	5806.29104	-0.566	0.57163	
nmeaning2	0.17072	0.02004	5806.22082	8.519	<2e-16	***
nlang2	-0.03553	0.10246	69.88126	-0.347	0.72981	
nlang1:nmeaning1	0.01452	0.03592	5806.46625	0.404	0.68602	
nlang1:nmeaning2	0.08512	0.03593	5806.36996	2.369	0.01786	*
nmeaning1:nlang2	-0.16116	0.05622	5806.26669	-2.866	0.00417	**
nmeaning2:nlang2	-0.05915	0.05560	5806.21470	-1.064	0.28745	

### Subject f0 max

```
fit_subject_max.c <- lmer(chaotone_max ~ nlang1*nmeaning1 +
nlang1*nmeaning2 + nlang2*nmeaning1 + nlang2*nmeaning2 + (1 | spkr) +
(1 | lexical), f0raw_syll12)
```

Model results for cross-linguistic comparison of f0 max on the subject

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	3.72247	0.06680	27.67881	55.725	<2e-16	***
nlang1	0.07003	0.06037	70.15205	1.160	0.249929	
nmeaning1	0.06963	0.01870	5806.22744	3.723	0.000199	***
nmeaning2	-0.03027	0.01853	5806.11537	-1.633	0.102428	
nlang2	0.21969	0.08635	69.87168	2.544	0.013164	*
nlang1:nmeaning1	-0.08136	0.03322	5806.41581	-2.449	0.014339	*
nlang1:nmeaning2	-0.01694	0.03322	5806.29765	-0.510	0.610103	
nmeaning1:nlang2	0.11377	0.05199	5806.18527	2.189	0.028673	*
nmeaning2:nlang2	-0.18612	0.05141	5806.10369	-3.620	0.000297	***

### Object f0 range

```
fit_object_range.c <- lmer(chaotone_range ~ nlang1*nmeaning1 +
nlang1*nmeaning2 + nlang2*nmeaning1 + nlang2*nmeaning2 + (1 | spkr) +
(1 | lexical), f0raw_syll15)
```

Model results for cross-linguistic comparison of f0 range on the object

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.39969	0.09918	24.04941	14.112	3.93e-13	***
nlang1	-0.07889	0.07661	70.35547	-1.030	0.3067	
nmeaning1	-0.20549	0.02927	4862.91918	-7.020	2.53e-12	***
nmeaning2	-0.01242	0.02793	4858.81737	-0.445	0.6567	
nlang2	-0.04314	0.10975	70.54841	-0.393	0.6955	
nlang1:nmeaning1	-0.11288	0.05021	4858.23705	-2.248	0.0246	*
nlang1:nmeaning2	-0.07252	0.05031	4858.47541	-1.442	0.1495	
nmeaning1:nlang2	-0.33968	0.08185	4863.67393	-4.150	3.38e-05	***
nmeaning2:nlang2	-0.03510	0.07752	4858.82227	-0.453	0.6507	

## Object f0 min

```
fit_object_min.c <- lmer(chaotone_min ~ nlang1*nmeaning1 +
nlang1*nmeaning2 + nlang2*nmeaning1 + nlang2*nmeaning2 + (1 | spkr) +
(1 | lexical), f0raw_syll15)
```

Model results for cross-linguistic comparison of f0 min on the object

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.456e+00	1.807e-01	1.804e+01	8.057	2.18e-07	***
nlang1	9.89e-02	6.045e-02	7.189e+01	1.636	0.106	
nmeaning1	-2.929e-01	2.944e-02	4.867e+03	-9.951	<2e-16	***
nmeaning2	3.204e-02	2.809e-02	4.86e+03	1.141	0.254	
nlang2	5.835e-03	8.663e-02	7.22e+01	0.067	0.946	
nlang1:nmeaning1	2.47e-02	5.05e-02	4.859e+03	0.489	0.625	
nlang1:nmeaning2	7.872e-02	5.06e-02	4.86e+03	1.556	0.120	
nmeaning1:nlang2	-1.004e+00	8.23e-02	4.868e+03	-12.202	<2e-16	***
nmeaning2:nlang2	-7.901e-02	7.796e-02	4.86e+03	-1.013	0.311	

## Object f0 max

```
fit_object_min.c <- lmer(chaotone_min ~ nlang1*nmeaning1 +
nlang1*nmeaning2 + nlang2*nmeaning1 + nlang2*nmeaning2 + (1 | spkr) +
(1 | lexical), f0raw_syll15)
```

Model results for cross-linguistic comparison of f0 max on the object

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	2.855e+00	1.53e-01	1.941e+01	18.668	7.31e-14	***
nlang1	2.052e-02	7.562e-02	7.095e+01	0.271	0.787	
nmeaning1	-4.986e-01	2.994e-02	4.864e+03	-16.654	<2e-16	***
nmeaning2	1.934e-02	2.856e-02	4.859e+03	0.677	0.498	
nlang2	-3.768e-02	1.083e-01	7.114e+01	-0.348	0.729	
nlang1:nmeaning1	-8.762e-02	5.135e-02	4.859e+03	-1.706	0.088	.
nlang1:nmeaning2	6.836e-03	5.145e-02	4.859e+03	0.133	0.894	
nmeaning1:nlang2	-1.344e+00	8.37e-02	4.864e+03	-16.056	<2e-16	***
nmeaning2:nlang2	-1.136e-01	7.928e-02	4.859e+03	-1.433	0.152	

## REFERENCES

- Ahn, E., & Chodroff, E. (2022). VoxCommunis: A Corpus for Cross-linguistic Phonetic Analysis. *Proceedings of the 13th Conference on Language Resources and Evaluation Conference (LREC 2022)*.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2018). *Gorillas in our Midst: Gorilla.sc, a new web-based Experiment Builder* (p. 438242). bioRxiv. <https://doi.org/10.1101/438242>
- Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2020). *Common Voice: A Massively-Multilingual Speech Corpus* (arXiv:1912.06670). arXiv. <https://doi.org/10.48550/arXiv.1912.06670>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Walter de Gruyter.

