CROSS-LINGUISTIC PROCESSING OF PROSODIC FOCUS IN L1 CANTONESE AND L3 MANDARIN BY TRILINGUAL ADULTS AND CHILDREN

BEI LI

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Cross-Linguistic Processing of Prosodic Focus in L1 Cantonese and L3 Mandarin by Trilingual Adults and Children

Bei Li

A thesis submitted in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

April 2023

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

Bei Li (Name of student)

Abstract

Speech prosody is an essential aspect of communication, together with co-speech visual information, aiding in the comprehension of new information in face-to-face conversations. Speech prosody can play a vital role in encoding information structure. Specifically, focus can be marked by duration lengthening, *fundamental frequency* (F_0) and intensity increment. In tonal languages, F_0 is employed both in conveying lexical meanings and marking focus, which may lead to some challenges in decoding focus. The degree to which tonal language listeners utilize linguistic and paralinguistic knowledge to decode prosodic focus in multimodal speech remains to be investigated. This dissertation seeks to investigate the decoding strategy for prosodic focus in multimodal speech in two tonal languages Cantonese and Mandarin by trilingual adults and children who speak Cantonese as their first language (L1) and Mandarin as their third language (L3). Their performance is also compared to native Mandarin-speaking adults and children.

This study conducted audio-only and audio-visual perceptual experiments in Cantonese and Mandarin, using a question-answer congruence paradigm with trilingual adults and school-aged children. Participants rated the naturalness of congruous and incongruous utterances elicited by different precursor questions involving focus categories, such as broad, narrow and contrastive focus on the verb. The utterances were designed to incorporate various tonal contexts, featuring both level and contour tones in various combinations.

Trilingual adults demonstrate the ability to effectively map narrow and contrastive focus onto the correct prosodic realizations, however, they often encounter challenges in identifying broad focus in their native Cantonese. They make probabilistic inferences

regarding category-form mapping, displaying biases in favor of narrow focus but against broad focus in decoding prosodic focus. The perceptual biases were also manifested in the processing of L3 prosodic focus by trilingual adults, but not in native Mandarin-speaking adults, who demonstrated different interpretations of all focus categories compared to trilingual adults. Overall, trilingual adults heavily rely on their prior knowledge and use probabilistic inference in processing the prosodic focus forms in relation to information structure across languages.

Trilingual children do not exhibit the ability comparable to trilingual adults in making probabilistic inferences when mapping between focus categories and prosodic forms in Cantonese, although their accuracy in assessing the naturalness of prosodic focus is on par with that of adults. In L3, their ability to rate prosodic naturalness is equivalent to that of their native-speaking peers, whereas trilingual children exhibit asymmetric expectations for prosodic prominence between contrastive focus and narrow focus in Mandarin. Like adults, the influence of tonal contexts and visual cues on trilingual children's focus processing in either L1 or L3 was found to be negligible. The trilingual children have yet to attain an entire adult-like or native-like competence to utilize prior knowledge and make probabilistic inferences when processing prosody in their L1 or L3.

This dissertation presents a comprehensive framework, the Prosody Processing Model in L1/L3, for probing the adoption of the probabilistic approach to focus decoding by trilingual children and adults in both their L1 and L3. By contributing to the field of probabilistic models in speech processing, this study highlights the need for further research on the developmental process of probabilistic inference in cross-linguistic contexts to enhance our understanding of prosody comprehension.

Acknowledgments

I would like to express my deepest appreciation and gratitude to my supervisor, Dr. Si Chen, for her unwavering support throughout my research journey at the Hong Kong Polytechnic University. Her expertise, patience, and inspiring personality have been instrumental in shaping my academic pursuits and personal growth. I am particularly grateful for her constant guidance and encouragement, which have been a source of motivation and inspiration for me. Her passion for research and her ability to generate new ideas have consistently inspired me to push my own limits and strive for progress in my academic work.

I would also like to extend my thanks to my co-supervisor, Dr. Angel Chan, and the members of the examination panel of Ph.D. registration, Dr. Roxana Fung and Dr. Yao Yao. I am grateful to the members of the examination committee, Professor Aoju Chen and Professor Ying Chen, for agreeing to be the external examiners and for their time in reviewing my work. Their comments and suggestions help me to improve the quality of my thesis and broaden my perspectives on my research topic.

Furthermore, I am deeply indebted to Professor John Wakefield, Professor Lian Hee Wee, Professor Yueyuan Huang and Professor Xiujuan Shi for their support and guidance during my undergraduate and graduate studies. Their encouragement, inspiration, and mentorship have played a pivotal role in igniting my passion for research and fostering my academic growth.

I would like to express my gratitude to all participants of my experiments, as well as the parents of child participants. Without their willingness to participate, this research would not have been possible. Their contribution to this study has been invaluable, and I

am deeply grateful for their cooperation. I would also like to thank all my friends who have helped in spreading the poster to recruit participants for my experiments.

I am also grateful to the faculty of Chinese and Bilingual Studies of PolyU, particularly to Professor Chu-Ren Huang, Dr. Stephen Politzer-Ahles and Dr. Emmanuele Chersoni, for their impressive lectures, insightful teaching, and guidance. I would also like to extend my heartfelt thanks to the staff of the department for providing me with a supportive research environment and resources that have made my PhD journey exceptionally smooth and rewarding. I am particularly grateful to my friend Dr. Yike Yang, who provided valuable guidance and suggestions for my research study, as well as to Dr. Yin Zhong and Ms. Xi Chen, who were always benevolent and assisted me with technical difficulties. I would also like to express my appreciation to my colleagues and friends, including Dr. Jing Shao, Ms. Xin Luo, Mr. Xuefeng Gao (d. 2020), Mr. Yubin Zhang, Dr. Yanlin Li, Ms. Yitian Hong, Ms. Yuzhe Meng, Ms. Xinyi Chen, Ms. Jing Chen, Ms. Viann Hui, Ms. Carmen Wu, Dr. Siyu Lei, Ms. Yingnan Li, and the members of Dr. Si Chen's lab. They all supported and encouraged me throughout my work and studies at PolyU.

Finally, I would like to extend my heartfelt thanks to my grandmother, my parents, my boyfriend, and my old friends outside of Hong Kong for their unwavering support, love, and encouragement. I am truly grateful for their presence in my life.

Table of Contents

Abstract I

Acknowledgments III

Table of Contents V

List of Figures VIII

List of Tables X

List of Abbreviations XI

Chapter 1. Introduction 1

1.1 Information structure 2

1.2 Prosodic realizations of information structure 4

1.3 Research questions of the current study 5

1.4 Dissertation outline 8

Chapter 2. Literature review 10

2.1 Prosodic focus processing 10

2.1.1 Acoustic variability and focus perception 10

2.1.2 Listener expectations on focus processing 13

2.1.3 The probabilistic modeling in speech comprehension 18

2.2 Prosodic focus acquisition in children 20

2.2.1 Prosodic focus-marking in children 20

2.2.2 Prosodic focus-processing in children 23

2.3 Cross-linguistic prosodic focus in Cantonese and Mandarin 28

2.3.1 Prosodic focus in Cantonese and Mandarin 28

2.3.2 Cross-linguistic prosodic focus 33

2.4 Multimodal focus processing 35

Chapter 3. Experiment 1: Cantonese focus processing in trilingual adults 39

3.1 Introduction 39

3.2 Method 42

3.2.1 Participants 42

3.2.2 Materials and design 43

3.2.3 Procedures 59

3.2.4 Data analyses 64

3.3 Results 65
3.3.1 Focus processing of congruous prosody in Cantonese by trilingual adults 65
3.3.2 Focus processing of incongruous prosody in Cantonese by trilingual adults 69
3.3.3 Comparison between congruous and incongruous focus conditions in Cantonese 71
3.4 Discussion 79
Chapter 4. Experiment2: L3 Mandarin focus processing in trilingual adults 89
4.1 Introduction 89
4.2 Method 91
4.2.1 Participants 91
4.2.2 Materials and design 93
4.2.3 Procedures 104
4.2.4 Data analyses 104
4.3 Results 105
4.3.1 Focus processing of congruous prosody in L3 Mandarin by trilingual adults 105
4.3.2 Focus processing of incongruous prosody in L3 Mandarin by trilingual adults 108
4.3.3 Comparison between congruous and incongruous focus conditions in L3 Mandarin by trilingual adults 111
4.4 Discussion 121
Chapter 5. Experiment 3: Cantonese focus processing in trilingual children 129
5.1 Introduction 129
5.2 Method 131
5.2.1 Participants 131
5.2.2 Materials and design 132
5.2.3 Procedures 132
5.2.4 Data analyses 132
5.3 Results 133
5.3.1 Focus processing of congruous prosody in Cantonese by trilingual children 134
5.3.2 Focus processing of incongruous prosody in Cantonese by trilingual children 136
5.3.3 Comparison between congruous and incongruous focus conditions in Cantonese by trilingual children 138
5.4 Discussion 146
Chapter 6. Experiment 4: L3 Mandarin focus processing in trilingual children 151
6.1 Introduction 151
6.2 Method 153

6.2.1 Participants 153
6.2.2 Materials and design 154
6.2.3 Procedures 154
6.2.4 Data analyses 155
6.3 Results 155
6.3.1 Focus processing of congruous prosody in L3 Mandarin by trilingual children 156
6.3.2 Focus processing of incongruous prosody in L3 Mandarin by trilingual children 158
6.3.3 Comparison between congruous and incongruous focus conditions in L3 Mandarin by trilingual children 160
6.4 Discussion 168
Chapter 7. Concluding Remarks 175
7.1 Summary of findings 175
7.2 General discussion 180
7.2.1 Cross-linguistic approach to focus processing in trilingual adults 180
7.2.2 Cross-linguistic approach to focus processing in trilingual children 185
7.3 The Prosody Processing Model in L1/L3 (PPM-L1/L3): A working model 190
7.4 Limitations and future directions 202
7.5 Conclusions 204
Appendices 205
Appendix 1. Target sentences in Cantonese for Experiments 1 and 3 205
Appendix 2. Filler sentences in Cantonese for Experiments 1 and 3 208
Appendix 3. Target sentences in Mandarin for Experiments 2 and 4 211
Appendix 4. Filler sentences in Mandarin for Experiments 2 and 4 214
Appendix 5. Statistical results of Experiment 1 216
Appendix 6. Statistical results of Experiment 2 219
Appendix 7. Statistical results of Experiment 3 222
Appendix 8. Statistical results of Experiment 4 225
References 228

List of Figures

Figure 3.1 F_0 time normalized contours of Cantonese target sentences 52
Figure 3.2 F_0 range by the focus of Cantonese target sentences..... 52
Figure 3.3 Duration by the focus of Cantonese target sentences..... 53
Figure 3.4 Intensity by the focus of Cantonese target sentences 53
Figure 3.6 Native adults' performance in incongruous Cantonese prosody..... 70
Figure 3.7 Adults' performance in Cantonese focus contrasts with different prosodic forms (Narrow-Broad). 73
Figure 3.8 Adults' performance in Cantonese focus contrasts with different prosodic forms (Contrastive-Broad). 74
Figure 3.10 Adults' performance in Cantonese focus contrasts with different information structures (Narrow-Broad). 77
Figure 3.11 Adults' performance in Cantonese focus contrasts with different information structures (Contrastive-Broad). 79
Figure 4.1 F_0 time normalized contours of Mandarin target sentences 99
Figure 4.3 Duration by the focus of Mandarin target sentences 100
Figure 4.4 Intensity by the focus of Mandarin target sentences 100
Figure 4.5 Adults' performance in congruous Mandarin prosody..... 106
Figure 4.6 Adults' performance in incongruous Mandarin prosody..... 109
Figure 4.7 Adults' performance in Mandarin focus contrasts with different prosodic forms (Narrow-Broad). 113
Figure 4.8 Adults' performance in Mandarin focus contrasts with different prosodic forms (Contrastive-Broad). 115
Figure 4.10 Adults' performance in Mandarin focus contrasts with different information structures (Narrow-Broad). 118
Figure 4.11 Adults' performance in Mandarin focus contrasts with different information structures (Contrastive-Broad). 120
Figure 5.1 Children and adults' performance in congruous Cantonese prosody..... 135
Figure 5.2 Children and adults' performance in incongruous Cantonese prosody..... 137
Figure 5.3 Children and adults' performance in Cantonese focus contrasts with different prosodic forms (Narrow-Broad). 140
Figure 5.4 Children and adults' performance in Cantonese focus contrasts with different prosodic forms (Contrastive-Broad). 142
Figure 5.6 Children's performance in Cantonese focus contrasts with different information structures (Narrow-Broad) 144
Figure 6.1 Native and non-native children's performance in congruous Mandarin prosody 157
Figure 6.2 Native and non-native children's performance in incongruous Mandarin prosody..... 159
Figure 6.3 Children's performance in Mandarin focus contrasts with different prosodic forms (Narrow-Broad). 162
Figure 6.4 Children's performance in Mandarin focus contrasts with different prosodic forms (Contrastive-Broad). 164
Figure 6.6 Children's performance in Mandarin focus contrasts with different information structures (Narrow-Broad) 166

Figure 7.1 The process of focus mapping by English listeners (Based on Roettger et al., 2019) 195
Figure 7.2 The process of focus mapping by adult listeners in L1 and L3 199
Figure 7.3 The process of focus mapping by child and adult listeners in L1 200
Figure 7.4 The process of focus mapping by child listeners in L1 and L3 201
Figure 3.9 Adults' performance in Cantonese focus contrasts with different prosodic forms (Narrow-Contrastive) 217
Figure 3.12 Adults' performance in Cantonese focus contrasts with different information structures (Narrow-Contrastive). 218
Figure 4.9 Adults' performance in Mandarin focus contrasts with different prosodic forms (Narrow-Contrastive) 220
Figure 4.12 Adults' performance in Mandarin focus contrasts with different information structures (Narrow-Contrastive). 221
Figure 5.5 Children and adults' performance in Cantonese focus contrasts with different prosodic forms (Narrow-Contrastive). 222
Figure 5.8 Children's performance in Cantonese focus contrasts with different information structures (Narrow-Contrastive). 224
Figure 6.5 Children's performance in Mandarin focus contrasts with different prosodic forms (Narrow-Contrastive) 225
Figure 6.8 Children's performance in Mandarin focus contrasts with different information structures (Narrow-Contrastive). 227

List of Tables

Table 1.1 Examples of focus categories elicited in Q-A dialogues 3
Table 1.2 Focus positions in SVO languages 4
Table 2.1 Six lexical tones in Hong Kong Cantonese 29
Table 2.2 Four lexical tones in Mandarin 29
Table 3.1 Tonal combinations of Cantonese target sentences 45
Table 3.2 Examples of Cantonese sentences in Experiment 1 46
Table 3.3 Focus categories elicited by Q-A paradigm in Cantonese 49
Table 3.4 Examples of matched and mismatched pairs used in Experiment 1 57
Table 3.5 Examples of matched and mismatched pairs for practice sessions in Experiment 1 61
Table 3.6 Post-hoc analyses of focus contrasts between broad focus and narrow-verb/contrastive-verb focus 68
Table 4.1 Tonal combinations of Mandarin target sentences 94
Table 4.2 Examples of Mandarin sentences in Experiment 2 94
Table 4.3 Focus categories elicited by Q-A paradigm in Mandarin 97
Table 4.4 Examples of matched pairs and mismatched pairs in Experiment 2 103
Table 3.7 Post-hoc analyses of mismatched focus conditions in Experiment 1 216
Table 3.8 Post-hoc analyses of focus contrasts (NV vs. NB) in Experiment 1 216
Table 3.9 Post-hoc analyses of mismatched focus conditions in Experiment 2 219

List of Abbreviations

BF	The matched pairs with **B**road **F**ocus: Questions of broad focus context - Answers of broad focus acoustic form
NV	The matched pairs with **N**arrow-**V**erb focus: Questions of narrow-verb focus context - Answers of narrow-verb focus acoustic form
CV	The matched pairs with **C**ontrastive-**V**erb focus: Questions of contrastive-verb focus - Answers of contrastive-verb focus acoustic form
NB	The mismatched pairs with broad focus: Questions of **N**arrow-verb focus context - Answers of **B**road focus acoustic form
CB	The mismatched pairs with narrow-verb focus: Questions of **C**ontrastive-verb focus context - Answers of **B**road focus acoustic form
NC	The mismatched contrastive-verb focus: Questions of **N**arrow-verb focus context - Answers of **C**ontrastive-verb focus acoustic form
LL	Sentences contexts of level-level lexical tones **L**evel tones on the subject and **L**evel tones on the verb phrase in Cantonese
RL	Sentences contexts of rising-level lexical tones **R**ising tones on the subject and **L**evel tones on the verb phrase in Cantonese
LR	Sentences contexts of level-rising lexical tones **L**evel tones on the subject and **R**ising tones on the verb phrase in Cantonese
RR	Sentences contexts of rising-rising lexical tones **R**ising tones on the subject and **R**ising tones on the verb phrase in Cantonese
F_0	Fundamental frequency
L1	First language
L2	Second language
L3	Third language
LILt	L2 Intonation Learning Theory
PAM-L2	Perceptual Assimilation Model of L2
PFC	Post focus compression
PPM-L1/L3	The Prosody Processing Model in L1/L3
SLM/SLM-r	Speech learning model/ Speech learning model revised

Chapter 1. Introduction

Prosodic modulation plays a vital role in speech communication, as it not only conveys speaker attitudes and emotions but also expresses linguistic meanings (e.g., speech act, information status: Krifka, 2008; Pierrehumber, 1980; Ladd, 1996; 2008; Gussenhoven, 2004). Notably, prosody can encode the information structure of utterances through phonetic implementation for effective communication. For instance, speakers utilize prosodic modulations (e.g., long duration, expanded pitch range on acoustic levels) to mark focus to emphasize new information to their interlocutors. By emphasizing the word "frog" in the sentence "Kevin caught a [frog]$_F$", the interlocutor emphasizes the object "frog" to the listener who probably asked, "What did Kevin catch?". The object "frog", indicating relatively new information to the preceding context, expresses the narrow focus of the utterance. In speech communication, adult listeners can process multiple linguistic or paralinguistic information to comprehend prosodic speech. Thus, prosodic competence in adults entails that children develop an efficient interpretation of prosodic focus in an adult-like manner during specific developmental periods. Similarly, learners of second or third languages have the potential to acquire comparable competence like native speakers in processing prosodic focus.

Tonal languages, such as Cantonese and Mandarin, use prosodic modulations to encode focus and discriminate lexical tones simultaneously. Consequently, the acoustic cues, particularly *fundamental frequency* (F_0), serve parallel functions in speech communication for tonal speakers. Despite this, the extent to which and how native tonal listeners can decode prosodic focus remains relatively unknown, particularly within

perplexing tonal contexts. Furthermore, the degree to which tonal listeners utilize visual cues, expressed by speakers during face-to-face communication to facilitate their processing of prosodic modulations has received little attention in current literature. This dissertation aims to investigate the perceptual mappings of information structure onto prosodic forms in audio-visual modalities for trilingual adults in their first language (L1) Cantonese and third language (L3) Mandarin.

Furthermore, it is still unknown whether school-aged children's have developed the ability to comprehend various prosodic focus in their native or non-native tonal languages. Cantonese-speaking children, in particular, are exposed to a multilingual environment and learn English and Mandarin as their second and third languages from an early age. Our knowledge of how trilingual children decode prosody in their L3 to guide their detection of speaker intentions is surprisingly limited. An interesting question arises as to how the co-existence of tones and prosodic focus influences the processing of focus by Mandarin learners who speak Cantonese as their L1. Moreover, it would be interesting to investigate whether trilingual children benefit from the visual cues in multimodal speech comprehension. This study also aims to contribute to research on the development of adult-like processing of prosodic focus in L1 Cantonese and native-like competence in L3 Mandarin, particularly in the context of multimodal inputs, for trilingual children.

1.1 Information structure

In spontaneous speech communication, focus is a crucial component of information structure that the speaker uses to convey their intention. It introduces changes in the common ground shared by the speaker and the listener by emphasizing a syntactic constituent, such as a phrase or a sentence, or an element of a syntactic constituent, such

as the noun of a phrase (Lambrecht, 1994; Rooth, 1992; Krifka, 2008; Chen, 2018). The information structure within a sentence is divided into two main components : the focus, which conveys the most important information and is emphasized, and the background, which provides contextual or given information and is deemphasized by the speaker. From the listener's perspective, focus refers to relatively new information that has not been mentioned in the previous context. In contrast, background refers to given information in the context. In the upcoming dialogues, for instance, the question "(1q) What did Kevin catch?" appropriately elicits the utterance "(1a) Kevin caught a $[\text{frog}]_F$ where the focus is on the object "frog". In this case, the speaker emphasized the object "frog", responding to the *Wh*-question. As the response solely highlights the object, it falls under the category of narrow focus, pertaining to its limited scope of emphasis. The response given in the second dialogue, specifically the answer "(2a) $[\text{Kevin caught a frog}]_F$", displays broad focus since it introduces fresh information to the listener regarding all components of the statement (e.g., Chafe, 1976; Ladd, 1980). When speakers intend to correct a specific part of the previous utterance asked by the interlocutor, they employ contrastive focus in their answer to emphasize the corrected item, as shown in Dialogue (3a). In essence, the term focus, as discussed in this dissertation, refers to the emphasized constituent - new information or correction - of the answer elicited by the precursor question in a question-answer (QA) dialogue.

Precursor Questions	**Answers**	**Focus Categories**
1q. What did Kevin catch?	1a. Kevin caught a $[\text{frog}]_F$	(Non-Contrastive) Narrow Focus
2q. What happened?	2a. $[\text{Kevin caught a frog}]_F$	Broad Focus
3q. Did Kevin catch a cat?	3a. Kevin caught a $[\text{frog}]_F$	Contrastive Narrow Focus

Table 1.1 Examples of focus categories elicited in Q-A dialogues

Besides the size of the focus, the position of the focus varies depending on different *Wh*-questions and corrections. In subject-verb-object (SVO) languages like English and Chinese, sentence-initial focus is attributed to new information about the subject, sentence-middle focus to new information about the verb, and sentence-final focus to new information about the object. Table 1.2 lists examples of focus in the three positions in English.

Precursor Questions	**Answers**	**Focus Types**
4q. Who caught a frog?	4a. [Kevin]$_F$ caught a frog.	Initial Subject Focus
5q. What did Kevin do to the frog?	5a. Kevin [caught]$_F$ a frog.	Medial Verb Focus
6q. What did Kevin catch?	6a. Kevin caught a [frog]$_F$.	Final Object Focus

Table 1.2 Focus positions in SVO languages

1.2 Prosodic realizations of information structure

Acoustic evidence on prosodic focus production shows that speakers employ acoustic measurements like F_0, duration, and intensity as crucial means of distinguishing between various focus types (e.g., Rump & Collier, 1996; Botinis et al., 1999; Xu, 1999; Breen, et. al., 2010). In English, for instance, longer duration, higher mean and maximum of F_0, and greater intensity are associated with narrow focus rather than broad focus (Breen, et. al., 2010). In Mandarin, which is a tonal language, speakers use a broader range of pitch on the focal constituent along with other acoustic correlates to mark focus (Xu, 1999). However, there is no consensus on the relationship between the acoustic signals of prosody and the meanings associated with focus in the literature. There are two main perspectives on the relationship between the acoustic signals of prosody and the meanings associated with focus. The first, known as the direct relationship approach, suggests that specific sets of acoustic parameters are directly associated with information structure (Cooper et al,

1985; Fry, 1955; Xu & Xu, 2005). However, on the other hand, individual differences in production have been observed when signaling the same focus category, and even within one speaker, the one-to-one mappings between acoustic cues and information structure do not always hold true (e.g., English: Cruttenden, 1997; German: Grice et al., 2017).

Although production studies on prosodic focus have received considerable attention in both phonetic and psycholinguistic literature, in comparison, research on the perception and comprehension of focus cues has received far less attention, especially in the areas of child development and cross-linguistic studies. A recent by Roettger et al. (2019) revealed that English-speaking adults can effectively map between prosodic forms and focus categories in dialogues. They utilize their prior knowledge of the mapping between information structures and prosodic forms, rather than solely depending on bottom-up cues of prosodic realizations during focus processing. However, for a comprehensive understanding of other languages, the processing mechanism of mapping between prosodic signals and information structure must be further investigated. Inspired by this study, the present research aims to investigate whether Cantonese adult listeners can distinguish prosodic forms in native Cantonese and non-native Mandarin, as well as the extent to which school-aged children of native Cantonese have developed the ability to comprehend Cantonese and Mandarin prosodic focus.

1.3 Research questions of the current study

The overarching objective of this dissertation is to explore the competence to process prosodic focus in multimodal speech acquired by trilingual adults and children in their L1 Cantonese and L3 Mandarin. Five more specific research questions have been operationalized for the present study, which are listed below:

1. Can native adult speakers of Cantonese map prosodic forms onto specific focus types in the complex tonal context in their native Cantonese?
2. Whether trilingual adults acquired a native-like competence to decode prosodic focus in their L3 Mandarin?
3. Have school-aged trilingual children developed an adult-like ability to process the prosodic focus in their native Cantonese?
4. Have school-aged trilingual children developed a native-like path to focus processing in L3 Mandarin?
5. Does the audio-visual modality facilitate listeners' comprehension of native or non-native prosodic focus?

To address the first research question, I conducted a Cantonese perceptual experiment (Experiment 1 in Chapter3) how discourse contexts are mapped to prosodic realizations in complex tonal environments in audio-only and audio-visual modalities. Ge et al. (2022) found that Cantonese adults are more likely to perceive inappropriate contrastive focus as appropriate prosody, in line with the findings in comprehension of prosodic focus in Western languages (e.g., German, Toepel et al., 2007; Zellin et al., 2011). Based on this, I hypothesize that Cantonese adults are likely to be less sensitive to prosodic incongruity in the current study, resulting in more errors in detecting focus conditions with incongruous prosody. Additionally, Cantonese adults are also likely to perform better in decoding focus of utterance bearing dynamic contour tones in utterances, compared to static level tones. This is consistent with Wu's (2012) findings on the identification of Cantonese focus position, where Cantonese adults have difficulties identifying the focused word in high-level tonal contexts.

Regarding the second research question, a Mandarin focus comprehension experiment (Experiment 2 in Chapter 4) was conducted to investigate whether native Cantonese-speaking adults process the prosodic focus of Mandarin in the same way as native speakers. Given that Cantonese and Mandarin differ in encoding prosodic focus (Cantonese: Wu & Xu, 2010, Wu, 2013; Yang, 2022; Mandarin: Xu et al., 2012; Yang, 2022), it is hypothesized that Cantonese-speaking adults cannot decode the prosodic focus of Mandarin in a fully native-like way, and will be less accurate in evaluating the prosodic naturalness, compared to native Mandarin-speaking adults. Meanwhile, Cantonese adults may transfer specific features of focus decoding in their L1 Cantonese to their L3 Mandarin.

To address the third research question, I conducted the same Cantonese experiment from Experiment 1 with trilingual adults in Experiment 3 (Chapter 5) to investigate the ability of Cantonese-speaking children to process prosodic focus in their native L1.Since phonological acquisition is a protracted process for Cantonese children (e.g., Mok et al., 2019; Mok et al., 2020), the hypothesis for this experiment is that school-aged trilingual children with L1 Cantonese cannot process prosodic focus as accurately as compared to adults. Additionally, another prediction is that Cantonese trilingual children may exhibit better performance in congruous prosody than in incongruous prosody, which is suggested by a previous study on the comprehension of contrastive focus in Cantonese children aged five to eight (Ge et al., 2022).

For the fourth research question, I replicated the Mandarin experiment from Experiment 2 in Experiment 4 (Chapter 6) with trilingual children to explore their ability to process L3 Mandarin focus in multimodal speech. Previous literature has provided evidence on how Mandarin children produce and comprehend prosodic focus (Yang, 2017;

Yang & Chen, 2018; Chen, 2022), whereas to the best of our knowledge, little is known about the comprehension of non-native Mandarin focus by native Cantonese-speaking children. This study hypothesizes that school-aged trilingual Cantonese children may not have developed a fully native-like competence to interpret Mandarin focus, given the challenging task of parallelly decoding lexical and prosodic functions in non-native speech.

The final research question seeks to unravel whether trilingual listeners would utilize the supplementary visual information during focus processing in both L1 and L3, and to what degree. In all four experiments, stimuli were presented in both audio-only (AO) and audio-visual (AV) modalities. The hypothesis is that the additional visual cues may enhance trilingual adults' ability to comprehend focus in non-native Mandarin, rather than native Cantonese, as they may rely more on visual input when auditory input is ambiguous. Additionally, this study predicts that Cantonese-speaking children would rely on less visual signals than adults in the Cantonese experiment (Experiment 3), due to the high cognitive demands of multimodal speech comprehension. In L3 Mandarin (Experiment 4), I hypothesize that Cantonese-speaking children would perform better in the AV modality than the AO modality, as the visual information may compensate for any difficulties in auditory speech comprehension.

1.4 Dissertation outline

The following provides a comprehensive review of previous literature on the processing of prosodic focus in adults and school-aged children in both tonal and non-tonal languages. Section 2.1 presents research on the comprehension of prosodic focus from three perspectives: acoustic variations, paralinguistic factors (i.e., listener expectations), and the model of probabilistic inference in focus processing. Section 2.2 outlines previous studies

on the encoding and decoding of prosodic focus manners employed by school-aged children. Section 2.3 discusses the prosodic phonology of Cantonese and Mandarin and investigates research on the acquisition of prosodic focus across languages. Section 2.4 reviews previous findings on multimodal input effects on focus processing, particularly the role of visual cues in the perception and comprehension of segments and suprasegments.

Chapter 3 (Experiment 1) examines the processing of Cantonese prosodic focus by trilingual adults with the input of both audio-visual and audio-only modalities in their L1. Chapter 4 (Experiment 2) investigates the trilingual adults' ability to process focus in their L3 Mandarin. Chapters 5 (Experiment 3) and 6 (Experiment 4) present the Cantonese and Mandarin experiments, respectively, conducted on trilingual children. Finally, Chapter 7 provides a comprehensive discussion encompassing the findings derived from all four experiments, an overview of the limitations encountered, proposals for future research directions, and the conclusive summary of this dissertation.

Chapter 2. Literature review

2.1 Prosodic focus processing

2.1.1 Acoustic variability and focus perception

In daily communication, listeners are adept at interpreting the prosodic signals in auditory input to discern a speaker's intentions, even in the face of individual variabilities across contexts. However, the extent to which listeners utilize data-driven inputs to process focus across different languages remains an area to be further explored. A handful of studies have been conducted on West Germanic languages, which investigate the interpretation of prosodic forms associated with discourse function, revealing the ubiquitous variability of focus realizations and no consistently unambiguous mapping between prosody and meaning in a language (e.g., Turnbull et al., 2017; Grice et al., 2017; Cangemi et al., 2015). Studies on speech production have revealed that speakers use prosodic modulations such as higher maximum pitch values, broader F_0 ranges, longer duration, and increased intensity on focused words or constituents to signal emphasis. However, the lack of an explicit mapping between prosodic parameters and different categories of focus, along with the potentially asymmetrical relationship between production and perception, results in considerable complexity in focus processing.

Breen et al. (2010) investigated the production and identification of prosodic focus in English, examining various focus categories based on focus breadth (broad and narrow), focus type (contrastive and non-contrastive), and focus position (initial, medial, and final). The authors found that English speakers may mark narrow object focus from broad focus differently in speech production, suggesting that the focus projection theory may not

always hold. According to this theory, emphasis on the narrow object in a subject-verb-object language like English can project to the verb phrase and then the entire utterance (Selkirk, 1984, 1995; Gussenhoven, 1984, 1999). In their subsequent perception experiment, Breen et al. (2010) instructed listeners to select one of seven precursor questions that could be an appropriate answer of the heard sentences. English listeners demonstrated only moderate accuracy in identifying both broad focus and narrow object focus. In particular, fewer than half of the participants were able to identify the broad focus beyond chance level.

Additionally, as the acoustic differences between narrow and contrastive focus in English are subtle, English speakers sporadically distinguish these two types of prosodic realizations when they are aware of the ambiguity and intend to distinguish them. Acoustic evidence revealed that contrastive focus was realized with longer duration and higher intensity, while the non-contrastive narrow focus was related to higher F_0. Previous studies (Turk & Sawusch, 1996; Kochanski et al., 2005) suggest that the increase in intensity and lengthening of duration may result in a more prominent prosody than the increase in the height of F_0. This indicates that contrastive focus may be perceived as more prominent than a narrow focus. However, listeners may encounter difficulties utilizing the subtle parametric differences between narrow and contrastive focus during identification tasks. Therefore, despite speakers using prosodic modulations to mark focus, listeners may struggle to effectively utilize these cues to comprehend the size and scope of focus. It is likely that the challenging nature of the tasks used in Breen et al.'s (2010) study, which required listeners to choose one from seven choices, contributed to this difficulty.

Previous studies suggest that while distinguishing the size and breadth of focus may be challenging for listeners, they can accurately identify the position of prosodic focus in utterances (Eady & Cooper, 1986; Rump & Collier, 1996; Breen et al., 2010). For instance, listeners of English can accurately identify various positions of prosodic focus, even when presented with a set of seven choices (Breen et al., 2010). However, previous studies have found that the syntactic roles of specific constituents, such as the subject and object in SVO sentences, may influence the corresponding positional prosody in utterances. Specifically, the initial focus is associated with the sentence topic and is realized by an expanded pitch range and other F_0-related correlates as a cross-language feature. The constituent on the phrase-final position is often lengthened even in non-focus utterances, yet listeners are sensitive to durational signals (Steffman, 2019; Baumann & Winter, 2018). Previous studies have shown that the identification accuracy of the sentence-final focus is dramatically low in specific languages, such as English, Swedish, and Mandarin (Botinis et al., 1999; Xu et al., 2012). Therefore, previous studies concentrating on sentence-initial or -final focus were entangled with other influence (e.g., word order) on interpretation rather than the exclusive relationship between prosodic realizations and information structure (Cangemi et al., 2015; Baumann et al., 2021; Calhoun et al., 2021).

Therefore, previous research on the production and perception of focus has shown evidence of acoustic cues derived from auditory input that listeners may exploit to decode prosodic focus in speech. However, in most cases, there is no explicit mapping between prosodic forms and focus categories, so listeners have to rely on signal-extrinsic factors to map between surface prosody and discourse functions successfully. The reliance on signal-extrinsic factors indicates a top-down approach to speech comprehension, by which

listeners can predict the incoming utterances they will hear based on their pre-existing knowledge (Wagner, 2005; Sohoglu et al., 2012; Bishop, 2016; Bishop et al., 2020). The previous literature has emphasized the significance of expectations regarding phonological sequences or prosodic patterns before disambiguous information is revealed at both the segmental (Warren, 1970; Samuel, 1981; Dupoux et al., 1999), and suprasegmental levels (Vainio & Järvikivi, 2006; Zheng & Pierrehumbert, 2010; Calhoun, 2010; Bishop, 2012; 2016; Roettger et al., 2019). Chapter 2.1.2 will focus on whether and to what extent listeners integrate the top-down approach to focus processing with the bottom-up cues of auditory input.

2.1.2 Listener expectations on focus processing

Recent studies on the comprehension of the interplay between prosodic focus and information structure have to pay more attention to the prior knowledge that listeners possess before encountering the new bottom-up information in focus processing. In general, the pre-existing knowledge of the world state in a listener's memory can influence how new information is received, processed, and retained in communication (for a review, see Brod et al., 2013). In speech comprehension, listeners make predictions about what will likely come next based on the contextual cues they have received, integrating this information of signal-based input with their prior knowledge. Empirical findings suggest that expectations of signal-extrinsic factors may prevail over acoustic cues of auditory inputs in prosodic perception (Norris et al., 2003; Clayards et al., 2008; Bishop, 2011, 2012; Turnbull et al., 2017; Roettger & Franke, 2019; Roettger et al., 2019; Roettger & Rimland, 2020; Toepel et al., 2007; Liu & Jaeger, 2018; Calhoun et al., 2021; Calhoun, 2010).

Specifically, in West Germanic languages, listeners' expectations of discourse demonstrate influence on their identification of focus positions and focus breadth. In Bishop (2012), American English listeners were instructed to judge whether the verb or object was prominent in the answers to various precursor questions. His study employed three preceding questions designed to elicit broad focus, narrow focus on the verb phrase, and narrow focus on the object, indicating three degrees of focus breadth, whereas the subsequent answers were acoustically identical. Results of Bishop (2012) suggested that listeners consistently perceived the object as more prominent when the question evoked a narrow object focus context (e.g., "What did Kevin catch?"), compared with the broad focus-inducing context (e.g., "What happened?"), despite the audio stimuli of the answers being the same production. That indicates that listeners integrated their expectation of the information structure with the signal-based bottom-up inputs to match focus in their prosodic comprehension. The way that listeners exploit contextual information was further explored by Bishop (2016), with the assertion that this is subject to individual differences in cognitive processing and may be associated with listeners' pragmatic skills. Although Bishop's studies contribute to our understanding of the role of listeners' expectations of information structures encountering various focus breadth conveyed by the prior discourse, it is limited to the effects of focus projection. It is worth exploring whether the top-down approach applied to listeners' prosodic comprehension facing different focus types (i.e., with or without correction).

Additionally, the expectation of correction (or contrastive focus) in the prior discourse may facilitate listeners' processing accuracy in quasi-natural dialogues. Toepel et al. (2007) examined German listeners' processing strategies for congruous or

incongruous prosody on the object (sentence-final) using an event-related brain potential (ERP) paradigm. They investigated the extent to which German listeners rely on pragmatic factors (e.g., focus) versus the actual prosodic features of utterances to resolve conflicts in information structure. Specifically, Toepel et al. (2007) investigated how German listeners process given focus and contrastive focus using question-answer congruence pairs. Each pair contained an incongruous half consisting of mismatched prosodic realizations. One half consisted of the context of the given focus followed by prosodic contrastive accentuation in the answer, and the other half consisted of the context of the contrastive focus followed by prosodic given focus accentuation in the answer. Behavioral results indicate that judging the correctness of the prosody is easier for sentences that contain information of contrastive focus, regardless of whether the prosodic emphasis is conveyed with correct accentuation. They proposed that listeners may not always be certain of the correct prosodic forms for specific information structures and thus rely more on the information structure to judge the prosodic realizations. Moreover, one possible explanation for the lower sensitivity observed in the context of the given focus is that listeners prefer to respond to their interlocutors with a simple "Yes" rather than repeating the preceding statements during conversations. Their ERP results also support the predominant role of contextual information structure, irrespective of the acoustic realizations of target utterances. There is still more research to be done to determine whether the benefit of contrastive focus over given focus can be applied to other focus categories, such as broad focus and contrastive focus.

To further understand the nonstationary mapping between prosodic form and discourse function in spoken language comprehension, Roettger and colleagues (2019)

proposed a probabilistic approach to prosodic focus processing in English. A series of perception experiments were designed to test the abilities of listeners to not only (Task 1) identify the prosodic form for a specific focus type but also (Task 2) recognize the information structure indicated by particular prosodic forms using question-answer (Q-A) congruence dialogues. Listeners were required to identify the natural response in two alternative forced-choice responses with or without appropriate prosody. In Task (1), listeners were presented with two dialogues containing the same questions, such as "Do you know who ripped the ledger?" but with answers differing in prosodic focus, (a) "Yes, [Mary]$_F$ ripped the ledger" and (b) "Yes, [Mary ripped the ledger]$_F$". Only the answer (a)was prosodically congruous to the question. For the two dialogues in Task (2), two types of focus were elicited in different precursor questions, (c) "Do you know who ripped the ledger?" (d) "Do you know what happened yesterday?". While the following answers shared the same prosodic production, like (a), with being the congruous response to the prior question (c) only. Results showed that English listeners are able to match one of two prosodic forms to a specific focus category in Task (1), although accuracy varied among focus categories. For example, the contrastive focus was easily distinguishable from the given and broad focus, while the accuracy was lower in discriminating narrow focus from given information and discriminating contrastive focus from narrow focus. Moreover, listeners showed difficulties in matching particular prosodic forms to the correct focus categories, especially in certain focus conditions in Task (2). That is, listeners showed bias against broad focus but toward narrow focus. However, the bias against broad focus was diminished in Task (3), in which listeners rated the prosodic acceptability of each dialogue

using a 5-point Likert scale instead of a two-forced choice in the first two tasks, though the bias towards narrow focus remained.

The various performance in the acoustic-distinguished prosodic forms was attributed to the influence of listeners' prior knowledge of discourse function, which integrates with the acoustic inputs in focus processing (Roettger et al., 2019). Their findings suggest that listeners possess probabilistic knowledge of the likelihood of a speaker using a specific prosodic form to convey a particular discourse function. This knowledge is based on the probabilities derived from listeners' prior experience and pre-activates their related interpretations of discourse functions when encountering specific question contexts. Specifically, biases in listeners' expectations impact their comprehension of what constitutes appropriate contexts, accounting for the discrepancy in mapping accuracy between broad focus and narrow focus. The context that induces broad focus is usually the question "What happened?" which is assumed to be followed by an 'out-of-the-blue' utterance, while the narrow focus is elicited by more specific *Wh*-questions that provide common ground between interlocutors (Dahan, 2015; Breen et al., 2010). This insight has important implications for understanding the stochastic co-occurrence of prosodic cues and information structure in discourse, highlighting the crucial role of listeners' prior knowledge and expectations in shaping their interpretation of prosodic cues. The study conducted by Roettger et al. (2019) contributes to our understanding of how listeners employ probabilistic inference to discern discourse functions based on prosodic distributions and their prior knowledge of the likelihood of the speaker's intentions. This finding adds to the existing probabilistic model of speech comprehension extended by Roettger and Franke (2019).

2.1.3 The probabilistic modeling in speech comprehension

Probabilistic models have increasingly been used to explain the cognitive science of language learning and processing (see a review by Chater & Manning, 2006). This computational model emphasizes the use of probabilistic models to analyze language data, enabling predictions and inferences about linguistic phenomena. According to Bayes' rule, the updated probability of a hypothesis (*posterior probability*) is proportional to the product of the *prior probability* (i.e., the probability of the hypothesis before considering the evidence) and the *likelihood* of the evidence (i.e., the probability of observing the evidence given the hypothesis). The probabilistic model has been employed in explaining aspects of linguistic processing in both syntax and pragmatics (e.g., Manning, 2003; Bod et al., 2003; Frank & Goodman, 2012; Goodman & Frank, 2016) and, more recently, to investigate speech comprehension in relation to prosodic cues (Roettger et al., 2019; Roettger & Rimland, 2020; Roettger et al., 2021; Xie et al., 2021; Kurumada & Roettger, 2021).

The probabilistic model in prosodic comprehension was formulated in Roettger and Franke's (2019) study. This model rationalizes how listeners interpret the possibility of prosodic distributions in comprehending a speaker's intentions in their utterances. The authors designed a mouse tracking task to test whether German listeners predict the referent based on the intonational cues in the speaker's speech and adapted to (un)reliable intonational cues. The results of their experiment supported the key concept of the probabilistic model, namely that listeners hold expectations about the probability of a speaker producing different utterances to express different meanings. Specifically, the probability that a speaker will produce a specific intonational contour when referring to a

particular referent is different, which depends on the likelihood of producing that utterance for that referent.

Several studies have clarified the approach to probabilistic inference under uncertainty in speech comprehension, in line with the results in Roettger and Franke (2019) (e.g., Roettger et al., 2019; Roettger & Rimland, 2020; Roettger et al., 2021; Xie et al., 2021; Kurumada & Roettger, 2021). The probabilistic nature of the sound-meaning mapping can be elucidated from two perspectives: identical acoustic forms of sounds can convey multiple meanings across talkers and contexts, and the same discourse meaning can be encoded in distinct acoustic patterns. For example, Kurumada and Roettger's (2021) framework sheds light on the probabilistic relationship in nonstationary mapping between prosodic distributions and focus categories (narrow focus vs. contrastive focus) regarding the talker variability in German. Their findings highlight the inherent uncertainty in perceptual and cognitive processes, prompting listeners to navigate it by inferring cue-category mapping. In this inference process, listeners integrate their existing knowledge to interpret the probability of prosodic cues being distributed across different focus categories based on their past experiences encountering the prosodic input. Furthermore, at least for contrastive focus in German, listeners appear to learn and store distributional statistics of prosodic cues specific to individual talkers and then make inferences, considering the variability of talkers in producing focus.