- Bauer, R. S., Cheung, K. H., Cheung, P. M., & Ng, L. (2004). Acoustic correlates of focus-stress in Hong Kong Cantonese. *11th Annual Meeting of the Southeast Asian Linguistics Society 2001, Mahidol, Bangkok, Arizona*, 29–49.
- Bei, X. (2020). 几种基频归一方法的比较 [Comparison of Several F0 Normalisation Methods]. *实验语言学 [Experimental Linguistics]*, 2.
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Bulus, M., & Jentschke, S. (2025). *pwrss: Statistical Power and Sample Size Calculation Tools* [Computer software]. <https://doi.org/10.32614/CRAN.package.pwrss>
- Chafe, W. L. (1974). Language and Consciousness. *Language*, 50(1), 111–133.  
<https://doi.org/10.2307/412014>
- Chao, Y. R. (1933). Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology*, 4, 121–134.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press.
- Chao, Y.-R. (1930). A system of tone letters. *Le Maître Phonétique*.

- Chen, H., & Li, L. (Eds). (2005). *广西语言文字使用问题调查与研究 [Investigation of the theory and practice of promoting Mandarin in Guangxi]* (Vols 1–513). Guangxi Education Publishing House.
- Chen, H., & Lin, Y. (Eds). (2009). *粤语平话土话方音字汇 Yueyu pinghua tuhua fangyin zihui* (Vol. 1, p. 427). Shanghai Educational Publishing House.
- Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge University Press.
- Chen, Y. (1991). “地方普通话”的性质特征及其他 [The nature and other features of ‘Local Mandarin’]. *Chinese Teaching in the World*, (01), 12–17.
- 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. (2014). *Prosodic realization of focus in second language speech: Effects of language experience* [PhD Thesis, University of Oregon]. <https://search.proquest.com/openview/01ec89f8cbc0386e4a5653c4f4637114/1?pq-origsite=scholar&cbl=18750>
- Chen, Y. (2015). Post-focus compression in English by Mandarin learners. *Proceedings of ICPHS 2015*, 1–5.

- Chen, Y. (2022). Mind the subtle f0 modifications: The interaction of tone and intonation in Sinitic varieties. *Stellenbosch Papers in Linguistics Plus*, 62(2), 113–136.
- Chen, Y., & Braun, B. (2006). Prosodic realization of information structure categories in standard Chinese. *Speech Prosody* 2006, paper 051-0.  
<https://doi.org/10.21437/SpeechProsody.2006-92>
- 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>
- Chen, Y., Xu, Y., & Guion-Anderson, S. (2015). Prosodic Realization of Focus in Bilingual Production of Southern Min and Mandarin. *Phonetica*, 71(4), 249–270.  
<https://doi.org/10.1159/000371891>
- Chen, Z. (1990). 关于普通话与方言的几个问题 [On several issues concerning Putonghua and dialects]. *Language Planning*, (04), 27–29.
- Chen Z. (2007). 语言的底层理论与底层分析方法 [Theory and analytical approaches to language substratum]. *Linguistic Sciences*, (6), 44–53.
- Comrie, B. (Ed.). (2018). *The World's Major Languages* (3rd edn). Routledge.  
<https://doi.org/10.4324/9781315644936>

- de Sousa, H. (2015). Language contact in Nanning: Nanning Pinghua and Nanning Cantonese. In H. M. Chappell (Ed.), *Diversity in Sinitic Languages* (pp. 157–189). Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780198723790.003.0007>
- de Sousa, H. (2022). On Pinghua and Yue: Some Historical and Linguistic Perspectives. *Crossroads*, 19(2), 257–295. <https://doi.org/10.1163/26662523-bja10004>
- de Sousa, H., Langella, F., & Enfield, N. J. (2015). Temperature terms in Lao, Southern Zhuang, Southern Pinghua and Cantonese. In M. Koptjevskaja-Tamm (Ed.), *Typological Studies in Language* (Vol. 107, pp. 594–638). John Benjamins Publishing Company.  
<https://doi.org/10.1075/tsl.107.19sou>
- DeFrancis, J. (1986). *The Chinese Language: Fact and Fantasy*. University of Hawaii Press.
- Ding, S. (1981). *古今字音对照手册 [A Handbook of Historical and Modern Chinese Pronunciations]*. Zhonghua Book Company.
- Duan, W., Jia, Y., & Ran, Q. (2013). Internal similarities and differences of phonetic realization of narrow focus in the Shandong dialect—A case study of Jinan, Liaocheng and Zibo dialects. *Journal of Tsinghua University (Science and Technology)*, 53(6), 37–40.
- Fan, F. (2020). *南宁普通话语音研究 [Nanning Mandarin Phonetic Research]* [Master]. Guangxi University.

- Féry, C. (Ed.). (2016a). 1 Introduction. In *Intonation and Prosodic Structure* (pp. 1–15). Cambridge University Press. <https://doi.org/10.1017/9781139022064.002>
- Féry, C. (Ed.). (2016b). 7 Tone and Stress at the Word Level. In *Intonation and Prosodic Structure* (pp. 178–224). Cambridge University Press. <https://doi.org/10.1017/9781139022064.008>
- Flaming, J. (2020). *Perception and production of Nanning Mandarin fourth tone* [Master of Arts (MA), University of North Dakota]. <https://commons.und.edu/theses/3269>
- Gu, W. (2016). 香港粤语人群的普通话声调语调偏误特征 [Tone and Intonation in Mandarin Utterances by HK Cantonese L2 Learners]. *南京师范大学文学院学报*, (02), 123–132.
- Gu, W., Hirose, K., & Fujisaki, H. (2006). Modeling the effects of emphasis and question on fundamental frequency contours of Cantonese utterances. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4), 1155–1170. <https://doi.org/10.1109/TASL.2006.876132>
- Gu, W., & Lee, T. (2007). Effects of tonal context and focus on Cantonese F0. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1033–1036.
- Gu, W., & Lee, T. (2009). Effects of tone and emphatic focus on F0 contours of Cantonese speech: A comparison with Standard Chinese. *Chinese Journal of Phonetics*, 133–147.

Gussenhoven, C. (2007). Types of Focus in English. In C. Lee, M. Gordon, & D. Büring (Eds), *Topic and Focus* (Vol. 82, pp. 83–100). Springer Netherlands.

[https://doi.org/10.1007/978-1-4020-4796-1\\_5](https://doi.org/10.1007/978-1-4020-4796-1_5)

Halliday, M. A. K. (1967). Notes on Transitivity and Theme in English: Part 2. *Journal of Linguistics*, 3(2), 199–244.

Han, W. (2013). *香港粤语语调初探 [Pilot Study in Hong Kong Cantonese]* [PhD, Nankai University].

[http://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFD1214&filename=1014166992.nh&uniplatform=OVERSEA&v=twzcc6c-9DuSd0nIVG9Djbc9kKY0G4m1vhGJD\\_Km\\_7C6jf3d1C86QlX0f6jCUuyO](http://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFD1214&filename=1014166992.nh&uniplatform=OVERSEA&v=twzcc6c-9DuSd0nIVG9Djbc9kKY0G4m1vhGJD_Km_7C6jf3d1C86QlX0f6jCUuyO)

Hinskens, F. (1998). Dialect Levelling—A Two-Dimensional Process. *Folia Linguistica*, XXII, 35–51. <https://doi.org/10.1515/flin.1998.32.1-2.35>

Hirst, D. (2011). The analysis by synthesis of speech melody: From data to models. *Journal of Speech Sciences*, 1(1), 55–83.

Hirst, D. (2022, March 16). *Pitch range*. Group.Io.

[https://groups.io/g/Praat-Users-List/topic/pitch\\_range/89818307](https://groups.io/g/Praat-Users-List/topic/pitch_range/89818307)

- Hong, J. (1989). 南宁白话的语音特点 [Phonetic Features of Nanning Cantonese]. *广西师范大学学报(哲学社会科学版)* [Journal of Guangxi Normal University (Philosophy and Social Science Edition)], (01), 47–54.
- Hsu, Y.-Y., Xu, A., & Ngai, H. (2018). Focus Prosody in Cantonese and Teochew Noun Phrases. *Speech Prosody 2018*, 961–965. <https://doi.org/10.21437/SpeechProsody.2018-194>
- Hu W., & Xu J. (2020). 地方普通话研究现状——基于 Citespace 的知识图谱分析 [Current state of research on Local Putonghua: A CiteSpace-based bibliometric analysis]. *品位·经典* *Pinwei Jingdian*, (9), 25–26, 30.
- Huang N. (2018). *A Study of the Use of Standard Chinese in Guangxi Zhuang Autonomous Region*. Social Sciences Academic Press (China). <https://gffhnc97f283abb114912s0uvwu6xnwwq56bx6ffgc.res.gxlib.org.cn/subject/30290169/>
- Huang, W. (2018). *南宁市汉语方言和壮语接触研究* [The Research on Language Contact about Chinese Dialects and Zhuang Language in Nanning City] [Master, Jilin University]. <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201901&filename=1018066127.nh&v=>
- Huo J. (2020). *A comparative study of Yue dialect modal particles in Nanning and Cenxi* [Master, Guangxi Minzu University]. <https://doi.org/10.27035/d.cnki.ggxmc.2019.000147>

Institute of Linguistics CASS (Ed.). (2012). *Language Atlas of China (Varieties of Ethnic Minority)* 中国语言地图集 少数民族语言卷 (2nd edn) [4 开]. The Commercial Press.  
<https://book.duxiu.com/bookDetail.jsp?dxNumber=000012813377&d=713472D9396C5E7635EB90DAB0BC4327&fenlei=080104&sw=%E4%B8%AD%E5%9B%BD%E8%AF%AD%E8%A8%80%E5%9C%B0%E5%9B%BE%E9%9B%86>

Institute of Linguistics CASS, 中国社会科学院民族学与人类学研究所, & 香港城市大学语言资讯科学研究中心 (Eds). (2012). *Language Atlas of China (Varieties of Chinese)* 中国语言地图集 汉语方言卷 (2nd edn) [4 开]. The Commercial Press.  
<https://book.duxiu.com/bookDetail.jsp?dxNumber=000012808987&d=54AA77B545F8231559879BB02DC29668&fenlei=080104&sw=%E4%B8%AD%E5%9B%BD%E8%AF%AD%E8%A8%80%E5%9C%B0%E5%9B%BE%E9%9B%86>

Jansen, N., Harding, E. E., Loerts, H., Başkent, D., & Lowie, W. (2023). The relation between musical abilities and speech prosody perception: A meta-analysis. *Journal of Phonetics*, 101, 101278. <https://doi.org/10.1016/j.wocn.2023.101278>