In summary, previous research has shown that listeners rely not only on the bottom-up input, in which they perceive the prominence conveyed by acoustic signals, but also on the top-down perceptual method to infer the speaker's intentions and prosodic characteristics in interpreting various foci. However, to my knowledge, the literature has

only covered prosody processing in West Germanic languages such as English and German (Roettger & Franke, 2019; Roettger et al., 2019). There have been limited studies on focus processing in tonal languages, where pitch variations distinguish between lexical tones. Intriguing questions arise in tonal languages, considering that the F_0 was parallelly used to discriminate lexical tones and convey prosodic focus (Liu & Xu, 2005; Xu, 2011). It is unrevealed to what extent native tonal listeners rely on prosodic cues to match particular information structures and to what extent they exploit prior knowledge and expectations in focus processing. Additionally, it remains unknown whether tonal language-speaking children can process prosodic focus in an adult-like manner. This highlights the need for further research to understand better how focus processing occurs in tonal languages and how it develops in children.

2.2 Prosodic focus acquisition in children

2.2.1 Prosodic focus-marking in children

This section reviews previous studies on how children encode and decode the interplay of pragmatic functions and prosodic variations. According to the literature on focus production, children can mark prosodic focus in an adult-like way in many aspects at an early age in both non-tonal languages (e.g., Finnish: Arnhold et al., 2016; Dutch: Chen, 2010; 2011; 2018; Chen & Höhle, 2018; Swedish: Romøren & Chen, 2017) and tonal languages (Mandarin: Yang, 2017; Yang & Chen, 2014; 2018). However, these findings suggest that language background, whether tonal or non-tonal, likely influences the acquisition order and efficiency of specific acoustic cues, such as pitch cues and duration, in child prosodic development.

In West Germanic languages, specifically, it has been found that children use accentuation to mark focus, but their use is not fully adult-like at an early age. According to Chen (2010), Dutch-speaking children aged four to five produce accented focal nouns in sentence-initial and sentence-final positions less frequently than adults when prompted with *Wh*-questions. By the age of seven to eight years old, however, Dutch children do not use duration as a means of marking initial focus (Chen, 2009). Finnish-speaking children aged four to five can use pitch-related cues to mark focus, although not in an adult-like manner (Arnhold et al., 2016). They emphasize words in contrastive focus with a broader F_0 range than adult speakers. Additionally, they tend to lengthen the duration of sentence-initial objects when in focus and shorten the duration of unfocused object words, considering Finnish is an OVS language. Similarly, English-speaking children exploit pitch variations to mark contrastive focus while the duration cue is not salient at three to four years old (Wonnacott & Watson, 2008).

In tonal languages, Yang (2017) and colleagues (Yang & Chen, 2018) examined the development of Mandarin focus realizations at three age ranges (four to five, seven to eight, and ten to eleven years old), which involved the comparisons among focus types (narrow focus, contrastive focus, and broad focus) as well as focus positions (initial, medial and final focus) using pictures-matching games previously used in production studies (e.g., Chen, 2011). Their study found that Mandarin-speaking children cannot use pitch-related variations (e.g., pitch-maximum or pitch-minimum) in a fully adult-like fashion to mark focus, even at ten to eleven years old. At the same time, they use the duration to distinguish narrow focus from broad focus and non-focus as early as four to five years old. As they proposed, the use of pitch cues in discriminating lexical meanings slows Mandarin

children's acquisition of pitch-related cues compared to the acquisition of duration. The research question at hand concerns whether children who speak tonal languages can acquire the ability to decode focus using the acoustic cues in their focus marking. If such an ability is present, it is important to investigate the extent to which these children rely on acoustic cues, and to explore other factors that may influence their decoding approach to prosodic focus. Clarifying these aspects will contribute to our understanding of how children acquire prosodic focus and how the unique features of tonal languages affect language acquisition.

Moreover, the acquisition of pitch-related cues varies among lexical tones, and the prosodic focus-marking in certain lexical tones is later acquired than others for Mandarin children at an early age in Yang (2017). Specifically, children aged seven to eight years old mark narrow focus with a higher pitch maximum in high-level Tone 1 and high-falling Tone 4 and a lower pitch minimum in Tone 2 and Tone 3 compared to adults. Yang (2017) explains that children probably associate a higher pitch with high lexical tones (i.e., high onsets in Tone1 and Tone 4) and a lower pitch with low lexical tones (i.e., low onsets in Tone2 and Tone 3) in focus-marking. This phenomenon has also been observed in Dutch, despite F_0 not discriminating meanings (Gussenhoven & Rietveld, 2000). At the age range of ten to eleven, Mandarin-speaking children differentiate between focus positions (medial vs. final) using means that are more similar to those of adults. Adults mark focus in Tone 1 and Tone 3 with a lower degree of pitch range expansion than the variations in dynamic lexical Tone 2 and Tone 4 to keep the identity of intact undynamic tones. However, when marking focus, children still produce different pitch-related features in Tone 1 (in terms of pitch range) and Tone 3 (in terms of pitch maximum) compared to adults. Yang's (2017)

findings imply that there is no strong correlation between the acquisition of prosodic focus-marking and lexical tones in Mandarin since young children generally acquire Tone 1 and Tone 4 earlier than Tone 2 and Tone 3 (e.g., Zhu, 2002; Wong, 2012). However, to date, there has been limited discussion regarding the impact of lexical tones on focus processing in child language acquisition, and even less research on the effects of tonal languages on comprehending prosodic focus in non-native speakers. Addressing this research gap is critical to understanding the interplay between lexical tones and prosodic focus in the development of tonal language acquisition in children and non-native listeners.

2.2.2 Prosodic focus-processing in children

From the perspective of prosodic focus comprehension, using prosodic information in an adult-like manner is a sophisticated task for children, as it requires perceptual mapping ability between information structures and prosodic forms. Child listeners' successful interpretation of information structure depends primarily on using multiple linguistic abilities and language-external cognitive competence rather than merely identifying acoustic forms. In a mature acquisition of prosody, listeners should integrate their prior-existing competence and knowledge with particular prosodic patterns they receive to understand speaker intentions, attitudes, and other discourse functions. Studies on the comprehension of prosodic focus are insufficient in the literature, probably due to its subtle and ambiguous contribution to pragmatic meaning beyond prosodic forms (Szendrői et al., 2018). The literature on child prosodic focus processing is limited to the picture-matching paradigm or auditory-only question-answer paradigm in either explicit behavioral or implicit online processing studies. The following section will mainly focus on children's interpretation of information structure without focus particles but with accentuations on

syntactic constituents (Chen, 2010; Chen et al., 2019; Szendrâi et al., 2018; Surányi & Pintér, 2021; Pannekamp et al., 2011).

Chen (2010) provides a question-answer paradigm with congruous and incongruous dialogues to test the sensitivity of four- to five-year-old Dutch-speaking children to accentuation in picture-matching games. In congruous dialogues, the accent was correctly placed on the subject or the object, whereas in incongruous dialogues, the accent was pragmatically inappropriate. Chen (2010)'s investigation suggests that children exhibit sensitivity to pragmatically incongruent prosody, as evidenced by their longer reaction time for congruent dialogues. Furthermore, her study finds that intonation plays less influence in the judgment of correctness compared with lexical-semantic errors and pronunciation errors in Dutch speech for both children and adults.

Szendrâi et al. (2018) conducted a cross-linguistic study to investigate the developmental comprehension of contrastive focus on the subject and object by children aged three to six in English, French, and German. In the simple picture-matching correction task, children displayed adult-like performance in correcting the subject and the object regardless of language, indicating that English, French, and German children at three years old could understand the prosodic focus on sentence-initial and sentence-final positions. They assert that children can understand contrastive focus at the same age of focus-marking (i.e., three years old) and attribute the divergences of previous studies to task effects.

Moreover, minor language-specific differences were found in focus comprehension among English, French, and German for both children and adults, although these languages use different syntactic focus-marking measures in addition to prosodic focus. No other cross-linguistic differences were found, except for French participants performing worse

in the accented condition compared to English participants. Surányi and Pintér (2021) used a similar method to show that the acquisition trajectory of prosodic focus comprehension is delayed in Hungarian children who achieve an adult-like manner at six years of age. Excluding task effects, they ascribe the procrastination in the comprehension of Hungarian prosodic focus to language-specific factors compared to the other three languages (Szendrâi et al., 2018).

Compared to English, French, and German children in the literature, Mandarin-speaking children acquire prosodic focus processing abilities at a later stage. Chen et al. (2019) followed the sentence-picture verification task (Szendrâi et al., 2018) to investigate focus comprehension in Taiwanese Mandarin-speaking children aged three to five years old. Their results indicated that in topic-dominant Mandarin, children and adults are adept at detecting contrastive focus on the object compared to the subject. However, children showed lower accuracy in the focus-congruence response on the subject than adults (16% for 3-4-year-olds, 14% for 5-year-olds, and 38% for adults). Chen et al. (2019) explained that the word order in child Mandarin might overweigh the prosodic cues in focus processing, as Mandarin has a relatively flexible word order in which the object is the default focus, and the subject is the topic of the sentence (Xu, 2004). This contradicts the findings of Chen (1998), who examined how Taiwanese Mandarin-speaking children aged five to thirteen interpret given and new information using a picture-verification task. Chen found that Mandarin children up to the age of 13 relied more on prosodic cues than word order in distinguishing between given and new information, which contrasts with adult performance. Compared to adults, Ge et al. (2022) found that Cantonese-speaking children aged 5 to 8 years old were less sensitive to prosodic emphasis on the subject or object,

regardless of matched or mismatched contexts. They performed better in syntactic cues such as the focus marker "*is*," though their accuracy did not reach adult-like levels. Despite these differences, overall, children in both stress and tonal languages tend to exhibit some degree of sensitivity to correcting the focal subject or object in sentence correction tasks, with tonal language children demonstrating this ability at a later developmental stage and continuing to improve with age.

In addition to behavioral studies, Pannekamp et al. (2011) investigated the development patterns of ERP in German-speaking children across three age groups (five, eight, and twelve years old) to explore the role of prosodic forms and contextual information in focus processing. The study focused on the narrow focus on the object, with or without contrastive emphasis, examined through question-answer dialogues with prosodically adequate or inadequate answers. The ERP results showed that 12-year-old children comprehended both types of focus in an adult-like, prosody-independent manner. Specifically, older children and adults demonstrated similar patterns between dialogues with contextual contrastive focus but differing in actual prosodic cues. Eight-year-olds responded to the inappropriate prosodic forms of contrastive focus but only exhibited awareness of narrow focus with either adequate or inadequate prosody. Five-year-old children did not demonstrate sensitivities to focus types (narrow or contrastive), but they were sensitive to inadequate prosody in the focal position, as indicated by the N400 response to the prosodic inadequacy. This study further concludes that as children mature, their strategies for focus processing in spoken language dialogues become less reliant on prosodic signals and more on contextual information, shifting toward an adult-like manner (e.g., French: Magne et al., 2005; German: Toepel et al., 2007).

Additionally, implicit evidence of children's online processing has confirmed their adult-like ability to use prosodic cues for establishing phrase boundaries (e.g., Snedeker & Yuan, 2008), disambiguating syntactic ambiguity (Zhou et al., 2012b) and resolving pragmatic ambiguity (Zhou et al., 2012a). For instance, in Zhou et al.'s (2012b) visual world paradigm study, Mandarin-speaking children between four to five years of age performed adult-like patterns of fixation proportion when presented with utterances containing accented or unaccented focus elements such as '*only*' (*Zhiyou*). Likewise, in another study by Zhou et al. (2012a), children within the same age range were able to use prosodic cues to distinguish between questions and statements in Mandarin Chinese. These findings support that at an early age, tonal language-speaking children are capable of utilizing prosodic cues to decode the interface between syntactic or pragmatic functions and surface prosodic realizations in online processing.

However, it is noteworthy that the implicit evidence does not consistently align with findings from previous behavioral studies. Specifically, behavioral research utilizing picture-matching games has revealed that even school-aged children experience difficulties in distinguishing between utterances with accented and unaccented focus-marking particles (Paterson et al., 2003; Hüttner et al., 2004; Bergsma, 2006). These behavioral results suggest that children either ignore focus markers, such as '*only*' and '*also*,' in sentences, or they may not process sentences differently with or without focus-marking particles. A plausible explanation for this phenomenon is that sophisticated metalinguistic skills such as parallel storage, comparison, and conscious decision-making likely influence behavioral results rather than a simple sensitivity to prosodic cues with or without accentuations. It is plausible that differences in linguistic competence requirements between explicit picture-

matching behavioral tasks and implicit eye-tracking or ERP experiments may account for the discrepancies in results.

In summary, previous behavioral studies examining prosodic focus comprehension have focused on whether children can respond to inappropriate prosody in the picture-matching task in a manner consistent with adult-like processing. However, the majority of research in this area has been concerned with the prosodic focus interpretation in West-Germanic language (e.g., Chen, 2010; Szendrâi et al., 2018), with only a limited number of studies investigating tonal languages such as Taiwan Mandarin (e.g., Chen, 1998; Chen et al., 2019). Moreover, previous research on this field has been primarily limited to the perception of acoustic accentuations on specific syntactic constituents of simple context-free sentences using sentence-correction paradigms rather than examining focus processing in consideration of contextual information emended in dialogues.

2.3 Cross-linguistic prosodic focus in Cantonese and Mandarin

2.3.1 Prosodic focus in Cantonese and Mandarin

What distinguishes the perceptual mappings of tonal languages from non-tonal languages to a large extent, is the parallel function of the fundamental frequency in both lexical meaning and prosody of speech production. Lexical tones in tonal languages like Mandarin and Cantonese are differentiated by variations in the fundamental frequency, which indicate differences in pitch height and contour. In Table 2.1, examples of Cantonese words are provided as illustrations of the five tone categories in the language. The tone categories are identified using tone numbers or descriptive names as defined by Bauer and Benedict (1997), and pitch height distinctions as described by Chao (1948). Tone 2 and Tone 5 in

Cantonese are characterized as dynamic, as they involve upward pitch movement. In contrast, Tone 1, Tone 3, Tone 4, and Tone 6 are considered static, with Tone 4 also having a downward pitch movement. Mandarin has four lexical tones, including High-Level Tone1, Mid-Rising Tone 2, Low-Dipping Tone 3, and High-Falling Tone4, listed in Table 2.2, along with pitch values and example words in Chinese.

Tone Numbers	Descriptive Names	Pitch Height	Example Words
Tone 1	High-Level	55	[ji55] “衣”
Tone 2	High-Rising	25	[ji25] “椅”
Tone 3	Mid-Level	33	[ji33] “意”
Tone 4	Mid-Low Falling	21	[ji21] “疑”
Tone 5	Mid-Low Rising	23	[ji23] “耳”
Tone 6	Mid-Low Level	22	[ji22] “二”

Table 2.1 Six lexical tones in Hong Kong Cantonese

Tone Numbers	Descriptive Names	Pitch Height	Example Words
Tone 1	High-Level	55	[ma55] “妈”
Tone 2	Mid-Rising	35	[ma35] “麻”
Tone 3	Low-Dipping	213	[ma213] “马”
Tone 4	High-Falling	51	[ma51] “骂”

Table 2.2 Four lexical tones in Mandarin

Studies have consistently demonstrated that changes in F_0 are a fundamental factor in prosodic focus in non-tonal languages, along with variations in duration and intensity (e.g., Rump & Collier, 1996; Botinis et al., 1999; Breen, et al., 2010). The literature has raised the intriguing question of whether and how variations in F_0 can be utilized concurrently to differentiate between lexical tones and highlight significant information in tonal languages (Xu, 1999; Man, 1999; Liu & Xu, 2005; Gu & Lee, 2007; Gu & Lee, 2008; Chen & Gussenhoven, 2008; Wu & Xu, 2010; Xu et al., 2012; Wu, 2013). Acoustic

evidence from Mandarin and Cantonese prosodic production suggests that marking focus significantly modulates the utterance's overall shape of the F_0 curve. However, the identity of the lexical tone within individual words remains the same, independent of the specific focus realization (e.g., Xu, 1999; Man, 2002). Specifically, in Cantonese, focusing on a particular element of a sentence causes a notable increase in both the range and level of pitch, but this change only occurs with dynamic tones (Man, 2002; Wu & Xu, 2010; Wu, 2013). The pitch contours of static tones (Tone 3, Tone 4, and Tone 6) in Cantonese are not usually affected by the different positions in which focus is realized, as observed by Wu in (2013). Moreover, findings in Mandarin suggest that only high-pitched tones (e.g., Tone 1 and Tone 2 in Mandarin) tend to be affected by focus, leading to changes in the F_0 range (Jia et al., 2006). However, studies investigating how native tonal listeners decode prosodic modulations in Cantonese and Mandarin speech production across different tonal contexts are currently insufficient.

Previous studies on focus perception in Cantonese and Mandarin have mainly concentrated on the identification of focus position and prosodic prominence while disregarding discourse context (Xu et al., 2012; Wu & Xu, 2012; Wu, 2013). Notably, the cross-linguistic feature of post-focus compression (PFC) was observed in Mandarin, indicating that speakers generally reduce the pitch range and intensity of elements that come after the focus, while expanding the pitch range and intensity, and lengthening the duration, of the focused element (English: Eady & Cooper, 1986; Liu & Xu, 2007; Mandarin: Xu et al., 2012; Chen et al., 2009; Japanese: Ishihara, 2002; Lee et al., 2022; Korean: Lee & Xu, 2010). For example, Xu et al. (2012) conducted a study to investigate the accuracy of focus position identification in five-syllable Mandarin declarative

sentences with the high-Level Tone 1, such as "[ma55] [ma55] [mɔ55] [mɑʊ55] [mi55]" (meaning "Mom touched the cat"), where the initial two syllables constituted the subject, the third syllable denoted the verb, while the last two syllables represented the object. In a context-free design, native Beijing Mandarin speakers were asked to identify which of the three positions (initial, medial, and final) was emphasized or if no word was emphasized. Results showed that native speakers had a high accuracy rate of over 90% for initial and medial focus, with overall performance exceeding the chance level for other focus types. The authors concluded that the use of PFC of pitch range and intensity facilitated identification accuracy for native Mandarin speakers. In contrast, the identification accuracy of focus in Taiwanese, a tonal language without PFC, was around chance level among its native listeners.

Following a similar paradigm to that used in Mandarin (Xu et al., 2012), Wu (2013) investigated the accuracy of identifying three focus positions and no focus in simple Cantonese sentences, each consisting of five syllables with identical tones from Tone 1 to Tone 6. Results showed that Cantonese listeners could identify focus positions with an accuracy rate of 70%-80%, except for High-Level Tone 1, where the accuracy rate was less than 60%. By incorporating results from Taiwanese, Taiwan Mandarin, and Beijing Mandarin in Xu et al. (2012), Wu (2013) concluded that the absence of PFC might affect the perception of focus, at least for words with a High-Level lexical tone. Regarding focus position, the identification rate of the initial focus was generally lower than that of medial or final focus in Cantonese, while a relatively low identification rate was found in the final focus in Mandarin statements (Liu & Xu, 2005; Xu et al., 2012).

In a recent study, Yang (2022) investigated the perceptual mapping between information structure and prosodic forms in Cantonese and Mandarin simple utterances embedded in semi-naturalistic question-answer dialogues using a paradigm similar to that used in English (Roettger et al., 2019). Native speakers of Cantonese and Mandarin demonstrated superior performance in mapping from prosodic cues to information structure compared to mapping from discourse function to acoustic signals. These findings suggest that listeners rely more on bottom-up inputs than top-down knowledge in focus processing. In addition, Cantonese and Mandarin speakers exhibited biases against broad focus in both mapping directions, consistent with English speakers in Roettger et al.'s (2019) study. However, Cantonese listeners showed lower perceptual accuracy than Mandarin listeners, which may be due to the fact that Cantonese does not heavily rely on prosody to mark focus compared to Mandarin.

Yang's (2022) research has broadened the literature by shifting the focus from the perception of prosodic prominence in previous studies to the consideration of contextual information in tonal languages. However, the investigation into the mapping mechanism in various tonal contexts remains insufficient, as Yang (2022) only examined high-level Tone 1 in Cantonese and Mandarin. Furthermore, previous studies on Cantonese and Mandarin have mainly focused on the size and position of focus, leaving us with limited knowledge about whether listeners can differentiate between various focus types, such as contrastive versus non-contrastive focus.

2.3.2 Cross-linguistic prosodic focus

A few studies have examined the acquisition of prosodic focus in a second language (L2) by bilingual adult speakers in both focus production (Nava & Zubizarreta, 2008; Chen & Guion-Anderson, 2012; Fung & Mok, 2014; Liu et al., 2016; Yang, 2022) and focus comprehension (Ge et al., 2018; Yang, 2022). In the literature on L2 focus production, Liu et al. (2016) found that sequential bilingual speakers of Bai-Mandarin were unable to mark Mandarin focus with PFC in the same way as native Mandarin monolingual speakers, as they did not use prosodic cues (duration and pitch-related cues) in a native-like manner to distinguish focus size and types. Bai is a Sino-Tibetan language spoken in Southwest China and has a similar tonal system to Cantonese. Similar to Cantonese, focus production in Bai also lacks the constant PFC feature and pitch variations play a minimal role in marking focus. Furthermore, even for speakers of native languages with PFC, transferring focus marking from L1 to L2 was challenging. Chen and Guion-Anderson (2012) discovered that advanced American learners of Mandarin were unable to mark Mandarin focus with either PFC or variations in intensity, despite marking English focus with PFC. Similarly, Mandarin learners of Dutch were also found not to mark focus with PFC in their L2 Dutch production (He et al., 2011).

In addition to Mandarin, the inability to mark focus with native-like prosodic cues has also been observed in the production of narrow focus in Hong Kong English by native Cantonese speakers (Fung & Mok, 2014). Acoustic analysis revealed that Cantonese speakers did not use PFC to mark focus in their English production, consistent with their use of prosody in native Cantonese. However, they did show an expansion of the pitch range of the focused words, a characteristic of English but not substantial in their native

Cantonese. These findings suggest that Cantonese speakers may employ distinct prosodic strategies for marking focus in different languages.

In studies on L2 focus perception, bilingual speakers of Cantonese have demonstrated difficulties in comprehending focus in L2 English. Ge and colleagues (2018) found that native Cantonese listeners showed insensitivity towards incongruent prosodic patterns with the focus marker “only” in L2 English. In contrast to Dutch learners of English, who interpreted L2 English focus similarly to native speakers, indicating positive transfer from L1 to L2, Cantonese learners' non-native-like performance implies negative transfer from Cantonese to L2 English. However, the comprehension of prosodic focus in L3 Mandarin by trilingual speakers (L1 Cantonese and L2 English) is not well understood.

To sum up, native Cantonese listeners have been shown to have the ability to identify the focus positions of simple context-free sentences, except for those with high-level tone (Wu & Xu, 2010; Wu, 2013), whereas their insensitivity to incongruous prosody in L2 English may be due to the limited use of pitch-related cues in their native language (Ge et al., 2018). Nonetheless, the topic of whether and how trilingual speakers of native Cantonese can process prosodic focus in their L3 Mandarin, which differs from Cantonese in focus marking, remains largely unexplored. Moreover, whether L2 or L3 experience influences the decoding process in native Cantonese is still unclear. It is, therefore, necessary to investigate how cross-linguistic focus-marking measures used by trilingual speakers affect the efficiency of focus processing.

2.4 Multimodal focus processing

In face-to-face conversations, people comprehend speech not only by relying on the acoustic signals in the auditory domain but also by exploiting the co-speech visual cues, such as body movements (e.g., head motion), facial expressions (e.g., eyebrow rising), and articulatory movements (e.g., lip and jaw movements) produced by interlocutors. Previous studies on speech production have found that speakers may use some of these visual cues to express or emphasize their intentions together with auditory cues in spoken language (Scarborough et al., 2009; Dohen & Lœvenbruck, 2005; Dohen et al., 2006; Beskow et al., 2006; Kim et al., 2014; Esteve-Gibert et al., 2017; Carignan et al., 2021). Specifically, the eyebrow raising and head nodding is associated with prosodic focus. For example, a study found that English speakers exhibited different eyebrow movements between the narrow and broad focus (Kim et al., 2014). Given that these visual cues are linked to prosodic focus, it is reasonable to assume that listeners can also benefit from visual cues in processing focus. Previous research has sought to investigate the extent to which listeners rely on additional visual cues in speech comprehension and whether such cues facilitate the perception of prosodic prominence in audio-visual modality input (Dohen & Lœvenbruck, 2005; Dohen & LAœvenbruck, 2009; Swerts & Krahmer, 2008; Swerts & Krahmer, 2005; Cvejic et al., 2010; Dohen & LAœvenbruck, 2009).

Dohen and Lœvenbruck (2005) found that French speakers exhibit variations in lip opening areas when expressing contrastive focus compared to broad focus. Furthermore, listeners were more successful in identifying contrastive focus than broad focus in a subsequent visual-only perception study. Their preliminary study on the perception and production of prosodic audio-visual cues also suggests that individual differences between

speakers in visually expressing prosodic focus and the salience of visual cues may facilitate perception. In their further study, Dohen and Lœvenbruck (2009) aimed to determine whether adding auditory-visual cues would enhance prosody perception compared to perception based on auditory cues alone. Their findings confirmed the ceiling effect of audio-only perception, previously observed in the perception of prosodic prominence in Dutch (Swerts & Krahmer, 2004), suggesting that the augmentation of additional visual information in prosodic perception may not be possible. Their results also supported the notion that, in prosody perception, visual information interacts with auditory signals rather than simply being superimposed on auditory cues. In other words, perceptual performance in the audio-visual modality may be comparable to the audio-only modality when the auditory information is sufficiently strong. However, if the perception accuracy in the visual-only modality is too poor, this may result in misleading perception and lead to worse performance in the audio-visual modality compared to the audio-only modality.

In the across-modals perception of lexical tones, the ceiling effect on audition was also found in the Mandarin tone identification task by native listeners (Hannah et al., 2017). Native Mandarin listeners demonstrated high accuracy in identifying isolated lexical tones in congruent auditory-visual modality, but the additional facial cues and gestures did not improve their performance compared to the auditory representations alone. However, the facial and gesture cues improved the performance of native English learners of Mandarin in the same study. Additionally, visual information aids naïve L2 tonal identification and discrimination in native tonal listeners (Burnham et al., 2011). The positive effect of visual cues was confirmed in discrimination tasks involving Thai and Mandarin lexical tones.

Dynamic tones were found to contain sufficient visual information compared to level tones, contributing to higher accuracy than auditory signals (Burnham et al., 2011).

Burnham et al. (2011) investigated the effect of adding visual cues on the identification of Cantonese lexical tones across three modalities: audio-only, video-only, and audio-video input. The study found that the inclusion of visual information had a negligible effect on the identification accuracy compared to auditory input alone. However, the participants demonstrated superior performance in audio-only and audio-visual modalities relative to visual-only input. Additionally, listeners exhibited greater identification accuracy of contour tones compared to level tones across all three modalities, with the difference being least pronounced in the visual-only modality. Burnham et al. (2022) further examined the effects of facial cues and head motions for non-native listeners on the discrimination of naïve Cantonese lexical tones as well as phones (vowels and consonants) in three modalities. The results revealed that in the video-only modality, more visual components (both face and head motions) were necessary for tone perception than phone perception. Furthermore, neither head nor face motions individually enhanced tone perception in the audio-video modality compared to audio information alone. These studies further proposed that complex and demanding cognitive load caused by multiple input modalities may impede the perceptual performance of non-native listeners.

In the literature on the development of audiovisual perception, previous studies have indicated that school-aged children are capable of using visual cues in conjunction with prosodic variations to understand speakers' beliefs and attitudes (Krahmer & Swerts, 2005; Visser et al., 2011; Rapin & Ménard, 2019; for a review, Esteve-Gibert & Guellaï, 2018). For instance, seven- to eight-year-old children may use visual information, such as

eyebrow movements or funny faces, to produce and comprehend speakers' signaling uncertainty, but both production and perceptual accuracy are lower than in adults (Krahmer & Swerts, 2005). Regarding the audiovisual perception of prosodic focus, Rapin and Ménard (2019) found that French nine-year-old children did not rely on visual cues (e.g., lip area) like adults, although both groups primarily relied on acoustic signals in prosodic focus perception. These findings are consistent with the production study in which eight- to nine-year-old children did not visually mark contrastive focus to the same extent as adults, relying primarily on acoustic cues to signal focus (Ménard et al., 2006).

Therefore, although some studies have explored the impact of visual cues on children's prosodic perception, there is still limited understanding of how children perform in complex tasks that involve decoding lexical and prosodic information from multimodal inputs. This knowledge gap is especially pronounced for native tonal language speakers who must decode lexical and prosodic information from multimodal inputs. Furthermore, beyond native language acquisition, the impact of audio-visual modality on complex prosodic comprehension from a cross-language perspective remains unclear.

Chapter 3. Experiment 1: Cantonese focus processing in trilingual adults

3.1 Introduction

This chapter aims to investigate whether and the extent to which native Cantonese-speaking adults employ congruous prosodic forms to map specific focus categories during multimodal speech comprehension. In terms of processing mechanisms for prosodic focus, a recent study on American English adults revealed that listeners adopt a probabilistic approach to infer the possibilities of specific prosodic focus realizations, demonstrating a bias in favor of narrow focus, while biased against broad focus (Roettger et al., 2019). Their findings suggest that listeners utilize prior knowledge, in conjunction with bottom-up cues, during their decoding of prosodic focus. However, it is still unclear whether tonal language listeners employ a similar approach to interpret focus, considering that F_0 variations parallelly discriminate lexical tones and express prosodic functions in tonal languages like Cantonese. Furthermore, the extent to which adult listeners utilize visual cues as facilitators during the cognitive demanding task of focus processing remains poorly explored.

Previous research on the perception and interpretation of Cantonese prosodic focus has been limited. Wu & Xu (2010) and Wu (2013) were the first to investigate the identification of focus positions utterances of five syllables with various lexical tones, wherein all syllables of an utterance bear the same tone. Their findings revealed that Cantonese listeners are capable of identifying the focal constituent, such as the subject, verb, object or no focus, except for the tonal environment of high-level Tone 1. More recently, Yang (2022) explored the comprehension of Cantonese focus using the paradigm

developed by Roettger et al. (2019). Yang's study aimed to examine both broad focus and narrow focus using varied focus sizes, but the target utterances were all designed in the tonal context of Tone 1. Yang's (2022) results indicate that Cantonese listeners experienced difficulty mapping the information structure of broad focus, as well as detecting the prosodic form of broad focus when competing with the narrow focus category or forms. Additionally, Ge et al.'s (2022) study provides limited evidence of focus comprehension, particularly in terms of adequate or inadequate contrastive focus, as their study focuses on comparing prosodic focus with syntactic focus markers.

The current body of literature on Cantonese focus perception and comprehension has primarily focused on the distinction between broad and narrow focus, with little investigation into the decoding of focus within different focus categories, such as narrow versus contrastive focus. This study aims to fill this gap by examining contrastive focus in both congruous and incongruous focus conditions and comparing it with broad and narrow focus categories. In addition, this study is the first to examine the impact of tonal contexts on the comprehension of prosodic focus in languages with complex tonal systems, such as Cantonese. Unlike previous studies that utilized Cantonese sentences with the same lexical tone (Wu & Xu, 2010, Wu, 2013; Yang, 2022), this study employs target sentences with various combinations of level and rising tones to elucidate the effects of lexical tones on prosodic focus in focus processing. Specifically, the pre-focal constituent (i.e., the subject) had either a level tone (Tone 1, Tone 3, or Tone 6) or a rising tone (Tone 2 or Tone 5), while the on-focus constituent (i.e., the verb) and post-focal constituent (i.e., the object) bore the same level or rising tone. Furthermore, incorporating co-speech visual and auditory cues, this study aims to explore the influence of audio-visual signals on the

comprehension of prosodic focus in complex tonal contexts. By using multimodal inputs, this study offers new insights into the processing of prosodic focus by native Cantonese-speaking adults.

This chapter is dedicated to addressing the following specific research questions:

(1) Can native Cantonese-speaking adults accurately map focus categories to appropriate prosodic forms in congruous prosody?

(2) Can they accurately map between prior discourse and prosodic realizations in incongruous prosody?

(3) Can they discriminate between congruous and incongruous focus conditions by comparing their performance in two different focus conditions?

(4) Do tonal contexts and additional visual information enhance trilingual adults' focus processing?

Based on the findings of previous studies, I hypothesize that Cantonese-speaking adults have the ability to map prosodic focus in congruous prosody, but they may perform worse in incongruous prosody. Moreover, they can distinguish matched and mismatched focus conditions, but their accuracy may vary for specific focus categories, such as broad focus, as supported by Yang (2022). Regarding the effects of tonal context, I predict that Cantonese adults will show better performance on rising tones than on level tones, in line with the results of Wu & Xu (2010) and Wu (2013). Additionally, they may display less reliance on visual cues since the acoustic cues of focus production are sufficiently informative.

3.2 Method

This study designed perceptual experiments to investigate the prosodic focus mappings in complex tonal contexts in Cantonese by native listeners. Following the question-answer dialogue paradigm (Chen, 2010; Büring, 2012; Roettger et al., 2019), the stimuli of this study consisted of dialogues where the prosody of the answer either matches or mismatches the focus elicited by the question. The prosodic naturalness of the pair of dialogues was evaluated using a 5-point Likert scale, where one stands for "very unnatural" and five stands for "very natural" (5 points). The current study created a set of five-syllable sentences with various tonal contexts in Cantonese (Chapter 3.2.2.1). This research empirically examined focus processing of broad focus, narrow-verb focus, and contrastive-verb focus with or without competitive acoustic forms/discourses by native Cantonese listeners. Details on the design of stimuli are illustrated in Chapter 3.2.2.3.

3.2.1 Participants

I recruited seventy-two adult participants (39 females and 33 males; age: 21.69 ± 2.61 years) of native Hong Kong Cantonese speakers for the online perception experiment. They were born in Hong Kong and had never lived outside Hong Kong for over half a year before they were eighteen. All participants are sequential trilingual whose first language is Cantonese, English as the second language, and Mandarin as the third language. Their parents were all native Cantonese speakers and spoke Cantonese with participants at home. They began to learn English at the mean age of 3.85 ± 1.37 years and Mandarin at 5.15 ± 1.35 years. Participants were asked to evaluate their language proficiency in listening, speaking, reading, and writing, using a 5-point scale from "not fluent at all" (one point) to "native or near-native" (five points). Native Cantonese adults evaluated their Mandarin proficiency

in listening, speaking, reading, and writing with scores of 3.89 ± 0.74, 3.39 ± 0.85, 4.25 ± 0.78, and 3.81 ± 0.87 points, and English of 3.78 ± 0.7, 3.36 ± 0.86, 3.81 ± 0.7, 3.46 ± 0.9 points. Besides, 54% of the participants had musical training experience in vocals or instruments with training years of 8.13 (SD± 5.18) on average, and a third of them kept training within the past five years. None of the participants reported hearing or speaking impairment. The current research has obtained human ethics approval from the Hong Kong Polytechnic University (HSEARS20200919002).

3.2.2 Materials and design

3.2.2.1 Target sentences

The target sentences of the current research are five-syllable declarative sentences with controlled lexical tones of syllables in Cantonese. Similar to the sentences used in previous studies (Xu et al., 2012, Wu and Xu, 2012 and Wu, 2013), the SVO sentences designed for this study have five syllables and consist of three words. The first word represents the subject and is composed of the first two syllables, while the third syllable represents the verb. Finally, the object is represented by the fourth and fifth syllables of the sentence. The lexical tone of each syllable was well controlled. In previous studies on tonal languages, the sentences they created were limited to a mono-tonal context in which all syllables of each sentence had the same lexical tone (e.g., Cantonese: Wu, 2013, Wu and Xu, 2012, Yang, 2022; Mandarin: Xu et al., 2012). In daily communication, however, people have to comprehend the prosodic prominence of utterance in which syllables hardly bear one specific tone throughout the whole sentence. This study created target sentences in both monotonal contexts, as tested in previous studies, as well as multi-tonal contexts to explore whether focus processing is robust in various tonal conditions in natural speech. For

sentences in the multi-tonal context in the current study, the lexical tone of the first two syllables (the subject of SVO sentences here) is different from that of the following three syllables (the verb and the object of a five-syllable SVO sentence). More specifically, the last three syllables composing the verb and the object keep the same lexical tone, which differs from the tone of the two syllables in the subject position in multi-tonal contexts. The first two syllables constituting the subject bear the same lexical tone.

The carryover and anticipatory effects on F_0 contours in tonal languages demonstrate that the offset of the preceding syllable tends to affect the early portion of the following syllable (carryover effect), while the pitch target of the following syllable only slightly influences the F_0 contour of preceding syllable (anticipatory effect) (e.g., Mandarin, Xu & Wang, 2001; Hong Kong Cantonese: Li et al., 2020; Wong, 2007). The design of tonal context in the current study included the combination of level tones (high-level Tone 1, middle-level Tone 3, and low-level Tone 6) and rising tones (high-rising Tone 2 and low-rising Tone 5). To consider the tonal context of the entire utterance, it was divided into two portions, specifically the subject and the verb phrase. Each portion is characterized by either a level tone or a rising tone, with either similar or distinct F_0 onsets and offsets. Specifically, we created four types of tonal combinations for a total of 20 target sentences as shown in below Table 3.1:

Type	**Tones on Subjects**	**Tones on Verb Phrases**	**Tonal Combination**	**Tonal Combination**	**Tonal Combination**
1	Level Tones (Tone1, Tone3), Tone6)	Level Tones (Tone1, Tone3)	Tone 1 + Tone 1	Tone 3 + Tone 1	Tone 6 + Tone 1
			Tone 1 + Tone 3	Tone 3 + Tone 3	Tone 6 + Tone 3
2		Rising Tones	Tone 1 + Tone 2	Tone 3 + Tone 2	Tone 6 + Tone 2
			Tone 1 + Tone 5	Tone 3 + Tone 5	Tone 6 + Tone 5

		(Tone2, Tone5)			
3	Rising Tones (Tone2, Tone5)	Level Tones (Tone1, Tone3)	Tone 2 + Tone 1		Tone 5 + Tone 1
			Tone 2 + Tone 3		Tone 5 + Tone 3
4		Rising Tones (Tone2, Tone5)	Tone 2 + Tone 2		Tone 5 + Tone 2
			Tone 2 + Tone 5		Tone 5 + Tone 5

Table 3.1 Tonal combinations of Cantonese target sentences

Type 1 is a level tone followed by another level tone, in which tones of the subject words are one of the level tones among Tone 1 (55), Tone 3 (33), or Tone 6 (22), followed by level tones of Tone 1 (55) or Tone 3 (33) on verb phrases. Six tonal combinations (3 tones on the subject *2 tones on the verb phrase) carry the level tones through the whole sentence. Type 2 is about sentences bearing a level tone followed by a rising tone. Lexical tones of the subject have three options Tone 1 (55), Tone 3 (33), or Tone 6 (22), and that of the verb phrase is a rising tone like Tone 2 (25) or Tone 5 (23). These tonal combinations also created six types of sentences. Type 3 refers to sentences where a level tone on verb phrases follows the rising tone on subject words, and four tonal contexts were designed for sentences. Type 4 stands for the tonal combination of two rising tones on the subject word and verb phrase, respectively.

Similarly, four sentences were created with this type of tonal context. Following the combinations of tones, the twenty target sentences we designed included four sentences in mono-tonal contexts, and the other sixteen had multi-tonal contexts. Moreover, checked tone syllables were excluded from all stimuli sentences. Details of all target sentences for the present Experiment 1 can be found in Appendix 1. Table 3.2 lists several examples of target sentences. In the column of sentences, the first line is traditional Chinese characters

of sentences used in the experiment, above International Phonetic Alphabet (IPA), Jyutping, and English transcriptions for reference. The tone of each syllable was marked after IPA and Jyutping symbols in Table 3.2.

Types	Tones of Subjects	Tones of VPs	Sentences
1	Tone 1 (55) High-Level	Tone 3 (33) Mid-Level	欣欣餵兔兔 /jɐn55 jɐn55 wɐi33 tʰou33 tʰou33/ Jan55 jan55wai33 tou33 tou33 Janjan fed the rabbit.
2	Tone 1 (55) High-Level	Tone 2 (25) High-Rising	坤坤洗水果 /kʰwɐn55 kʰwɐn55 sɐi25 sɵy25 kwɔ25/ Kwan55 kwan55 sai25 seoi25 gwo25 Kwankwan washed fruits.
3	Tone 2 (25) High-Rising	Tone 1 (55) High-Level	丙仔吹風車 /pɪŋ25 tsɐi25 tsʰɵy55 fɔŋ55 tsʰɛ55/ bing25 zai25 ceoi55 fung55 ce55 Bingzai blew a windmill.
4	Tone 5 (23) Low-Rising	Tone 2 (25) High-Rising	敏敏剪海藻 /mɐn23 mɐn23 tsin25 hɔi25 tsou25/ Man23 man23 zin25 hoi25 zou25 Manman cut seaweeds.

Table 3.2 Examples of Cantonese sentences in Experiment 1

In addition, I created sixteen sentences as fillers in which the lexical tones were controlled in different tonal combinations of target sentences in Experiment 1. They followed the pattern of tones in that the first two syllables bear the same lexical tone and the last three syllables on verb phrases share one lexical tone. Instead of combinations of level and rising tones on target sentences, filler sentences involved the mid-low falling Tone 4 (21) and the other five lexical tones. Fourteen filler sentences were designed,

including fourteen sentences with multi-tonal combinations and two sentences in the mono-tonal context. Details of the fillers can be found in Appendix 2.

3.2.2.2 Audiovisual recordings

Following the question-answer paradigm in Breen et al. (2010), we used utterances with various focus categories elicited by different questions in conversation. First, *Wh*-questions elicit answers varying in focus size (broad vs. narrow) and focus positions (initial, medial, and final). For example, precursor *Wh*-questions (1), (2), (4), and (6) in Table 3.3 respectively question the whole sentence, the subject, the verb, and the object of the target sentence and elicit answers with a broad focus (1) and narrow focus on different constituents (2), (4), (6). Besides, like precursor questions (3), (5), and (7) in the same table, each yes-no question contains a wrong constituent that elicits a contrastive-narrow focus in answer to correct the false information in the preceding context. Dialogues (1) (2) (3) listed the categories of focus that were examined in the current study, which involved broad focus on the whole target sentence as well as (non-contrastive) narrow focus and contrastive focus on the medial constituent (the verb).