Jia, Y., Xiong, Z., & Li, A. (2006). Phonetic and Phonological Analysis of Focal Accents of Disyllabic Words in Standard Chinese. In Q. Huo, B. Ma, E.-S. Chng, & H. Li (Eds), *Chinese Spoken Language Processing* (Vol. 4274, pp. 55–66). Springer Berlin Heidelberg.  
[https://doi.org/10.1007/11939993\\_10](https://doi.org/10.1007/11939993_10)

- Jin, S. (1996). An acoustic study of sentence stress in Mandarin Chinese [Master, The Ohio State University]. In *ProQuest Dissertations and Theses*.  
<http://www.proquest.com/docview/304288041/abstract/832F83D8588D45F8PQ/1>
- Jin S. (2004). 中介语“僵化”的语言学意义 [The linguistic significance of fossilisation in interlanguage]. *Minority Languages of China*, (2), 32–37.
- Katz, J., & Selkirk, E. (2011). Contrastive Focus Vs. Discourse-New: Evidence from Phonetic Prominence in English. *Language*, 87(4), 771–816. <https://doi.org/10.1353/lan.2011.0076>
- Kerswill, P. (2003). Dialect levelling and geographical diffusion in British English. In D. Britain & J. Cheshire (Eds), *IMPACT: Studies in Language, Culture and Society* (Vol. 16, pp. 223–243). John Benjamins Publishing Company. <https://doi.org/10.1075/impact.16.16ker>
- Kerswill, P. (2010). Contact and new varieties. In R. Hickey (Ed.), *The handbook of language contact* (pp. 230–251). Wiley-Blackwell.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3–4), 243–276. <https://doi.org/10.1556/aling.55.2008.3-4.2>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26.  
<https://doi.org/10.18637/jss.v082.i13>

- Kwok, B.-C. (2010). 語言接觸中的語法變化：論南寧粵語「述語 + 賓語 + 補語」結構的來源 / Grammatical Change in Language Contact: On the Origin of the ‘Verb + Object + Complement’ Construction in Nanning Yue. *Journal of Chinese Linguistics Monograph Series*, (24), 201–216.
- Kwok, B.-C. (2014). The morpheme GO in three languages of the Nanning region: Paths of grammaticalization and contact-induced ‘replication’. *Language and Linguistics*, 15(5), 663–697. <https://doi.org/10.1177/1606822X14528640>
- Kwok, B.-C., Chin, A. C., & Tsou, B. K. (2011). Poly-functionality of the preverbal “acquire” in the Nanning Yue dialect of Chinese: An areal perspective. *Bulletin of the School of Oriental and African Studies*, 74(1), 119–137. <https://doi.org/10.1017/S0041977X10000431>
- Ladd, D. R. (2008). *Intonational Phonology* (2nd edn). Cambridge University Press. <https://doi.org/10.1017/CBO9780511808814>
- Ladd, D. R., Jr. (1980). *The Structure of Intonational Meaning: Evidence from English*. Indiana University Press. [https://muse.jhu.edu/pub/3/oa\\_monograph/book/84691](https://muse.jhu.edu/pub/3/oa_monograph/book/84691)
- Lan, L. (2007). 南宁普通话的句末语气词研究 [Research on the sentence-final particles in Nanning Mandarin]. *Journal of Guangxi University (Philosophy and Social Science)*, (S1), 112–113.

- Lau, M., Chen, Q., Fang, Y., Xu, T., Chen, T., & Golik, P. (2025). Data Quality Issues in Multilingual Speech Datasets: The Need for Sociolinguistic Awareness and Proactive Language Planning. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 7466–7492). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.acl-long.370>
- Lee, J. L., Chen, L., Lam, C., Lau, C. M., & Tsui, T.-H. (2022). PyCantonese: Cantonese Linguistics and NLP in Python. *Proceedings of the 13th Language Resources and Evaluation Conference*.
- Lee, W.-S. (2004). The Effect of Intonation on the Citation Tones in Cantonese. *Proceedings of TAL*, 107–110.
- Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33(1), 109–112. <https://doi.org/10.1017/S0025100303001208>
- Lee, W.-S., & Zee, E. (2008). Prosodic characteristics of the neutral tone in Beijing Mandarin/北京话轻声的韵律特征. *Journal of Chinese Linguistics*, 1–29.
- Lenth, R. V., & Piaskowski, J. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version R package version 2.0.0) [Computer software]. <https://rvlenth.github.io/emmeans/>

- Li, K. (2024). *Focal prominence in tone languages: A case study of three Chinese languages* [PhD, University of Cambridge].  
<https://www.repository.cam.ac.uk/items/e183e68b-98d9-4020-b2fe-b536f6129e5c>
- Li, K., Nolan, F., & Post, B. (2023). Variations of focus prominence in three tone languages. In R. Skarnitzl & J. Volín (Eds), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1420–1424). Guarant International.
- Li, R. (1988). 论方言和普通话之间的过渡语 [On the transitional varieties between dialects and Putonghua]. *Journal of Fujian Normal University (Philosophy and Social Sciences Edition)*, (02), 62–70.
- Li, S. (2002). The Comparative Study of the Intonation System of Cantonese in Nanning. *JOURNAL OF GUANGXI UNIVERSITY FOR NATIONALITIES (Philosophy and Social Science Edition)*, (3), 149–153.
- Li, T. (2012). 浅谈南宁白话对南宁方言普通话的声调影响 [A brief discussion of the impact of Nanning Cantonese on the tones of Nanning dialect Mandarin]. *剑南文学 (经典阅读)* [*Jiannan Literature (Classics Reading)*], (2), 97–98.
- Li, X. (2015). 在邕留学生对南宁普通话的认同研究 [International students' identity alignment towards Nanning Mandarin]. *Peony*, (18), 96–98.

- Li, X. (2018). *The acoustic experimental study on mono-syllable and bi-syllable tone in Cantonese in Guangxi Yongning District* [Master, Northeast Normal University].  
<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201901&filename=1018706010.nh&v=>
- Li, X., & Xiang, M. (2020). *汉语方言学基础教程 [Basic tutorials in Chinese dialectology]* (2nd edn). Peking University Press.
- Li, Y. (2012). *A STUDY OF THE MOTIVATION, THE VALUE, THE FUTURE TRENDS ABOUT DIALECT MANDARIN --- TAKE NANNING PEOPLE'S ORAL MANDARIN FOR INSTANCE* [Master, Guangxi University].  
<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201301&filename=1012495374.nh&v=>
- Liang G. (1994). 浅论南宁市多语现象 [A preliminary discussion of multilingualism in Nanning]. *Journal of Guangxi Teachers Education University (Philosophy and Social Sciences Edition)*, (3). <https://doi.org/10.16601/j.cnki.issn1002-5227.1994.03.015>
- Liang, Y. (2005). “南宁普通话”韵母研究 [A Study of the Finals in Nanning People's Oral Mandarin] [Master]. Guangxi University.
- Lin, C. (2015). 广西南宁语言使用情况调查分析——以青秀区长垌村與江南區白沙村為例 [A survey of language use in Nanning, Guangxi: A case study of Changgang Village (Qingxiu

*District) and Baisha Village (Jiangnan District)] [Master, Yunnan Normal University].  
<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201601&filename=1015903922.nh&v=>*

Lin, Y. (2008). 粤语（南宁）~普通话中介语 [Cantonese (Nanning) ~ Standard Mandarin interlanguage]. In J. Yu (Ed.), *广西推广普通话的理论与实践研究 [A Study of the theory and practice of promoting Putonghua in Guangxi]* (pp. 179–232).

Lin, Y., & Qin, F. (2008). *广西南宁白话研究 [A study of Cantonese in Nanning, Guangxi]*. Guangxi Normal University Press.

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>

Liu, J., Xu, Y., & Lee, Y. C. (2019). Post-focus compression is not automatically transferred from Korean to L2 English. *Phonetics and Speech Sciences*, 11(2), 15–21.

Liu, X. (2024). 南宁白话与广州话是非问句比较研究 [A Comparative Study of Yes–No Questions in Nanning Cantonese and Guangzhou Cantonese]. *今古文创 [Modern and Ancient Cultural Creation]*, (03), 133–136.

Lu X. (2012). 南宁普通话中介谓词探析 [An analysis of intermediate predicative structures in Nanning Mandarin]. *广西社会科学 [Social Sciences in Guangxi]*, (7), 152–154.

- Luo, R., Xu, J., Zhang, Y., Zhang, Z., Ren, X., & Sun, X. (2022). *PKUSEG: A Toolkit for Multi-Domain Chinese Word Segmentation* (arXiv:1906.11455). arXiv. <https://doi.org/10.48550/arXiv.1906.11455>
- Luo, T. (2015). *南宁白话单字调实验语音研究 [An Experimental Study on the Mono-Syllable of Nanning Cantonese]* [Master]. Guangxi University.
- Lyu, J., & Yu, J. (2021). 多语接触与通语变异——广西“南普”形成及变异研究之考察 [Multi-language Contact and Standard Language Variation: A Study on the Form and Variation of ‘Nanpu’ in Guangxi]. *西南交通大学学报(社会科学版) [Journal of Southwest Jiaotong University (Social Science)]*, 22(01), 47–55.
- Mai, Y. (2012). 粤语的形成、发展与粤语和平话的关系 Yueyu de xingcheng, fazhan yu Yueyu he Pinghua de guanxi. In *著名中年语言学家自选集 麦耘卷 Zhuming Zhongnian Yuyanxuejia Zixuanji, Maiyun Juan* (pp. 250–279). Shanghai Educational Publishing House.
- Mair, V. H. (1991). *What Is a Chinese ‘Dialect/Topolect’? —Reflections on Some Key Sino-English Linguistic Terms.*
- Man, C. H. V. (1999). *An acoustic study of the effects of sentential focus on Cantonese tones* [Master of Arts (MA), University of Victoria]. <https://library-archives.canada.ca/eng/services/services-libraries/theses/Pages/item.aspx?idNumber=52373486>

- Man, V. C. H. (2002). Focus effects on Cantonese tones: An acoustic study. *Speech Prosody 2002*, 467–470. <https://doi.org/10.21437/SpeechProsody.2002-102>
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Proceedings of the 18th Conference of the International Speech Communication Association*, 498–502. <https://doi.org/10.21437/Interspeech.2017-1386>
- Mok, P. K. P., & Zuo, D. (2012). The separation between music and speech: Evidence from the perception of Cantonese tones. *The Journal of the Acoustical Society of America*, 132(4), 2711–2720. <https://doi.org/10.1121/1.4747010>
- Nanning Municipal Bureau of Statistics & Nanning Survey Office of the National Bureau of Statistics. (2025). *Statistical Bulletin on Nanning's National Economic and Social Development in 2024*. Guangxi Zhuang Autonomous Region Bureau of Statistics. [https://english.nanning.gov.cn/Government\\_2025/Reports\\_2025/t6350719.html](https://english.nanning.gov.cn/Government_2025/Reports_2025/t6350719.html)
- Pater, J. (2009). Weighted Constraints in Generative Linguistics. *Cognitive Science*, 33(6), 999–1035. <https://doi.org/10.1111/j.1551-6709.2009.01047.x>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), Article 1. <https://doi.org/10.3758/s13428-018-01193-y>