Dialogues	**Question-Answer**		**Focus Categories**	**Focus Positions**
1	Q-1	發生咗咩事啊？ fat33 sɐŋ55 tsɔ25 mɛ55 si22 a55? faat33 sang55 zo25 me55 si22 aa55? What happened?	Broad	Broad
	A-1	[坤坤洗水果]$_F$ [$k^hwɐn55$ $k^hwɐn55$ sɐi25 sɵy25 kwɔ25]$_F$ [kwan55 kwan55 sai25 seoi25 gwo25]$_F$ Kwankwan washed the fruit.		
2	Q-2	坤坤乜嘢水果？ $k^hwɐn55$ $k^hwɐn55$ mɐt55 jɛ23 sɵy25 kwɔ25? Kwan55 kwan55 mat55 je23 seoi25 gwo25?	Narrow	Medial

		What did Kwankwan do to the fruit?		
	A-2	坤坤[洗]$_{F}$水果 k^{h}wɐn55 k^{h}wɐn55 [sɐi25]$_{F}$ sɵy25 kwɔ25 kwan55 kwan55 [sai25]$_{F}$ seoi25 gwo25 Kwankwan [washed]$_{F}$ the fruit.		
3	Q-3	坤坤切水果? k^{h}wɐn55 k^{h}wɐn55 ts^{h}ɐi33 sɵy25 kwɔ25? kwan55 kwan55 cit33 seoi25 gwo25? Did Kwankwan cut the fruit?	Contrastive	Medial
	A-3	坤坤[洗]$_{\text{Contrastive-F}}$水果 k^{h}wɐn55 k^{h}wɐn55 [sɐi25]$_{\text{Contrastive-F}}$ sɵy25 kwɔ25 kwan55 kwan55 [sai25]$_{\text{Contrastive-F}}$ seoi25 gwo25 Kwankwan [washed]$_{\text{Contrastive-F}}$ the fruit.		
4	Q-4	邊個洗水果? bin55 kɔ33 sɐi25 sɵy25 kwɔ25? bin55 go33 sai25 seoi25 gwo25? Who washed the fruit?	Narrow	Initial
	Ar-4	[坤坤]$_{F}$洗水果 [k^{h}wɐn55 k^{h}wɐn55]$_{F}$ sɐi25 sɵy25 kwɔ25 [kwan55 kwan55]$_{F}$ sai25 seoi25 gwo25 [Kwankwan]$_{F}$ washed the fruit.		
5	Q-5	媽媽洗水果? ma55 ma55 sɐi25 sɵy25 kwɔ25? maa55 maa55 sai25 seoi25 gwo25? Did Mom wash the fruit?	Contrastive	Initial
	A-5	[坤坤]$_{\text{Contrastive-F}}$洗水果 [k^{h}wɐn55 k^{h}wɐn55]$_{\text{Contrastive-F}}$ sɐi25 sɵy25 kwɔ25 [kwan55 kwan55]$_{\text{Contrastive-F}}$ sai25 seoi25 gwo25 [Kwankwan]$_{\text{Contrastive-F}}$ washed the fruit.		
6	Q-6	坤坤洗乜嘢啊? k^{h}wɐn55 k^{h}wɐn55 sɐi25 mɐt55 jɛ23 a55? Kwan55 kwan55 sai25 mat55 jɛ23 aa55? What did Kwankwan wash?	Narrow	Final

	A-6	坤坤洗[水果]]F kʰwɐn55 kʰwɐn55 sɐi25 [sɵy25 kwɔ25]F kwan55 kwan55 sai25 [seoi25 gwo25] F Kwankwan washed [the fruit] F.		
7	Q-7	坤坤洗蔬菜? kʰwɐn55 kʰwɐn55 sɐi25 sɔ55 tsʰɔi33? Kwan55 kwan55 sai25 so55coi33? Did Kwankwan wash the vegetable?	Contrastive	Final
	A-7	坤坤洗[水果]Contrastive-F kʰwɐn55 kʰwɐn55 sɐi25 [sɵy25 kwɔ25] Contrastive-F kwan55 kwan55 sai25 [seoi25 gwo25] Contrastive-F Kwankwan washed [the fruit] Contrastive-F.		

Table 3.3 Focus categories elicited by Q-A paradigm in Cantonese

We recruited two native Hong Kong Cantonese speakers (one female, 20 years; one male: 21 years) to record the stimuli using the question-answer paradigm in a semi-spontaneous conversation. Both speakers have similar language experiences as the participants of the perception experiment, who were born and live in Hong Kong and sequentially learned English and Mandarin as their second and third languages. Two speakers were first asked to familiarize themselves with target sentences (answers) and the questions, following the procedures of production research in previous studies (e.g., Breen et al., 2010). Speakers were then instructed to read all target sentences naturally for familiarization. The content of question-answer dialogues was represented by images without any text of target sentences and displayed on E-prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). The male speaker asked a question according to the presented picture shown on the screen, and the female speaker answered the question using the target sentence indicated by the picture. All images were created for the target sentences in the current study, and the questioning constituent (e.g., subject, verb, or object) was covered using a gray square to elicit the focus of each answer. Images without covered

areas would be exposed to speakers after the male speaker was questioned. For example, in each trial, when Image 3.1 was displayed on the screen, the male speaker was instructed to ask, “What did Kwankwan do to the fruit?”, as the part of the picture representing the verb of the target sentence was covered. The experimenter immediately displayed Image 3.2, and the female speaker answered with the target sentence, “Kwan kwan [sai]$_F$ seoi gwo”. Two speakers were not informed about the research purpose. There are 252 pairs of dialogues (36 target sentences * 7 focus categories) with three repetitions randomly displayed on E-prime 3.0. We selected one repetition as the stimuli used in the following experiments. The conversations were recorded by a Handycam Sony HDR-JP670, and the video screen was adjusted to film the female speaker from the shoulder to her head. They were also recorded by Audacity with a sampling rate of 44.1 kHz at the Speech and Language Sciences Laboratory of the Hong Kong Polytechnic University.

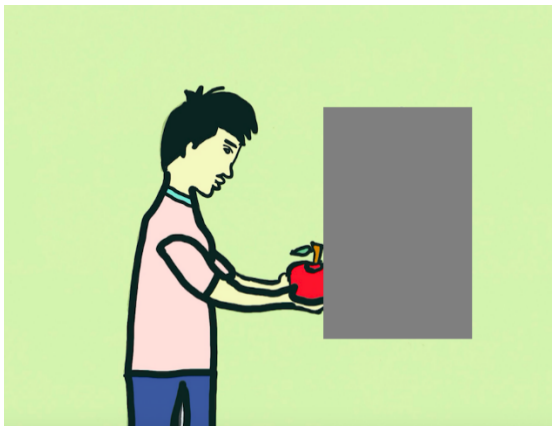


Image 3.1

Image 3.2

For the target sentences with three repetitions, I segmented the auditory recordings, labeled them with syntactic positions and focus categories, and extracted several acoustic measurements by the script of ProsodyPro (Xu, 2013) in Praat (Boersma & Weenink, 2022).

Figure 3.1 describes the time-normalized F_0 contours of broad focus, narrow-verb focus, and contrastive-verb focus of the stimuli examined in the perception experiment. Figure 3.2 shows the female speaker's F_0 range of each constituent under the three focus categories: broad focus (green color), narrow-verb focus (orange color), and contrastive-verb focus (purple color). Figures 3.3 and 3.4 plot the mean duration (ms) and intensity (dB) of broad focus, narrow and contrastive focus on verbs divided by syntactic positions in three facets. Moreover, I submitted the measures of F_0 correlates, duration, and intensity extracted from the SVO utterances to the linear discriminant analyses (LDA) by *MASS* package in R. The LDA results demonstrate that mean duration obtains the highest accuracy among all measurements in Cantonese. The LDA of mean duration discriminates broad focus from narrow-verb focus (63.33%) and contrastive-verb focus (66.12%). The accuracy of LDA in distinguishing narrow-verb focus from contrastive-verb focus was 51.37%. For F_0 relatives, the LDA accuracies on the F_0 range were 52.78% and 52.19% in discriminating broad focus from narrow-verb and contrastive-verb focus and 50.27% in differentiating narrow-verb and contrastive-verb focus. The accuracy of LDA on mean F_0 was 53.89% (broad focus vs. narrow-verb focus), 51.91% (broad focus vs. contrastive-verb focus), and 52.73% (narrow-verb focus vs. contrastive-verb focus), respectively. The LDA results of mean intensity were 54.17%, 53.55%, and 53.01% in the discrimination of the three contrasts of focus types.

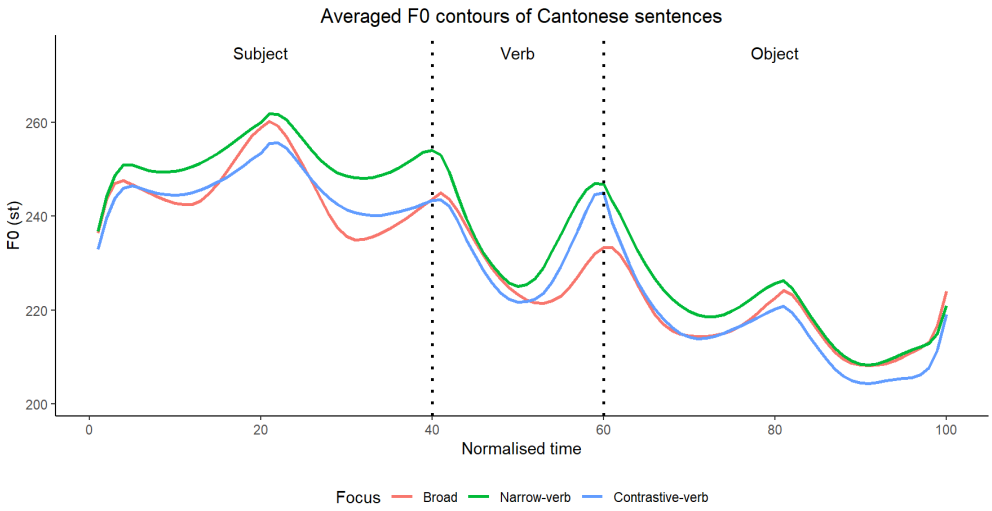


Figure 3.1 F_0 time normalized contours of Cantonese target sentences

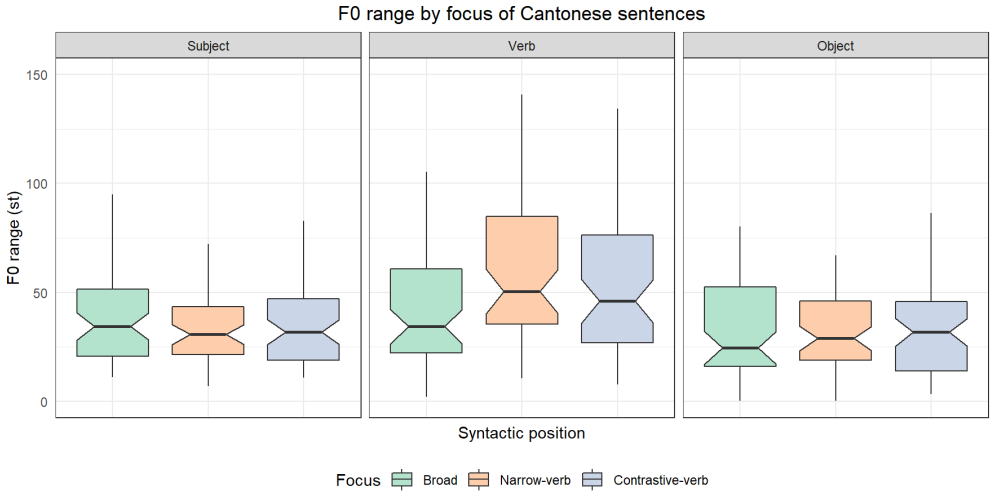


Figure 3.2 F_0 range by the focus of Cantonese target sentences

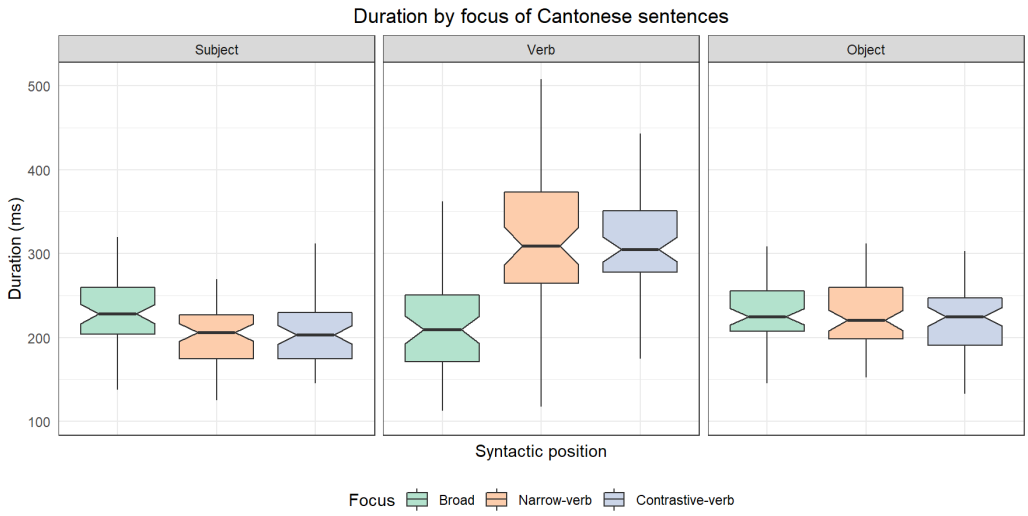


Figure 3.3 Duration by the focus of Cantonese target sentences

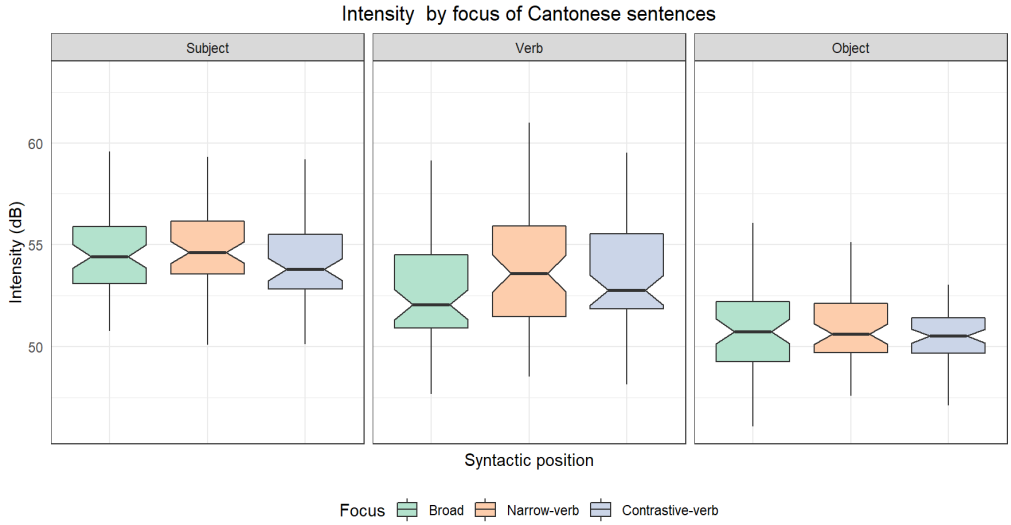


Figure 3.4 Intensity by the focus of Cantonese target sentences

3.2.2.3 Stimuli

Following Roettger et al. (2019)'s question-answer paradigm, I adopted congruous and incongruous pairs of dialogues in the perception experiment to examine native adults' focus comprehension in Cantonese. All stimuli were blocked by modality: audio-only

modality and audio-visual modality in terms of the answer utterances. In each trial, the precursor question was audio-only for both two modalities, and the elicited target sentence was either audio-only presented, or audio-visual displayed to participants by E-prime Go, the online platform provided by E-prime 3.0 software. Participants were instructed to rate how natural the prosody of the answer (target sentence) matched the precursor question using a 5-Likert scale, in which one point referred to the least natural prosody and five points presented the most natural prosody. I developed a 3×2×2 design to manipulate Focus Categories (broad, narrow, contrastive), Pairs (match, mismatch), and Modalities (audio-only, audio-visual). For focus categories, the experimental stimuli tested the narrow focus, contrastive focus on verbs, and broad focus on the whole sentences. Dialogues with narrow and contrastive focus on subjects and objects were used as fillers and practice trials.

Matched pairs indicated the real and natural dialogues recorded by the two native speakers in semi-spontaneous question-answer conversations. Mismatched pairs referred to dialogues whose answers were not elicited by the current precursor questions. For instance, in Table 3.4, pair (a) demonstrated the congruous prosody of a target sentence with a narrow focus on the verb; pair (b) showed an example of mismatched pair of dialogue used in the current study. The answer "[Kwankwan washed the fruit] $_{F}$" should be elicited by questions like "What happened?" instead of the question "What did Kwankwan do to the fruit?" in dialogue (b).

a. Matched Pair with BF focus condition	**Focus category**

Question	發生咗咩事啊？ fat3 sɐŋ1 tsɔ2 mɛ1 si6 a1? faat3 sang1 zo2 me1 si6 aa1? What happened?	Broad
Answer	[坤坤洗水果] $_{F}$。 [k^{h}wɐn1 k^{h}wɐn1 sɐi2 sœy2 kwɔ2] $_{F}$. [kwan1 kwan1 sai2 seoi2 gwo2] $_{F}$. [Kwankwan washed F the fruit] $_{F}$.	Broad
b. Matched Pair with NV focus condition		
Question	坤坤乜嘢水果? k^{h}wɐn1 k^{h}wɐn1 mɐt1 jɛ5 sœy2 kwɔ2? kwan1 kwan1 mat1 je5 seoi2 gwo2? What did Kwankwan do to the fruit?	Narrow Verb
Answer	坤坤[洗] $_{F}$ 水果。 k^{h}wɐn1 k^{h}wɐn1 [sɐi2] $_{F}$ sœy2 kwɔ2. kwan1 kwan1 [sai2] $_{F}$ seoi2 gwo2. Kwankwan [washed] $_{F}$ the fruit.	Narrow Verb
c. Matched Pair with CV focus condition		
Question	坤坤切水果? k^{h}wɐn1 k^{h}wɐn1 ts^{h}it3 sœy2 kwɔ2? kwan1 kwan1 cit3 seoi2 gwo2? Did Kwankwan cut the fruit?	Contrastive Verb
Answer	坤坤[洗] $_{\text{contrastive-F}}$ 水果。 k^{h}wɐn1 k^{h}wɐn1 [sɐi2] $_{\text{contrastive-F}}$ sœy2 kwɔ2. kwan1 kwan1 [sai2] $_{\text{contrastive-F}}$ seoi2 gwo2. Kwankwan [washed] $_{\text{contrastive-F}}$ the fruit.	Contrastive Verb
d. Mismatched Pair with NB focus condition		

Question	坤坤乜嘢水果？ $k^hwɐn1\ k^hwɐn1\ mɐt1\ jɛ5\ sœy2\ kwɔ2?$ kwan1 kwan1 mat1 je5 seoi2 gwo2? What did Kwankwan do to the fruit?	Narrow Verb
Answer	[坤坤洗水果]$_F$ [$k^hwɐn1\ k^hwɐn1\ sɐi2\ sœy2\ kwɔ2$]$_{F.}$ [kwan1 kwan1 sai2 seoi2 gwo2]$_{F.}$ [Kwankwan washed the fruit]$_F$.	Broad
e. Mismatched Pair with CB focus condition		
Question	坤坤切水果？ $k^hwɐn1\ k^hwɐn1\ ts^hɐi3sœy2\ kwɔ2?$ kwan1 kwan1 cit3 seoi2 gwo2? Did Kwankwan cut the fruit?	Contrastive Verb
Answer	[坤坤洗水果]$_F$ [$k^hwɐn1\ k^hwɐn1\ sɐi2\ sœy2\ kwɔ2$]$_{F.}$ [kwan1 kwan1 sai2 seoi2 gwo2]$_{F.}$ [Kwankwan washed the fruit]$_F$.	Broad
f. Mismatched Pair with NC focus condition		
Question	坤坤乜嘢水果? $k^hwɐn1\ k^hwɐn1\ mɐt7\ jɛ5\ sœy2\ kwɔ2?$ kwan1 kwan1 mat1 je5 seoi2 gwo2? What did Kwankwan do to the fruit?	Narrow Verb

Answer	坤坤[洗]$_{\text{Contrastive-F}}$ 水果。 $k^hwɐn1$ $k^hwɐn1$ [sɐi2]$_{\text{Contrastive-F}}$ sɵy2 kwɔ2. kwan1 kwan1 [sai2]$_{\text{Contrastive-F}}$ seoi2 gwo2. Kwankwan [washed]$_{\text{Contrastive-F}}$ the fruit.	Contrastive Verb

Table 3.4 Examples of matched and mismatched pairs used in Experiment 1

In the current research, I designed three conditions of matched pairs and three conditions of mismatched pairs with a balanced number in audio-only and audio-visual modalities. In matched pairs, the real and natural dialogues of broad focus (BF), narrow-verb focus (NV), and contrastive-verb focus (CV) constituted three types of focus conditions with felicitous congruous prosody. In mismatched pairs, the answer utterances produced with broad focus or contrastive-verb focus were used as the infelicitous answers (competitors) to inappropriate precursor questions. Specifically, the first condition of mismatched pairs was that prosodic broad focus competed with target narrow-verb focus (NB) in answers, like exampled dialogue (b) in Table 3.4. The second condition of mismatched pairs involved answers with prosodic broad focus competing with target contrastive-verb focus (CB) like dialogue (c). The last condition of mismatched dialogues examined the competition between the contrastive-verb focus (competitor) and narrow-verb focus (target), where the *Wh*-question is followed by an answer with a contrastive focus on the same constituent (NC) like dialogue (d). I separately extracted questions and answers from recordings and then combined them into matched or mismatched pairs of dialogues according to above-mentioned six types of focus conditions (Matched: BF, NV, CV; Mismatched: NB, CB, NC).

The audio-only (AO) and audio-visual (AV) modalities indicated different input modes of the answering utterances of each pair of dialogues. At the same time, the precursor questions kept a consistent audio-only context in both AO and AV modalities. In the AO modality, the audios of questions and answers were separately extracted from auditory recordings to be used as the stimuli in which a 750-millisecond interval between the question and the answer was set by E-prime 3.0 software. In the AV modality, I imported the auditory files used in AO condition, synchronized them with videos in Adobe Premiere Pro, and substituted the original sounds of videos recorded by the camera. Thus, both modalities' auditory inputs were acoustically consistent, and each dialogue's answer was extracted into individual videos for the AV modality. Again, the extracted sound files of questions and video files of answers were presented with an interval of 750 ms on E-Prime 3.0.

Experiment 1 included 216 trials of dialogues (20 target sentences * 6 focus categories + 16 fillers * 6 focus categories) evenly distributed into three sets. Each trial of dialogues was only displayed once to avoid task effects on participants' cognitive comprehension of prosody. Three groups of participants were recruited and randomly assigned to one of the three sets. Thus, each participant was randomly presented with 72 trials of dialogues (20 target sentences * 2 focus categories + 16 fillers * 2 focus categories) in one list. The two focus categories included the matched (one of BF, NV or CV condition) and mismatched (one of NB, CB or NC condition) pairs of one target utterance in different modalities. The order of AO and AV modalities was random in one list, and each modality contained 36 trials of dialogues in random order.

3.2.3 Procedures

3.2.3.1 Practice sessions

The purpose of practice sessions was to familiarize participants with perceptual evaluations and the understanding of prosodic naturalness in question-answer congruences. Each participant started with practice sessions, in which they were asked to judge how natural the prosody of answers produced by the female speaker was matched with precursor questions asked by the male speaker. Particularly, only two degrees of naturalness, "very natural" and "very unnatural", were tested in practice sessions to help participants understand the task efficiently. When the prosody of answers is congruous with the preceding questioning content, participants were instructed to press "5" on their keyboards, indicating that the answer had a "very natural" prosody that matches the elicited focus by the question. Practice sessions also included "very unnatural" prosody and participants were asked to press "1" on the keyboard. They were expected to discriminate between "very natural" and "very unnatural" prosody in practice sessions.

Following the design of the experimental stimuli, the practice stimuli used two filler sentences and were designed with four focus conditions for each sentence, of which half involved matched pairs and half were mismatched. The four focus conditions were excluded from experimental trials so that the research purpose was not revealed to the participants. Matched pairs included question-answer conversations in which answers have a matched prosody with a narrow focus or contrastive focus on objects of utterances. For example, in Table 3.5, the dialogues we used were as follows: question (1) "What did Jiujiu draw?" elicited the felicitous answer (1) "Jiujiu drew [an elephant]$_{F}$", or answer (2) "Jiujiu drew [an elephant]$_{\text{Contrastive- F}}$" had appropriate prosody preceded by question (2) like "Did

Jiujiu draw a tiger?". In mismatched pairs, answers (3) and (4) were acoustically identical to answers (1) and (2) in Table *3.5*, but their lead-in questions did not appropriately match with the answers. For instance, in dialogues (3) the answer "Jiujiu drew [an elephant]$_{F}$" was preceded by the question "Who drew an elephant?"; (4) the answer "Jiujiu drew [an elephant] $_{\text{Contrastive-F}}$" followed a mismatched (4) question "Did Waawaa draw an elephant?". Practice sessions presented the four exampled dialogues in an audio-only modality, and the same focus conditions designed with another filler sentence were displayed in an audio-visual modality. In the current study, eight trials of dialogues (half with matched pairs) were used in the practice sessions, and the order of AO or AV modalities was randomized.

a. Matched Pairs		**Focus types**
Question (1)	耀耀畫乜嘢啊? jiu6 jiu6 wak6 mɐt1 jɛ5 a1? jiu6 jiu6 waak6 mat1 je5 aa1? What did Jiujiu draw?	Narrow Object
Answer (1)	耀耀畫[大象]$_{F}$. jiu6 jiu6 wak6 [tai6 tsœŋ6] $_{F.}$ jiu6 jiu6 waak6 [daai6 zoeng6] $_{F.}$ Jiujiu drew an [elephant] $_{F}$.	Narrow Object
Question (2)	耀耀畫老虎? jiu6 jiu6 wak6 lou5 fu2? jiu6 jiu6 waak6 lou5 fu2? Did Jiujiu draw a tiger?	Contrastive Object
Answer (2)	耀耀畫[大象] $_{\text{Contrastive-F}}$. jiu6 jiu6 wak6 [tai6 tsœŋ6] $_{\text{Contrastive-F}}$ jiu6 jiu6 waak6 [daai6 zoeng6] $_{\text{Contrastive-F}}$ Jiujiu drew an [elephant] $_{\text{Contrastive-F}}$.	Contrastive Object

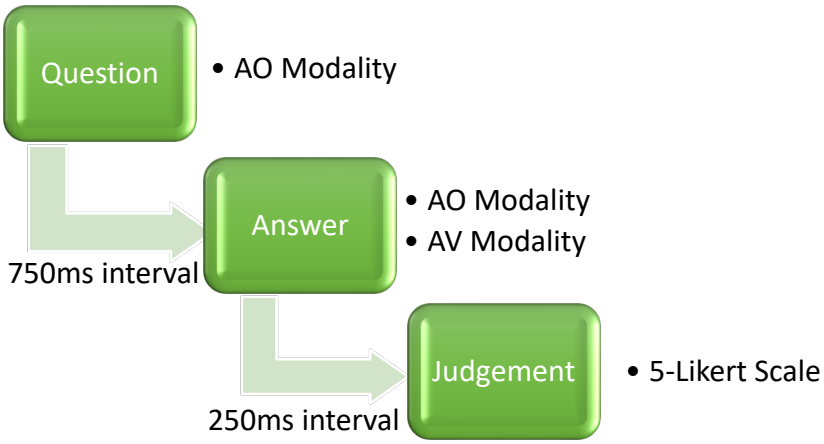
b. Mismatched Pairs

Question (3)	邊個畫大象啊？ pin1 kɔ3 wak6 tai6 tsœŋ6? bin1 go3 waak6 daai6 zoeng6? Who drew an elephant?	Narrow Subject
Answer (3)	耀耀畫[大象]$_{F}$ jiu6 jiu6 wak6 [tai6 tsœŋ6]$_{F}$ jiu6 jiu6 waak6 [daai6 zoeng6]$_{F}$ Jiujiu drew an [elephant]$_{F}$.	Narrow Object
Question (4)	華華畫大象？ Wa4 wa4 wak6 tai6 tsœŋ6? Waa4 waa4 waak6 daai6 zoeng6? Did waawaa draw an elephant?	Contrastive Subject
Answer (4)	耀耀畫[大象]$_{Contrastive\text{-}F}$ jiu6 jiu6 wak6 [tai6 tsœŋ6]$_{Contrastive\text{-}F}$ jiu6 jiu6 waak6 [daai6 zoeng6]$_{Contrastive\text{-}F}$ Jiujiu drew an [elephant]$_{Contrastive\text{-}F.}$	Contrastive Object

Table 3.5 Examples of matched and mismatched pairs for practice sessions in Experiment 1

In the first session of practice, participants were informed of the prosodic naturalness of answers regarding specific questions and asked to familiarize with correct choices of each trail depending on the instructions on screens. Each trial included three scenarios on screens, as the below flow chart demonstrated. The first step was to display the auditory question with a text reminder of "Please listen to the question" in traditional Chinese characters. After a 750ms interval, the second scenario showed the video/audio of

the answer with the text “Please listen to the answer”. The last scene raised the 5-Likert rating selection screen together with the text guidance, and participants should make judgments of the prosodic naturalness. In the current session, participants were presented with the correct choice of naturalness degree on the third scenario for each trial. For example, the answer for congruous trials was like “the prosody of this answer was very natural, please press ‘5’ on the keyboard”. For mismatched trials like exampled dialogues (3) and (4) in Table 3.5, the suggested choice was to select “very unnatural”, on the third scenario, which indicated the most inappropriate prosody in the current study. The eight trials were sequentially displayed with two repetitions on audio-only and audio-visual modalities by E-Prime Go.



Participants rated the prosody of answers with no answers presented in the second session, and immediate feedback was presented after each trial. The practice trials used in the first session were examined in the second practice session. Participants were instructed to evaluate whether the trials were matched or mismatched in prosody by pressing the “5” or “1” button on the keyboard. All trials were randomly displayed with two repetitions in each modality (audio-only or audio-visual), and participants were presented with their accuracy of the perceptual judgments at the end of each modality. Two practice sessions

were in succession; participants can practice more than once if their accuracy did not reach 60% or if they would like to practice more to familiarize themselves with the task. According to the performance in practice sessions, we excluded six participants who had difficulties discriminating between congruous and incongruous prosodic realizations. Seventy-two participants who obtained an accuracy of or above 60% in practice sessions were included in the experiment. Participants joined the practice sessions by Zoom, during which the experimenter explained the procedures of practice sessions for participants and supervised the whole session of each participant.

3.2.3.2 Experimental sessions

Following the paradigm used in the practice sessions, the experimental session contained one task of Cantonese perceptual judgments adopting a 5-Likert scale designed by E-Prime 3.0 software and conducted on the online platform E-Prime Go. Each participant joined a Zoom meeting with the experimenter before starting the online experiment. They were first informed that experimental trials included relatively complex conditions of focus, and they should rate the prosodic naturalness of the answers using five degrees of naturalness (from "1" and "5" points) instead of two degrees in the practice sessions. Besides, no feedback was offered in the experimental sessions. Participants who passed practice sessions with an accuracy of or above 60% were invited to the experiment and all tasks were completed online within one day. Each listener participated the experimental session only once and completed all tasks (including the practice sessions) within 30 minutes in average.

In the experimental session, each participant perceptually evaluated 72 trials of question-answer dialogues, including 40 pairs of experimental stimuli and 32 pairs of fillers without repetition. Half of them had congruous prosody of answer utterances and half incongruous prosody of answers in terms of precursor questions. Details on stimuli used in

experimental sessions can be found in Chapter 3.2.2.3. All trials of experimental trials and fillers were evenly divided into AO and AV modalities. Participants were randomly assigned to start with the AO or AV modality and notified about the modality of each block. They were instructed to complete the two modalities of all trials in succession without a break. The display of each trial was similar to the three scenarios presented in the practice sessions, in which questions in auditory forms were first displayed and followed by answers in audio or videos with an interval of 750ms. Then the options of the 5-Likert scale (digits from "1" to "5") of naturalness were presented with a 250ms-interval after answers. In the last scenario, five numbers indicating prosodic naturalness degrees were horizontally listed from smallest to largest together with the text "very unnatural", "a little bit unnatural", "neutral", "a little bit natural" and "very natural" in traditional Chinese characters. Their responses on the 5-point scale of prosodic naturalness were recorded.

3.2.4 Data analyses

I collected responses on the 5-Likert scale in the experimental trials from native Cantonese adults in Experiment 1. The cumulative link mixed effects model (CLMM) was fitted to examine how *Focus* (BF, NV, CV, BN, BC, NC), *Tones* (LL, LR, RR, RL), and *Modalities* (AO, AV) influenced Cantonese adults' focus processing in complex tonal contexts with the *ordinal* package (Christensen, 2015) in R (R Core Team, 2022). This study applied the CLMM because the dependent variables were ordered factors: $1 < 2 < 3 < 4 < 5$ on the degree of prosodic naturalness for each trial. In ordinal regression models, responses from 1 to 5 points were dependent variables (*Response*), *Focus*, *Tones,* and *Modalities* factors were predictors, and *Subjects* and *Trials* of dialogues were involved as random effects. Post-hoc tests for factors were conducted with the *emmeans* package (Lenth R, 2023) in R.

3.3 Results

Results from the cumulative link mixed model showed main effects on *Focus* ($\chi^2(5)$ = 392.10, $P < .001$), *Tone* ($\chi^2(3) = 10.17$, $P = 0.017$), and interaction of *Focus* and *Tone* ($\chi^2(15) = 23.91$, $P = 0.067$), but the effect of *Modality* was not significant ($\chi^2(1) = 0.72$, $P = 0.396$). There was no significant effect of the interaction between *Focus* and *Modality* ($\chi^2(5) = 4.72$, $P = 0.453$), the two-way interaction between *Tone* and *Modality* ($\chi^2(3) = 1.12$, $P = 0.771$), or the three-way interaction among *Focus*, *Tone,* and *Modality* ($\chi^2(15) = 16.11$, $P = 0.375$). Next, I closely examined the perceptual responses in various focus conditions and tonal contexts to test native adult listeners' perceptual sensitivity to prosodic felicitousness by post hoc analyses.

3.3.1 Focus processing of congruous prosody in Cantonese by trilingual adults

For congruous prosody, I examined the perceptual mapping on broad focus (BF), narrow-verb focus (NV), and contrastive-verb focus (CV) of matched pairs of dialogues. Figure 3.5 provide an overview of perceptual patterns of congruous prosody in AO and AV modalities grouped by focus categories. The stacked bar plots present the percentage of each response obtained from a 5-Likert scale, where the prosodic naturalness increases from top to bottom in each bar. As Figure 3.5 shows, the pattern of responses from broad focus (the left) is different from that of narrow focus (the middle) and contrastive focus (the right) on verbs. In the dissertation, "natural" and "very natural" responses are regarded as successful identification of congruous prosody, "unnatural" and "very unnatural" responses as incongruous prosody, and "neutral" as unsure answers.

In broad focus, Cantonese adults obtained an accuracy of 34.2% and 33.4% of the natural prosody in the AO and AV modalities, respectively. At the same time, they

performed better in the narrow-verb (NV) focus and contrastive-verb (CV) focus of congruous prosody. Listeners were inclined to accept congruous pairs with narrow-verb focus with an accuracy of 75.8% (in AO) and 79.6% (in AV) as natural prosody. Similarly, they identified the matched acoustic forms of contrastive-verb focus with an accuracy of 77.6% (in AO) and 71.3% (in AV) of congruous prosody in the CV condition. Moreover, the overall perception between AO and AV modalities did not differ significantly among any focus condition, consistent with the statistical results described above.

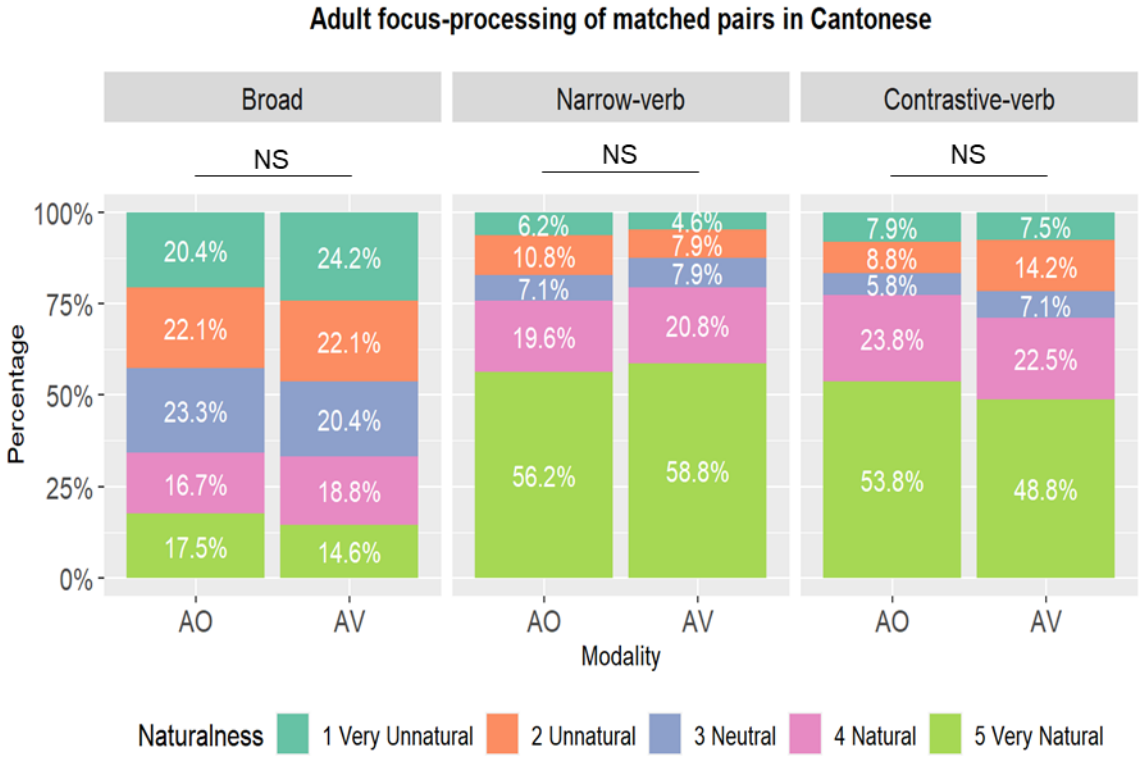


Figure 3.5 Native adults' performance in congruous Cantonese prosody

The post-hoc analysis further provided statistical evidence on Cantonese adult listeners' perceptual mapping of congruous prosody in various tonal contexts. Results showed that they significantly discriminated broad focus from narrow-verb focus and contrastive-verb focus regardless of modalities in most tonal contexts. More specifically, Cantonese adults evaluated the matched broad focus with a significantly lower degree of

naturalness than the ratings for the matched narrow-verb focus in all tonal contexts, as the top half in Table 3.6 demonstrates. The estimated values confirmed that the matched pairs with narrow-verb focus were constantly perceived as more natural than that with the broad focus. Besides, the congruous broad focus's low identification accuracy was maintained compared to the matched pairs with contrastive-verb focus. That is, the matched contrastive-verb focus tended to obtain significantly higher scores than broad focus, except for specific tonal contexts in the AV modality (*Tone*: RL, *SE*=0.386, *P*=0.687; *Tone*: RR, *SE*=0.376, *P*=0.16). The bottom half in Table 3.6 lists the comparison results between broad and contrastive-verb focus in congruous speech prosody. In addition, Cantonese adult participants indiscriminately processed the congruous narrow-verb (NV) focus and contrastive-verb (CV) focus without significance. NV and CV conditions obtained comparable ratings on the prosodic naturalness suggested by the estimated values in the post-hoc analyses. Moreover, neither tonal contexts nor modalities had significant effects on distinguishing between narrow focus and contrastive focus on the medial constituent of utterances.

Focus Contrasts	Modality	Tone	*Estimate*	*SE*	*P*. value
Broad Focus (BF) ~ Narrow-Verb Focus (NV)	AO	LL	-1.822	0.319	<.001***
		RL	-1.985	0.395	<.01**
		LR	-1.525	0.319	<.001***
		RR	-1.926	0.378	<.001***
	AV	LL	-2.211	0.331	<.001***
		RL	-1.633	0.369	<.001***
		LR	-1.857	0.317	<.01**
		RR	-2.501	0.416	<.001***

Broad Focus (BF) ~ Contrastive-Verb Focus (CV)	AO	LL	-2.140	0.326	<.001***
		RL	-1.537	0.377	<.05 *
		LR	-1.318	0.317	<.05 *
		RR	-1.653	0.385	<.05 *
	AV	LL	-1.829	0.319	<.001***
		RL	-1.125	0.386	0.687
		LR	-1.601	0.310	<.001***
		RR	-1.358	0.376	0.160

Table 3.6 Post-hoc analyses of focus contrasts between broad focus and narrow-verb/contrastive-verb focus

This study also compared native listeners' performance in congruous prosody influenced by the two-way interaction of *Focus* and *Tone*, considering that *Modality* insignificantly affected the processing of matched pairs. Results showed that the matched broad focus was significantly judged as less natural than the matched narrow-verb (NV) in all tonal contexts (*Tone*: LL, *Estimate*=-2.016, *SE*=0.231, *P*<.001; *Tone*: RL, *Estimate*=-1.809, *SE*=0.271, *P*<.001; *Tone*: LR, *Estimate*=-1.691, *SE*=0.226, *P*<.001; *Tone*: RR, *Estimate*=-2.214, *SE*=0.281, *P*<.001). The difference between matched broad focus and contrastive-verb focus was also significant in all tonal environments (*Tone*: LL, *Estimate*=-1.984, *SE*=0.229, *P*<.001; *Tone*: RL, *Estimate*=-1.331, *SE*=0.269, *P*<.001; *Tone*: LR, *Estimate*=-1.459, *SE*=0.223, *P*<.001; *Tone*: RR, *Estimate*=-1.505, *SE*=0.269, *P*<.001). Also, matched pairs with narrow-verb focus and contrastive-verb focus obtained comparable and insignificant responses suggested by the post-hoc analyses.

In short, Cantonese listeners found it easier to process matched pairs with a narrow or contrastive focus on the verb, compared to those with a broad focus on the entire

utterance. Furthermore, their ability to comprehend the focus was not impacted by the modality of the stimulus inputs or the tone settings of the target utterances.

3.3.2 Focus processing of incongruous prosody in Cantonese by trilingual adults

For incongruous prosody, the mismatched pairs with broad focus were examined in the narrow-verb focus context (NB) and contrastive-verb focus context (CB); the mismatched pairs with contrastive-verb focus were tested in the prior context of narrow-verb focus (NC). Figure 3.6 plots the overview of the perceptual patterns on focus mapping from information structure onto prosodic signals in which acoustic forms in answers were mismatched with the discourse contexts. First, in the left panel, the mismatched pairs with broad focus in narrow-verb focus context (NB) referred to dialogues whose questions were "what did somebody do to something" followed by answers with incongruous prosody of broad focus. The exampled Q-A pair of the NB condition has been illustrated in the mismatched pair (b) in Table 3.4 in Chapter 3.2.2.3. The stacked bar graphs in this panel showed that 25.3% (in AO) and 27.9 % (in AV) of responses were rated as having unnatural prosody (including "unnatural" and "very unnatural" responses) in comprehending incongruous prosodic forms of broad focus. The second condition (the middle panel) of mismatched pairs with broad focus in contrastive-verb focus context (CB) tested whether listeners can distinguish incongruous prosodic broad focus in a preceding context of "contrastive question" like "Did somebody do [*a competitive action verb*] to something?". The mismatched pair (c) in Table 3.4 in Chapter 3.2.2.3 is an example of the CB focus condition. In this condition, visualizations of responses indicated that listeners reached an accuracy of 37.5% (in AO) and 38.8% (in AV), which accurately identified the incongruous broad focus as unnatural prosody. The last comparison (the right panel) examined the perceptual judgements of

mismatched pairs with contrastive-verb focus in narrow-verb focus context (NC), like the mismatched pair (d) in Table 3.4 in Chapter 3.2.2.3. I used mismatched dialogues of which the preceding context was like "what did somebody do to something," eliciting non-contrastive narrow-verb focus, whereas the verb of the answer was with a contrastive focus. Responses of the NC focus condition manifested that listener considered the prosodic contrastive focus and narrow focus on the same constituent in the answers as indistinguishable. As the graph shows, only 15.4% (in AO) and 15.9% (in AV) of responses were identified as having unnatural prosody, whereas 77.1% (in AO) and 67.1% (in AV) of responses indicated that the mismatched contrastive-verb focus naturally answered to the precursor question.