- Qie, Y. (2015). 三十年来“地方普通话”研究综述 [A review of ‘Local Putonghua’ studies in the past 30 years]. *现代语文 (语言研究版) [Modern Chinese (Language Studies Edition)]*, (06), 4–6.
- Qin, F. (2005). 广西人的普通话面貌 [The Characteristics of Putonghua Spoken by Guangxi People]. In H. Chen & L. Li (Eds), *广西语言文字使用问题调查与研究 [Investigation of the theory and practice of promoting Mandarin in Guangxi]* (pp. 106–118). Guangxi Education Publishing House.
- Qin P. (2021). Intonation Evidence in Distinguishing Information Focus and Contrastive Focus in Chinese. *Contemporary Linguistics*, 23(1), 74–86.
- Selinker, L. (1969). Language Transfer. *General Linguistics*, 9(2), 67–92.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(1–4), 209–232. <https://doi.org/10.1515/iral.1972.10.1-4.209>
- Selkirk, E. O. (2008). Contrastive Focus, Givenness and the Unmarked Status of ‘Discourse-New’. *Acta Linguistica Hungarica*, 55(3–4), 331–346.
- Senin, P. (2008). Dynamic Time Warping Algorithm Review. *CSDL Technical Report*. [https://www.researchgate.net/publication/228785661\\_Dynamic\\_Time\\_Warping\\_Algorithm\\_Review](https://www.researchgate.net/publication/228785661_Dynamic_Time_Warping_Algorithm_Review)

- Shen, C., & Xu, Y. (2016). Prosodic focus with post-focus compression in Lan-Yin Mandarin. *Speech Prosody 2016*, 340–344. <https://doi.org/10.21437/SpeechProsody.2016-70>
- Shen F. (2005). 从南宁人的“不懂”漫谈地方普通话 [A linguistic reflection on Local Putonghua through the usage of ‘bù dǒng’ (‘don’t understand’) by Nanning people]. *Journal of Guangxi College of Education*, (3), 117–119.
- Shen, J. (1985). 北京话声调的音域和语调 [Pitch range of tone and intonation in Beijing dialect]. In T. Lin & L. Wang (Eds), *北京语音实验录 [Working papers in experimental phonetics]* (Vol. 84). Peking University Press.
- Shen X. (2000). 南宁人说普通话拾零 [Notes on how Nanning people speak Putonghua]. *Journal of Guangxi College of Education*, (5), 93–97.
- Shi, F. (1986). 天津方言双字组声调分析 [Analysis of bi-syllabic tones in Tianjin dialect]. *语言研究 [Studies in Language and Linguistics]*, (01), 77–90.
- Shi, F. (1991). 北京话的声调格局 [The Tonal Patterns of Beijing Mandarin]. *语言研究 [Studies in Language and Linguistics]*, 17–24.
- Shi, F. (1999). The manifestations of Chinese intonation pattern in different velocities. In F. Shi & W. Pan (Eds), *The new developing of Chinese linguistics* (pp. 381–394). City University of Hong Kong Press.

- Shi, F. (2008). *语音格局——语音学与音系学的交汇点* [*Phonetic Patterns: The Interface between Phonetics and Phonology*]. The Commercial Press.
- Shi, F., & Ran, Q. (2011). 普通话上声的本质是低平调——对《汉语平调的声调感知研究》的再分析 [A Comment on “Mandarin Tone Perception: A Report on a Low Level Tone”]. *中国语文* [*Studies of the Chinese Language*], (6), 550-555+576.
- Shi, F., Wang, P., & Liang, L. (2009). 汉语普通话陈述句语调的起伏度 [The Undulating Scale of the Intonation of Declarative Sentences in Standard Chinese]. *Nankai Linguistics*, 2.
- Shi F., & Xia Q. (2022). 试论汉语普通话语句的焦点原理 [The Principles of Sentence Focus in Mandarin Chinese]. *语言科学 Linguistic Sciences*, 21(6), Article 6.  
<https://doi.org/10.7509/j.linsci.202208.035209>
- Shi S. (2007). *《现代汉语方言音库》单字调实验研究* [*An Experimental Study of Monosyllabic Tone in Phonetic Database of Modern Chinese Dialects*] [Master, Nanjing Normal University].  
<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2007&filename=2007107915.nh>
- Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 83–109.

- Siegel, J. (1985). Koines and koineization. *Language in Society*, 14(3), 357–378.  
<https://doi.org/10.1017/S0047404500011313>
- Siegel, J. (2001). 7. Koine formation and creole genesis. In N. Smith & T. Veenstra (Eds), *Creole Language Library* (Vol. 23, pp. 175–198). John Benjamins Publishing Company.  
<https://doi.org/10.1075/cll.23.08sie>
- So, K. K., Xu, C., Cao, G. W., & Mok, P. (2025). *Performance of Montreal Forced Aligner on Cantonese Spontaneous Speech*. 5398–5402. <https://doi.org/10.21437/Interspeech.2025-1882>
- Sonderegger, M., & Sóskuthy, M. (2025). *Advancements of Phonetics in the 21st century: Quantitative data analysis*. OSF. [https://osf.io/mc6a9\\_v2](https://osf.io/mc6a9_v2)
- Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction* (arXiv:1703.05339). arXiv. <https://doi.org/10.48550/arXiv.1703.05339>
- Tang, C., van Heuven, V. J., Heeringa, W., & Gooskens, C. (2025). Chinese “Dialects” and European “Languages”: A Comparison of Lexico-Phonetic and Syntactic Distances. *Languages*, 10(6), 127. <https://doi.org/10.3390/languages10060127>
- Tang, Q. (2020). *广西汉语方言概要 [An Overview of Chinese Dialects in Guangxi]* (p. 333). World Publishing Guangdong Co., Ltd.