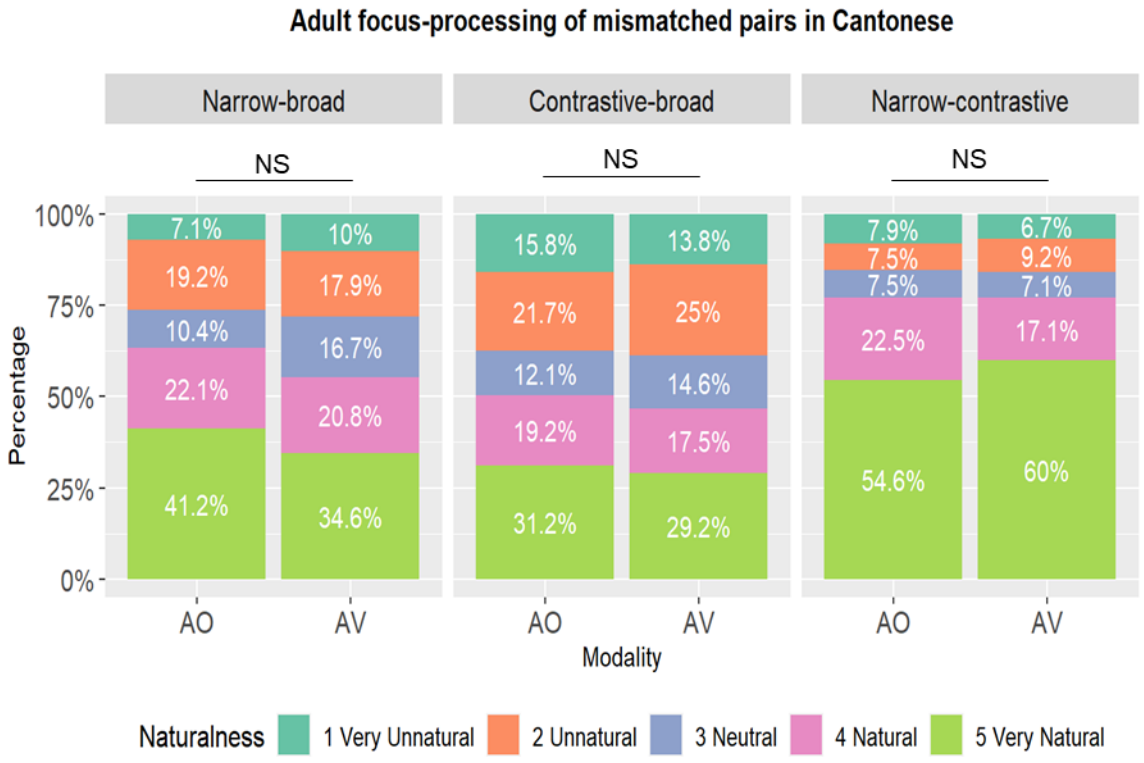


Figure 3.6 Native adults' performance in incongruous Cantonese prosody

The post-hoc analysis of the two-way interaction between *Focus* and *Modalities,* indicated that in the AO modality, the accuracy for the NB condition was significantly

lower than that for the CB condition (*Estimate*=0.654, *SE*=0.173, *P*<0.01). Additionally, the mismatched NC focus conditions were perceived as much more natural than the mismatched NB and CB conditions. The lower accuracy of the NC condition was consistently observed across the three-way and two-way interaction effects of *Focus, Tone,* and *Modality*. The statistical analyses for the NC condition are detailed in Table 3.7 in Appendix 5.

In summary, Cantonese adult listeners employed a variety of perceptual strategies to process three types of mismatched prosodic foci in the current study. Mismatched pairs with a broad focus were found to be more distinguishable from a contrastive-verb focus context (CB) than from a narrow-verb focus context (NB). At the same time, it can be challenging for native listeners to recognize incongruous pairs where the context had a narrow-verb focus that was followed by a contrasting-verb focus. The AO or AV modalities and surrounding tones displayed little influence on the focus processing of incongruous Cantonese prosody.

3.3.3 Comparison between congruous and incongruous focus conditions in Cantonese

In this study, the interpretation of prosodic focus was analyzed by comparing matched (BF, NV, and CV) and mismatched (NB, CB, and NC) pairs of focus conditions. Chapter 3.3.3.1 examined how trilingual adult listeners processed specific focus realizations in two pairs of Q-A dialogues. These pairs shared the same prosodic form but were preceded by different information structures, including NV vs. NB, CV vs. CB, and NV vs. NC focus conditions. In Chapter 3.3.3.2, the study compared trilingual adults' processing of focus between matched and mismatched pairs, which consisted of different information structures in the preceding contexts but were followed by the same prosodic form. The

comparison included BF vs. NB, BF vs. CB, and CV vs. NC focus conditions were included. In this dissertation, the focus in the matched pairs was referred to as the "target", while its contrast in the mismatched pairs was marked as the "competitor" in the following figures.

3.3.3.1 Same information structure-different prosodic forms

Matched pairs with narrow-verb (NV) focus and mismatched pairs with broad focus in the narrow-verb focus context (NB) shared the same precursor questions, which both elicited utterances with a narrow focus on the verb. The following answers in NV and NB focus conditions varied in the prosodic realizations of focus. That is, the answers were produced with a matched narrow-verb focus in the NV condition and a mismatched broad focus in the NB condition. Figure 3.7 describes the perceptual mappings of congruous and incongruous prosody in different tonal contexts. In each tonal context, the left stacked bar indicates responses of congruous prosody (NV, target), while the right one plots the perception in the incongruous prosody (NB, competitor). Responses were grouped by the level-level, rising-level, level-rising, and rising-rising tonal contexts and were horizontally presented in four panels.

In the target narrow-verb (NV) focus condition, trilingual adult listeners correctly identified congruous prosody at rates of 77.8% in LL, 73.9% in RL, 77.7% in LR, and 81.4% in RR contexts including responses of "natural" and "very natural". Mismatched broad focus (NB, competitor) was identified by responses of "unnatural" and "very unnatural," with accuracy rates of 36.4% in RR, 27.7% in LR, 23.6% in LL, and 21.9% in RL. Post hoc analyses of the two-way interaction between *Focus* and *Modality* showed that Cantonese adults rated the prosody in the NV condition as much natural than that in the NB condition (*Modality*: AO, *Estimate*=0.716, *SE*=0.181, *P*<.01; *Modality*: AV, *Estimate*=1.04, *SE*=0.183, *P*<.001). When considering the tonal context, the difference in

evaluations between NV and NB were only significant in the rising-rising tonal context (*Tone*: RR, *Estimate*=1.486, *SE*=0.285, *P*<.001). Results of other tonal contexts are listed in Table 3.8 in Appendix 5.

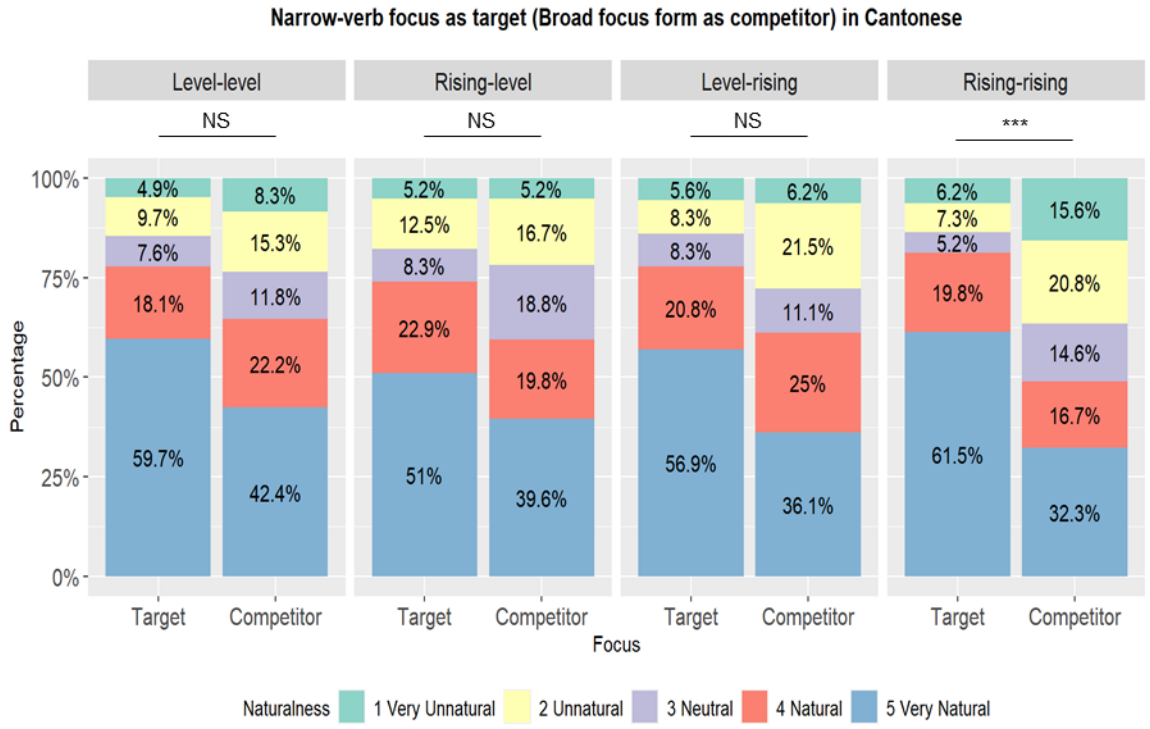


Figure 3.7 Adults' performance in Cantonese focus contrasts with different prosodic forms (Narrow-Broad).

Targets refer to pairs with narrow-verb focus context and matched narrow-verb focus forms; Competitors refer to pairs with narrow-verb focus context and mismatched broad focus forms.

A similar pattern was found between matched pairs with contrastive-verb (CV) focus and mismatched pairs with broad focus (CB). In this contrast, dialogues of the matched contrastive-verb (CV) focus and the mismatched broad (CB) focus shared the same precursor questions which elicited contrastive-verb focus; and answers of the matched CV focus condition had congruous prosody with the contrastive-verb focus, while the CB focus condition contained answers with the mismatched broad focus. Responses

obtained from the target contrastive-verb focus and the competitive broad focus are presented in Figure 3.8.

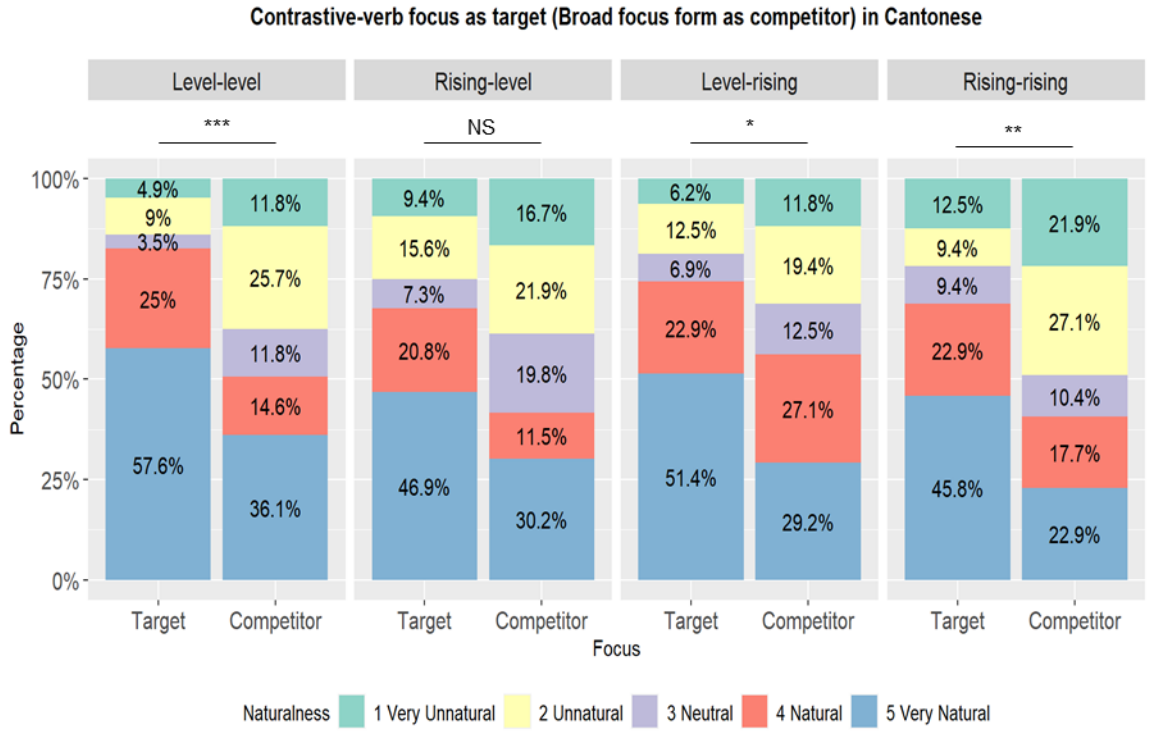


Figure 3.8 Adults' performance in Cantonese focus contrasts with different prosodic forms (Contrastive-Broad).

Targets refer to pairs with contrastive-verb focus context and matched focus forms; Competitors refer to pairs with contrastive-verb focus context and mismatched broad focus forms.

In this comparison, more target contrastive-verb focus (CV) trials were identified as natural prosody than the mismatched broad focus in the CB focus condition. In Figure 3.8, it is apparent that the congruous target prosody was dominantly rated as natural prosody (82.6% in LL, 66.7% in RL, 74.3% in LR, and 68.7% in RR of "natural" and "very natural" responses). Meanwhile, more mismatched pairs in the CB condition were recognized as unnatural prosody compared with the matched pairs in the CV condition. In the mismatched CB condition, moreover, the highest accuracy of prosodic naturalness

identification occurred in the context of rising-rising tones (RR: 49%, RL: 38.6%, LL: 37.5%, and LR: 31.2% of "unnatural" and "very unnatural" responses).

In addition, closer examinations of focus contrasts (CV vs. CB) provided us with statistical evidence on the effects of tonal contexts. Post-hoc results revealed that the prosodic realizations of matched pairs (CV) were processed as more natural prosody than mismatched pairs (CB) with significantly higher evaluations of naturalness degree in most tonal contexts (*Tone*: LL, *Estimate*=1.227, *SE*=0.231, *P*<.001; *Tone*: LR, *Estimate*=0.902, *SE*=0.225, *P*=0.013; *Tone*: RR, *Estimate*=1.222, *SE*=0.274, *P*<.001), except for the rising-level tonal environment (*Tone*: RL, *Estimate*=0.958, *SE*=0.275, *P*=0.081).

At last, the matched pairs with narrow-verb focus (NV) were contrasted with the mismatched pairs with contrastive-verb focus (NC). The preceding discourse was identical for each pair of the two focus conditions (NV vs. NC), eliciting an answer with a narrow focus on the verb. The perceptual patterns for the NV and NC conditions were statistically similar, and there were no effects of tonal contexts on the comparison (See Figure 3.9 in Appendix 5).

In a nutshell, when the information structures in prior discourses were the same, it was possible to distinguish mismatched broad focus from matched narrow-verb and contrastive-verb focus in specific tonal contexts. Previous discourse and tonal context had an impact on the accuracy of comprehending incongruent broad focus. In narrow-verb focus discourse (NV vs. NB), the distinction between incongruous broad focus and congruous narrow-verb focus was significant in the rising-rising tonal context. In the case of contrastive-verb focus discourse (CV vs. CB), differences between targets and competitors were significant in most tonal contexts. Moreover, no substantial difference

was observed in the perceptual results between matched narrow focus and mismatched contrastive focus on the verb when preceded by the same information structure as narrow-verb focus.

3.3.3.2 Different information structures-the same prosodic form

This section compared the perceptual accuracy of prosodic focus in situations when the same prosodic realization of answers was preceded by different preceding contexts, including the focus conditions of BF vs. NB, BF vs. CB, and CV vs. NC. In BF (matched) and NB (mismatched) conditions, each pair of dialogues contained different precursor questions but acoustically identical utterances as answers. The question was "What happened?" in the matched BF condition and "What did somebody do to something?" in the mismatched NB condition. The BF and NB focus conditions had identical answers with broad focus, which matched the former condition but was mismatched with the latter.

Figure 3.10 describes the perceptual patterns of the BF (target) focus condition in the left stacked bars and NB (competitor) focus condition in the right stacked bars in each tonal context. Surprisingly, mismatched broad focus preceded by the narrow-verb focus context was rated as more natural prosody than the matched pairs with broad focus in all tonal contexts. Fewer pairs of the matched broad focus (BF, target) were correctly comprehended as natural prosody (35.5% in LL, 34.3% in RL, 37.5% in LR, and 25% in RR of "natural" and "very natural" degrees), compared with the identification accuracy of the mismatched broad focus in the NB focus condition (64.6% in LL, 59.4% in RL, 61.1% in LR, and 49% in RR of "natural" and "very natural" degrees). The relatively low evaluation of matched broad focus (BF) was consistent with the results of congruous prosody reported in Chapter 3.3.1.

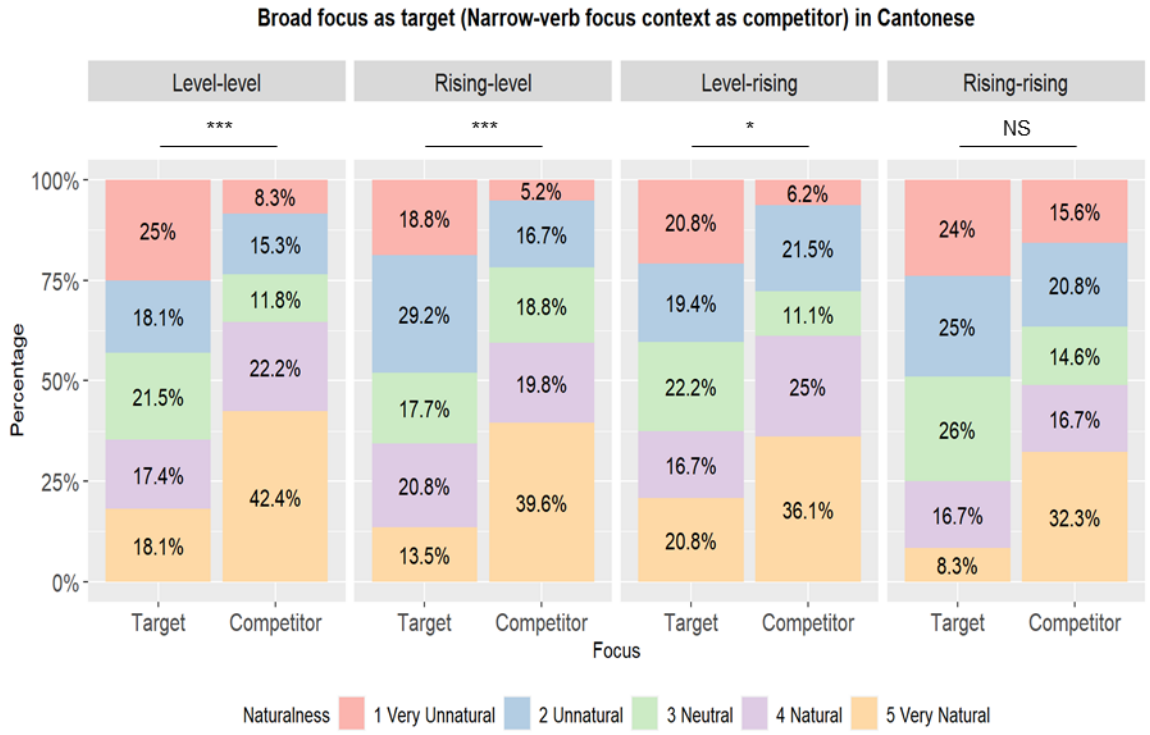


Figure 3.10 Adults' performance in Cantonese focus contrasts with different information structures (Narrow-Broad).

Targets refer to pairs with matched broad focus contexts and forms; Competitors refer to pairs with mismatched narrow-verb focus contexts and broad focus forms.

The estimated values of focus contrasts revealed that the ratings of prosodic naturalness were consistently lower for the matched broad focus (target, BF) than for the mismatched broad focus (NB). Moreover, the differences were significant in most tonal contexts (*Tone*: LL, *Estimate*=-1.303, *SE*=0.224, *P*<.001; *Tone*: RL, *Estimate*=-1.314, *SE*=0.267, *P*<.001; *Tone*: LR, *Estimate*=-0.874, *SE*=0.219, *P*=0.014), except for the tonal environment of continuous rising tones on utterances (*Tone*: RR, *Estimate*=-0.728, *SE*=0.263, *P*=0.463). Since the highest identification accuracy of the NB condition was obtained in the rising-rising tonal context, the rating gap between the competitor (NB, with a higher score) and the NV target (BF, with a lower score) was narrowed and statistically insignificant.

Besides, the pairs of BF (target, matched) and CB (competitor, mismatched) conditions shared acoustically identical response utterances, but the preceding contexts differed in their information structures. For instance, the preceding question, "what happened?" elicited an answer with a matched broad focus in the target BF condition. In the competitive CB condition, the prior context was "Did someone do (*a contrastive action verb*) to something?" which was followed by an utterance with a mismatched broad focus form. Figure 3.11 presents the overview perceptual pattern of congruous (the left stacked bar) or incongruous (the right stacked bar) prosodic broad focus preceded by various information structures in four tonal contexts. Details on distributions of BF and CB conditions have been separately illustrated above and in Chapters 3.3.1 and 3.3.3. In comparing BF and CB focus conditions, listeners performed similar perceptual evaluations of the prosodic naturalness degrees. Notably, they recognized both matched and mismatched broad focus as unnatural prosody in the rising-rising tonal environment with a high percentage (49% for both BF and CB conditions) of "unnatural" and "very unnatural" responses. However, the difference between BF and CB was insignificant in any tonal environment.

Regarding the final contrast CV vs. NC, trilingual adults exhibited similar patterns of focus processing for contrastive-verb focus between matched and mismatched lead-in contexts, as visualized in Figure 3.12 in Appendix 5.

In summary, the information structure in prior discourses demonstrated a complex impact on how native Cantonese adults processed broad focus. The perceptual accuracy for broad focus was found to be the lowest when the prior discourse had a matched information structure with congruous prosody. Interestingly, incongruous broad focus

following a narrow-verb focus context was perceived as having more natural prosody in most tonal environments than congruous broad focus, with a statistically significant difference. In contrast, when preceded by an incongruous contrastive-verb focus context, the perceptual accuracy of incongruous broad focus was not significantly higher compared to that of the congruous broad focus. Moreover, the comprehension of contrastive-verb focus was not affected by the mismatched narrow-verb focus context.

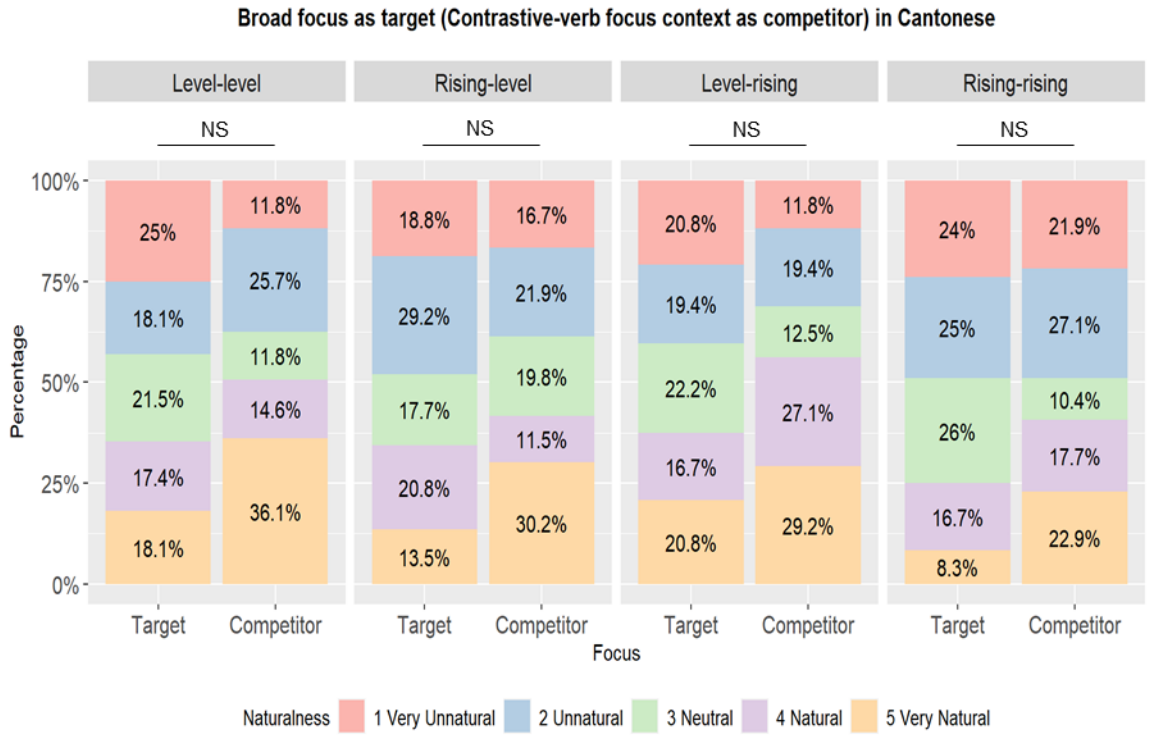


Figure 3.11 Adults' performance in Cantonese focus contrasts with different information structures (Contrastive-Broad).

Targets refer to pairs with matched broad focus contexts and forms; Competitors refer to pairs with mismatched contrastive-verb focus contexts and broad focus forms.

3.4 Discussion

The objective of this experiment was to investigate the ability of native adult Cantonese speakers to map various prosodic forms onto specific focus categories within complex tonal contexts. This question was addressed through five main perspectives: (1) the

perceptual accuracy of congruous prosody in matched focus conditions, including broad focus, narrow-verb focus, and contrastive-verb focus, (2) the mapping perception in incongruous prosody involving these three focus conditions, (3) the comparison of perceptual accuracy of matched and mismatched pairs with the same preceding context following various prosodic forms (4) the comparison of mapping perception of matched and mismatched pairs with various precursor questions following the same prosodic form and (5) the impact of lexical tonal environments and additional visual information on focus processing in Cantonese.

Congruous prosody: The present study investigated the strategies used by native Cantonese-speaking adults to comprehend different focus categories (broad, narrow, and contrastive focus) of congruous prosody in the Q-A dialogue-congruence paradigm. The results indicate that Cantonese adults were able to identify narrow and contrastive focus with greater ease than broad focus. Specifically, the identification accuracy for narrow focus was 77.7%, and 74.45% for contrastive focus, whereas only 33.8% of congruous broad focus was rated as having natural prosody. On the one hand, Cantonese adults could match the acoustic forms of specific focus categories (i.e., the narrow-verb and contrastive-verb focus) to their intended corresponding prior discourses. The present study confirms the role of duration in Cantonese listeners' prosodic processing, as narrow and contrastive focus were acoustically distinguished by duration only, which was indicated by the linear discrimination analysis (LDA) results of stimuli (see Chapter 3.2.2.2). The finding that Cantonese adults were relatively accurate in identifying congruous narrow focus in Q-A dialogues is consistent with previous research on positional focus identification in context-free utterances (Wu & Xu, 2010; Wu, 2013). Furthermore, the present study complements

the identification of contrastive focus, which has not been examined in previous research on Cantonese focus comprehension (Wu & Xu, 2010; Wu, 2013; Yang, 2022). Listeners obtained comparable accuracy between contrastive focus and narrow focus in congruous prosody, which suggests a symmetrical relation with their LDA results in terms of acoustic similarities. The current study's perceptual patterns on prosodic naturalness contribute to a better understanding of how Cantonese-speaking adults use prosodic cues to comprehend different focus categories in natural language dialogues.

On the other hand, listeners' prior knowledge of specific discourse may influence their focus processing, leading them to predict the prosody of incoming utterances based on their previous experience with the information structure (Roettger et al., 2019). Prior knowledge in focus mapping pertains to the pre-assigned probability associated with a specific focus category prior to receiving any specific acoustic input. This knowledge is acquired through the listener's prior experience in focus processing and can be further adjusted across discourse based on the linguistic contexts preceding it (Roettger et al., 2019; Kurumada & Roettger, 2021). The difficulty that Cantonese listeners experienced in interpreting congruous broad focus suggests that they may rely more heavily on their expectations regarding the information structure of broad focus rather than acoustic cues from the input. The acoustic analysis of stimuli in Chapter 3.2.2.2 revealed that broad focus was produced with a significantly shorter duration on the verb compared to narrow and contrastive verb focus. This less salient prosody for broad focus was erroneously perceived as unnatural when answering "What happened?". The insensitivity to broad focus is in line with findings from previous studies on how listeners map broad focus forms onto specific focus categories (e.g., select the appropriate precursor question), both in Cantonese (Yang,

2022) and English (Breen et al., 2010; Roettger et al., 2019). The challenges that listeners encountered in decoding broad focus are likely to be common across languages, irrespective of methodology variations in experimental settings. Specifically, listeners showed a bias (erroneously) against broad focus in two-forced choice (Roettger et al., 2019; Yang, 2022), one-of-seven choice (Breen et al., 2010), and 5-point Likert scale in matching the target broad focus forms to felicitous prior questions.

Incongruous prosody: When interpreting the information structures of narrow or contrastive focus of incongruous prosody of answers, Cantonese adults were generally able to identify the mismatched prosodic forms of broad focus (i.e., NV vs. NB and CV vs. CB), though the rating accuracy of unnatural prosody fell within a range of 25% to 35%. It is consistent with Yang's (2022) finding that Cantonese native listeners can accurately identify narrow-verb focus with an accuracy of more than 80% when the broad focus form serves as a competitor in a two-forced choice task. Both studies indicate that Cantonese listeners are sensitive to at least positional medial focus in the prosodic processing of Q-A dialogues. One possible explanation is that listeners have greater certainty in their probabilistic inference of narrow focus derived from their prior knowledge of the information structure related to narrow focus. Another possibility, from the perspective of acoustic cues, is that experienced listeners can exploit the only available duration cue in decoding a particular focus category, even though Cantonese uses less prosody compared to syntactic cues for focus-marking (Matthews & Yip, 2011).

Nevertheless, trilingual adult listeners often struggle to maintain accurate perceptual mapping when the target is in the form of broad focus and when it follows mismatched prior discourses, such as narrow or contrastive focus contexts. Preceding

discourses may significantly influence focus comprehension more than acoustic signals. As the present study has shown, it is challenging for trilingual adult listeners to discriminate between broad focus forms preceded by matched (broad focus) or mismatched (narrow focus) information structures. They may even mistakenly perceive the congruent broad focus forms as less natural than acoustically identical forms preceded by incongruous interrogative eliciting narrow focus. This challenge of mapping broad focus forms to one of two discourses was also observed in Yang's (2022) study, where listeners' identification accuracy of the broad focus context was below the chance level for most focus categories. Consistent with Yang's findings (2022), Roettger et al. (2019) identified a consistent bias against broad focus independent of task designs, in which English listeners displayed decreased accuracy in their sensitivity to detect broad focus forms and tended to mistakenly assign them to a narrow focus context. Similar to English listeners, the present study found that Cantonese listeners also exhibited a tendency to interpret utterances with prosodic focus as the narrow focus rather than broad focus, indicating a preference for the former over the latter.

The prior knowledge of the stochastic relationship between information structure and surface prosodic signals that listeners have obtained before hearing utterances may explain the cross-linguistic biases against broad focus and towards narrow focus. Roettger and colleagues (2019) proposed an ad-hoc explanation that the process of focus decoding could be conceptualized as a process of probabilistic inference, in which listeners rely on the probability of speakers using particular prosodic patterns to signal the discourse status of a referent. In the current case, the broad focus is elicited by the context "what happened," which provides rarely given information for listeners to pre-activate any related

expectations of discourse functions. Roettger et al. (2019) assumed that the out-of-the-blue scenarios are rarely encountered in actual dialogues without any grounded prior information for the interlocutors, thus, listeners have low expectations for the discourse function of broad focus. In contrast, they proposed that listeners have more prior experience with the likelihood of speakers producing specific prosodic patterns to mark narrow focus, considering that narrow focus on a constituent can be seen as a widely applicable pragmatic function in daily communications.

It is important to note that the explanation for probabilistic inference under uncertainty is still speculative and has been suggested as an ad-hoc explanation for focus mapping in English in the study by Roettger et al. (2019). They proposed that listeners utilize their prior knowledge not only of the category-cue mapping but also the likelihood of a specific speaker producing prosodic forms in given contexts. However, there is a lack of corpus evidence supporting the differential frequency of producing various focus categories in English, let alone generalizing to other languages. Additionally, their current examination of English focus mapping remains deficient in the examination of the role of a listener's prior knowledge of specific speakers in various contexts. Therefore, the present study serves as a further investigation into the application of a probabilistic approach to focus processing in tonal languages.

According to Roettger et al. (2019) and Roettger and Franke (2019), the rational predictive processing of prosodic focus allows listeners to update their beliefs about the probability of an event occurring based on new evidence or information. In focus processing, it informs listeners that the updated (*posterior*) *probability* of an event (i.e., the interpretation of the prosodic cues as specific focus) is proportional to the *prior probability*

of the event and the *likelihood* of the observed evidence (i.e., prosodic cues themselves) given the event. The *prior probabilities* of broad and narrow focus are denoted as *P*(*Broad*) and *P*(*Narrow*), respectively. The likelihood of the evidence given a broad focus is denoted as *P*(*Prosody*|*Broad*), and the likelihood of the evidence given a narrow focus is denoted as *P*(*Prosody*|*Narrow*). Assuming that any prosodic cue (*Prosody*) has an equal *probability* (*P*) of conveying either broad focus or narrow focus, the *likelihood ratio* between them (i.e., *P*(*Prosody*|*Broad*)/ *P*(*Prosody*|*Narrow*)) equals 1. In addition, if listener' prior experience with the probability of narrow focus is higher than that of broad focus (i.e., *P*(*Narrow*) > *P*(*Broad*)), the resulting posterior odds are less than 1, which suggests that listeners are more likely to interpret prosodic cues as indicating a narrow focus rather than a broad focus. Therefore, the findings of the present study appear to support the idea that Cantonese listeners, similar to English listeners, employ probabilistic inference in their native language to interpret prosodic focus, although the presupposition *P*(*Narrow*) > *P*(*Broad*) remains speculative until further corpus studies on Cantonese are conducted.

Furthermore, the present findings highlight the crucial role of expectations in Cantonese listeners' decoding of contrastive prosodic modulation for focus. Specifically, it was found that listeners were able to accurately judge mismatched broad focus forms as unnatural when preceded by a contrastive focus context, indicating their relatively precise expectation of contrastively focused prosodic cues. The mastery of contrastive prosodic distributions was manifested in comparison with listeners' insensitivity to congruous broad focus. In other words, utterances with broad focus were rated as equally natural in prosody when answering precursor questions that elicited broad focus or contrastive focus. However, the study also found that listeners mistakenly rated broad focus forms preceded

by a narrow focus context as more natural than those preceded by a matched information structure. These results suggest an implicit distinction between expectations for narrow focus and contrastive focus in focus processing, despite the fact that their acoustic cues are ambiguous in focus marking. Specifically, narrow focus was elicited by *Wh*-questions targeting a specific syntactic position (i.e., the verb in the medial position in the current study), and listeners anticipated new information from a wide variety of entities. In contrast, contrastive focus was elicited through the correction of previously mentioned information in the precursor question, achieved through the use of interrogative intonation. In this case, listeners were given information about the alternative and thus anticipated the use of prosodic accentuation on the corrected information within the target utterance. Similar to the findings of the present study with Cantonese listeners, Toepel et al. (2007) observed higher accuracy in German listeners' understanding of contextual contrastive focus. Their study found that listeners were more accurate in making prosody-match judgments on German dialogues when the preceding question elicited contrastive focus, as opposed to an interrogative eliciting given information in a prior context, regardless of whether or not there was the actual accentuation of the utterances. These results suggest that listeners' accuracy in understanding contextual contrastive focus may be a across-linguistic features in focus processing. The present study highlights the importance of listeners' expectations in prosodic decoding cues for focus, suggesting that this ability is not solely dependent on explicit prosodic cues but is also shaped by implicit expectations based on the prior context.

Tonal contexts and audiovisual effects: The addition of visual cues did not enhance the prosodic processing beyond the use of acoustic signals alone for Cantonese listeners. The minimal contribution of visual cues observed in my study is in line with previous research

on focus identification in whispered speech in French (Dohen & LAœvenbruck, 2009), where visual information was shown to have a greater impact than in normal speech. It is worth noting that, in normal speech, listeners were able to almost perfectly identify the focal position in the auditory-only condition, indicating a ceiling effect in previous studies (Dohen & LAœvenbruck, 2009; Swerts & Krahmer, 2004). This may have limited the potential for audio-visual improvement. Interestingly, in the present study, Cantonese-speaking adults did not display such a ceiling effect in the auditory-only condition, but also showed only modest improvements in the audio-visual modality. According to the "tonal hypothesis" (Sekiyama et al., 2003), tonal language speakers are influenced more by auditory information than visual information as auditory information is sufficient to identify and discriminate lexical tones. Thus, it is plausible that acoustic cues primarily signal various prosodic features in the present study, leaving little room for additional visual information to contribute. These findings of limited role of visual input are consistent with previous research on Cantonese lexical tone perception, which found no difference between native listeners' perception in audio-visual versus audio-only modalities (Burnham et al., 2001). The strength of the auditory signals in Cantonese is also in line with Burnham et al.'s (2022) study, in which more visual components are required to augment Cantonese tone perception beyond auditory information alone. Moreover, it is possible that participants engaged in a cognitively demanding task may not fully utilize the supplementary visual cues. This is due to the fact that they have to simultaneously process both their prior knowledge of information structures and the auditory cues, which serve a dual function by conveying both lexical and prosodic information.

Furthermore, the current results showed that dynamic lexical tones possibly facilitated the comprehension of focused elements in specific focus categories and positions. This study is the first to design target sentences in a controlled complex tonal context in Cantonese, examining the effects of lexical tones on prosodic focus processing. The results of the study indicate that Cantonese listeners are more sensitive to prosodic naturalness in the tonal context of continuous rising tones (Tone 2 and Tone 5) compared to the tonal contexts of continuous level tones (Tone 1, Tone 3, and Tone 6) and combinations of rising and level lexical tones. This finding is consistent with a previous study on Cantonese prosodic focus (Wu & Xu, 2010), which also reported high accuracy in the identification of focus position in Tone 2 and Tone 5 by Cantonese adults. Moreover, Wu and Xu's (2010) found that when producing focus, Cantonese speakers expanded the pitch range of the focused constituents exclusively in the dynamic tones. Considering the tonal characteristics of Cantonese, it is reasonable that rising contours are more dynamic in nature, allowing speakers to add accentuation on the lexical tones to mark focus, and listeners are, in turn, more accurate in detecting phonetic modulations.

Chapter 4. Experiment2: L3 Mandarin focus processing in trilingual adults

4.1 Introduction

As has been demonstrated in previous experiments investigating the processing of prosodic focus by native Cantonese-speaking adults, an important question arises: do trilingual adult speakers exhibit the same processing abilities for focus in their L3 Mandarin as their native peers? Mandarin, like Cantonese, is a tonal language, but has a different tonal system as described in Chapter 2.3.1. The F_0 cues in Mandarin are used not only to differentiate between lexical tones, but also to indicate focus in prosody. However, Cantonese learners of Mandarin often experience difficulty in identifying specific tonal contrasts, such as Tone 1 vs. Tone 4 and Tone 2 vs. Tone 3 (e.g., Hao, 2012). Therefore, it is important to investigate whether the tonal context can impact the processing abilities of trilingual adults in L3 Mandarin.

There have been limited studies on the comprehension of prosodic focus in non-native tonal languages by Cantonese listeners (Yang, 2022). However, previous research has examined focus production in both Mandarin and Cantonese languages (Wu & Xu, 2010; Wu, 2013; Xu et al., 2012; Chen, 2014; Yang, 2022), which suggests that the differences between these two languages may not be limited to encoding, but may also extend to decoding prosodic focus in their native language. In Mandarin, focus is marked by accentuation of the on-focus constituent, as well as by a compression of pitch range and intensity for the post-focus constituent. This is referred to as post-focus compression (PFC) in Xu (2011). However, there is no consensus on whether Cantonese uses PFC like Mandarin to mark focus. For instance, Wu and Xu (2010) found that Cantonese speakers

mark narrow focus with changes in F_0 excursion size on the focused constituent, as well as with PFC on mean F_0 rather than intensity, limited to specific tones on certain focus positions. Yang (2022) also found that Cantonese adults occasionally mark narrow focus with PFC. Moreover, Xu and colleagues (2012) found that the lack of PFC results in lower accuracy in focus position identification among Taiwanese compared with native Mandarin speakers.

Although studies on focus production have indicated that the feature of focus encoding that is lacking in their L1 is accessible and can be acquired in L2 Mandarin (Chen, 2014; Chen et al., 2012; Chen et al., 2015; Yang, 2022), it is still unclear whether learners of Mandarin can decode prosodic focus like native speakers. This study aims to investigate whether Cantonese-speaking adults have the ability to match prosodic realizations with specific information structures in L3 Mandarin, and whether they possess native-like competence. If acoustic cues are ambiguous for non-native listeners, it remains unclear whether there are any other cues that Cantonese adults can utilize to aid their focus processing. The Mandarin experiment will follow the Cantonese experiment to explore whether Cantonese adults will be influenced by specific tonal contexts of utterances, as well as by additional visual information during the decoding of focus in multimodal speech.

The focus of this chapter is to specifically address the following research questions:

(1) Can trilingual adults comprehend congruous prosodic focus in their L3 Mandarin with the same proficiency as Mandarin-speaking adults?

(2) Can trilingual adults process incongruous prosodic focus in their L3 Mandarin as nativelike as L1 Mandarin speakers did?

(3) Can trilingual adults differentiate between two sets of Mandarin focus conditions (one congruous, one incongruous) as nativelike as L1 Mandarin speakers did?

(4) Do tonal context and additional visual cues influence the focus processing of trilingual adults in L3 Mandarin?

Based on the aforementioned research findings, this study proposes a hypothesis that Cantonese-speaking adults may experience difficulty in comprehending Mandarin focus with a level of accuracy comparable to that of native speakers. Specifically, similar to their perception in Cantonese, these adults may exhibit better performance in congruous focus conditions than in incongruous focus conditions. With respect to tonal context, it is predicted that trilingual adults will show higher accuracy in comprehending focus in tonal contexts with contour tones (Tone 2 and Tone 4) than in contexts with level tones (Tone 1). Furthermore, it is expected that trilingual adults will rely more on visual information than native Mandarin-speaking adults, since the supplementary visual cues may compensate for their non-native ability to utilize auditory cues.

4.2 Method

4.2.1 Participants

Experiment 2 was designed for adult Cantonese learners of L3 Mandarin and examined the Mandarin focus processing by an online perception experiment. The seventy-two native Cantonese speakers (age 21.93 ± 2.61 years) recruited in Experiment 1 (L1 Cantonese) also took part in Experiment 2 as experienced L3 learners of Mandarin. As the details on Cantonese adult participants in Chapter 3.2.1, they learned English as the second language and Mandarin as the third language. The mean age that they started to learn English is 3.85

± 1.37 years and 5.15 ± 1.35 years for Mandarin. *T-Test* results prove that native Cantonese trilingual adults started to learn English at a significantly earlier age than Mandarin (P<.001). Cantonese-speaking participants had an average of 9.49 ± 2.21 years of formal training in Mandarin throughout primary school to secondary school, with an average of 1.63 ± 2.37 hours of formal training per week. They were instructed to self-report the language competence in listening, speaking, reading, and writing L3 Mandarin. Using a five-Likert scale, one point stood for "not fluent at all" and five points indicated "native or near-native". Cantonese adults evaluated their Mandarin proficiency with scores of 3.89 (SD =± 0.74), 3.39 (SD =± 0.85), 4.25 (SD =± 0.78), and 3.81 (SD =± 0.87) points in listening, speaking, reading, and writing respectively.

Moreover, seventy-two adults (35 females and 37 males) of native Mandarin speakers participated in the online experiment as the control group. The mean age of native Mandarin speakers is 23.89 ± 2.89 years. Mandarin-speaking participants were born and brought up in Northern Mainland China and spoke Mandarin as their first language. As bilinguals, they started to learn English at the age of 6.78 ±1.98 years in schools. Participants of native Mandarin speakers had no experience learning Cantonese or other Chinese dialects. 37.5% of Mandarin-speaking participants had musical training on vocals or instruments, with a mean of 4.8 ± 3.22 years of the training period. 22.2% (6 participants) of the musically experienced participants kept training within the past five years. None of the participants reported any history of speech or language difficulties. A written consent has been obtained from all participants before the experiments, and their participation was paid for online.

4.2.2 Materials and design

4.2.2.1 Target sentences

Consistent with the target sentences in Cantonese used in Experiment 1, Mandarin subject-verb-object (SVO) target sentences consisted of five syllables with controlled lexical tones. The one-syllable verb is in the middle position, preceded by the two-syllable subject. The object was composed of the last two syllables. Unlike sentences limited to the same high-level tone (Tone 1) in previous studies (e.g., Xu et al., 2012, Yang, 2022), sentences in the current study varied the tonal contexts between subjects and verb phrases. Specifically, the lexical tone on the first two syllables (the subject) was controlled to be either the same as or different from the tone on the last three syllables (the verb phrase). To exclude tone sandhi effects in Mandarin, target sentences only involved the high-level tone (Tone 1), high-rising tone (Tone 2), and high falling tone (Tone 4), and did not include the low-dipping tone (Tone 3). As Table 4.1 listed, nine tonal combinations were created in the target sentences. I designed two different sentences for each type of tonal combination, and a total of eighteen target sentences in Mandarin were created for Experiment 2. The various tonal contexts of target sentences were labeled by the lexical tone of their focused constituent, the verb, as the rightest column listed in Table 4.1. Thus, Mandarin target sentences involved the level, rising, and falling tonal contexts in which each context contained six target sentences. Table 4.2 lists three tonal contexts when the preceding tone is the high-level Tone 1. In the last column, Chinese characters of example sentences were transcribed into IPA, Pinyin, and English. The entire target sentences were listed and transcribed in Appendix 4.

Type	Tones on Subjects	Tones on Verb Phrases	Tonal Contexts
1	High level Tone 1	High level Tone 1	Level
2		High rising Tone 2	Rising
3		High falling Tone 4	Falling
4	High Rising Tone 2	High level Tone 1	Level
5		High rising Tone 2	Rising
6		High falling Tone 4	Falling
7	High Falling Tone 4	High level Tone 1	Level
8		High rising Tone 2	Rising
9		High falling Tone 4	Falling

Table 4.1 Tonal combinations of Mandarin target sentences

Types	Tones of Subjects	Tones of VPs	Sentences
1	Tone 1 (55) High-Level	Tone 1 (55) High-Level	軍官背書包 /tɕyn55 kwan55 peɪ55 ʂu55 pɑʊ55/ jun55 guan55 bei55 shu55 bao55 The officer carried a schoolbag.
2	Tone 1 (55) High-Level	Tone 2 (35) High-Rising	汪叔提茶壺 /wɑŋ55 ʂu55 tʰi35 ʈʂʰa35 xu35/ Wang55 shu55 ti35 cha35 hu35 Uncle Wang held a teapot.
3	Tone 1 (55) High-Level	Tone 4 (51) High-Falling	周歡買木凳 /ʈʂoʊ55 xwan55 maɪ51 mu51 tɤŋ51/ Zhou55 huan55 mai51 mu51 deng51 Zhouhuan sold a wooden stool.

Table 4.2 Examples of Mandarin sentences in Experiment 2

I also created ten similar five-syllable SVO sentences as fillers. They followed the tonal combinations in which the first two syllables (the subject) carried the same lexical tone, and the last three syllables (the verb phrase) had the same tone. Low-dipping Tone 3

was considered and combined with the other three lexical tones in fillers. Details of filler sentences are listed in Appendix 5.

4.2.2.2 Audiovisual recordings

Following the method in Experiment 1, this study employs the question-answer paradigm to produce various categories of Mandarin prosodic focus for the perceptual stimuli in Experiment 2. The recording session involved seven focus categories and exampled in Table 4.3. In dialogue (1), the precursor question elicits the broad focus in the answering utterance, in which the focused position is the whole sentence. Dialogues (2), (4), and (6) exploit *Wh*-questions to elicit answers with a narrow focus on various syntactic constituents. By questioning SVO, the focal positions are produced in the initial, middle, and final parts, respectively, listed in the rightest column described in Table 4.3. Questions in dialogue (3), (5), and (7) contain misleading information about specific syntactic constituents, which elicit contrastive focus on the subject, verb, or object. The current study concentrates on broad focus and contrastive and narrow focus on the middle position, as dialogues (1) to (4) exampled. Dialogues produced with the other four focus types are used as fillers or practice trials.

Dialogues		**Question-Answer**	**Focus Categories**	**Focus Positions**
1	Question-1	發生了什麼事啊？ What happened?	Broad	Broad
	Answer-1	[周歡賣木凳]$_F$ [ʈʂou55 xwan55 maɪ51 mu51 tɤŋ51]$_F$ [Zhou55 huan55 mai51 mu51 deng51]$_F$ [Zhouhuan sold a wooden stool]$_F$.		
2	Question-2	周歡怎麼木凳啊？ What did Zhouhuan do to the wooden stool?	Narrow	Medial

	Answer-2	周歡[賣]$_{F}$木凳 ʈʂoʊ55 xwan55 [maɪ51]$_{F}$ mu51 tɤŋ51 Zhou55 huan55 [mai51]$_{F}$ mu51 deng51 Zhouhuan [sold]$_{F}$ a wooden stool.		
3	Question-3	周歡做木凳？ Did Zhouhuan make a wooden stool?	Contrastive	Medial
	Answer-3	周歡[賣]$_{Contrastive\text{-}F}$木凳 ʈʂoʊ55 xwan55 [maɪ51] $_{Contrastive\text{-}F}$ mu51 tɤŋ51 Zhou55 huan55 [mai51] $_{Contrastive\text{-}F}$ mu51 deng51 Zhouhuan [sold] $_{Contrastive\text{-}F}$ a wooden stool.		
4	Question-4	誰買木凳啊啊？ Who sold the wooden stool?	Narrow	Initial
	Answer-4	[周歡] $_{F}$ 賣木凳 [ʈʂoʊ55 xwan55]$_{F}$ maɪ51 mu51 tɤŋ51 [Zhou55 huan55] $_{F}$ mai51 mu51 deng51 [Zhouhuan] $_{F}$ sold a wooden stool.		
5	Question-5	王鵬賣木凳？ Did Wangpeng sell a wooden stool?	Contrastive	Initial
	Answer-5	[周歡] $_{Contrastive\text{-}F}$ 賣木凳 [ʈʂoʊ55 xwan55] $_{Contrastive\text{-}F}$ maɪ51 mu51 tɤŋ51 [Zhou55 huan55] $_{Contrastive\text{-}F}$ mai51 mu51 deng51 [Zhouhuan] $_{Contrastive\text{-}F}$ sold a wooden stool.		
6	Question-6	周歡賣什麼啊? What did Zhouhuan sell?	Narrow	Final
	Answer-6	周歡賣[木凳]$_{F}$ ʈʂoʊ55 xwan55 maɪ51 [mu51 tɤŋ51]$_{F}$ Zhou55 huan55 mai51 [mu51 deng51]$_{F}$ Zhouhuan sold a [wooden stool]$_{F}$		
7	Question-7	周歡賣鮮花？ Did Zhouhuan sell fowlers?	Contrastive	Final

	Answer-7	周歡賣[木凳]Contrastive-F ʈʂoʊ55 xwan55 maɪ51 [mu51 tɤŋ51] Contrastive-F Zhou55 huan55 mai51 [mu51 deng51] Contrastive-F Zhouhuan sold a [wooden stool] Contrastive-F		

Table 4.3 Focus categories elicited by Q-A paradigm in Mandarin

Two native Mandarin speakers, one female (23 years) and one male (25 years), were recruited to be video recorded of question-answer dialogues for Mandarin stimuli. They were born and brought up in Beijing, mainland China, and speak Mandarin as their first language. They lived in Hong Kong for less than one year and had no experience learning Cantonese. No participant reported any history of hearing or speaking difficulties. Speakers were asked to familiarize themselves with the dialogues before the recording. Then the male speaker asked questions, and the female speaker answered using target sentences. Instead of displaying the texts of dialogues, all dialogues were depicted by cartoon images and shown for the speakers on the screen by E-prime 3.0. The image with covers on the questioning constituent was first presented to speakers in each Q-A dialogue. Reminded by the covered image, like the Image 4-1 example, the male speaker asked, “Who sold the wooden stool?”. Then the experimenter displayed the image without covered areas like Image 4-2 as hints to the answer. The female speaker spontaneously answered the question using target sentences reminded by each uncovered image. A total of 196 conversations (28 sentences * 7 focus categories) were recorded in random order with three repetitions, and only one repetition would be selected as the stimuli for the perception experiment. In the recording session, videos of the female speakers’ facial expressions were recorded with a Handycam Sony HDR-JP670 at the Speech and

Language Sciences Laboratory of the Hong Kong Polytechnic University. Simultaneously, auditory conversations were recorded by Audacity with a sampling rate of 44.1 kHz.

Image 4-1

Image 4-2

This study further measured several acoustic characteristics of target stimuli, including F_0, F_0 range, duration, and intensity. Extracting the answers from auditory Q-A conversations, I segmented and labeled the voiced parts of each syllable in the five-syllable target utterances and extracted the measurements by ProsodyPro in Praat. Note that the acoustic analyses were conducted on three repetitions of target utterances produced by the female speaker, of which only one repetition was used in the perception experiment. Figure 4.1 plots the F_0 contours of broad focus (red line), narrow-verb focus (green line), and contrastive-verb focus (blue line) of five-syllable SVO Mandarin utterances in spontaneous Q-A conversations. The *X*-axis refers to the time-normalized 20 points of each syllable and continuously presents SVO constituents in three syntactic positions. The *Y*-axis indicates the F_0 (Hz) value of each time-normalized point. The pitch contours prove that the pitch values on verbs under narrow-verb and contrastive-verb focus were dramatically improved compared with the broad-focus contour. The box plots in Figure 4.2 describe the F_0 range of subjects (the left), verbs (the middle), and objects (the right) on broad focus (green plots), narrow-verb focus (orange plots), and contrastive-verb focus (purple plots). The duration

(dB) and intensity (Hz) of subjects, verbs, and objects of target utterances on the three focus conditions are plotted in Figures 4.3 and 4.4.

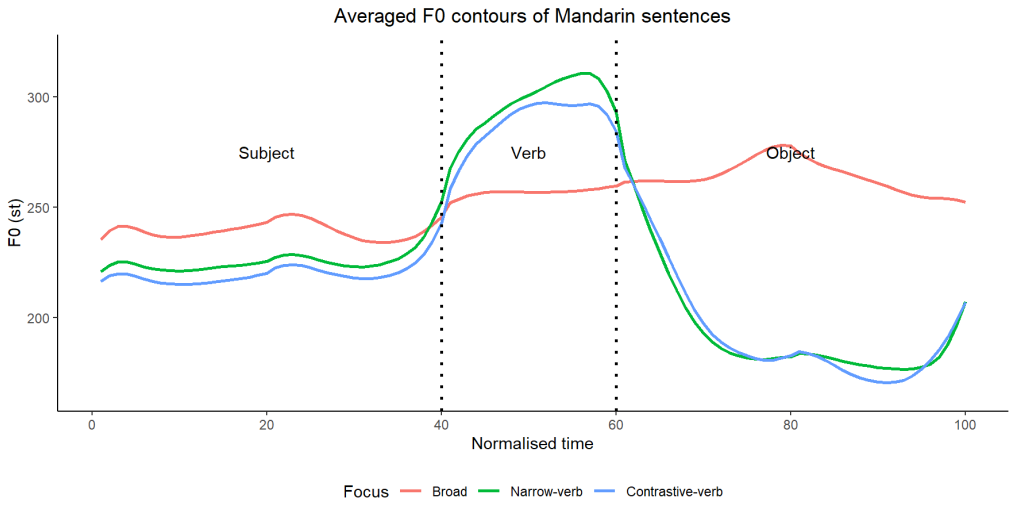


Figure 4.1 F_0 time normalized contours of Mandarin target sentences

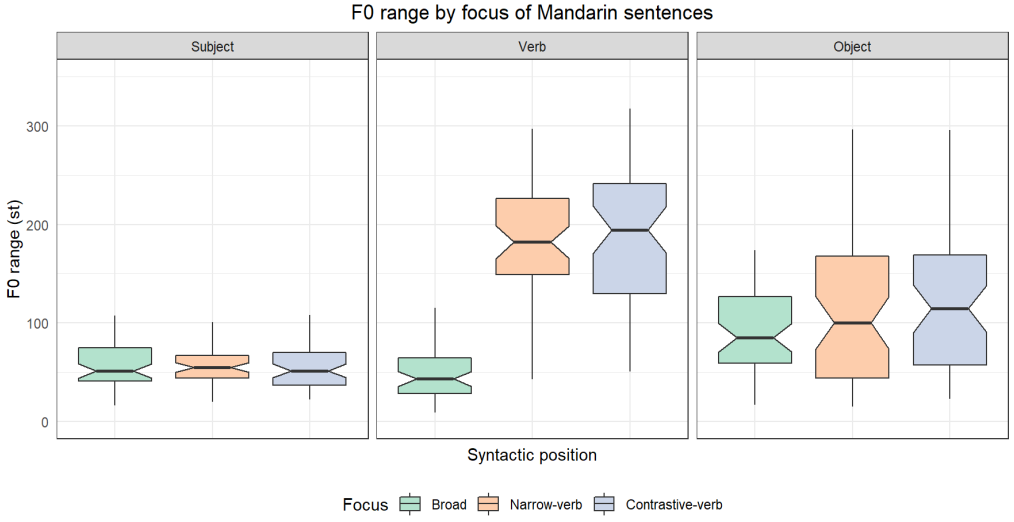


Figure 4.2 F_0 range by the focus of Mandarin target sentences

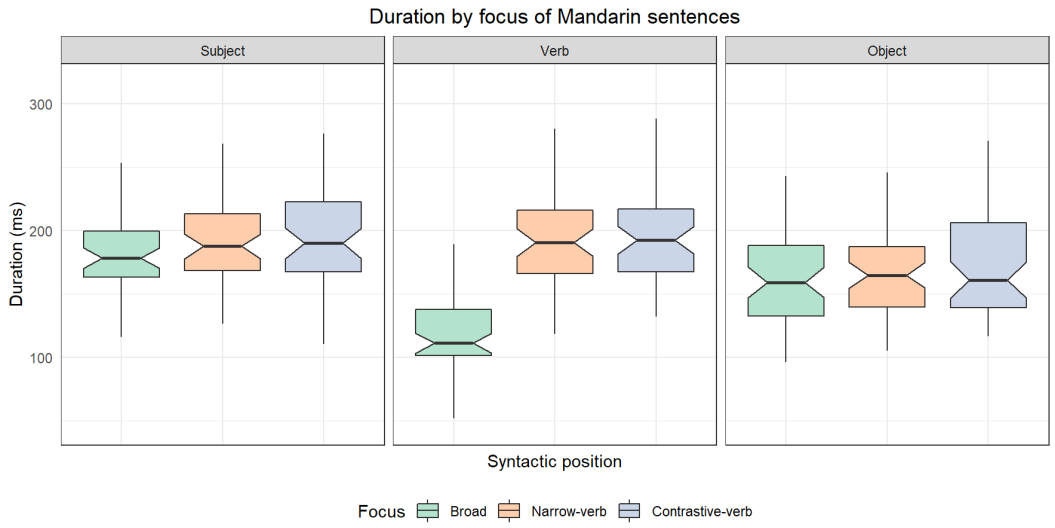


Figure 4.3 Duration by the focus of Mandarin target sentences

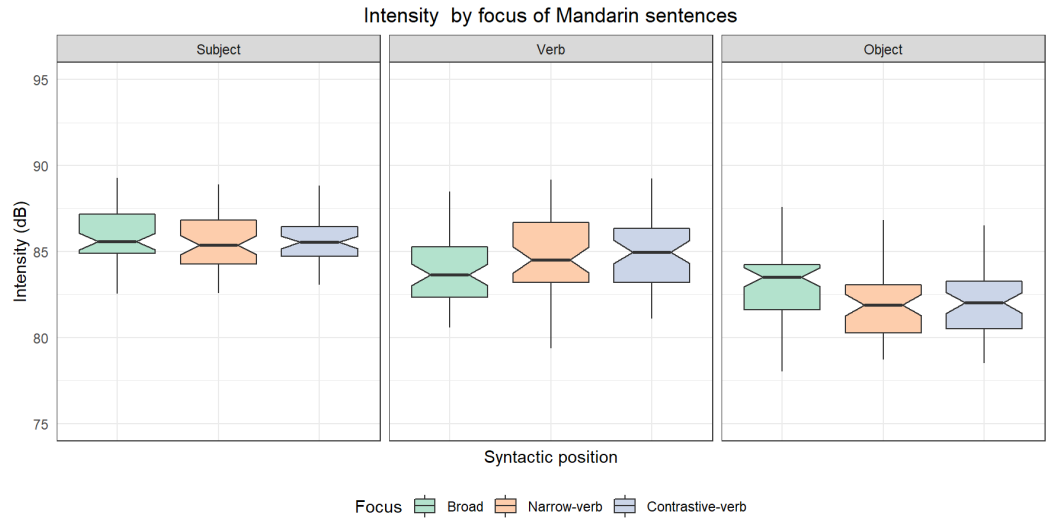


Figure 4.4 Intensity by the focus of Mandarin target sentences

Then, the linear discriminant analyses (LDA) were conducted to the acoustic measurements by *MASS* package in R. For F_0, the accuracy of LDA on F_0 range was 65.43% in the discrimination between broad focus and narrow-verb focus, 67.25% between broad focus and contrastive-verb focus, and 52.16% between and narrow-verb focus and contrastive-verb focus. The LDA of mean F_0 obtained an accuracy of 71.3% and 70.06% in differentiating broad focus from narrow-verb focus and contrastive-verb focus. It was

53.09% of LDA on mean F_0 between the narrow and contrastive verbs. Results on the measurement of duration were 64.51% (broad focus vs. narrow-verb focus) and 64.51% (broad focus vs. contrastive-verb focus), and 50.31% (narrow-verb focus vs. contrastive-verb focus). The LDA results on intensity were 56.48%, 56.17%, and 51.85% on the three contrasts.

4.2.2.3 Stimuli

Like the stimuli design in Experiment 1 (Chapter 3.2.2.3), Experiment 2 included matched and mismatched dialogues in Mandarin by question-answer congruence paradigm, in which the stimuli were presented in audio-only or audio-visual modalities. In the 3×2×2 design of the Mandarin perception experiment, three focus categories (broad focus, narrow-verb focus, and contrastive-verb focus) in matched or mismatched pairs were investigated in two modalities. In matched pairs, the study included broad focus (BF), narrow-verb (NV) focus, and contrastive-verb (CV) focus, as exampled pairs (a-c) in Table 4.4. In mismatched pairs, the question expected to elicit a narrow-verb focus of the following utterance was answered with an inappropriate broad focus in the narrow-verb (NB) focus condition like pair (d) in Table 4.4. The mismatched contrastive-broad (CB) focus condition refers to dialogues whose precursor question should answer with contrastive-verb focus, but the broad focus competes with it in the answer, as the exampled pair (e) shows. Pair (f) exampled the mismatched focus condition (NC) of narrow and contrastive focus on the verb. Following this design of focus conditions, I extracted the individual questions and answers from dialogues and combined them into matched (BF, NV, CV) and mismatched (NB, CB, NC) focus conditions. Details of the stimuli organization are illustrated in Experiment 1 in Chapter 3.2.2.3.

a. Matched Pair with BF focus condition		**Focus category**
Question	發生了什麼事啊？ Fa sheng le shen me shi a? What happened?	Broad Focus
Answer	[周歡賣木凳]$_{F}$ [ʈʂoʊ1 xwan1 maɪ4 mu4 tɤŋ4]$_{F}$ [Zhou1 huan1 mai4 mu4 deng4]$_{F}$ [Zhouhuan sold a wooden stool]$_{F}$.	Broad Focus
b. Matched Pair with NV focus condition		
Question	周歡怎麼木凳啊？ Zhou huan zen me mu deng a? What did Zhouhuan do to the wooden stool?	Narrow Focus
Answer	周歡[賣]$_{F}$ 木凳 ʈʂoʊ1 xwan1 [maɪ4]$_{F}$ mu4 tɤŋ4 Zhou1 huan1 [mai4]$_{F}$ mu4 deng4 Zhouhuan [sold]$_{F}$ a wooden stool.	Narrow Focus
c. Matched Pair with CV focus condition		
Question	周歡做木凳？ Zhou huan zuo mu deng? Did Zhouhuan make a wooden stool?	Contrastive Focus
Answer	周歡[賣]$_{\text{Contrastive-F}}$ 木凳 ʈʂoʊ1 xwan1 [maɪ4] $_{\text{Contrastive-F}}$ mu4 tɤŋ4 Zhou1 huan1 [mai4] $_{\text{Contrastive-F}}$ mu4 deng4 Zhouhuan [sold] $_{\text{Contrastive-F}}$ a wooden stool.	Contrastive Focus
d. Mismatched Pair with NB focus condition		

Question	周歡怎麼木凳啊？ Zhou huan zen me mu deng a? What did Zhouhuan do to the wooden stool?	Narrow Focus
Answer	[周歡賣木凳]$_{F}$ [ʈʂoʊ1 xwan1 maɪ4 mu4 tɤŋ4]$_{F}$ [Zhou1 huan1 mai4 mu4 deng4]$_{F}$ [Zhouhuan sold a wooden stool]$_{F}$.	Broad Focus
e. Mismatched Pair with CB focus condition		
Question	周歡做木凳？ Zhou huan zuo mu deng? Did Zhouhuan make a wooden stool?	Contrastive Focus
Answer	[周歡賣木凳]$_{F}$ [ʈʂoʊ1 xwan1 maɪ4 mu4 tɤŋ4]$_{F}$ [Zhou1 huan1 mai4 mu4 deng4]$_{F}$ [Zhouhuan sold a wooden stool]$_{F}$.	Broad Focus
f. Mismatched Pair with NC focus condition		
Question	周歡怎麼木凳啊？ Zhou huan zen me mu deng a? What did Zhouhuan do to the wooden stool?	Narrow Focus
Answer	周歡[賣]$_{\text{Contrastive-F}}$木凳 ʈʂoʊ1 xwan1 [maɪ4] $_{\text{Contrastive-F}}$ mu4 tɤŋ4 Zhou1 huan1 [mai4] $_{\text{Contrastive-F}}$ mu4 deng4 Zhouhuan [sold] $_{\text{Contrastive-F}}$ a wooden stool.	Contrastive Focus

Table 4.4 Examples of matched pairs and mismatched pairs in Experiment 2

The present experiment included 168 trials of dialogues (18 target sentences * 6 focus categories + 10 fillers * 6 focus categories) which were evenly blocked into three sets. Each set contained two focus conditions for every target utterance, in which one is matched, and the other was mismatched. Participants were randomly assigned to one of the three sets; thus, each target utterance of felicitous and infelicitous prosody was displayed only once to individual participants. Therefore, 56 trials of Q-A conversations were randomly displayed for each participant. Half trials had answers in the audio-only (AO) modality and half in the audio-video (AV) modality.

4.2.3 Procedures

Following the procedures in Experiment 1 (Chapter 3.2.3), native Cantonese and native Mandarin listeners started with practice sessions and then the experimental session in the present experiment of Mandarin perceptual mappings. The order of the Cantonese experiment (Experiment 1) or the Mandarin experiment (Experiment 2) was counterbalanced among trilingual Cantonese listeners to avoid any task effect.

4.2.4 Data analyses

All responses obtained from native Cantonese listeners and native Mandarin listeners were fitted to the ordinal regression model using the *ordinal* package (Christensen, 2015) in R (R Core Team, 2022). The dependent variable (Response) examined in the current study was the 5-Likert scale of prosodic naturalness. I submitted the ordinal data to the cumulative link mixed effects models (CLMM) by the *clmm* function. The fixed effects included *Group* (Cantonese, Mandarin), *Focus* (BF, NV, CV, BN, BC, NC), *Tones* (TL, TR, TF), and *Modalities* (AO, AV). Factors of *Subjects* and *Trials* were examined as

random effects. Post hoc analyses by the *emmeans* package (Lenth R, 2023). in R were conducted to analyze responses to various focus contrasts in different tonal environments.

4.3 Results

I first conducted the likelihood ratio test on the interaction of *Group* and *Modality*, and native Cantonese and native Mandarin adult listeners showed insignificant differences (*Group* * *Modality*: $\chi^2(3) = 1.41$, $P = 0.703$) in perceptual responses between audio-only (AO) and audio-video (AV) modalities. This study finally fitted the optimal regression model for the data involving the interaction of *Group*, *Focus,* and *Tone* as fixed effects and the factor *Subject* and *Trial* as random effects. Results showed that the factor *Group* was insignificant ($\chi^2(1) = 0.06$, $P = 0.8$) in the perceptual mappings, but *Focus* ($\chi^2(5) = 687.35$, $P<.001$) and *Tone* ($\chi^2(2) = 17.53$, $P<.001$) significantly affected listeners' responses. The interaction of *Group* and *Focus* ($\chi^2(5) = 288.81$, $P<.001$) and interaction among *Group*, *Focus, and Tone* ($\chi^2(10) =136.62$, $P<.001$) also showed significance in the Mandarin focus processing. The perception of Cantonese and Mandarin native listeners were compared from three perspectives in the following subchapters. The overall identification accuracy of congruous and incongruous speech prosody in Mandarin dialogues is respectively presented in Chapters 4.3.1 and 4.3.2. Following that, Chapter 4.4.3 reports the comparison between matched and mismatched pairs, either with the same information structure followed by various prosodic realizations, or with different information structures preceding the specific prosodic form.

4.3.1 Focus processing of congruous prosody in L3 Mandarin by trilingual adults

For congruous prosody, this study examined native and non-native Mandarin listeners' comprehension of broad focus (BF), narrow-verb (NV) focus, and contrastive-verb (CV)

focus in Q-A dialogues. Figure 4.5 shows response distributions of the three types of matched focus conditions evaluated with five levels of prosodic naturalness. In each focus type, the perceptual patterns of non-native listeners are presented on the left and those of native listeners on the right. Consistent with Experiment 1 in Chapter 3, "natural" and "very natural" responses were counted as identification of congruous prosody, "unnatural" and "very unnatural" as identification of incongruous prosody, and "neutral" as unsure responses throughout the study.

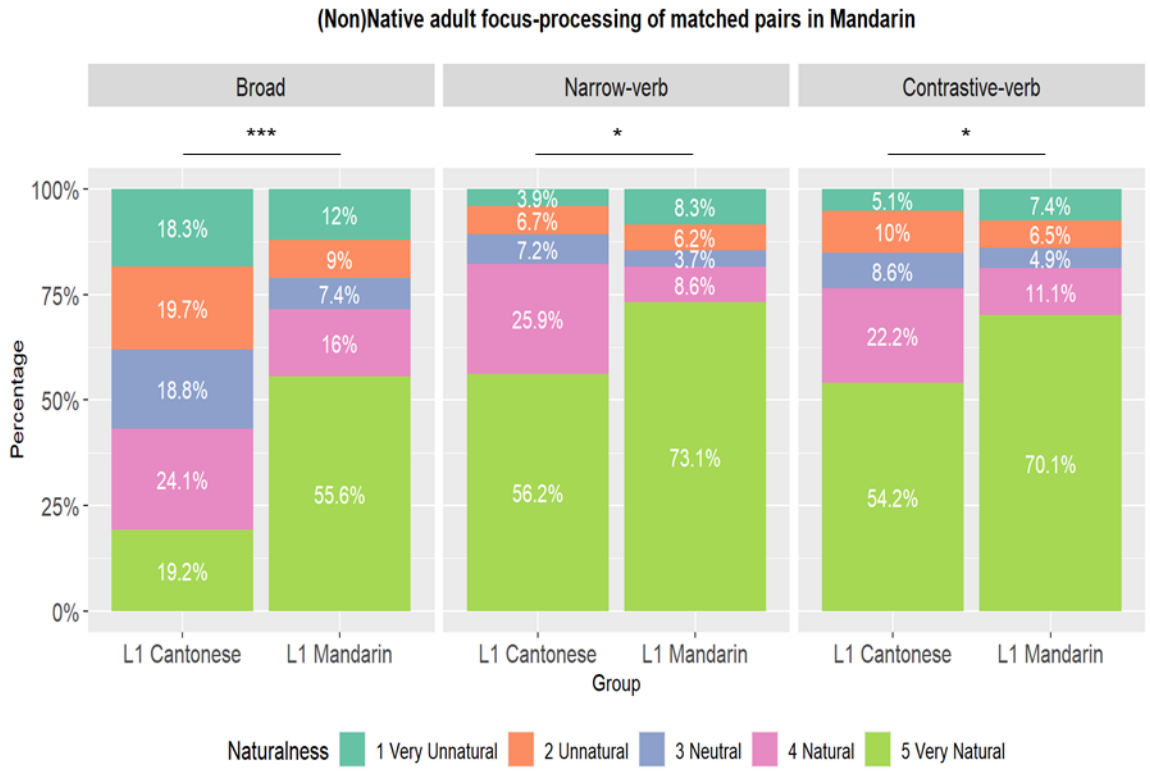


Figure 4.5 Adults' performance in congruous Mandarin prosody

First, native Cantonese and native Mandarin adult listeners demonstrated different perceptual patterns in processing the matched broad focus. Non-native listeners correctly identified only 43.3% of natural broad focus while native listeners had a correct identification rate of 71.6% ("natural" and "very natural" responses). Statistical results

showed that Cantonese listeners judged the broad focus as less natural prosody than native listeners (*Estimate*=-1.251, *SE*=0.175, *P*<.001). Considering the effects of tonal contexts, post-hoc analyses revealed differences in identifying the matched broad focus between groups was significant in the level tonal context (*Estimate*=-1.758, *SE*=0.277, *P*<.001).

Then, in the other two matched focus conditions (NV and CV), non-native listeners correctly identified 82.1% of pairs with narrow-verb focus and 76.4% of pairs with contrastive-verb focus as having natural prosody. The proportion of pairs with the two focus conditions for native listeners was 81.7% and 81.2%, respectively. Post hoc analyses indicated that non-native listeners rated the narrow-verb focus as less natural prosody than Mandarin listeners with a significant difference (*Estimate*=-0.711, *SE*=0.208, *P*<.001), and only the rising tonal context demonstrated significance (*Estimate*=-1.522, *SE*=0.314, *P*<.001). Similar to this, non-native participants evaluated matched dialogues under contrastive-verb focus with slightly lower naturalness ratings compared to native participants (*Estimate*=-0.658, *SE*=0.198, *P*=0.043), and a significance was found in the context of falling tones (*Estimate*=-1.419, *SE*=0.289, *P*<.001).

Comparing the matched focus conditions, post-hoc results proved that both groups significantly differed in distinguishing the broad focus from the narrow-verb and contrastive-verb focus. Specifically, Cantonese listeners evaluated the broad focus with lower degrees of naturalness than narrow-verb focus (*Estimate*=-1.623, *SE*=0.129, *P*<.001), which was significant in all tonal contexts (*Tone* level: *Estimate*=-2.029, *SE*=0.301, *P*<.001; *Tone* rising: *Estimate*=-1.173, *SE*=0.283, *P*=0.016; *Tone* falling: *Estimate*=-1.667, *SE*=0.3, *P*<.001). Similarly, Mandarin listeners discriminated the matched broad focus from the matched narrow-verb focus (*Estimate*=-1.084, *SE*=0.175, *P*<.001), and the differences

were significant in the rising and falling tonal contexts (*Tone* Rising: *Estimate*=-1.697, *SE*=0.314, *P*<.001; *Tone* Rising: *Estimate*=-2.082, *SE*=0.393, *P*<.001).

Besides, non-native listeners of Mandarin distinguished the broad focus (BF) from the contrastive-verb focus (CV) with lower evaluations of naturalness (*Estimate*=-1.482, *SE*=0.128, *P*<.001). The differences showed significance in the level and rising tonal environments (*Tone* Level: *Estimate*=-1.495, *SE*=0.291, *P*<.001; *Tone* Rising: *Estimate*=-1.584, *SE*=0.3, *P*<.001). Similarly, native Mandarin listeners discriminated the broad focus from the contrastive-verb focus with lower ratings of prosodic naturalness (*Estimate*=-0.89, *SE*=0.164, *P*<.001). Broad focus was shown to be considerably less natural than contrastive-verb focus in the context of falling tones (*Tone* Falling: *Estimate*=-1.551, *SE*=0.275, *P*<.001). Moreover, both groups equally processed the congruous narrow and contrastive focus on the verb regardless of tonal contexts.

Therefore, non-native listeners showed differences in processing the matched focus conditions from native Mandarin listeners, as their identification accuracy was significantly lower than that of native listeners. At the same time, both groups evaluated matched pairs with broad focus as less natural prosody than the congruous prosody under the narrow-verb or contrastive-verb focus.

4.3.2 Focus processing of incongruous prosody in L3 Mandarin by trilingual adults

The mismatched broad focus in the NB and CB conditions and mismatched contrastive-verb focus in the NC condition were examined in the focus processing of incongruous prosody. Examples of each focus condition are illustrated in Table 4.4 in Chapter 4.2.2.3. Distributions of perceptual judgments for the mismatched focus conditions are presented in Figure 4.6. Responses obtained from native Cantonese listeners (the left) and native

Mandarin listeners (the right) are grouped by different focus conditions. The degree of prosodic naturalness decreases from the bottom to the top in each stacked bar chart. Responses of "unnatural" and "very unnatural" degrees are regarded as successful identification of the incongruous prosody throughout the dissertation.

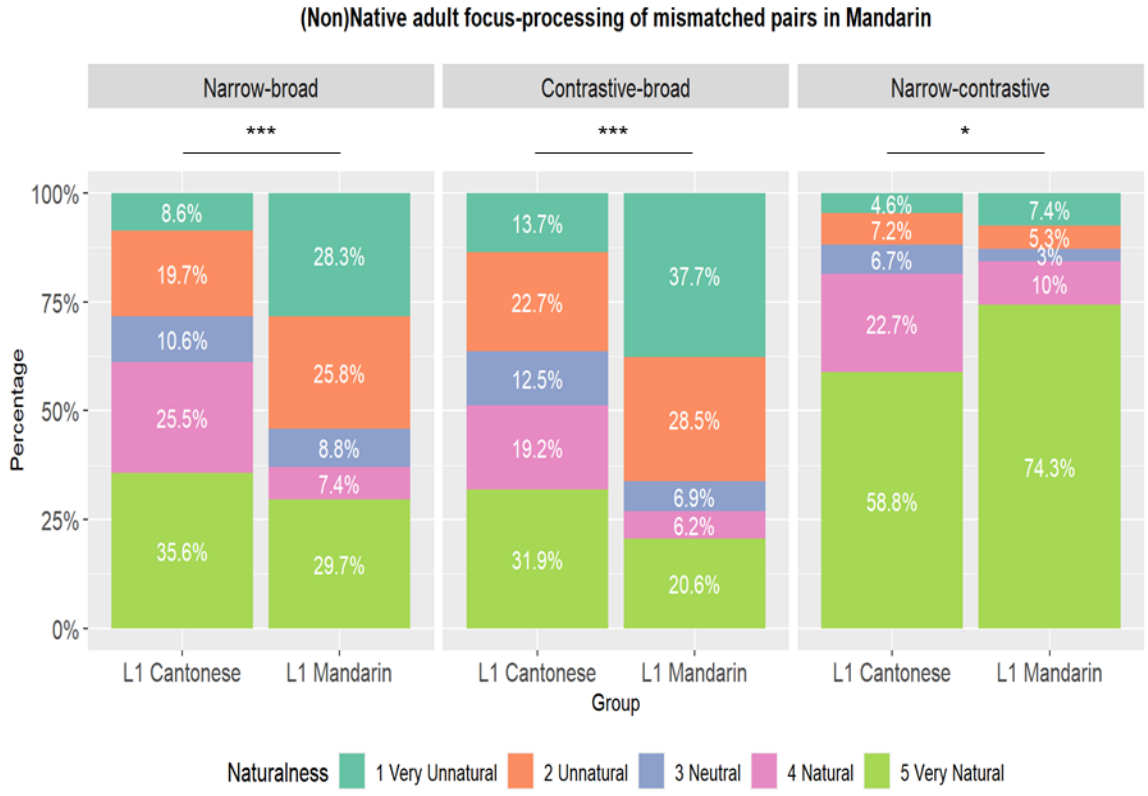


Figure 4.6 Adults' performance in incongruous Mandarin prosody

In the mismatched NB focus condition, the identification accuracy of the incongruous prosody was 28.3% of Cantonese listeners and 54.1% of Mandarin listeners. Statistic results supported that non-native listeners processed the incongruous broad focus with significantly lower accuracy than native listeners (*Estimate*=1.09, *SE*=0.179, *P*<.001). The difference was significant in the rising and falling tonal contexts (*Tone* Rising: *Estimate*=1.192, *SE*=0.265, *P*<.01; *Tone* Falling: *Estimate*=1.686, *SE*=0.331, *P*<.001).

In the mismatched CB condition, 36.4% of dialogues with incongruous broad focus were rated as having unnatural prosody by Cantonese listeners, while 66.2% of responses were rated similarly by Mandarin listeners. The significant differences in their performance were evident in the post-hoc analysis (*Estimate*=1.337, *SE*=0.179, *P*<.001), and the variations between groups were specifically found in the level and falling tonal environments (*Tone* Level: *Estimate*=2.148, *SE*=0.328, *P*<.001; *Tone* Falling: *Estimate*=1.332, *SE*=0.262, *P*<.001).

In the mismatched contrastive-verb focus (NC) condition, non-native listeners identified 11.8% of dialogues as having unnatural prosody, while native listeners had an accuracy rate of 12.7% in identifying such dialogues. The stacked bar graphs clearly showed that more mismatched pairs in the NC focus condition were wrongly rated as having natural prosody by native listeners compared to non-native listeners. Two groups of listeners showed significantly different performances, with Cantonese listeners having a higher accuracy in identifying incongruous prosody than Mandarin listeners (*Estimate*=-0.705, *SE*=0.198, *P*=0.019). A closer examination of the effects of tonal contexts demonstrated that better performance of non-native listeners was significant in the level tonal environment (*Tone* Level: *Estimate*=-1.331, *SE*=0.325, *P*=0.019).

Apart from the comparison results of individual focus conditions between two groups of listeners, this subchapter also compared the identification accuracy of three types of mismatched pairs within each group. First, native listeners processed the mismatched broad focus in the NB as more natural than that of the CB focus conditions (*Estimate*=0.553, *SE*=1.478, *P*=0.01), while non-native listeners followed a similar preference but without significance. Then, both groups rated the prosody of mismatched contrastive-verb focus in

the NC condition as more natural than the mismatched broad focus in NB or CB focus conditions. Statistical results are presented in Table 3.9 in Appendix 6.

In summary, the processing of mismatched pairs with broad focus in both narrow-verb focus (NB) and contrastive-verb focus (CB) contexts differed between native and non-native Mandarin listeners. Native listeners outperformed non-native listeners in accurately identifying the incongruous broad focus, with the level of significance varying depending on the tonal environment. Additionally, both groups of listeners predominantly perceived the mismatched pairs with contrastively emphasized verbs as having prosody that was congruent with the preceding discourse's narrow-verb focus.

4.3.3 Comparison between congruous and incongruous focus conditions in L3 Mandarin by trilingual adults

The present study analyzed the perceptual results of focus processing between matched and mismatched pairs, as introduced in Experiment 1 (Chapter 3.3.3). These pairs were either comprised of the same information structure followed by different prosodic forms (NV vs. NB, CV vs. CB, and NV vs. NC focus conditions), or different information structures preceding the same prosodic realization (BF vs. NB, BF vs. CB, and CV vs. NC focus conditions). To clarify, this dissertation employs the term "target" to refer to the appropriate focus in matched pairs, and the term "competitor" to describe the incongruous focus in mismatched pairs.

4.3.3.1 Same information structure-different prosodic forms

This subchapter examined contrasts of focus conditions in which the different prosodic forms of answers were preceded by questions containing the same information structure in Q-A dialogues. Table 4.4 in Chapter 4.2.2.3 has exampled all types of focus conditions

examined in the current experiment. The first comparison was the matched narrow-verb (NV) focus condition and the mismatched broad focus in a narrow-verb focus context (NB). Precursor queries in the NV and NB conditions were expected to produce responses with a narrow-verb focus, but in the mismatched NB focus condition, the answer was broadly focused on the entire utterance. Congruous narrow-verb focus was the target in the answers in the matched NV focus condition. Figure 4.7 displays perceptual patterns of target conditions (left) and competitive prosodic forms of broad focus (right). Three tonal contexts (Level, Rising, and Falling) classify the responses for native Cantonese listeners (top) and native Mandarin listeners (bottom). Each stacked bar represents the proportions of five degrees of prosodic naturalness, decreasing from the bottom (blue color) to the top (green color).

The response distributions of the NV condition (target) have been reported in Chapter 4.3.1 and the NB condition (competitor) in Chapter 4.3.2 without details on each tonal context. In targets, the identification accuracy of the matched narrow-verb focus was 84.9%, 82%, and 79.4% of Cantonese listeners in the level, rising, and falling tonal environments. Mandarin listeners identified the matched narrow focus on verbs with an accuracy of 65.8%, 93.4%, and 96% in the three tonal contexts. In competitors, Cantonese listeners recognized the mismatched broad focus with the percentage of 20.6%, 25.5%, and 39% “unnatural” and “very unnatural” responses in the three tonal situations, respectively. Mandarin listeners obtained higher accuracy of 38.6%, 62.3%, and 74.7% in the level, rising, and falling tonal contexts in detecting the mismatched broad focus as unnatural prosody.