- Tang, Q., & Liu, Y. (2016). 南宁普通话陈述句语调实验研究 [An experimental investigation of the intonation of declarative sentences in Nanning Mandarin]. *Journal of Mudanjiang University*, 25(09), 12–15.
- Tang, Z. (2016). 音韻學教程 [A Course in Phonology] (5th edn). Peking University Press.  
<https://book.douban.com/subject/26838113/>
- Teng, Z. (2018). 南宁市与桂平市粤方言比较研究 [The Comparative Study of Cantonese Dialect in Nanning and Guiping] [Master, Guangxi Normal University].  
[https://kns.cnki.net/kcms/detail/frame/list.aspx?dbcode=CMFD&filename=1018850873.nh&dbname=CMFD201901&RefType=1&vl=xSe5IINbGr7-ebWRQHhJihXih9puRsZxKjEetwmGF15OT5gA2h7\\_YP6pI2p26fOI](https://kns.cnki.net/kcms/detail/frame/list.aspx?dbcode=CMFD&filename=1018850873.nh&dbname=CMFD201901&RefType=1&vl=xSe5IINbGr7-ebWRQHhJihXih9puRsZxKjEetwmGF15OT5gA2h7_YP6pI2p26fOI)
- Trudgill, P. (1986). *Dialects in Contact*. Wiley-Blackwell.
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, D. (2015). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. <https://cran.r-project.org/package=itsadug>
- Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, 33(5), 368–392. <https://doi.org/10.1159/000259793>
- Visceglia, T., Su, C., & Tseng, C. (2012). Comparison of English narrow focus production by L1 English, Beijing and Taiwan Mandarin speakers. *2012 International Conference on Speech Database and Assessments*, 47–51. <https://doi.org/10.1109/ICSDA.2012.6422454>

- Wan, X. (2012). *音韵学要略 [Essentials of Phonology]* (2nd edn). Wuhan University Press.
- Wang, B., Li, C., Wu, Q., Zhang, X., Wang, B., & Xu, Y. (2012). Production and Perception of Focus in PFC and non-PFC Languages: Comparing Beijing Mandarin and Hainan Tsat. *Interspeech 2012*, 663–666. <https://doi.org/10.21437/Interspeech.2012-204>
- Wang, B., Wang, L., & Kadir, T. (2011). Prosodic encoding of focus in six languages/dialects in China. *Proceedings of the 17th International Congress of Phonetic Sciences*, 144–147.
- Wang, S. (2015). 南宁白话(新派)单字调声学实验研究 [An acoustic experimental study on the monosyllabic tones in Nanning Cantonese from the younger generation]. *语文学刊 [Journal of Language and Literature Studies]*, (04), 37–39.
- Wang, Z. (2023). *南宁市城中村语言生态研究 [Research on Language Ecology of the Urban Villages in Nanning]* [PhD, 广西民族大学]. <http://61.175.198.136:8083/rwt/288/http/GEZC6MJZFZZUPLSSG63B/kcms/detail/detail.aspx?recid=&FileName=1023746704.nh&DbName=CDFDLAST2024&DbCode=CDFD&uid=WEF2dXFFeHgrYVFFLySwa2xqNGkvYUsvYXZsMGRvSXA1bXJjM3hNaXkyRmdYL0lj>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>

- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, 70, 86–116. <https://doi.org/10.1016/j.wocn.2018.03.002>
- Wood, S. N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(1), 3–36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>
- Wu W. (2001). 论桂南平话的粤语系属 [On the Cantonese Affiliation of Guinan Pinghua]. *方言 [Fangyan]*, (2), 133–141.
- Wu, W. L. (2013). *Cantonese prosody: Sentence-final particles and prosodic focus* [PhD Thesis, UCL (University College London)]. <https://discovery.ucl.ac.uk/id/eprint/1400569/>
- Wu, W. L., & Chung, L. (2011). POST-FOCUS COMPRESSION IN ENGLISH-CANTONESE BILINGUAL SPEAKERS. *Proceedings of ICPHS 2011*, 148–151.
- Wu, W. L., & Xu, Y. (2010). Prosodic Focus in Hong Kong Cantonese without Post-focus Compression. *Proceedings of Speech Prosody 2010*, 1–4.
- Wu, Y. (2017). *An experimental investigation of the intonation of interrogative sentences in Nanning Mandarin* [Master, Guangxi University].

<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201801&filename=1017228848.nh&v=>

- Xian W., & Yang B. (2022). Tone-merger from the perspective of perception and production: Focusing on Nanning Cantonese. *Linguistic Sciences*, 21(4), Article 4. <https://doi.org/10.7509/j.linsci.202201.034918>
- Xie, J. (1994). 南宁白话同音字汇 [Nanning Cantonese homophone glossary]. *方言 [Fangyan]*, (04), 286–303.
- Xie, J. (2007). *广西汉语方言研究 [Research on Sinitic Dialects in Guangxi]* (p. 1065). 广西人民出版社 [Guangxi People's Press].
- Xu, C., & Zhang, C. (2024). A cross-linguistic review of citation tone production studies: Methodology and recommendations. *The Journal of the Acoustical Society of America*, 156(4), 2538–2565. <https://doi.org/10.1121/10.0032356>
- 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(3–4), 220–251. <https://doi.org/10.1016/j.specom.2005.02.014>