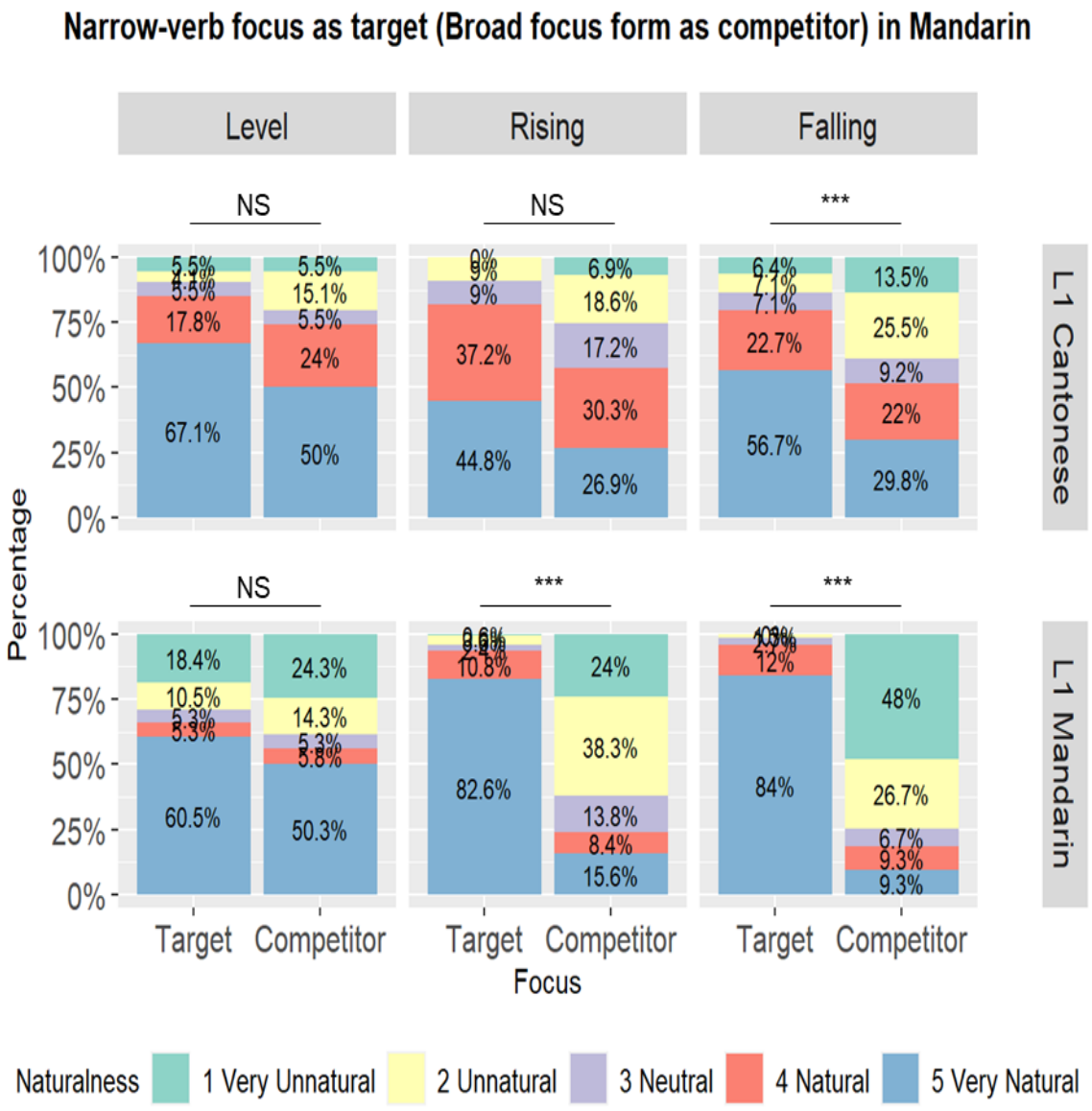


Figure 4.7 Adults' performance in Mandarin focus contrasts with different prosodic forms (Narrow-Broad).

Targets refer to pairs with narrow-verb focus context and matched focus forms; Competitors refer to pairs with narrow-verb focus context and mismatched broad focus forms.

Post-hoc analyses supported the comparison results between the matched narrow-verb focus (NV) and mismatched broad focus (NB) within groups. Non-native listeners significantly distinguished the mismatched broad focus with lower rating scores than the matched narrow-verb focus (*Estimate*=0.918, *SE*=0.132, *P*<.001), which was only

significant in the falling tonal context (*Tone* Falling: *Estimate*=1.25, *SE*=0.23, *P*<.001). For native listeners, dialogues with an incongruous broad focus in the NB focus condition were less natural than the prosody of congruous narrow-verb focus (NV) with statistical significance (*Estimate*=2.719, *SE*=0.182, *P*<.001). Specifically, they significantly differed the matched NV pairs from the mismatched NV pairs in the rising and falling tonal contexts (*Tone* Rising: *Estimate*=3.499, *SE*=0.261, *P*<.001; *Tone* Falling: *Estimate*=4.317, *SE*=0.416, *P*<.001).

The second comparison was the matched pairs with a contrastive-verb focus in the CV condition versus the mismatched pairs with a broad focus in the CB condition. The questions in CV and CB focus conditions were similarly designed to elicit responses with a contrastive emphasis on the verb. In contrast, answers had congruous prosody in the CV (target) focus condition but incongruous broad focus in the CB (competitor) focus condition. Figure 4.8 displays the perceptual distributions of these two focus conditions in each tonal context. The top shows the patterns of Cantonese adult listeners, and the bottom of Mandarin adult listeners. Different areas of each color in stacked bars represent the proportions of one degree in the prosodic naturalness.

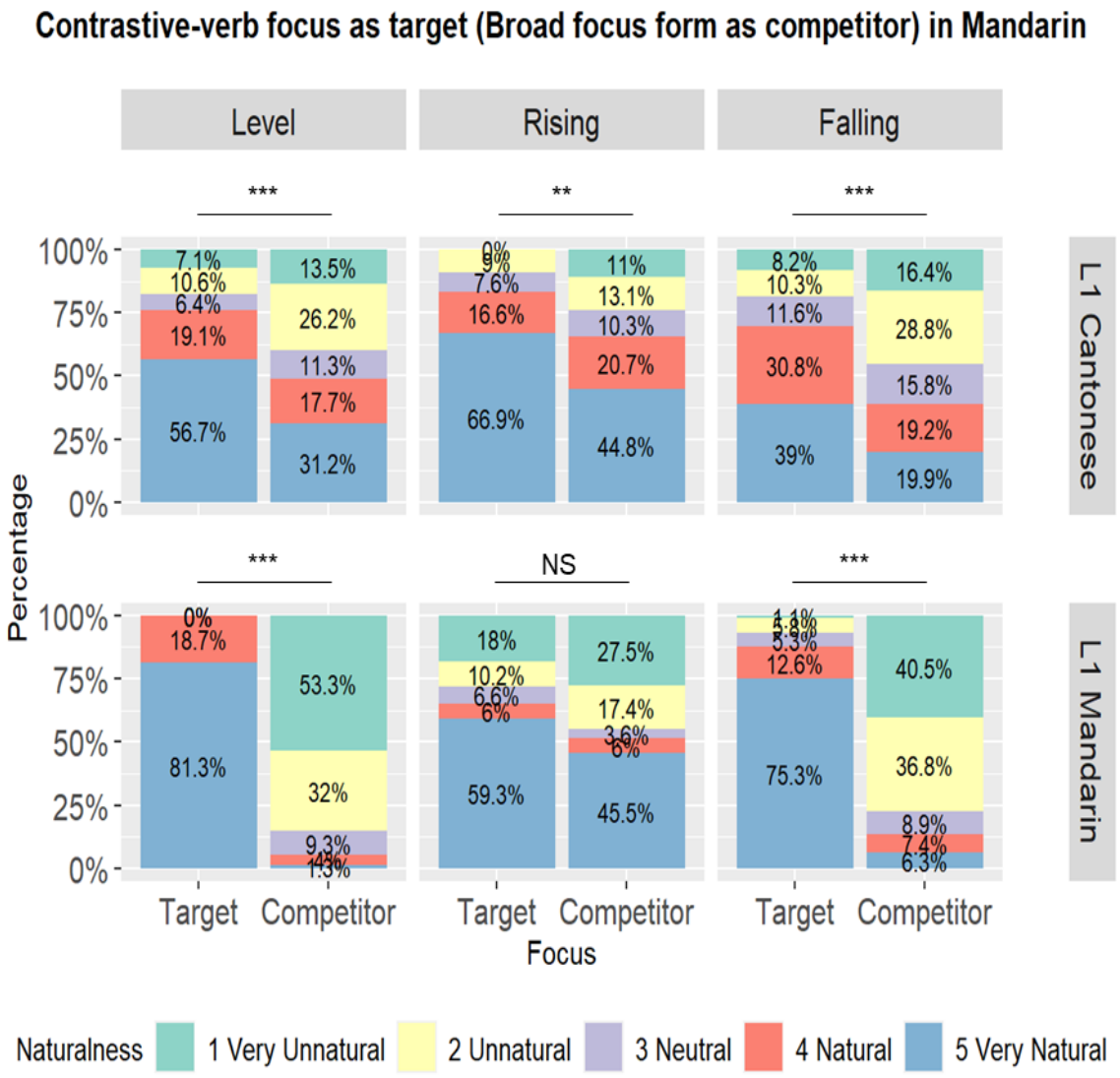


Figure 4.8 Adults' performance in Mandarin focus contrasts with different prosodic forms (Contrastive-Broad).

Targets refer to pairs with contrastive-verb focus context and matched focus forms; Competitors refer to pairs with contrastive-verb focus context and mismatched broad focus forms.

In the target CV condition, Cantonese adult listeners processed the matched contrastive-verb focus with high percentages of natural responses in various tonal contexts (75.8%, 83.5%, and 69.8% in the level, rising, and falling tonal contexts). The accuracy of Mandarin adult listeners was 100% in the level, 65.3% in the rising, and 87.9% in the falling tonal environments. For the competitor CB condition, Cantonese listeners

comprehended the mismatched broad focus in the competitor condition with ratios of 39.7%, 24.1%, and 45.2% of unnatural prosody in the level, rising, and falling tonal contexts. In contrast, native Mandarin participants identified the unnatural prosody with an accuracy of 85.3%, 44.9%, and 77.3% in the three tonal environments.

Listeners of the two groups significantly differed in processing the matched contrastive-verb focus in the CV focus condition, as illustrated in Chapter 4.3.1. Similarly, the comprehension of mismatched broad focus in the contrastive-verb focus context (CB) showed significant differences between non-native and native listeners, as reported in Chapter 4.3.2. Thus, the current subchapter reported the comparison between CV and CB conditions within each group of listeners. Post-hoc analyses showed that both Cantonese and Mandarin adult listeners significantly discriminated the matched contrastive-verb focus in the CV focus condition from the mismatched broad focus in the CB focus condition (*Group:* Cantonese, Estimate=1.083, *SE*=0.132, *P*<.001; *Group:* Mandarin, *Estimate*=3.078, *SE*=0.173, *P*<.001). Taking the tonal contexts into consideration, Cantonese listeners demonstrated significance in differing the congruous and incongruous prosody in all tonal conditions (*Tone*: Level, *Estimate*=1.201, *SE*=0.23, *P*<.001; *Tone*: Rising, *Estimate*=1.022, *SE*=0.239, *P*<.01; *Tone*: Falling, *Estimate*=1.027, *SE*=0.211, *P*<.001). Mandarin listeners showed differences in rating scores between CV and CB conditions in the level and falling tonal contexts (*Tone* Level: *Estimate*=4.609, *SE*=0.391, *P*<.001; *Tone* Falling: *Estimate*=3.778, *SE*=0.234, *P*<.001).

The last contrast was the perceptual patterns of the matched narrow-verb focus in the NV focus and the mismatched contrastive-verb focus in the NC focus condition. The pairs of NV and NC conditions shared the same preceding question eliciting a narrow focus

on the verb of answers, but the following answers in NV and NC focus conditions had different prosodic forms. Responses to the target NV condition and the competitor NC condition in each tonal condition are presented in Figure 4.9 in Appendix 6. The top chart refers to the proportions of Cantonese listeners, and the bottom indicates patterns of Mandarin listeners. Post hoc results indicated that Cantonese listeners indifferently distinguished the matched and mismatched prosody of dialogues between NV and NC focus conditions. In contrast, Mandarin listeners discriminated the mismatched contrastive-verb focus with higher evaluations than the matched narrow-verb focus in the level and falling tonal contexts (*Tone* Level: *Estimate*=-1.791, *SE*=0.308, *P*<.001; *Tone* Falling: *Estimate*=1.594, *SE*=0.397, *P*=0.026).

To conclude, both groups of adult listeners were able to differentiate between the mismatched broad focus and the matched narrow-verb focus or matched contrastive-verb focus, despite nuanced differences in perception across tonal contexts. Besides, it was challenging to distinguish between mismatched contrastive-verb focus and matched narrow-verb focus when they were both preceded by the information structure of narrow focus.

4.3.3.2 Different information structures-the same prosodic form

This subchapter investigated focus contrasts that shared the same acoustic forms of answers but different information structures in the preceding discourses in each pair of dialogues. The first contrast was the matched broad focus in the BF condition versus the mismatched broad focus in the NB condition. Precursor questions were expected to elicit the answer with a broad focus in the BF condition and a narrow-verb focus in the NB condition. The following answers, however, were generated with a broad focus, which matched the BF focus condition (target) but mismatched the NB focus condition (competitor). Figure 4.10

shows the perceptual results of the matched BF focus condition (on the left) and the mismatched NB focus condition (on the right).

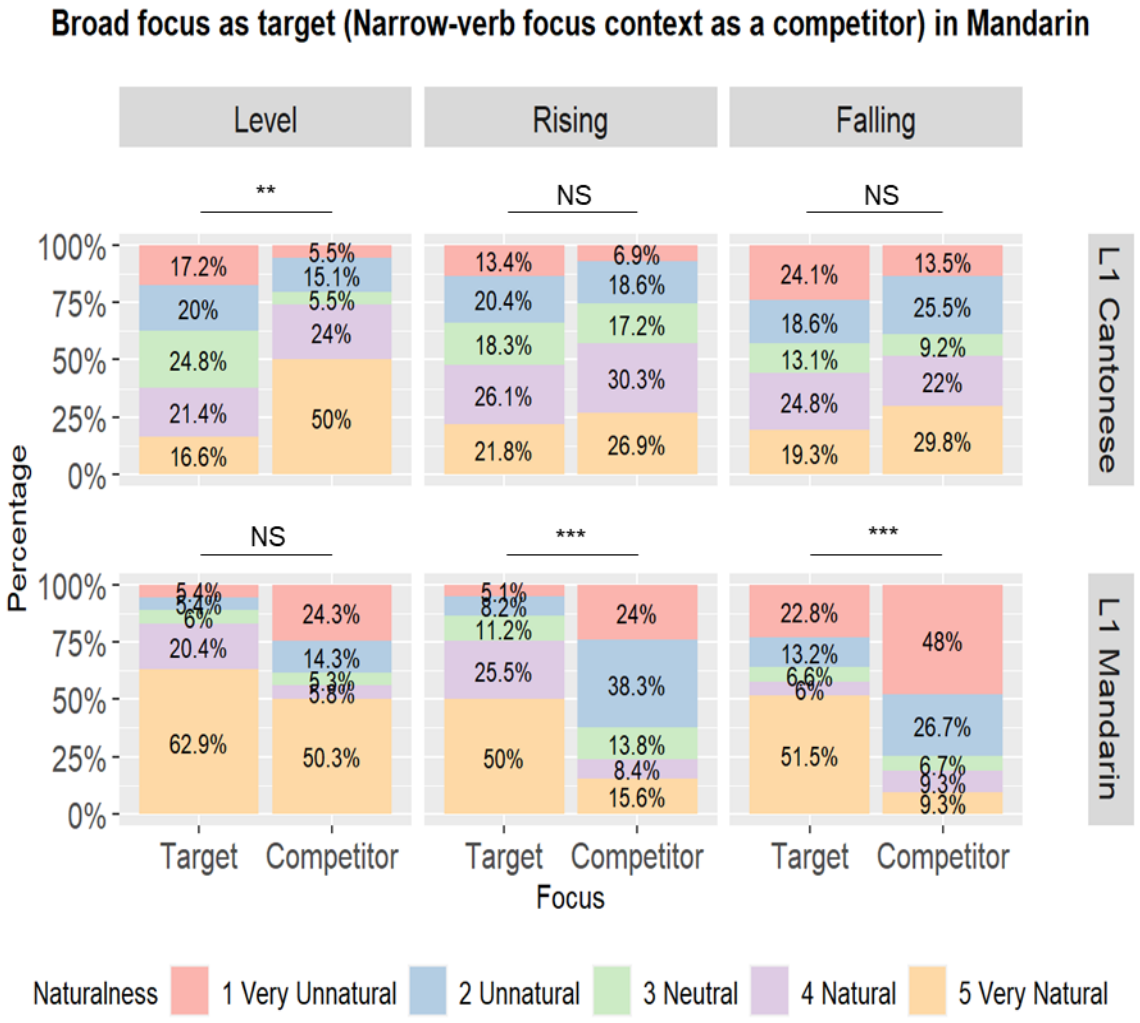


Figure 4.10 Adults' performance in Mandarin focus contrasts with different information structures (Narrow-Broad).

Targets refer to pairs with matched broad focus contexts and forms; Competitors refer to pairs with mismatched narrow-verb focus contexts and broad focus forms.

According to Figure 4.10, in BF conditions (targets), non-native listeners evaluated the matched broad focus with 38%, 47.9%, and 44.1% of natural responses in the level, rising, and falling tonal contexts. The accuracy was 83.3%, 75.5%, and 57.5% of native listeners in the three tonal surroundings. In NB conditions (competitors), the perceptual

patterns of Cantonese and Mandarin listeners' responses to the mismatched broad focus have been reported in Chapter 4.3.3.

Comparing the BF and NB conditions within group, Cantonese-speaking adults rated the matched broad focus as less natural than the mismatched broad focus (*Estimate*=-0.706, *SE*=0.125, *P*<.001). The deviation between the BF and NB conditions was increased in the context of level tones (*Tone* Level: *Estimate*=-1.28, *SE*=0.291, *P*<.01). As opposed to this, Mandarin adults significantly differed the matched and mismatched broad focus with higher evaluations of the prosodic naturalness for the congruous prosody (*Estimate*=1.636, *SE*=0.149, *P*<.001). When the focused constituents carried the rising and falling tones, native listeners demonstrated significance between BF and NB conditions (*Tone* Rising: *Estimate*=1.801, *SE*=0.27, *P*<.001; *Tone* Falling, *Estimate*=2.235, *SE*=0.311, *P*<.001).

The current study also examined how native and non-native listeners process broad focus in answering utterances that are preceded by the mismatched information structure of contrastive-verb focus. The answers in the BF and CB conditions had the same prosodic forms of broad focus in each pair. However, only the precursor question in the matched BF condition was "what happened?" in order to elicit the matched broad focus in the answering utterance. In the mismatched CB condition, the information structure of the preceding discourse was the contrastive focus on the verb, while the mismatched answer was delivered with a broad focus on the entire utterance. Figure 4.11 shows the contrast in focus processing between the BF (target) and CB (competitor) conditions in native Cantonese (top) and Mandarin (bottom) participants.

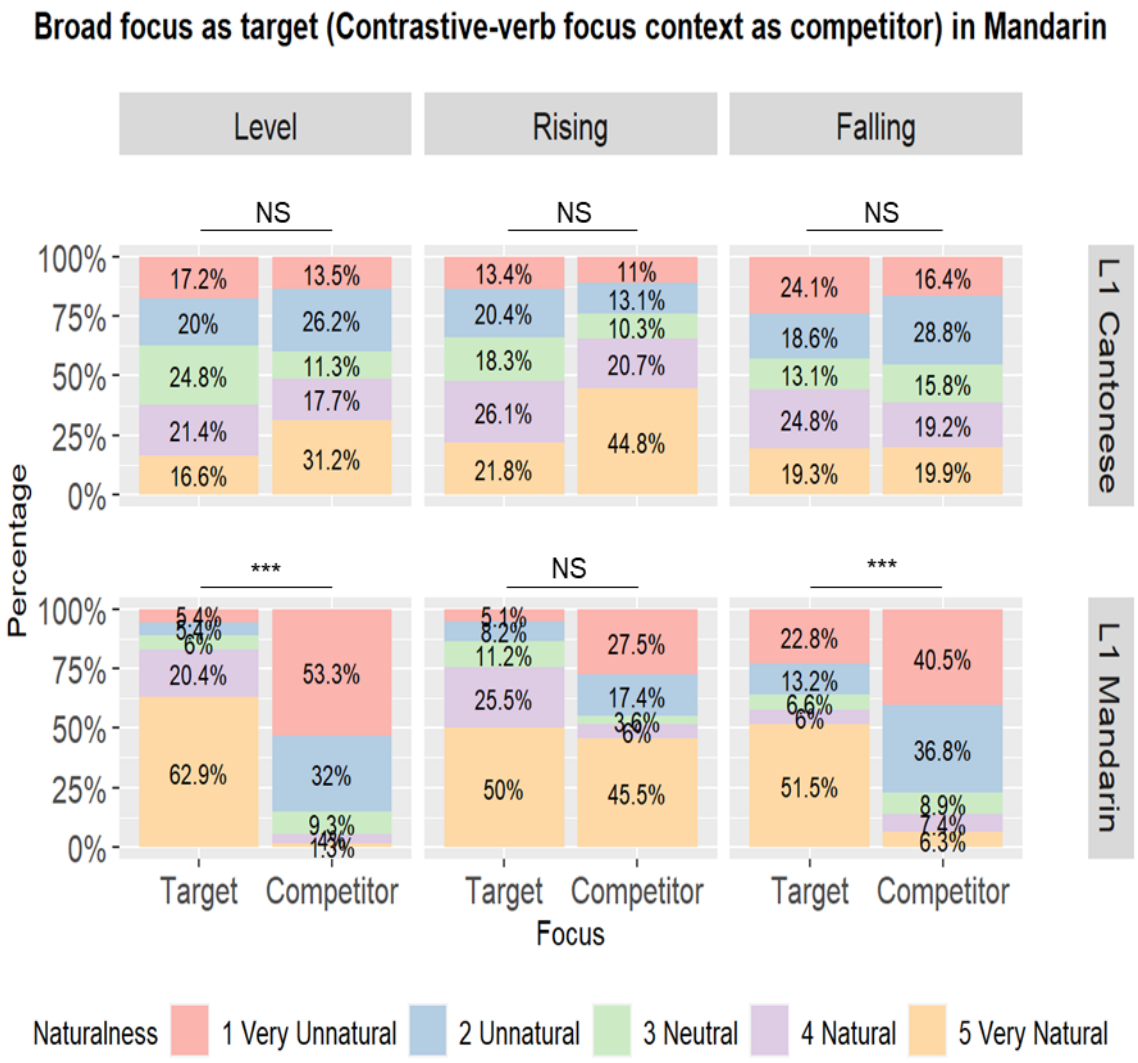


Figure 4.11 Adults' performance in Mandarin focus contrasts with different information structures (Contrastive-Broad).

Targets refer to pairs with matched broad focus contexts and forms; Competitors refer to pairs with mismatched contrastive-verb focus contexts and broad focus forms.

According to statistical findings, Cantonese adult listeners insignificantly differed the matched broad focus (BF) from the mismatched broad focus (CB) with lower evaluations of prosodic naturalness for the former condition. However, Mandarin adult participants evaluated the matched broad focus (BF) with significantly higher scores of naturalness degrees than the mismatched broad focus in the CB condition (*Estimate*=2.188, *SE*=0.151, *P*<.001). Closer examinations of the tonal effects demonstrated that the

differences in perceptual evaluations between BF and CB conditions were significant in the level and falling tonal contexts (*Tone* Level: *Estimate*=3.612, *SE*=0.31, *P*<.001; *Tone* Falling: *Estimate*=2.227, *SE*=0.255, *P*<.001).

The final comparison, the CV versus the NC focus conditions, was to probe the perceptual discriminations between narrow-verb focus and contrastive-verb focus. In the target contrastive-verb (CV) focus condition, the precursor question was to elicit an utterance with a contrastive focus on the verb, and its following answer was produced with the matched prosodic focus. In the competitive NC conditions, each pair had the same answers as the CV condition, but the preceding discourses were to elicit answers with a narrow focus on the verb. Participants' perceptions of congruous CV and incongruous NC conditions under various tonal settings are shown in Figure 4.12 in Appendix 6. Post-hoc analyses showed that non-native listeners, like native listeners, indifferently processed the contrastive-verb focus either preceded by matched information structures or not.

In summary, native Cantonese-speaking adults were unable to distinguish between congruous and incongruous broad focus forms in response to preceding discourses that elicited narrow or contrastive focus on the verb. In contrast, native Mandarin-speaking adults were able to accurately differentiate between incongruous broad focus forms of broad focus. Also, when the preceding context elicited narrow-verb focus, both groups of adult listeners equally rated the mismatched contrastive-verb focus as natural as the mismatched prosodic form.

4.4 Discussion

This study explored whether trilingual adults, whose native language is Cantonese, have acquired a native-like ability to decipher prosodic focus in their third language, Mandarin.

This question was approached through analyses of five aspects: (1) a comparison of how non-native adults process congruous prosody of broad focus, narrow-verb focus, and contrastive-verb focus with how native adults process it, (2) a comparison between native and non-native adults in comprehending these three focus categories in incongruous prosody, (3) a comparison between matched and mismatched pairs with the same preceding context but different prosodic forms, (4) a comparison between matched and mismatched pairs with the same prosodic realization but different information structures, and (5) an investigation of whether lexical tonal environments and audio-visual modalities impact the processing of focus differently between native and non-native adults.

Congruous prosody: In summary, the results of this study showed that Cantonese adult listeners were less proficient in comprehending congruous prosody regardless of focus categories compared to native listeners and tended to rate the congruous focus conditions as less natural than native listeners. The differences between the two groups were slightly significant in narrow and contrastive focus on the verb, in which Cantonese listeners did not rate the congruous narrow and contrastive focus as natural as Mandarin listeners did. Although their accuracy is not entirely native-like, Cantonese listeners have acquired the ability to process these two focus categories with an accuracy rate of nearly 80%.

However, the two groups had a notable difference in processing broad focus. Cantonese adults evaluated congruous broad focus with an accuracy below chance, whereas Mandarin listeners achieved over 70% of natural prosody accuracy. As discussed in Chapter 3.4, Cantonese listeners also demonstrated insensitivity to congruous broad focus in their native Cantonese, potentially attributed to their prior knowledge of the infrequent usage of broad focus in conversations. According to Roettger et al. (2019), the

precursor question “what happened?”, which typically elicits a broad focus, rarely occurs between interlocutors and provides limited contextual information. Consequently, listeners may lack experience in anticipating and interpreting the prosodic features of subsequent utterances when encountering the information structure of broad focus. Thus, it is possible that Cantonese listeners employ their prior knowledge of broad focus in their L1 and apply a similar probabilistic inference approach in their processing of prosodic focus in their L3. In non-native Mandarin, Cantonese listeners may exhibit greater uncertainty in broad focus production due to limited exposure to appropriate prosodic distributions that accurately answer questions like "What happened?". The consistent performance of Cantonese listeners in both L1 and L3 suggests that their focus processing strategies may be maintained across languages.

A subsequent issue raised is whether the uncertainty of broad focus occurs and influences the probabilistic inference to Mandarin prosodic focus by native adults. On one hand, there was a significant difference in the ratings of naturalness between broad focus and narrow-verb or contrastive-verb focus. The ratings for broad focus were noticeably lower, although the accuracy for broad focus still achieved above 70%. This discrepancy suggests that native adult listeners may have more accumulated experience with the probability of narrow and contrastive focus production compared to broad focus production. On the other hand, the accuracy of more than 70% for natural prosody in the current study aligns with the focus identification accuracy observed in context-free settings (Xu et al., 2012), indicating that Mandarin broad focus production may have provided sufficient bottom-up signals to detect the focus category. Therefore, these findings propose that Mandarin adults can process relatively definite prosodic representations of broad focus,

despite a possibility of differences in pre-existing experiences between broad focus and other focus categories (narrow and contrastive focus). In contrast, non-native listeners with different prosodic representations of broad focus from native listeners, may differ in the strategies by employing more top-down cues than actual prosodic signals in Mandarin focus processing.

Incongruous prosody: The results of the present study indicate that Cantonese listeners were less sensitive than Mandarin listeners to incongruous prosody of broad focus, regardless of the preceding context. Specifically, Cantonese listeners achieved an accuracy of only around 30% in detecting unnatural prosody, while Mandarin listeners were able to evaluate incongruous prosody with an accuracy above the chance level. The study also compared the listeners' perception between matched and mismatched pairs from two perspectives: perceptual accuracy in mapping particular focus categories to one of two prosodic forms and mapping the same prosodic realization to one of two information structures.

In terms of the first perspective addressing sub-question (3), Cantonese listeners were able to discriminate competitively incongruous broad focus forms from the target narrow or contrastive focus, although their accuracy in identifying incongruous prosody was lower than that of native Mandarin listeners. This ability to detect narrow and contrastive focus prosodic forms is consistent with their performance in native Cantonese, suggesting that Cantonese listeners can acquire and retain relatively accurate knowledge of prosodic distributions of narrow and contrastive focus on specific constituents in both L1 Cantonese and L3 Mandarin.

In the second perspective addressing sub-question (4), the study compared listeners' ability to discern broad focus forms preceded by matched or mismatched (narrow and contrastive focus) prior discourse. Results showed that it was challenging for Cantonese listeners to distinguish between congruous and incongruous broad focus when it was mistakenly elicited by narrow-verb and contrastive-verb focus. Specifically, the study revealed a reverse pattern in the perceived prosodic naturalness between congruous and incongruous broad focus pairs when preceded by a mismatched narrow focus context, in which the congruous pairs were rated as having lower prosodic naturalness compared to the incongruous pairs. This finding provides support for the hypothesis that the difficulty of Cantonese adult listeners in detecting broad focus is likely attributed to their pronounced inclination to reject congruous prosody associated with broad focus, as well as their biased preference for narrow focus contexts. Notably, this observation is not limited to their native Cantonese, but also extends to their L3 Mandarin. However, it is important to note that the control group consisting of native Mandarin-speaking adult listeners did not exhibit the same perceptual biases. This observation suggests that the perceptual biases observed in trilingual adult listeners for Mandarin are likely influenced by the prior knowledge regarding focus mapping in their L1 Cantonese.

Additionally, when an incompatible prior context was used to elicit a contrastive-verb focus, the subsequent broad focus answers were interpreted as having equally natural prosody compared to the prosodically congruous broad focus, irrespective of their congruity or incongruity. These findings suggest that Cantonese adult listeners can to some extent exploit their prior knowledge of the information structure of contrastive focus to anticipate the salient prosody in L3 Mandarin. This implies that they possess a relatively

sufficient level of experience in discerning prosodic cues of contrastive focus, which is realized by a prosodic accentuation on the correction information, such as the verb used in the target utterances in the current study. Their familiarity with the context of contrastive focus likely explains their sensitivity to prosodic incongruity when confronted with prosodically non-prominent broad focus instances.

In contrast, native Mandarin listeners were able to detect the incongruous prosody of broad focus forms preceded by both narrow and contrastive focus contexts, rating them as less natural prosody compared to congruous pairs. These results align with a previous study by Yang (2022), in which Mandarin listeners accurately (above 80% acc.) mapped the broad focus forms onto the correct information structure when the competitive information structure is the narrow-verb focus. In Yang's (2022) study, Mandarin listeners achieved the highest identification accuracy of target broad focus when the mismatched prior question elicited a narrow focus on the verb among focused elements of subject, verb, object, and verb phrase. The consistent findings from both studies imply that Mandarin adult listeners appear to possess a stable comprehension of prosodic distributions associated with broad focus, as evidenced by the results obtained from both two-forced choice and 5-point Likert scale tasks. In overall performance, when preceding questions elicit new information or require the correction of a specific constituent, Mandarin adult listeners demonstrate an ability to predict the presence of prosodic focus in the subsequent utterances. However, in cases where the target utterance is in broad focus, deviating from the expectations set by the preceding question, they display accurate sensitivity to the prosodic incongruity.

Furthermore, the current study expands the understanding of Mandarin focus processing from various focus breadths (broad focus vs. narrow focus) in Yang (2022) to focus types (narrow focus vs. contrastive focus). The present study yielded insights into the perceptions of Mandarin listeners, revealing that they rated the incongruous broad focus as significantly more unnatural in a context that elicited contrastive-verb focus compared to narrow-verb focus. Notably, Mandarin listeners demonstrated variations in their expectations of narrow focus versus contrastive focus, despite rating the prosody of narrow-verb and contrastive-verb focus as similarly natural. Specifically, native Mandarin listeners anticipated a more prominent prosody for contrastive focus when it was applied to the same syntactic constituent as narrow focus. Interestingly, this distinction in prosodic expectations was not observed in the processing of focus in L3 Mandarin by native Cantonese-speaking adult listeners. It is worth noting that this inclination towards intensive expectations of contrastive focus in semi-natural dialogues was also evident in both Cantonese (Experiment 1 in the present study) and German (Toepel et al., 2007) native listeners.

Tonal contexts and audiovisual effects: The fingdings of this experiment showed that the influence of lexical tones on perceptual patterns was only evident in particular focus categories of incongruous prosody, with no notable difference in perceptual judgments between different lexical tones in congruous prosody. Specifically, when the narrow focus was the target, Cantonese listeners were able to discriminate the congruous narrow-verb focus from the incongruous broad focus forms significantly in the falling tones, as shown in Figure 4.7 in Chapter 4.3.3. The post hoc analysis revealed that the significant advantage was mainly attributed to better perception in the mismatched pairs with broad focus forms

rather than the matched pairs with narrow-verb focus in the tonal context of falling contours. Cantonese listeners' perceptual patterns of the mismatched pairs demonstrated a tendency towards decreasing mapping accuracy from falling tones to rising tones and then to level tones. It is possible that Cantonese listeners' anticipation of prosody is more prominent in the syllables carrying falling contours (Tone 4) compared to the rising contours (Tone 2) and level tone (Tone 1). However, this minor finding cannot be generalized to Cantonese listeners' focus processing strategy, as this tendency was only evident in the specific focus condition.

In addition, the incorporation of extra visual stimuli appeared to have a negligible impact on the processing of focus in Mandarin by both native and non-native listeners. This outcome aligns with earlier studies examining the audiovisual perception of Mandarin lexical tones (Hannah et al., 2017; Burnham et al., 2001), demonstrating that auditory signals alone are robust enough for perception, with little potential for improvement through the addition of visual cues. Given the multifaceted role of pitch in the modulation of prosody in tonal languages, it is logical to assume that these languages provide adequate signals for effective modulation, with visual cues as a supplementary aid for native tonal listeners.

Chapter 5. Experiment 3: Cantonese focus processing in trilingual children

5.1 Introduction

The This study aims to investigate whether school-aged trilingual children possess adult-like abilities to decode prosodic focus in their native Cantonese language. To achieve this, the trilingual Cantonese experiment conducted with adults will be replicated with trilingual children. Currently, the available literature on the comprehension of prosodic focus in tonal languages mainly focuses on contrastive focus in picture-matching tasks. These tasks involve children judging whether subject or object stimuli match with the auditory information (e.g., Chen, 2010; Ge et al., 2022; Tang et al., 2022). As an illustration, Ge et al. (2022) discovered that Cantonese children between the ages of five to eight exhibit a higher degree of accuracy in identifying suitable contrastive focus on the subject, compared to inaccurate prosody. However, this experimental paradigm only tests children's sensitivity to adequate or inadequate prosody concerning specific correcting constituent of target utterances. Thus, it remains unclear whether Cantonese-speaking children can process prosody in a variety of focus categories in dialogues. Such dialogues require them to consider the entire prosody of the utterance, which is elicited by different precursor questions, and involves a high cognitive demand in processing the information structures of discourse.

An ERP study by Pannekamp et al. (2011) examined Q-A pair in German and found that children display sensitivity to incongruent prosody yet are unable to distinguish narrow focus from contrastive focus at the age of five. With increasing age, their ERP responses demonstrated differences in processing various focus categories (narrow vs. contrastive)

by age eight and achieved adult-like performance by age twelve. However, it is currently unknown how tone languages-speaking children process different prosodic focus in semi-naturalistic Q-A dialogues. For example, it is not clear if school-aged Cantonese children also acquire a developmental advantage of contrastive focus over narrow focus, comparable to the developmental trajectory of German children at a similar age. Furthermore, Yang (2017) found that Mandarin school-aged children are unable to use acoustic cues to mark focus in a manner similar to adults until they reach the ages of ten to eleven, rather than between seven to eight years old. To the best of my knowledge, there is currently no available behavioral research that investigates whether school-aged native speakers of tonal languages, such as those at the age of nine, have developed the adult-like ability to process focus during middle and late childhood.

Furthermore, this study will address this question by examining whether trilingual children show different accuracies in decoding focus across tonal contexts. The development of Cantonese tonal inventory is known to be a lengthy process (Mok et al., 2019), raising the question of whether children are capable of processing lexical and prosodic functions in an adult-like manner simultaneously. As decoding Cantonese focus is cognitively demanding, it remains unclear whether children will rely on co-speech visual signal, such as facial expressions and head nodding, to aid their focus processing. To address these issues, the present study will examine the following questions:

(1) Can trilingual children process prosodic focus in congruous prosody in their native Cantonese in an adult-like manner?
(2) Can trilingual children interpret incongruous prosodic focus in their L1 Cantonese with an adult-like accuracy?

(3) Can trilingual children differentiate congruous from incongruous pairs that share either the same information structure or the same prosodic form in Cantonese, similar to native adults?

(4) To what extent do tonal context and supplementary visual cues impact the focus processing of trilingual children in Cantonese?

The hypothesis of the current study for trilingual children is that they might not have developed an adult-like competence to interpret various prosodic focus in their native Cantonese. Moreover, they are likely to demonstrate higher accuracy in detecting congruous focus conditions than incongruous focus conditions. During the process of focus decoding, trilingual children may encounter difficulties with specific lexical tones, such as the high-level Tone 1. Furthermore, it is probable that they will exhibit a greater reliance on auditory cues than visual cues when comprehending multimodal speech.

5.2 Method

5.2.1 Participants

In this experiment, forty-seven native Cantonese-speaking trilingual children were recruited for the online perceptual tasks. The children were twenty-nine boys and eighteen girls whose mean age was 9.2 ± 0.93 years (range, 7.42-11 years). All children were born and lived in Hong Kong, and their parents are both Hong Kong residents who speak Cantonese with the children at home. As sequential trilingual children whose first language is Cantonese, they learned English and Mandarin as their second and third languages, respectively. The average starting age for learning English is 3.69 ± 1.64 years and 4.72 ± 1.65 years for Mandarin, and they kept learning English and Mandarin at school when they participated in the experiment. All participants have not reported any speech or language

impairment. The informed consent for each participant was obtained from their parents, who also received the payment for participation on behalf of their children. The current project has been approved by the human ethics committee of the university (HSEARS20200919002). Besides, adult participants recruited in Experiment 1 (in L1 Cantonese) were the comparison group of the current Experiment 3 (in L1 Cantonese) and completed the same task as the children.

5.2.2 Materials and design

The current experiment adopted the same materials and experiment design in Experiment 1 (Chapter 3). Specifically, the target sentences in Cantonese are illustrated in Chapter 3.2.2.1, the audiovisual recordings in Chapter 3.2.2.2, and the design of stimuli in Chapter 3.2.2.3.

5.2.3 Procedures

Participant began the experiment with practice sessions, followed by the experimental session. Listeners of trilingual children followed the procedures illustrated in the Cantonese experiment for adults in Experiment 1 (Chapter 3.2.3). This study recruited forty-seven trilingual children who passed the practice session with an accuracy of 60% or above to following experimental sessions. Like adults, trilingual children participants completed the online experiment on E-prime Go, and all sessions were finished in one day.

5.2.4 Data analyses

I submitted the responses obtained from children and adults to the cumulative link mixed effects models (*CLMM*) with the *ordinal* package (Christensen, 2015) in R (R Core Team, 2022). As the responses on 5-Likert scale are ordinal data, dependent variables of the

current analysis, I fitted them to the ordinal regressions with random effects by the *clmm* function. The fixed effects are *Group* (Child, Adult), *Focus* (BF, NV, CV, BN, BC, NC), *Tones* (LL, LR, RR, RL), and *Modalities* (AO, AV), and the factors of *Subjects* and *Trials* are tested for random effects. Closer examinations of focus contrasts were realized by post hoc analysis using the *emmeans* package (Lenth R, 2023) in R.

5.3 Results

To define the optimal regression model, I first conducted likelihood ratio tests for the fixed factors. Results proved that the factor *Modality* (AO, AV) was insignificant to both groups (*Group* * *Modality*: $\chi^2(3) = 1.411$, $P = 0.703$) in identifying various prosodic focus forms. Thus, the final model involved the main effects of three-way interaction among *Group* (Child, Adult), *Focus* (BF, NV, CV, BN, BC, NC), *Tones* (LL, LR, RR, RL), together with random effects of *Subject* and *Trial* of stimuli. Statistical results demonstrated the significant effects of *Focus* ($\chi^2(5) = 434.76$, $P<.001$), and interaction of *Group* and *Focus* ($\chi^2(5) = 48.07$, $P<.001$). No significant effect was found in the factor of *Group* ($\chi^2(1) = 0.37$, $P = 0.541$), *Tone* ($\chi^2(3) = 6.13$, $P = 0.106$), or the interaction between *Group* and *Tone* ($\chi^2(3) = 6.26$, $P = 0.1$). The three-way interaction among *Group, Focus,* and *Tone* ($\chi^2(15) = 18.59$, $P = 0.233$) was insignificant either. Chapter 5.3.1 first compares the comprehension of congruous prosodic focus by trilingual children and adults, followed by their perception in focus processing of incongruous prosody in Chapter 5.3.2. Chapters 5.3.3 compares children's and adults' perceptual patterns of focus contrasts (matched vs. mismatched) in various tonal contexts.

5.3.1 Focus processing of congruous prosody in Cantonese by trilingual children

The congruous dialogues in the current study included matched prosodic broad focus (BF), narrow-verb focus (NV), and contrastive-verb focus (CV), similar to Experiment 1 in Chapter 3.3.1. Figure 5.1 presents an overview of the perceptual patterns of these three focus conditions from children and adults. The stacked bar plots compare the perception of children and adults in congruous prosody categorized by focus categories. The areas on each scale indicate the ratios of the 5-Likert scale on prosodic naturalness, and the naturalness degrees decrease from the bottom to the top of the stacked bars. This dissertation used the responses of "natural" and "very natural" to identify congruous prosody and "unnatural" and "very unnatural" responses to identify incongruous prosody. Responses marked as "neutral," which indicate ambiguity, were not considered in the study.

For broad focus (BF, the left), trilingual children had a higher accuracy than adults in comprehending congruous prosodic realizations. Children evaluated the matched broad focus with an accuracy of 52.9% of “natural” and “very natural” responses, whereas the accuracy of trilingual adults was only 33.7%. However, post hoc analyses showed that adults evaluated the congruous broad focus with an insignificant lower score than children (*Estimat*e=-0.69, *SE*=0.221, *P*=0.075).

In terms of other pairs of congruous prosody, trilingual children and adults processed the narrow-verb focus (NV) and contrastive-verb focus (CV) in a similar manner, as illustrated in Figure 5.1. Both groups demonstrated a relatively high accuracy in identifying congruous narrow-verb and contrastive-verb focus. In the narrow-verb (NV) focus condition, 72.8% and 77.7% of responses from children and adults, respectively, identified the natural prosody. Similarly, in the contrastive-verb (CV) focus, 69.3% of

children and 75.3% of adults accurately identified the congruous prosody. Statistically, there was no significant difference between children and adults in comprehending matched narrow-verb or contrastive-verb focus.

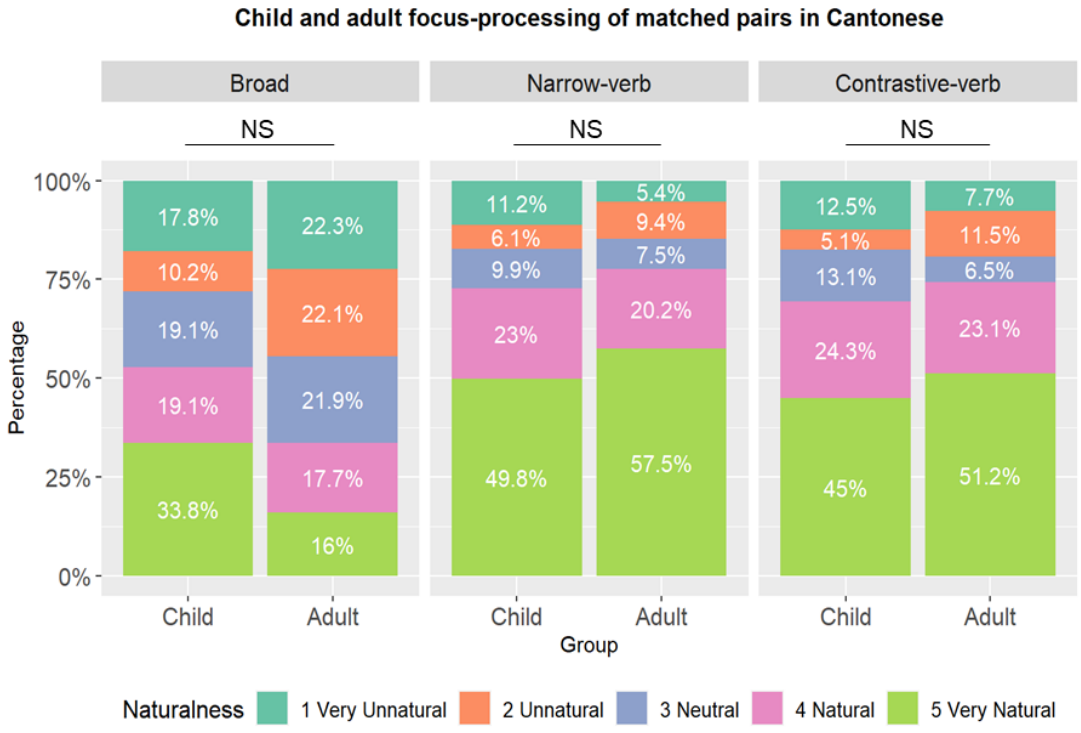


Figure 5.1 Children and adults' performance in congruous Cantonese prosody

Although BF, NV, and CV focus conditions were all dialogues with congruous prosody, children and adults both comprehended the broad focus as a less natural prosody than the narrow-verb or contrastive-verb focus. Post hoc analyses demonstrated that children rated the matched broad focus with significantly lower scores of prosodic naturalness than the narrow-verb (*Estimate*=-0.855, *SE*=0.159, *P*<.001) and contrastive-verb focus (*Estimate*=-0.679, *SE*=0.154, *P*<.001). Similarly, adults significantly distinguished the matched broad focus from the narrow-verb (*Estimate*=-1.931, *SE*=0.128, *P*<.001) and contrastive-verb focus (*Estimate*=-1.565, *SE*=0.125, *P*<.001) with assessments of low degrees in prosodic naturalness.