- Xu, Y. (2011). POST-FOCUS COMPRESSION: CROSS-LINGUISTIC DISTRIBUTION AND HISTORICAL ORIGIN. *Proceedings of ICPHS 2011*, 152–155.
- Xu, Y. (2019). Prosody, Tone, and Intonation. In W. F. Katz & P. F. Assmann (Eds), *The Routledge Handbook of Phonetics* (pp. 314–256). Routledge.  
<https://doi.org/10.4324/9780429056253-13>
- 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).  
<https://doi.org/10.1515/tlr-2012-0006>
- 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>
- Yang, H. (1997). *南宁话音档 [Sound archive of Nanning languages]* (H. Yang, Ed.). Shanghai Educational Publishing House.
- Yang, H. (2003). 南宁普通话的语调、语气词及其声学表现 [Intonation and Model Particles Used in Nanning Putonghua and Their Manifestations in Acoustics]. *语言研究 [Studies in Language and Linguistics]*, 23(04), 104–109.
- Yang, H. (2008). 12 广西话~普通话中介语的韵律特征 [Prosodic Features of the Guangxi Languages ~ Standard Mandarin Interlanguage]. In J. Yu (Ed.), *广西推广普通话的理论*

- 实践研究 [A Study of the theory and practice of promoting Putonghua in Guangxi] (pp. 451–470). Shanghai Educational Publishing House.
- Yang, Y. (2022). *First language attrition and second language attainment of Mandarin-speaking immigrants in Hong Kong: Evidence from prosodic focus* [Ph.D., Hong Kong Polytechnic University]. <https://theses.lib.polyu.edu.hk/handle/200/11661>
- Yang, Y., Chen, S., & Li, K. (2018). Pitch realization of post-focus components in Chongming Chinese. *The Journal of the Acoustical Society of America*, 144(3\_Supplement), 1938–1938. <https://doi.org/10.1121/1.5068477>
- Yao, Y. (1989). 应该开展对“地方普通话”的研究 [Research should be conducted on ‘Local Standard Mandarin’]. *语文建设 [Language Planning]*, (03), 20–23.
- Yip, M. (2002). *Tone*. Cambridge University Press.
- Yu, J. (Ed.). (2008). *广西推广普通话的理论与实践研究 [A Study of the theory and practice of promoting Putonghua in Guangxi]*. Shanghai Educational Publishing House.
- Yu, J., & Qin, F. (2008). 1 广西推广普通话的实践 [The practice of promoting Putonghua in Guangxi]. In J. Yu (Ed.), *广西推广普通话的理论与实践研究 [A Study of the theory and practice of promoting Putonghua in Guangxi]* (pp. 1–40). Shanghai Educational Publishing House.

- Yuan J. (1960). *汉语方言概要 [An Outline of Chinese Dialects]*. 文字改革出版社 [Chinese Character Reform Press].
- Yuan, J. (2004). Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling [Ph.D., Cornell University]. In *ProQuest Dissertations and Theses*.  
<http://www.proquest.com/docview/305205747/abstract/E4A16AB8103D43A7PQ/1>
- Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin Broadcast News Speech. *Speech Communication, 65*, 67–74. <https://doi.org/10.1016/j.specom.2014.06.001>
- Zhang, C., Jepson, K., Lohfink, G., & Arvaniti, A. (2021). Comparing acoustic analyses of speech data collected remotely. *The Journal of the Acoustical Society of America, 149*(6), 3910.  
<https://doi.org/10.1121/10.0005132>
- Zhang, J., Duanmu, S., & Chen, Y. (2020). 22 China and Siberia. In C. Gussenhoven & A. Chen (Eds), *The Oxford Handbook of Language Prosody*. Oxford University Press.  
<https://doi.org/10.1093/oxfordhb/9780198832232.013.21>
- Zhang, L. (2013). *Intonation effects on Cantonese lexical tones in speaking and singing* [PhD, Hong Kong Polytechnic University]. <https://theses.lib.polyu.edu.hk/handle/200/7275>
- Zhang, L. (2022). 南宁地区语码转换现象的语用功能探析——以《小南宁爱南宁》为例 [A Pragmatic Analysis of Code-Switching in the Nanning Area: A Case Study of ‘Xiao Nanning, Ai Nanning’ ]. *汉字文化 [Sinogram Culture]*, (S2 期 ).

<http://61.175.198.136:8083/rwt/288/http/GEZC6MJZFZZUPLSSG63B/kcms/detail/detail.aspx?recid=&FileName=HZWZ2022S2019&DbName=CJFDLAST2023&DbCode=CJFQ&uid=WEF2dXFFeHgrYVFFLyswa2xqNGkvYUsvYXZsMGRvSXA1bXJjM3hNaXkyRmdYL0lj>

Zhao, L. (2024). *PRODUCTION AND PERCEPTION OF LEXICAL TONE VARIATION IN MANDARIN DIALECTS* [PhD Thesis]. University of York.

Zhou L. (2017). The “Reference Realization” of Tone 3 and the Identification of Tone Contour Realization. *Modern Linguistics*, 05(03), 254–260. <https://doi.org/10.12677/ML.2017.53035>

Zhu, X. (2003). 基频归一化——如何处理声调的随机差异? [F0 Normalization: How to Deal with Between speaker Tonal Variations?]. *语言科学 [Linguistic Sciences]*, 3(2), 3–19.

北京大学中文系现代汉语教研室 [Teaching and Research Section of Modern Chinese, Department of Chinese Language and Literature, Peking University] (Ed.). (2012). *现代汉语 [Modern Chinese]*. The Commercial Press.

南宁市地方志编纂委员会 [Nanning Municipal Local Records Compilation Commission]. (1998). *南宁市志 文化卷 [Nanning Gazetteer: Culture Volume]* (p. 800). 广西人民出版社 [Guangxi People’s Press].

广西壮族自治区地方志编纂委员会 [Guangxi Zhuang Autonomous Region Local Documentation  
Compilation Committee] (Ed.). (1998). *广西通志·汉语方言志* [*Comprehensive  
documentation of Guangxi: Sinitic dialects*]. Guangxi People's Press.