Furthermore, a closer examination also examined the effect of tonal contexts on the comprehension of congruous prosodic focus by children and adults. Results showed that children performed no significant differences between matched broad focus and narrow-verb focus in all tonal contexts except for the rising-level tonal context (*Tone* RL: *Estimate*=-1.447, *SE*=0.364, *P*=0.049). For adults, as reported in Chapter 3.3.1, they significantly rated congruous broad focus as less natural than congruous narrow-verb focus in all tonal contexts (*Tone* LL: *Estimate*=-1.987, *SE*=0.23, *P*<.001; *Tone* RL: *Estimate*=-1.797, *SE*=0.27, *P*<.001; *Tone* LR: *Estimate*=-1.705, *SE*=0.225, *P*<.001; *Tone* RR: *Estimate*=-2.237, *SE*=0.28, *P*<.001). Their evaluations of contrastive focus on verbs were also significantly higher than that of the matched broad focus (*Tone* LL: *Estimate*=1.943, *SE*=0.228, *P*<.001; *Tone* RL: *Estimate*=1.306, *SE*=0.269, *P*<.01; *Tone* LR: *Estimate*=1.473, *SE*=0.222, *P*<.001; *Tone* RR: *Estimate*=1.538, *SE*=0.269, *P*<.001). As for the comparison of matched narrow-verb and contrastive-verb focus, like adults, children indifferently processed the narrow and contrastive focus on the verbs in all tonal environments.

To sum up, trilingual children and adults demonstrated statistically equivalent comprehension of the three matched focus conditions in Cantonese. Moreover, both groups rated the narrow and contrastive focus on the verb as more natural than the broad focus, despite all three focus conditions having congruous prosody.

5.3.2 Focus processing of incongruous prosody in Cantonese by trilingual children

The incongruous prosody was examined in mismatched conditions of narrow-broad focus (NB), contrastive-broad focus (CB), and narrow-contrastive focus (NC), as investigated in Experiment 1 in Chapter 3.3.2. Figure 5.2 compares the response distributions of three mismatched focus conditions obtained from trilingual children and adults. In general,

listeners of children and adults similarly comprehended the incongruous prosody of mismatched broad focus (NB and CB) and contrastive-verb focus (NC). The identification accuracy of unnatural prosody (including "unnatural" and "very unnatural" responses) was 27.5% of children and 27% of adults in the NB focus condition, 31.3% of children and 38.1% of adults in the CB condition, and 17.2% of children and 15.6% of adults in the NC condition.

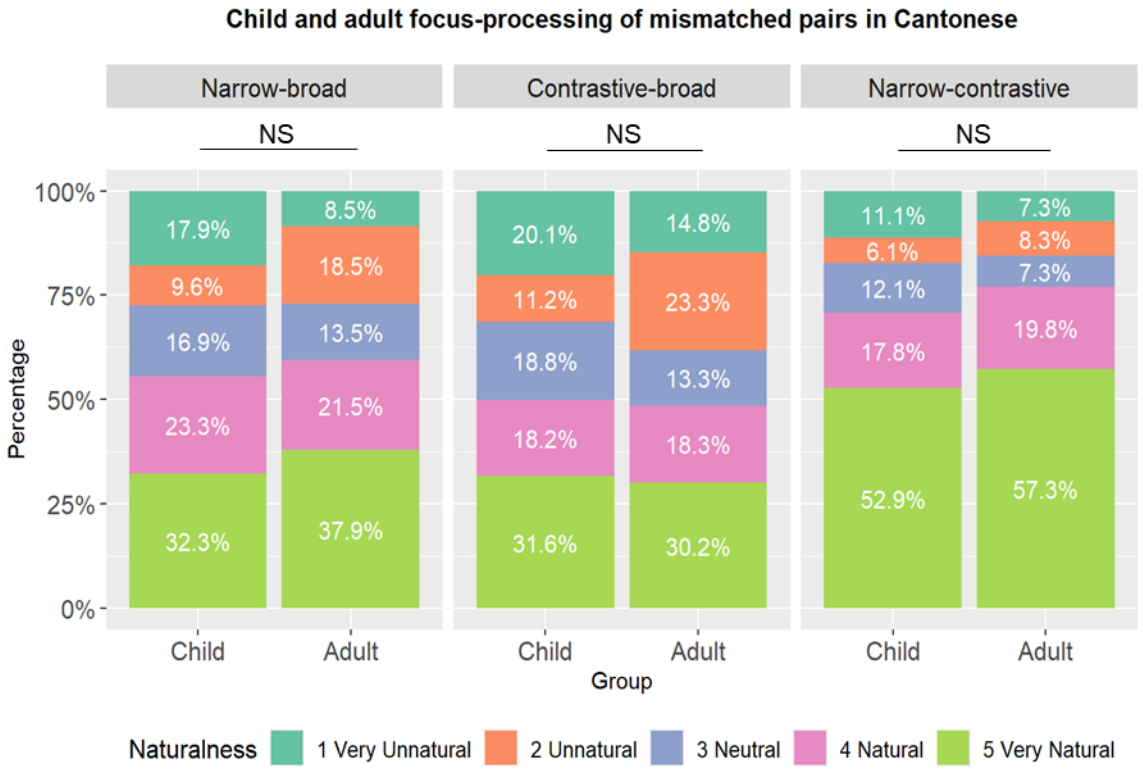


Figure 5.2 Children and adults' performance in incongruous Cantonese prosody

Post hoc analysis revealed no statistical significance between groups in comprehending any mismatched focus conditions. However, trilingual children insignificantly outperformed adults in identifying the incongruous broad focus (NB) in all tonal contexts. When comparing mismatched focus conditions, differences were observed between the narrow-verb (NB) and contrastive-verb (CB) focus contexts. Children

demonstrated equivalent processing of the incongruous broad focus when preceded by narrow or contrastive focus contexts. In contrast, adults exhibited difficulties in the NB condition with significantly lower accuracy compared to the CB condition (*Estimate*=0.559, *SE*=0.122, *P*<.001). Both children and adults rated the mismatched contrastive-verb focus in the NC condition as more natural prosody than the mismatched broad focus in NB (*Group*: child, *Estimate*=0.864, *SE*=0.156, *P*<.001; *Group*: adult, *Estimate*=0.891, *SE*=0.129, *P*<.001) and CB (*Group*: child, *Estimate*=1.03, *SE*=0.159, *P*<.001; *Group*: adult, *Estimate*=1.45, *SE*=0.129, *P*<.001) conditions.

Overall, there were no significant differences between children and adults in their ability to identify incongruous focus conditions. Both groups rated the mismatched broad focus as having less natural prosody compared to the mismatched contrastive-verb focus. Furthermore, trilingual adults showed greater accuracy in identifying incongruous broad focus in the context of contrastive-verb focus compared to narrow-verb focus. Trilingual children displayed a similar trend, but the difference was not statistically significant.

5.3.3 Comparison between congruous and incongruous focus conditions in Cantonese by trilingual children

Similar to Experiment 1, the current study investigated trilingual children's focus processing in Cantonese by comparing two focus conditions (matched vs. mismatched). In Section 5.3.3.1, pairs of Q-A dialogues that shared the same precursor questions were compared but differed in the prosodic realization of answers. These included NV vs. NB, CV vs. CB, and NV vs. NC focus conditions. Section 5.3.3.2 reported a comparative analysis of two focus conditions with different information structures preceding the same prosodic forms, including BF vs. NB, BF vs. CB, and CV and NC focus conditions. The

focus conditions used in this experiment can be found in Table 3.4 of Chapter 3.2.2.3 (Experiment 1).

5.3.3.1 Same information structure-different prosodic forms

This section compares listener performance in focus contrasts between matched narrow-verb focus (NV, target) and mismatched narrow-broad focus (NB, competitor) conditions. Figure 5.3 displays the perceptual evaluations of targets and competitors in four tonal contexts by both children (top) and adults (bottom). The left panels show target pairs with congruous narrow-verb focus, while the right panels show competitors with incongruous broad focus.

In general, children and adults performed similarly in processing matched narrow-verb (NV, target) focus and mismatched broad (NB, competitor) focus. Across all tonal contexts, participants identified a higher number of targets with matched narrow-verb focus as having natural prosody compared to competitors with mismatched broad focus. Children's accuracy in identifying matched narrow-verb focus was 76.6% in LL, 76.1% in RL, 69.2% in LR, and 69.3% in RR tonal contexts. For the NB condition, children obtained a significantly lower accuracy rate of 26.6%, 22.2%, 29.7%, and 30.7% in LL, RL, LR, and RR tonal contexts, respectively. Within the groups, children were able to significantly distinguish between matched narrow-verb focus and mismatched broad focus (*Estimate*=0.857, *SE*=0.161, *P*<.001), although the difference was not significant in any tonal environment. Adult listeners also had significantly higher ratings for matched narrow-verb focus compared to mismatched broad focus (*Estimate*=0.886, *SE*=0.131, *P*<.001), especially in the RR tonal context (*Tone:* RR, *Estimate*=1.482, *SE*=0.289, *P*<.001).

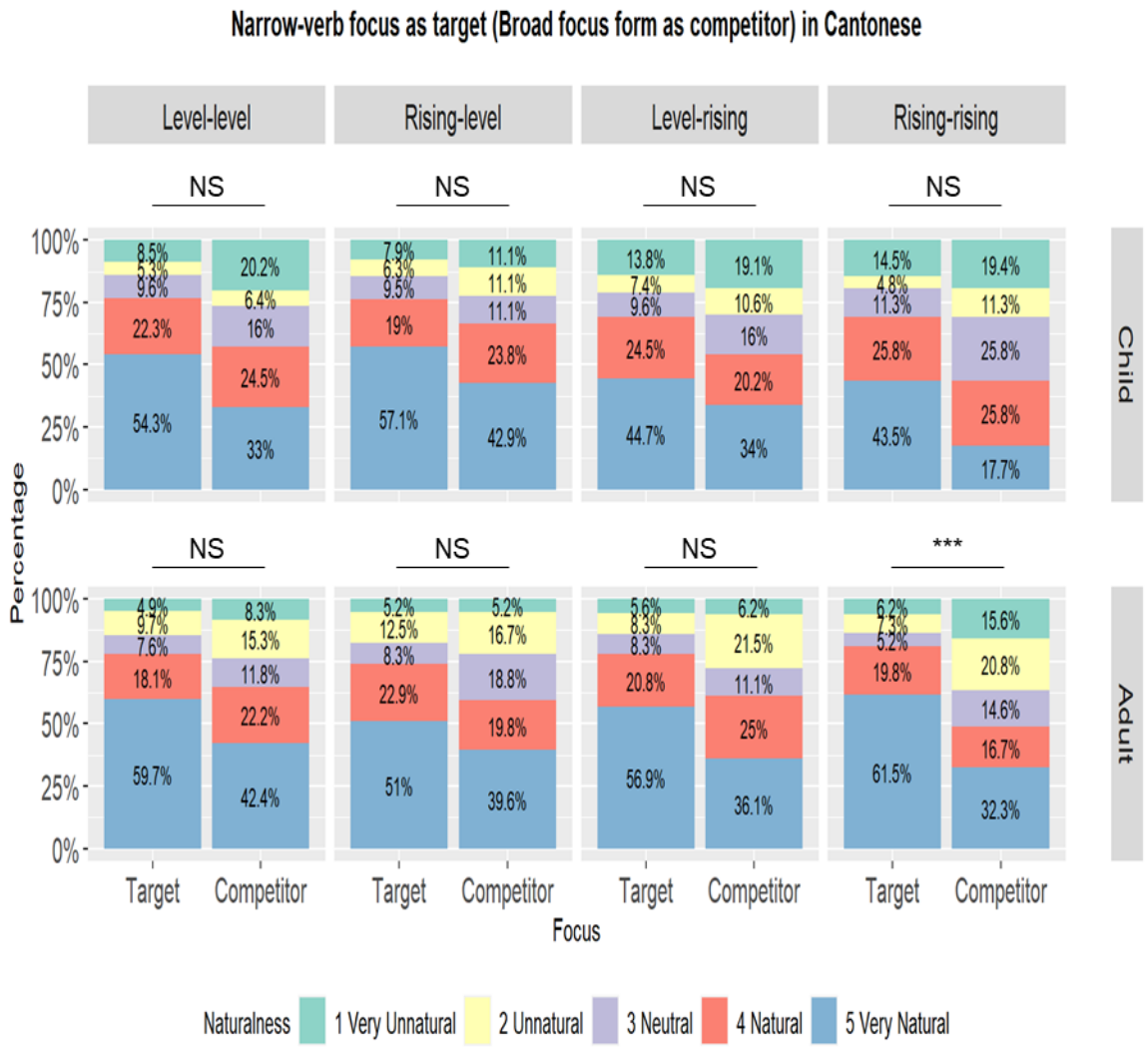


Figure 5.3 Children and adults' performance in Cantonese focus contrasts with different prosodic forms (Narrow-Broad).

Targets refer to pairs with narrow-verb focus context and matched narrow-verb focus forms; Competitors refer to pairs with narrow-verb focus context and mismatched broad focus forms.

Additionally, inn comparing CV and CB conditions, pairs with the matched contrastive-verb focus (CV) in answers were the targets, while answers with the mismatched broad focus (CB) were the competitors. Responses from children (the top) and adults (the bottom) are visualized in Figure 5.4. Children and adults demonstrated similar perceptual patterns in processing congruous contrastive-verb focus (CV) and incongruous broad focus (CB). In the matched CV condition, the identification accuracy of natural

prosody reached 69.2% in LL, 66.2% in RL, 68.1% in LR, 74.6% in RR tonal contexts by children, and 82.6% in LL, 67.7%in RL, 74.3% in LR, and 68.7% in RR tonal contexts by adults. In the mismatched CB condition, children identified the unnatural prosody with ratios of 31.9% in LL, 33.9% in RL, 29.8% in LR, and 30.2% in RR tonal contexts; adults comprehended the unnatural prosody with the percentage of 37.5% in LL, 38.6% in RL, 31.2% in LR, and 49% in RR tonal environments.

The comparison between the CV and CB conditions showed a significant difference in ratings from both children and adults. Specifically, children rated the mismatched broad focus as less natural compared to the matched contrastive-verb focus (*Estimate*=-0.846, *SE*=0.159, *P*<.001). However, this difference was not found to be significant in any tonal contexts. Among adults, a significant difference was found between the mismatched CB and matched CV conditions (*Estimate*=-1.078, *SE*=0.129, *P*<.001), especially in most tonal contexts (*Tone:* LL, *Estimate*=-1.223, *SE*=0.234, *P*<.001; *Tone:* LR, *Estimate*=-0.91, *SE*=0.229, *P*=0.048; *Tone:* RR, *Estimate*=-1.226, *SE*=0.278, *P*<.01).

Finally, this study examined the perception of participants in processing matched narrow-verb focus (NV) and mismatched narrow-contrastive focus (NC) conditions. In both conditions, the precursor questions were designed to elicit answers with a narrow-verb focus. However, the answers were produced with an incongruous contrastive-verb focus in the mismatched NC condition. Figure 5.5 in Appendix 7 displays the responses of children (top) and adults (bottom) to this focus contrast (NV vs. NC) in various tonal contexts.

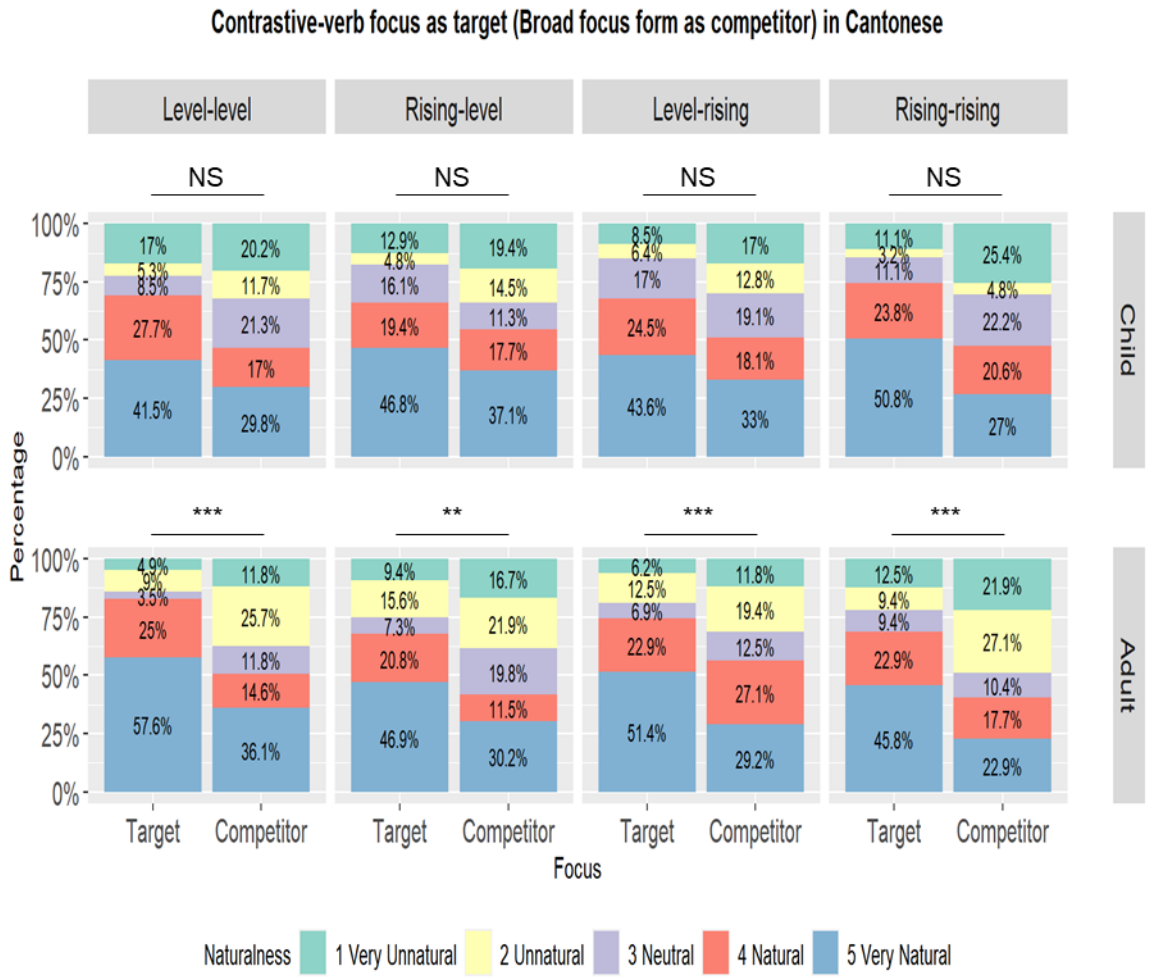


Figure 5.4 Children and adults' performance in Cantonese focus contrasts with different prosodic forms (Contrastive-Broad).

Targets refer to pairs with contrastive-verb focus context and matched focus forms; Competitors refer to pairs with contrastive-verb focus context and mismatched broad focus forms.

To summarize, the findings of this study suggest that native Cantonese children, like adults, are able to distinguish incongruous broad focus (NB and CB) from congruous narrow-verb (NV) and contrastive-verb focus (CV). However, children and adults showed different degrees of sensitivity to prosody in different tonal contexts, with lexical tones having only a limited impact on children's comprehension. Additionally, adults relied more heavily on prior discourse to map prosodic forms, resulting in significantly better processing of the information structure of contrastive focus than that of narrow focus, a

contrast which was not as apparent in children. Moreover, both groups were unable to discriminate between the contrastive-verb focus (NC) and the narrow-verb focus (NV).

5.3.3.2 Different information structures-the same prosodic form

This section evaluated how well children processed a particular prosodic focus when the prior information structures differed. The study first examined the comprehension of congruous and incongruous broad focus in various lead-in discourses (BF vs. NB and BF vs. CB). Firstly, results of the matched BF (the left) and mismatched NB (the right) conditions from children (the top) and adults (the bottom) are presented in Figure 5.6.

In the BF condition, children achieved natural prosody responses with an accuracy of 52.2%, 44.5%, 54.2%, and 60.3% in LL, RL, LR, and RR tonal environments, respectively. For adults, the accuracy of matched broad focus reached 35.5% in LL, 34.3% in RL, 37.5% in LR, and 25% in RR tonal environments. In the competitor (NB) condition, detailed response distributions for both groups were presented in Chapter 5.3.3. Post-hoc analysis showed that for children, there was no significant difference in their ratings of matched broad focus (BF) compared to mismatched ones in the NB condition, regardless of the tonal context. In contrast, adults significantly distinguished matched broad focus with lower evaluations than the mismatched pairs (Estimate=-1.046, SE=0.124, P<.001), particularly in specific tonal contexts (*Tone* LL: *Estimate*=-1.253, *SE*=0.227, *P*<.001; *Tone* RL: *Estimate*=-1.294, *SE*=0. 272, *P*<.01).

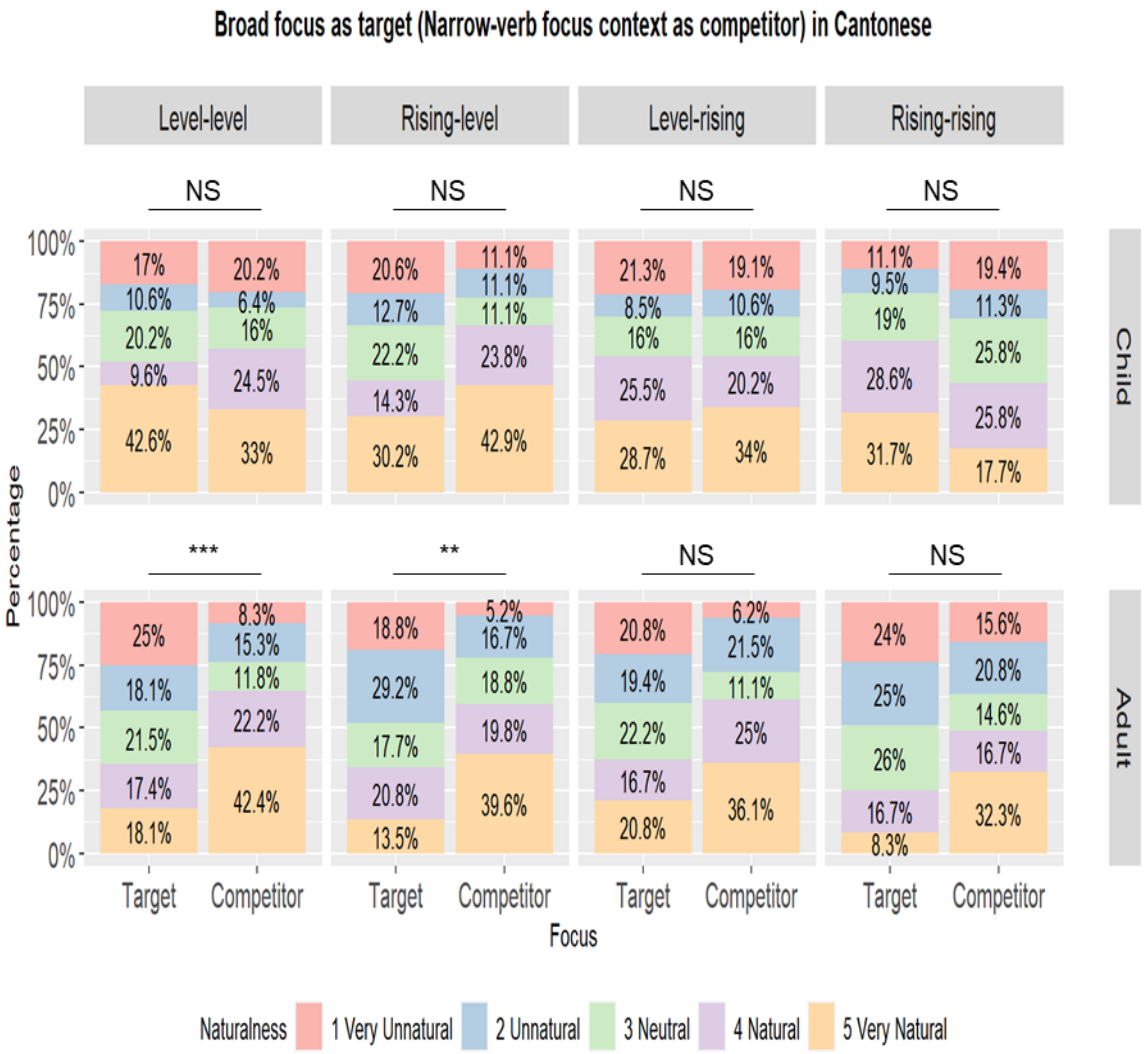


Figure 5.6 Children's performance in Cantonese focus contrasts with different information structures (Narrow-Broad)

Targets refer to pairs with matched broad focus contexts and forms; Competitors refer to pairs with mismatched narrow-verb focus contexts and broad focus forms.

This study also compared the matched broad focus forms in the BF condition with the mismatched broad focus forms in the CB condition. In each comparison, the acoustic realizations of the answers had the same broad focus forms as in the BF and CB conditions, but the preceding question varied in these two focus conditions. In the mismatched CB condition, however, the answering utterance with a broad focus form incongruously followed the lead-in question that elicited a contrastive-verb focus. Figure 5.7 in Appendix

7 presents the focus processing of children (top) and adults (bottom) in congruous broad focus (left) and incongruous broad focus (right). Post-hoc analyses have shown that children equally identified the mismatched broad focus (CB) and the matched broad focus, although the former received insignificantly lower scores for prosodic naturalness. In contrast, adults rated the congruous broad focus as significantly less natural in prosody than the incongruous broad focus in the CB focus condition (*Estimate*=-0.487, *SE*=0.124, *P*<.01), but this variation was not significant in any tonal environments.

The final comparison was made between the matched CV and mismatched NC conditions. These conditions shared the same answers with a contrastive-verb focus, while their preceding discourse differed in information structure. The questions in the mismatched NC condition (competitor) were expected to elicit answers with a narrow-verb focus, but the mismatched answers were produced with a contrastive-verb focus. Figure 5.8 in Appendix 7 presents the responses in the five-point scale for prosodic naturalness obtained from children and adults. Despite the variation in preceding discourse, both children and adults equally comprehended the matched and mismatched contrastive-verb focus as having natural-sounding prosody. Post-hoc analyses revealed no statistically significant differences between these two focus conditions for both groups.

In summary, trilingual children appeared to outperform the adults in distinguishing between congruous and incongruous broad focus forms, particularly when the information structures were intended to elicit narrow-verb or contrastive-verb focus in mismatched pairs. The children were potentially capable of rating matched broad focus as having a more natural prosody compared to the mismatched pairs, although the difference was statistically insignificant. However, trilingual adults exhibited reverse perceptual mappings,

rating the incongruous broad focus as having more natural prosody than congruous prosody. Furthermore, both groups of listeners rated congruous (CV) and incongruous contrastive-verb focus forms (NC) similarly with respect to natural-sounding prosody.

5.4 Discussion

This experiment aims to investigate whether trilingual children have attained a proficient level of decoding various prosodic focus in their native Cantonese language. The research question was addressed through five sub-questions: (1) Can children process congruous prosody, including broad, narrow-verb, and contrastive-verb focus, similarly to adults? (2) Have children acquired an adult-like ability to decode incongruous prosody in these three focus categories? (3) How does the perception of trilingual children and adults in focus processing differ between matched and mismatched pairs with the same informational structure but different prosodic forms? (4) What are the differences in focus processing for trilingual adults and children in comparison between matched and mismatched pairs, where the prosodic realization is the same but the informational structures differ? and (5) What is the impact of lexical tones and audiovisual modality on the processing of focus for native children?

Congruous prosody: In an overview, Cantonese-speaking children exhibited a level of accuracy in the identification of congruous prosody in three distinct focus conditions that was comparable to that of adults. Specifically, they correctly identified narrow-verb (NV) focus and contrastive-verb (CV) focus with an accuracy of approximately 70% of natural prosody. Consistent with adults, children also rated congruous narrow and contrastive focus on the verb as more natural compared to congruous broad focus. However, unlike Cantonese adults who struggle to comprehend broad focus, children were able to rate broad

focus (BF) with an accuracy rate above chance level, albeit with a slight difference. These findings suggest that trilingual children have developed an adult-like strategy for processing focus in congruous prosody, at least in laboratory-based settings. It is likely that Cantonese-speaking children's ability to discern contrastive focus has been acquired at an early age, prior to the age range of the current study. For instance, in a picture-verification Yes/No task, Ge et al. (2022) discovered that Cantonese-speaking children aged 5 to 8 performed with comparable accuracy as adults in correctly identifying the contrastive emphasis on the initial position (i.e., subject) of Cantonese utterances. The present experiment improved our understanding that school-aged trilingual children can process congruous prosodic focus on the medial syntactic position regardless of the focus types (narrow vs. contrastive).

Furthermore, an interesting observation is that children at nine can interpret congruous broad focus with an accuracy above the chance level, whereas Cantonese adults' accuracy is below the chance level. Supposing that the probabilistic assumption on focus processing is confirmed in Cantonese adults, the slight difference between children and adults may imply that children have not yet acquired the probabilistic approach to interpreting focus and instead primarily rely on the bottom-up acoustic signals. Linear discriminant analyses (LDA) reported in Chapter 3.2.2.2 show that the duration of the medial syntactic constituent is statistically shorter for broad focus compared to narrow and contrastive focus. This acoustic parameter can be exploited by native tonal-language children to mark focus at an early age (e.g., Mandarin: Yang, 2017; Yang & Chen, 2018).

Incongruous prosody: The observations of the present study indicate that trilingual children possess a sensitivity similar to trilingual adults in detecting prosodic incongruity

in mismatched focus conditions, although there is a subtle difference in the way that they interpret signal-extrinsic factors during focus processing. Like adults, children demonstrated an ability to differentiate between incongruous broad focus forms (i.e., mismatched NB and CB) and congruous narrow-verb/contrastive-verb focus realizations. Based on the aforementioned findings regarding congruous prosody, it is reasonable to suggest that children have developed a proficient understanding of prosodic cues related to narrow and contrastive focus. This ability allows them to detect prosodic incongruity in utterances that deviate from the expected prosodic patterns for a narrow or contrastive focus.

Despite similarities in detecting prosodic incongruity in mismatched focus conditions, children and adults differ in their dependence on information structures for narrow and contrastive focus. In Experiment 1 (Chapter 3), trilingual adults identified the incongruous broad focus, preceded by the contrastive focus context, more accurately than the narrow focus context in Cantonese. This suggests that trilingual adults expect a greater prosodic emphasis on contrastive focus compared to narrow focus. However, this tendency was not observed in trilingual children, who interpreted the incongruous prosodic forms of broad focus equally following the mismatched information structures of narrow or contrastive focus. Pannekamp et al. (2011) reported that only quasi-adolescents displayed entirely adult-like ERP responses to adequate or inadequate narrow and contrastive focus, while 8-year-old German-speaking children still relied more on prosodic cues than information structures. Consistent with this, the present behavioral study revealed that trilingual children of L1 Cantonese, around nine years old, did not exhibit the advantage of specific contextual information observed in trilingual adults. This finding provides

additional support for Ge et al.'s (2022) proposition that focus contexts have minimal influence on the comprehension of focus among Cantonese children aged between 5 and 8 years, as indicated by reaction time evidence. However, the advantage of contrastive information over new information (narrow focus) in online processing, exhibited in German 8-year-old children in Pannekamp et al.'s (2011) study, was not found in the current behavioral study. Further studies on the neurophysiological or physiological development of focus processing in Cantonese are necessary to gain a better understanding of whether L1 Cantonese trilingual children develop focus types (narrow vs. contrastive) asynchronously.

Furthermore, the reduced dependence of trilingual children on information structure may account for their equivalent sensitivity to broad focus forms, regardless of the preceding contexts. Trilingual children rated congruous and incongruous broad focus forms in Cantonese as equally natural, in which the acoustical form is indistinguishable. In contrast to adults, it is evident that children primarily relied on the prosodic cues themselves rather than their expectations of the information structure conveyed by precursor questions. Thus, it is plausible to suggest that by the age of nine, trilingual children may encounter challenges in acquiring expectations of prior contexts and integrating them with prosodic realizations in Cantonese. There is a prerequisite that children have already acquired the ability to process prosodic violations extending beyond a single sentence's boundaries. This requirement is supported by Pannekamp et al.'s (2011) ERP study, which revealed that German children at the age of five exhibited this ability through brain responses elicited during question-answer dialogues.

Tonal contexts and audiovisual effects: Trilingual children use similar methods to decode prosodic focus surrounded by different tonal contexts, consisting of combinations of rising and level tones. However, the advantage of prosodic focus mapping observed in native adults for continuous rising lexical tones (Tone 2 and Tone 5) did not occur in children, suggesting that dynamic contour tones did not significantly facilitate focus marking for children up to the age of nine. Previous studies on Cantonese focus production have revealed that native adults utilize an expansion of the pitch range on focused syllables when producing utterances with dynamic tones, rather than with other lexical tones, to mark focus (Wu & Xu, 2010). This advantage of dynamic tones was further demonstrated in their subsequent perception tasks that aimed to identify focus positions, which is consistent with the findings of the present study on focus processing. The absence of advantages in contour tones for school-aged children may have resulted from their inclination to avoid extreme answers, such as "very natural" and "very unnatural". Post-hoc analyses revealed that there was no between-group difference in any tonal context for the specific focus category. Furthermore, trilingual children did not rely on additional visual information for perceptual augmentation like adults. This expected phenomenon indicates that the auditory signals alone were sufficient for both native children and adults in the current task, leaving little room for supererogatory visual cues in cognitively demanding tasks.

Chapter 6. Experiment 4: L3 Mandarin focus processing in trilingual children

6.1 Introduction

The Mandarin experiment that trilingual adults completed in Chapter 4 will be replicated with trilingual children in the present study, with the purpose of investigating whether they have developed a native-like ability to comprehend various prosodic focus in L3 Mandarin with multimodal inputs. The trilingual children involved in this study have Hong Kong Cantonese as their first language and have subsequently learned English and Mandarin as their second and third languages within the multilingual community of Hong Kong. This study aims to deepen our understanding of how school-aged children from multilingual communities acquire the ability to decode prosodic in tonal language.

Several studies have examined the comprehension of prosodic focus in tonal languages like Mandarin (Chen, 1998; Chen, 2022; Chen et al., 2019; Tang et al., 2022), revealing that native speakers of tonal languages typically develop the ability to interpret focus at a later age than children who speak Germanic or Romance languages.. Specifically, a cross-linguistic study investigating focus interpretation among English, French, and German revealed that three-year-old children can interpret contrastive focus on the subject or object similarly to adults (Szendrői et al., 2018). In contrast, a similar picture-matching paradigm in Szendrői et al. (2018) conducted with Taiwan Mandarin speaking children aged between three and five years old showed that they are not able to interpret contrastive focus on the subject with adult-like proficiency (Chen, 2022; Chen et al., 2019). Earlier research on Taiwan Mandarin children revealed that even children at the

age of thirteen cannot discriminate given information and new information with the same ability as adults (Chen, 1998).

In addition, a recent study employed the "visual-world" paradigm to investigate Mandarin focus processing by testing whether adequate prosody of contrastive focus facilitates children's identification of referents (Tang et al., 2022). Their findings revealed that Mandarin-speaking children exhibit adult-like sensitivity to congruous prosodic forms of contrastive focus at the age of seven, and they acquire the adult-like ability to form-function mapping in incongruous prosody at the age of ten. However, the existing research on processing Mandarin prosodic focus mainly focuses on the developmental trajectory of native Mandarin-speaking children, and it remains uncertain whether Cantonese learners of Mandarin, at school-age, can acquire a native-like ability to map between focus categories and prosodic forms. This study aims to investigate this issue by examining the following research questions:

(1) Can trilingual children comprehend congruous prosodic focus in their L3 Mandarin with the same proficiency as Mandarin-speaking peers?

(2) Can trilingual children exhibit a native-like ability in processing incongruent prosodic focus in their L3 Mandarin?

(3) Can trilingual children differentiate between two sets of Mandarin focus conditions, one congruent and the other incongruent, similar to native children?

(4) To what extent are trilingual children's processing of focus in their L3 Mandarin influenced by tonal context and additional visual cues?

The hypothesis of this study is that trilingual children at school age have not developed a native-like ability to process focus in Mandarin dialogues, regardless of

prosodic congruity. In particular, native Cantonese-speaking children may exhibit a smaller discrepancy compared to their Mandarin-speaking peers, consistent with their performance in their native Cantonese (Ge et al., 2022). However, their accuracy in incongruous focus conditions may be much lower than that of native Mandarin-speaking children. Moreover, the decoding process of non-native prosodic focus is challenging due to the interaction between lexical tones and prosodic modulations. This study predicts that trilingual children will experience difficulties in detecting prosodic focus when the tonal context is level tones (Tone 1). Trilingual children are likely to rely on additional visual information during multimodal speech, resulting in better accuracy in the audio-visual modality than in the audio-only modality.

6.2 Method

6.2.1 Participants

Forty-seven trilingual children who participated in Experiment 3 of their L1 Cantonese were also recruited in Experiment 4 of L3 Mandarin. The mean age of the eighteen girls and twenty-nine boys is 9.20 ± 0.93 years. They are native speakers of Cantonese and learn English and Mandarin sequentially as their second and third languages. The learning experience of Mandarin started at the mean age of 4.72 ± 1.65 years in kindergarten, significantly later than the age of the beginning of learning English (age 3.69 ± 1.64 years, P<.01). Parents helped in evaluating the children's language competence in Mandarin using a 5-Likert scale. For example, 1 - 5 points stood for fluency in Mandarin from the lowest to the highest. Collected information showed that the scores of listening, speaking, reading, and writing abilities were respectively 3.87 ± 0.71 points, 3.61 ± 0.77 points, 3.64 ± 0.82 points, and 3.32± 0.84 points.

Besides, we recruited thirty-one children who are native speakers of Mandarin as the comparison group in the study of Mandarin focus processing. In the comparison group, the children are thirteen girls and eighteen boys at a mean age of 9 ± 0.92 years (range, 7.67-11). They were born and live in Beijing or nearby areas and speak Mandarin with their parents at home. All participants in the comparison group are bilinguals who started to learn English at the mean age of 5 ± 1.48 years at school. Children in both groups self-reported with nothing impairment in speaking or listening. All participants' parents have signed the informed consent of the research and received the payment for participation once they completed the online tasks. As mentioned in previous chapters, the human ethics committee of the Hong Kong Polytechnic University has approved the present research (HSEARS20200919002)

6.2.2 Materials and design

The materials and stimuli design used in Experiment 2 (Chapter 4) were also exploited in the current experiment. The detailed target sentences and audiovisual recordings are presented in Chapters 4.2.2.1 and 4.2.2.2. The design of the Mandarin stimuli used in Experiments 2 and 4 was described in Chapter 4.2.2.3.

6.2.3 Procedures

The practice session and experimental task followed the Mandarin experiment for adults in Experiment 2 (Chapter 4). Native Cantonese and Mandarin children who passed the practice sessions with an accuracy of 60% or above were instructed to complete the following experimental sessions. Forty-seven trilingual children with Cantonese as their native language and thirty-one bilingual children with Mandarin as their native language

participated in the perceptual experiment on E-prime Go. To prevent task effects, the trilingual Cantonese children's participation in the Cantonese experiment (Experiment 3) or the Mandarin experiment (Experiment 4) was counterbalanced.

6.2.4 Data analyses

The study fitted all responses from listeners who were both non-native Mandarin speakers and native Mandarin speakers to the ordinal regression model in R (R Core Team, 2022). Specifically, the dependent variable, responses of the 5-Likert prosodic naturalness scales, were examined on the cumulative link mixed effects models (*CLMM*) by the *clmm* function using the ordinal package (Christensen, 2015). The fixed effects comprised *Group* (Cantonese, Mandarin), *Focus* (BF, NV, CV, BN, BC, NC), *Tones* (TL, TR, TF), and *Modalities* (AO, AV). The random effects involved *Subjects* and *Trials*. For a more in-depth analysis of the perception of different focus types within and across groups, post-hoc analyses using the *emmeans* package (Lenth R, 2023) were carried out.

6.3 Results

The likelihood ratio test demonstrated that there was no significance in the interaction of *Group* and *Modality* ($\chi^2(3) = 2.795$, $P = 0.424$), indicating that non-native and native children of Mandarin listeners similarly performed between audio-only (AO) and audio-video (AV) modalities. Thus, the optimal regression model was fitted to the three-way interaction among *Group*, Focus, and *Tone*, with *Subject* and *Trial* as random factors. In the ordinal regression model, the fixed factors *Group* ($\chi^2(1) = 0.75$, $P= 0.389$) and Tone ($\chi^2(2) = 3.29$, $P= 0.193$) were insignificant, whereas *Focus* ($\chi^2(5) = 329.29$, $P<.001$) was significant. As for two-way interactions, significances were found in the interaction of *Group* and *Focus* ($\chi^2(5) = 55.44$, $P<.001$) rather than the interaction of *Group* and *Tone*

($\chi^2(2) = 0.65$, $P = 0.723$). The three-way interaction among *Group*, *Focus*, and *Tone* ($\chi^2(10) = 30.15$, $P < .001$) was significant in the perception in Mandarin focus processing by native or non-native children. Chapter 6.3.1 details the perceptual mapping outcomes of congruous Mandarin prosody for two groups of children, while Chapter 6.3.2 summarizes their perception in incongruous prosody. Chapter 6.3.3 examines focus contrasts, which include congruous and incongruous focus conditions, and share either the same information structure or the identical prosodic form.

6.3.1 Focus processing of congruous prosody in L3 Mandarin by trilingual children

In this section, I explore the comprehension of prosody in broad focus (BF), narrow-verb focus (NV), and contrastive-verb focus (CV) in congruous dialogue pairs by both native and non-native Mandarin-speaking children. The dialogue pairs for each focus condition can be found in Table 4.4 in Chapter 4.2.2.3. The response distributions for the three types of matched focus situations, rated on a scale of five degrees of prosodic naturalness, are displayed in Figure 6.1. The perceptual patterns of non-native children whose L1 is Cantonese are shown on the left stacked bar, while those of native children are presented on the right in each focus type. Consistent with previous experiments, responses categorized as "natural" and "very natural" were considered indicative of congruous prosody, while responses classified as "unnatural" and "very unnatural" were regarded as indicating incongruous prosody. It is noteworthy that "neutral" responses, indicating ambiguity, were not discussed in the study.

In general, both native and non-native children demonstrated similar comprehension of the matched prosodic focus in Mandarin dialogues. Specifically, for broad focus (BF), native Cantonese-speaking children evaluated 57.5% of the pairs as

having natural prosody, while native Mandarin-speaking children assessed 59.2% as having natural prosody. The perceptual accuracy of natural prosody under narrow-verb focus (NV) was 71.7% for Cantonese children and 87.1% for Mandarin children. Additionally, for contrastive-verb focus (CV), Cantonese children judged 69.9% of responses as having natural prosody, whereas Mandarin children judged 82.8% of responses as having natural prosody. Statistical results confirmed that native Mandarin-speaking children performed insignificantly better than non-native children in comprehending matched focus conditions.

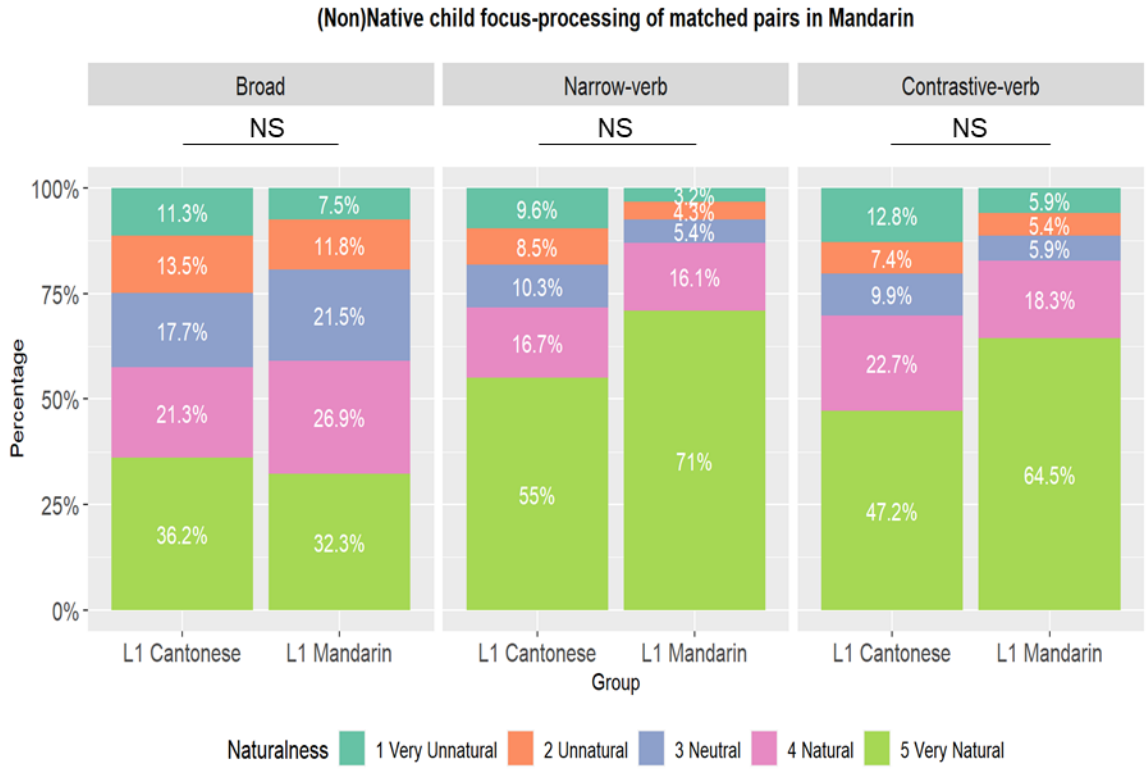


Figure 6.1 Native and non-native children's performance in congruous Mandarin prosody

The comparison of matched focus conditions showed that both groups of participants distinguished between broad focus and narrow-verb focus, with the latter receiving significantly higher scores in terms of perceived naturalness (L1 Cantonese:

Estimate=-0.763, *SE*=0.162, *P*<.001; L1 Mandarin: *Estimate*=-1.587, *SE*=0.213, *P*<.001). In addition, both groups perceived the broad focus as less natural than the contrastive-verb focus, but this was only significant for Mandarin children (*Estimate*=-1.361, *SE*=0.215, *P*<.001). The effects of lexical tonal contexts were only significant for level tones in discriminating broad focus and contrastive-verb focus for Mandarin children (*Tone:* Level, *Estimate*=-2.376, *SE*=0.552, *P*<.01).

In summary, both groups of children exhibited similar processing of congruous prosody. Specifically, they rated broad focus as less natural compared to narrow-verb focus in a similar manner. However, significant differences between broad focus and contrastive-verb focus were only observed for native speakers.

6.3.2 Focus processing of incongruous prosody in L3 Mandarin by trilingual children

In the processing of incongruent prosody, mismatched broad focus in the NB and CB conditions and mismatched contrastive-verb focus in the NC condition were addressed. Figure 6.2 illustrates the overall responses of the two groups of children to the three types of mismatched focus conditions. Both groups of children exhibited similar perceptual patterns in the mismatched focus conditions, consistent with their performance in the matched focus conditions. The accuracy of identifying the mismatched broad focus in the NB focus condition was 30.2% for Cantonese listeners and 34.9% for Mandarin listeners. In the mismatched CB focus condition, Cantonese and Mandarin children correctly identified 39% and 43.6% of unnatural prosody responses, respectively. The percentage of unnatural responses in the mismatched contrastive-verb focus (NC) was 17% for L1 Cantonese children and 6.5% for L1 Mandarin children. Post-hoc analyses did not find any

significant differences between the two groups of listeners in processing any of the three types of mismatched focus conditions.

Furthermore, Cantonese children and Mandarin adults demonstrated similar discrimination between the mismatched NB and mismatched CB conditions. Meanwhile, both groups displayed different comprehension of the mismatched contrastive-verb focus (NC) from the mismatched broad focus (NB and CB). They rated the mismatched contrastive-verb focus in the NC condition as significantly more natural prosody than the mismatched broad focus in the NB (*Group*: Cantonese, *Estimate*=0.979, *SE*=0.165, *P*<.001; *Group*: Mandarin, *Estimate*=2.28, *SE*=0.216, *P*<.001) and CB (*Group*: Cantonese, *Estimate*=1.383, *SE*=0.164, *P*<.001; *Group*: Mandarin, *Estimate*=2.685, *SE*=0.218, *P*<.001).

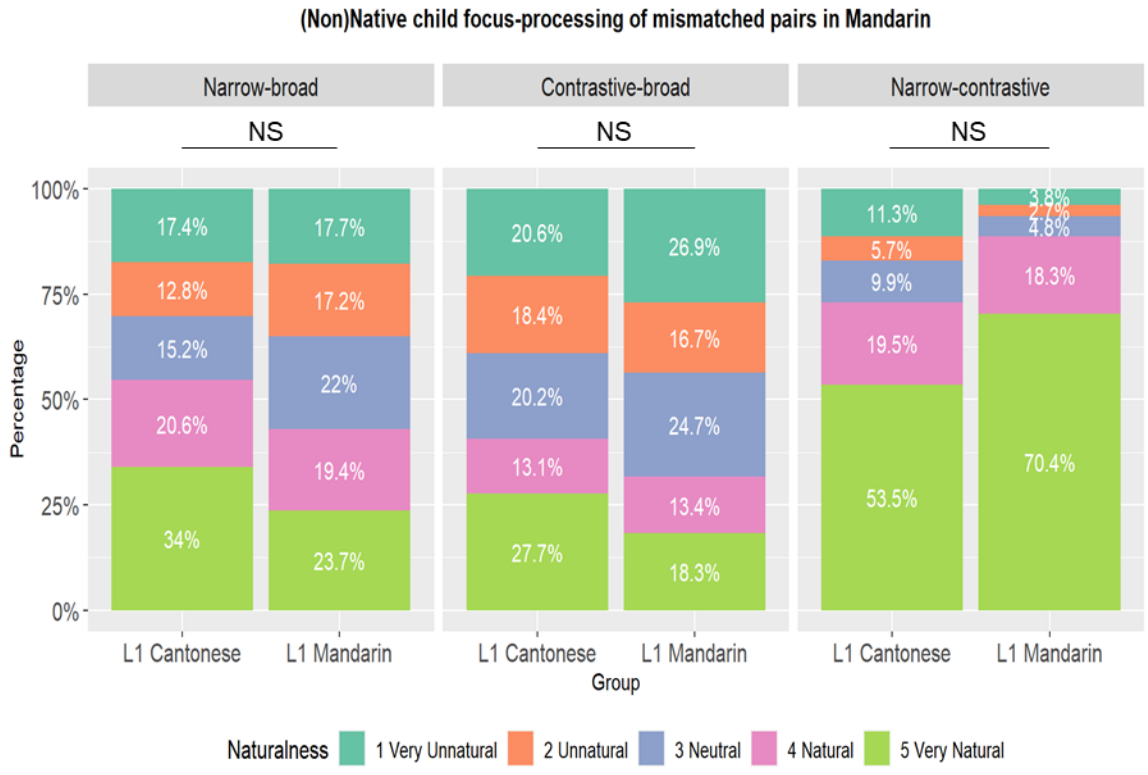


Figure 6.2 Native and non-native children's performance in incongruous Mandarin prosody

The findings of this study suggest that there is no significant difference in perceptual accuracy between native and non-native children in the mismatched focus conditions. However, non-native children showed a slightly lower accuracy in identifying incongruous broad focus, although this difference was not statistically significant when compared to their native counterparts. When presented with narrow-verb focus as the information structure, both native and non-native groups perceived the mismatched pairs with a contrastive-verb focus as natural prosody.

6.3.3 Comparison between congruous and incongruous focus conditions in L3 Mandarin by trilingual children

Consistent with Experiment 2 for trilingual adults, this section compares the responses of trilingual children to focus contrasts (including a matched pair and an unmatched pair) from two perspectives. Firstly, it analyzes the two Q-A dialogue pairs that consist of the same precursor questions but have different prosodic realizations in the NV vs. NB, CV vs. CB, and NV vs. NC conditions, as illustrated in Chapter 6.3.3.1. Secondly, this section is based on pairs that consist of two different discourses, but the following answers have the same prosodic form. Chapter 6.3.3.2 compares focus contrasts in this regard, such as the BF vs. NV, BF vs. CB, and CV vs. NC conditions. Each focus condition has been illustrated in Table 4.4 in Chapter 4.2.2.3 (Experiment 2).

6.3.3.1 Same information structure-different prosodic forms

This section presents the contrastive results of focus conditions, where various prosodic realizations in the answers followed a specific information structure in the preceding discourse. In the comparison between NV and NB conditions, the same information structure was shared, while the NB focus condition featured a mismatched broad focus that

answered the questions unnaturally. Figure 6.3 shows the perceptual mapping results of native Cantonese children on the top and native Mandarin children on the bottom. For each tonal context, the left stacked bar represents the congruous NV focus condition (target), and the right one indicates the incongruous NB focus condition (competitor). As the Y-axis increases, the degree of prosodic naturalness decreases, and the proportions of each scale are displayed.

For the matched targets (NV), native Cantonese children accurately comprehended the target narrow-verb focus as natural prosody with an accuracy of 67.8% in level tones, 72.9% in rising tones, and 74% in falling tones. The ratios of matched narrow-verb focus were 86.6%, 84.9%, and 90% in the three tonal contexts for native Mandarin children. For the mismatched competitors (NB), the identification accuracy of unnatural prosody was 32.2%, 26%, and 32.3% in surroundings of level, rising, and falling lexical tones for Cantonese children. The listeners of native Mandarin-speaking children recognized 41.7%, 16.7%, and 48.3% of mismatched pairs as unnatural prosody in the level, rising, and falling tonal environments.

Post-hoc results confirmed that Cantonese children were able to differentiate between the matched narrow-verb focus (NV) from the mismatched broad focus (NB), assigning a significantly higher rating to the former in terms of prosodic naturalness (*Estimate*=0.994, *SE*=0.17, *P*<.001). When the impact of lexical tones was analyzed, it was found that the difference between NV and NB conditions was insignificant across all tonal contexts. Additionally, native Mandarin-speaking children also exhibited substantial dissimilarities between the targets and competitors (*Estimate*=2.278, *SE*=0.219, *P*<.001) and achieved remarkable proficiency in identifying targets in level and falling tonal

surroundings (*Tone:* Level, *Estimate*=2.57, *SE*=0.405, *P*<.001; *Tone:* Falling, *Estimate*=3.096, *SE*=0.392, *P*<.001).

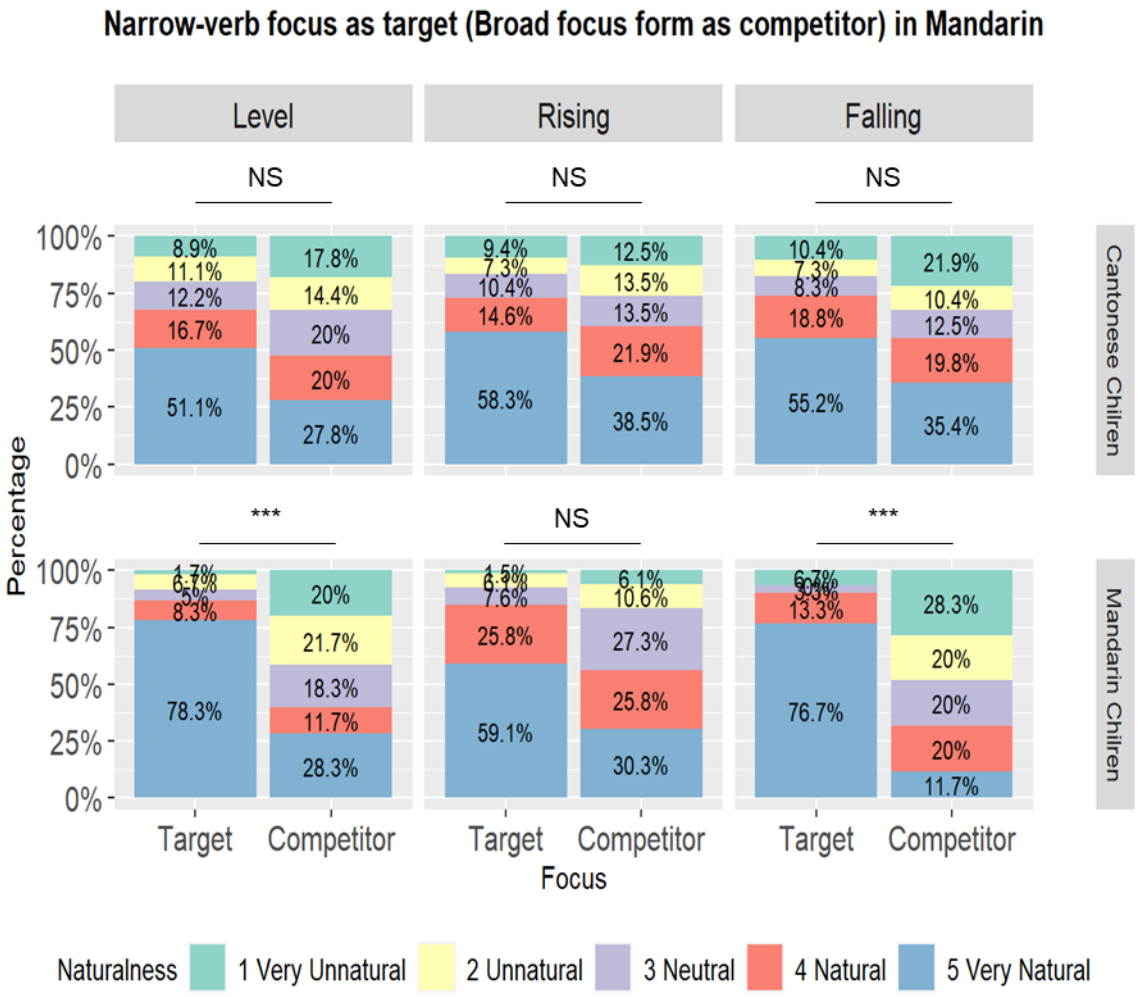


Figure 6.3 Children's performance in Mandarin focus contrasts with different prosodic forms (Narrow-Broad).

Targets refer to pairs with narrow-verb focus context and matched focus forms; Competitors refer to pairs with narrow-verb focus context and mismatched broad focus forms.

The second comparison aimed to analyze the results of perceptual mapping between the matched contrastive-verb focus (CV) and the mismatched broad focus (CB) following the same discourse with a contrastive-verb focus context. Both the CV and CB focus conditions were designed to elicit a contrastive-verb focus, whereas the answer in the CB

condition incongruously answered with a mismatched broad focus form. Figure 6.4 illustrates the perceptions of Cantonese children (at the top) and Mandarin children (at the bottom) regarding the two focus conditions in each tonal context.

For targets (CV), native Cantonese-speaking children rated 73.9%, 67.8%, and 67.7% of matched pairs with contrastive-verb focus as having natural prosody in level, rising, and falling tonal contexts, respectively. Native Mandarin-speaking children rated 98.3%, 86.7%, and 65.1% in the same tonal contexts. For competitors (CB), Cantonese children comprehended the mismatched broad focus as unnatural prosody in level, rising, and falling tonal contexts with ratios of 42.7%, 40%, and 34.4%, respectively. Native Mandarin listeners correctly detected the unnatural prosody with an accuracy rate of 50%, 36.6%, and 43.9% in the three tonal environments.

Post-hoc analyses comparing the two pairs of CV and CB conditions revealed that Cantonese children significantly distinguished between the matched and mismatched pairs (*Estimate*=1.104, *SE*=0.165, *P*<.001), particularly in the tonal contexts of the level and rising lexical tones (*Tone* Level: *Estimate*=1.43, *SE*=0.294, *P*<.001; *Tone* Rising: *Estimate*=1.14, *SE*=0.276, *P*=0.017). Similarly, for native Mandarin children, the matched contrastive-verb focus received considerably higher evaluations compared to the mismatched broad focus (*Estimate*=2.457, *SE*=0.223, *P*<.001). The observed significance was substantial in the context of level and rising lexical tones (*Tone* Level: *Estimate*=3.847, *SE*=0.441, *P*<.001; *Tone* Rising: *Estimate*=2.288, *SE*=0.372, *P*<.001).

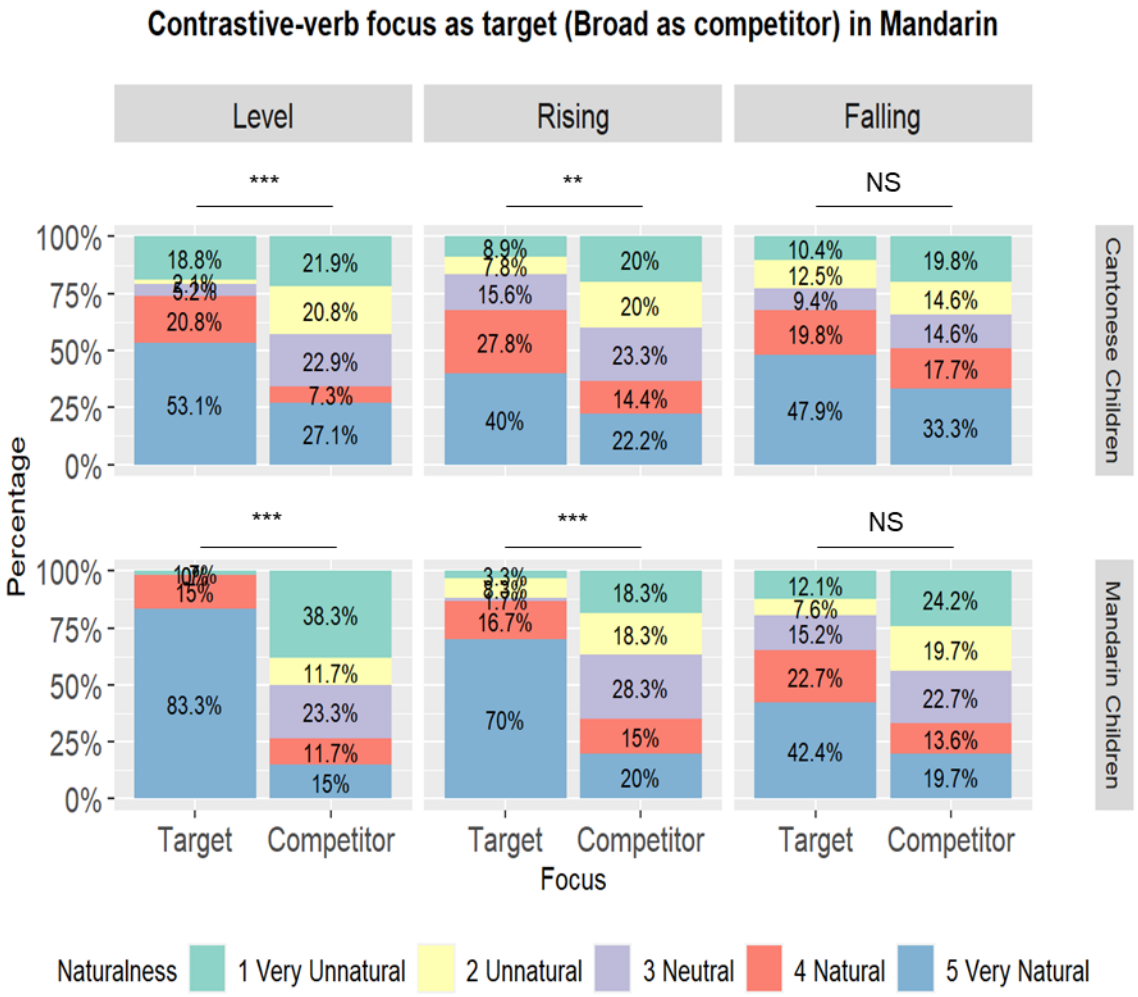


Figure 6.4 Children's performance in Mandarin focus contrasts with different prosodic forms (Contrastive-Broad).

Targets refer to pairs with contrastive-verb focus context and matched focus forms; Competitors refer to pairs with contrastive-verb focus context and mismatched broad focus forms.

The final comparison compared the matched pairs with a narrow-verb focus (NV) to the mismatched pairs with a contrastive-verb focus (NC). Specifically, each pair of NV and NC conditions used the same precursor questions to elicit a narrow focus on the verb. In the mismatched NC focus condition, the answers were inappropriately produced with a contrastive-verb focus. Figure 6.5 in Appendix 8 presents the responses obtained from native Cantonese-speaking children (top) and native Mandarin-speaking children (bottom)

in the contexts of level, rising and falling lexical tones, for both the matched (target) and mismatched (competitor) focus conditions. The statistical results confirmed that both groups of children processed the matched narrow focus and the mismatched contrastive focus on the verb equally, when the precursor questions elicited the narrow-verb focus identically.

In conclusion, the comparative results revealed that non-native children, like their native peers, could distinguish between mismatched broad focus forms and matched narrow-verb/contrastive-verb focus realizations. Additionally, both groups rated the mismatched pairs with contrastive-verb focus as having natural prosody, when the target was utterances with narrow-verb focus.

6.3.3.2 Different information structures-the same prosodic form

This section compared children's perception on the same prosodic forms when the preceding discourse was designed to elicit different types of prosodic focus in their responses. The first comparison was between the matched broad focus in the BF condition and the mismatched broad focus in the NB condition. The pairs of BF and NB conditions had two different leading questions that were expected to elicit broad focus and narrow-verb focus in responses, respectively. Both dialogues had answers with broad focus, which matched the BF condition, but were mismatched with the NB focus condition. Figure 6.6 illustrates the perceptual responses of the matched BF focus condition (target, on the left) and the mismatched NB focus condition (competitor, on the right).

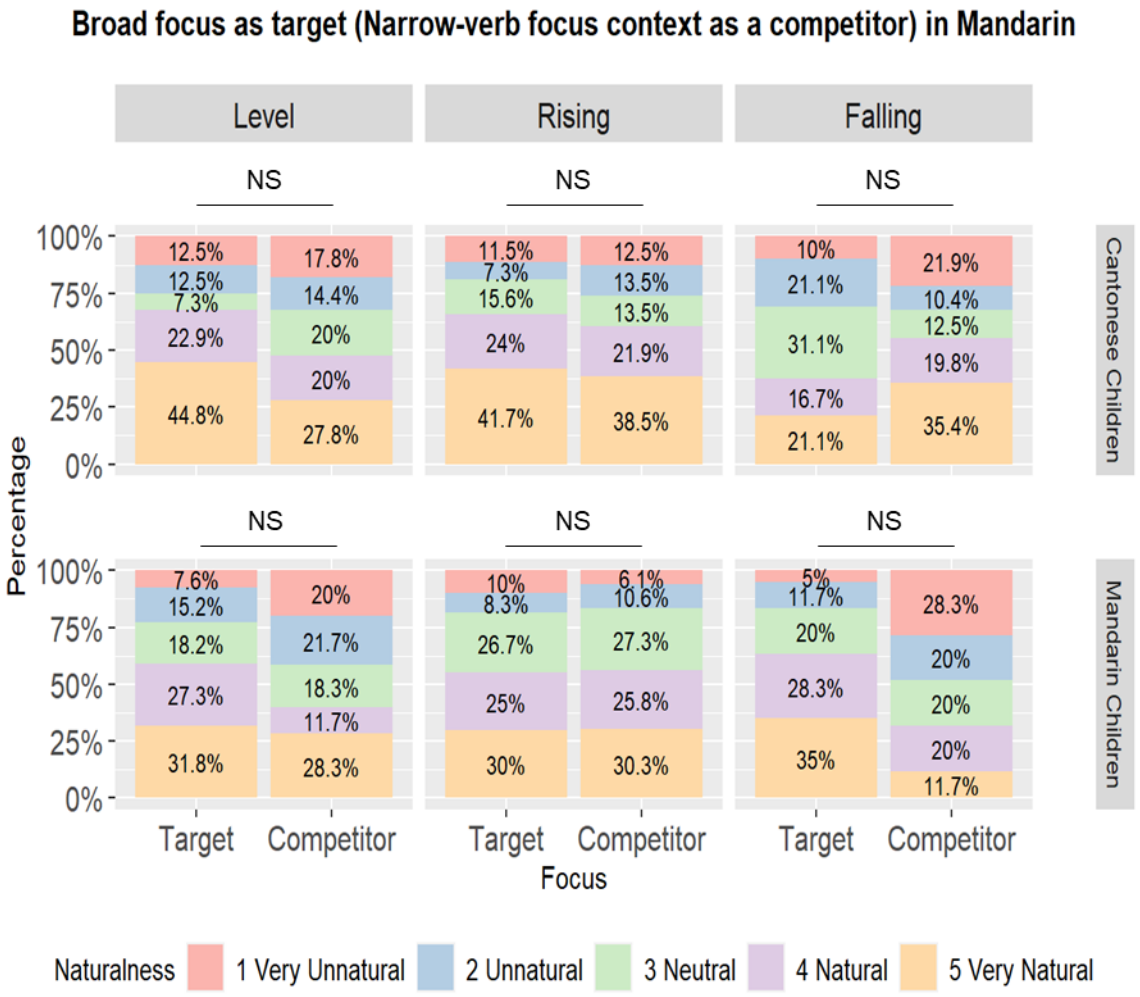


Figure 6.6 Children's performance in Mandarin focus contrasts with different information structures (Narrow-Broad)

Targets refer to pairs with matched broad focus contexts and forms; Competitors refer to pairs with mismatched narrow-verb focus contexts and broad focus forms.

Cantonese children evaluated the matched broad focus in the BF conditions (targets) with 67.7%, 65.7%, and 37.8% accuracy in the level, rising and falling tonal contexts, respectively. The accuracy among Mandarin children was 59.1%, 55%, and 63.3% in the respective tonal surroundings. The perceptual mapping results of the incongruous broad focus in the NB condition have been reported in Chapter 6.3.3. The post-hoc results showed that Cantonese children processed matched and mismatched broad focus similarly in BF and NB focus conditions. In contrast, Mandarin children perceived the matched broad

focus as more natural prosody compared to the mismatched broad focus in the NB condition (*Estimate*=0.69, *SE*=0.189, *P*=0.014), despite without significant influence of tonal context.

The current study also examined how children process broad focus forms when preceded by mismatched information structure of contrastive-verb focus (CB). The pair of BF and CB conditions shared the same answer with an acoustic identical broad focus. However, the precursor question in the mismatched CB condition was designed to elicit a contrastive-verb focus, mismatching with the broad focus form in answers. Figure 6.7 in Appendix 8 displays the comprehension of broad focus in the matched BF (target) and mismatched CB (competitor) conditions between native Cantonese (top) and Mandarin (bottom) speaking children. The post-hoc analysis revealed that Cantonese-speaking children differentiated between congruent broad focus and incongruent broad focus, giving higher evaluations of prosodic naturalness to the former (*Estimate*=0.635, *SE*=0.16, *P*<.001). This finding was consistent with the Mandarin-speaking children's results (*Estimate*=1.096, *SE*=0.191, *P*<.001). The results indicated no significant differences in a tonal context upon examination of the three-way interaction among *Group*, *Focus,* and *Tone*.

Finally, In the comparison of the CV and NC focus conditions, both pairs of answers had an identical contrastive focus on the verb. The precursor question in the matched CV condition was intended to elicit a contrastive-verb focus, while the precursor question in the mismatched NC condition was expected to produce a narrow-verb focus. In Appendix 8, Figure 6.8 compares the target (CV) and competitor (NC) in level, rising, and

falling tonal contexts. Post-hoc results showed that Cantonese and Mandarin-speaking children evaluated matched and mismatched contrastive-verb focus similarly.

In conclusion, the study found that Cantonese-speaking children were unable to distinguish incongruous broad focus when the mismatched information structure evoked a narrow-verb focus, while Mandarin-speaking children demonstrated the ability to discriminate between the BF and NB conditions. However, both groups were able to discern the differences between congruous and incongruous broad focus when a contrastive-verb focus was elicited in a mismatched prior context. Moreover, both groups rated the prosodic forms of incongruous contrastive-verb focus in the NC condition as equally natural as the congruous prosodic realization in the CV condition.

6.4 Discussion

This study investigated whether school-aged native Cantonese children have acquired a native-like pattern for focus processing in their third language, Mandarin. To this end, five sub-questions were posed: (1) Do non-native Mandarin-speaking children process matched focus categories (broad, narrow-verb, and contrastive-verb focus) in a manner consistent with native child speakers? (2) Are non-native Mandarin-speaking children capable of decoding the incongruous prosody of these three focus categories similarly to their native counterparts? (3) How do native and non-native Mandarin-speaking children perform in processing focus when presented with matched and mismatched Q-A pairs, which share the same information structure but differ in their prosodic forms? (4) How do they perform in focus processing when comparing matched and mismatched pairs with the same prosodic realization but preceded by different information structures? (5) To what extent

do lexical tones and visual information impact the focus decoding ability of both native and non-native children?

Congruous prosody: Native Cantonese-speaking children have been found to exhibit comparable abilities to native Mandarin-speaking children in decoding the prosodic focus of congruous prosody in their third language, Mandarin. Specifically, both groups have acquired the ability to accurately detect narrow and contrastive focus on the verb with an accuracy exceeding 70%, as well as interpret broad focus of natural prosody with an accuracy above the chance level. Additionally, both groups rated congruous broad focus as less natural than narrow focus on the verb, a finding consistent with the perceptual patterns observed in native and non-native adults. Furthermore, like adults, children have demonstrated less certainty in processing entirely new information elicited by the question "What happened?" as compared to narrowly new information on the verb evoked by questions like “What did A do to B?”. These findings suggest that both native and non-native Mandarin-speaking children, at least by the age of nine, have shown adult-like proficiency in decoding the prosodic patterns of congruous Mandarin focus, thus filling a gap in the literature regarding the extent to which school-aged children, both native and non-native, capitalize on the linguistic and para-linguistic factors involved in decoding focus in tonal languages.

However, there were notable variations between Cantonese and Mandarin children in their acceptance of congruous broad focus (BF) and contrastive-verb focus (CV) in Mandarin. Specifically, Cantonese children rated them as conveying similar prosodic congruity, whereas Mandarin children evaluated the broad focus as less natural than the contrastive focus. Considering that Cantonese children rated congruous narrow focus as

more natural than broad focus, they exhibited different accuracy levels when distinguishing between narrow-verb and contrastive-verb focus with congruous prosody. One potential explanation for this finding is that they have higher expectations regarding the prosodic naturalness for the information structure of contrastive focus. Nevertheless, it is crucial to note that narrow focus and contrastive focus positioned syntactically similarly are acoustically indistinguishable. This holds true not only for the stimuli used in the present study but also for the focus production of native peers, as reported in previous studies (Yang, 2017; Yang & Chen, 2018). As such, if it is assumed that Cantonese children possess the capacity to anticipate information structures, it is likely that they expect greater prosodic salience when encountering contrastive information compared to non-contrastive new information (narrow focus) on the medial focal position in Mandarin. Consequently, Cantonese children may demonstrate a propensity to rely on contextual contrastive information when processing focus in their L3 Mandarin, even in situations where this feature is not present in their native Cantonese (See Experiment 3).

Incongruous prosody: Both groups of children demonstrated similar proficiency levels in identifying unnatural prosody, with an accuracy rate of approximately 30%. This suggests that Cantonese children have acquired a native-like competence in processing prosodic violations with incongruities. Moreover, the children of both groups demonstrated comparable proficiency in discerning congruous narrow-verb/contrastive-verb focus (NV/CV) and incongruous broad focus prosodic realizations (NB or CB) when preceded by the same information structures, thereby addressing sub research question (3) introduced in Chapter 6.4. This outcome was anticipated, given that the acoustic features distinguishing broad focus from narrow/contrastive focus were statistically evident in the

LDA analysis of the stimuli employed in the investigation (Chapter 4.2.2.2). These findings indicate that Cantonese children have demonstrated proficient use of acoustic cues in acquiring L3 Mandarin prosodic focus and have exhibited a well-developed understanding of the identification of narrow and contrastive focus in L3 Mandarin, closely resembling their performance in their L1 Cantonese. Therefore, they showed sensitivity to unexpected and incongruous prosodic cues conveyed by broad focus when the prior discourse was to elicit narrow (NB) or contrastive focus (CB).

Regarding sub research question (4), Cantonese children have exhibited an incomplete ability to integrate their knowledge of information structures, if any, with acoustic forms of broad focus in mismatched NB and CB focus conditions. In contrast, native Mandarin-speaking children have demonstrated accurate proficiency in rating incongruous broad focus forms (NB and CB conditions) as being less natural than congruous broad focus forms (the matched BF condition), independent of preceding information structures. That is, Cantonese children could only make the distinction when the information structure conveys a contrastive focus rather than a narrow focus in terms of incongruous prosody. In the case of a mismatched context eliciting narrow focus, Cantonese children rated the following incongruous broad focus forms (NB) as equally natural as the congruous broad focus (BF). It could be posited that the challenges encountered in associating the prosodic form of broad focus with matched (broad focus) or mismatched (narrow verb) focus categories may stem from the influence of focus processing from their L1, considering that the identical cohort of children struggles to differentiate between BF and NB focus conditions in their native Cantonese (Experiment 3). The findings from Experiment 1 and Experiment 2 revealed that trilingual adults, being

native speakers of Cantonese, encountered difficulties in focus mapping across these specific focus conditions, both within their L1 Cantonese and their L3 Mandarin. Recall that trilingual adults achieved an accuracy of around 40% in detecting congruous broad focus in L3 Mandarin, as shown in Experiment 2, while native Mandarin speakers scored above 70% accuracy. For trilingual adults, the failure to differentiate between BF and NB focus conditions is caused by two mistaken biases: one against broad focus and the other towards narrow focus. However, in Experiment 4, both native and non-native Mandarin-speaking children rated congruous broad focus (BF) as natural prosody with an accuracy of nearly 60%. Therefore, the inaccurate discrimination between BF and NB focus conditions by trilingual children is mainly due to their insensitivity to mismatched broad focus forms preceded by the information structure of narrow focus (NB). The mistaken bias towards narrow focus seems to be limited to native Cantonese speakers, including both school-aged children and adults, rather than Mandarin speakers.

Similar to their native-speaking peers, however, native Cantonese children can accurately differentiate between the incongruous broad focus (CB) and congruous broad focus (BF), with the lower ratings of prosodic naturalness for broad focus forms incongruously elicited by contrastive focus (CB). Since Cantonese children did not display the same level of proficiency as Mandarin children in differentiating between broad focus (BF) and narrow focus (NB) conditions, it is evident that they exhibit distinct variations in processing incongruous prosody that follows prior discourse functions associated with narrow focus and contrastive focus. One possible explanation for the superior ability of Cantonese children to decode contrastive focus compared to narrow focus is their pre-existing expectation of more pronounced prominence associated with contrastive focus in

L3 Mandarin. This anticipation of salient prosody subsequently contributes to their heightened accuracy in identifying the incongruous form of broad focus within the context of the contrastive focus (CB) condition. Moreover, the sensitivity of Cantonese children towards prosodic incongruity elicited by the information structure of contrastive focus was evident not only in instances of incongruous prosody (CB) but also in cases of congruous prosody (CV), as previously discussed. The developmental advantage in decoding contrastive focus exhibited by trilingual children at the age of nine has been found to be comparable to that observed in German-speaking children at eight years old in their native language (Pannekamp et al., 2011). Interestingly, trilingual children did not show a developmental advantage in employing contrastive focus context in their L1 Cantonese, whereas trilingual adults exhibited an ability to anticipate contextual contrastive information in their native Cantonese. Hence, it is plausible to propose that, by the age of nine, trilingual children have not fully acquired the adult-like or native-like ability to use their prior knowledge of information structure in conjunction with acoustic forms both in their L1 Cantonese and L3 Mandarin. Consequently, additional language exposure and further development are deemed necessary for trilingual children in the school-age range to attain these skills more comprehensively.

Tonal contexts and audiovisual effects: Native Cantonese-speaking children displayed a preference for neutral options, which suggested that the effects of lexical tones only appeared in specific focus conditions, such as discriminating between congruous (CV) and incongruous (CB) contrastive-verb focus (refer to Figure 6.4 in Chapter 6.3.3). The results indicated significant differences between congruous and incongruous prosody in the tonal surroundings of level and rising tones, consistent with the perceptual patterns in tonal

contexts observed in Mandarin children. The fact that Cantonese children demonstrated native-like proficiency only in comprehending contrastive focus further suggests that trilingual children have an advantage in processing contrastive focus over narrow focus in their L3 Mandarin.

Additionally, the study found no audiovisual effects on focus processing by either native or non-native school-aged children. Previous studies have found that school-aged children are unable to efficiently utilize the additional visual information in focus perception (Rapin & Ménard, 2019; Ménard et al., 2006). The original study on the McGurk effect found evidence that preschool children (ages 3-5) and school-aged children (ages 7-8) rely less on visual information when identifying English syllables than adults do (McGurk & MacDonald, 1976). Moreover, researchers found that the McGurk effect, whereby visual information influences listeners' perception of auditory cues when they are discrepant, shows a developmental trend in English speakers but not in Japanese speakers (Sekiyama et al., 2003). Given that trilingual adults did not exhibit the McGurk effect in processing Mandarin focus, it is plausible that trilingual children's reduced reliance on visual cues is not solely due to age. In this regard, Cantonese-speaking listeners resemble Japanese listeners, who exhibit weaker McGurk effects not only in childhood but also in adulthood. Furthermore, this finding adds to our understanding of how children process multimodal information during cognitively demanding tasks. The current data reveal the limited role of supplementary visual cues in focus decoding.

Chapter 7. Concluding Remarks

7.1 Summary of findings

The present dissertation explores the mapping of prosodic forms onto information structures in L1 Cantonese and L3 Mandarin by trilingual adults and children. A question-answer congruence paradigm (Roettger et al., 2019) was employed. Six focus conditions were examined and half of which had congruous prosody, and the other half had incongruous prosody in both audio-only and audio-visual modalities. The focus conditions for the Cantonese and Mandarin experiments were illustrated in Table 3.4 and Table 4.4, respectively. Recall that the dissertation proposed the following five research questions to investigate the focus processing mechanism used by trilingual adults and children, which were addressed through four perception experiments.

Research question 1: Can native adult speakers of Cantonese map prosodic forms onto specific focus categories in various tonal contexts in their native Cantonese?

Research question 2: Can trilingual adults acquire a native-like competence in decoding prosodic focus in their L3 Mandarin?

Research question 3: Have school-aged trilingual children developed an adult-like ability to process the prosodic focus in their native Cantonese?

Research question 4: Have school-aged trilingual children developed a native-like path in focus processing in L3 Mandarin?

Research question 5: Does the audio-visual modality facilitate listeners' focus processing in their native or non-native language?

To answer research question 1, Experiment 1 (Chapter 3) was designed for trilingual adults. The results of the study indicate that trilingual adults had higher

perceptual sensitivity to congruous prosody, except for the broad focus, compared to incongruous prosody. They employed a probabilistic inference approach to process prosodic focus, with biases against broad focus and towards narrow focus. Furthermore, the study revealed that trilingual adults demonstrated a strong anticipation of contrastive focus forms, which facilitated their comprehension of incongruous prosody. These observations indicate that native adult listeners may incorporate their prior knowledge of information structure and bottom-up signals in focus decoding in their L1 Cantonese.

To address research question 2, the same participants who participated in Experiment 1 was recruited to participate in Experiment 2 (Chapter 4), which aimed to examine trilingual adults' focus processing in L3 Mandarin using a similar paradigm. The results demonstrate that native Cantonese-speaking trilingual adults have not acquired native-like proficiency in all focus conditions, whether congruous or incongruous prosody, compared to native Mandarin-speaking adults. Parallel to their perceptual biases in decoding prosodic focus in Cantonese, trilingual adults exhibited a predisposition towards narrow focus while displaying a tendency to disregard broad focus in their acquisition of L3 Mandarin, a phenomenon not observed among their native Mandarin-speaking counterparts. Moreover, their expectations for a strong prosodic prominence of contrastive focus facilitated their comprehension of incongruous prosody, although the accuracy was still not fully native-like. The findings from this study provide evidence that trilingual adults are probably influenced by their L1 focus processing and employ their prior knowledge of focus categories in L3 Mandarin, suggesting a cross-linguistic approach to focus decoding.

To address research question 3, Experiment 3 (Chapter 5) investigated whether school-aged trilingual children have developed an adult-like ability to interpret congruous and incongruous prosody focus in their L1 Cantonese. Trilingual children, with a mean age of nine, comprehend focus with a comparable level of accuracy to that of native adults in evaluating the prosodic naturalness of each focus condition. In contrast to adults who tend to rely on their prior knowledge, children tend to utilize acoustic signals regardless of the focus categories, resulting in a perceptual pattern that does not exhibit the biases observed in native adult speakers during the mapping process. One possibility is that trilingual children, at the age of nine, have not yet acquired the adult-like competence to utilize prior knowledge of information structures. It is also likely that the acquisition of focus mapping in their L3, Mandarin, may influence trilingual children's pre-existing knowledge of the likelihood for cue-category mappings in their first language. In summary, this study provides evidence that trilingual children, although achieving comparable accuracy to adults, have not yet fully acquired the adult-like ability to incorporate prior knowledge and prosodic cues during speech comprehension.

With regard to research question 4, Experiment 4 (Chapter 6) recruited the same participants as Experiment 3 to explore how trilingual children map prosodic cues to focus categories in L3 Mandarin. Trilingual children generally comprehend congruous and incongruous focus with similar accuracy to their native peers. However, they exhibit differences in their expectations for prosodic prominence between narrow focus and contrastive focus. Trilingual children anticipated a greater prominence for contrastive focus, facilitating their decoding of focus under incongruous prosody. In contrast, native Mandarin children process incongruous prosody equally well regardless of the information

structures of the discourse. This finding suggests that school-aged trilingual children acquire the ability to employ both expectations of information structures and acoustic cues for focus processing in their L3 Mandarin at an earlier stage, albeit not to the same extent as native peer speakers.

Regarding research question 5, all experiments utilized stimuli designed in both audio-only and audio-visual modalities to investigate whether visual input can facilitate the proficiency of trilingual adults and children in focus processing. Both groups demonstrated similar performance between the two modalities, indicating that the restricted augmentation of visual information providing little benefit to decode prosodic focus in tonal languages. No celling effects were observed for trilingual listeners, however, they remained insusceptible to the supplementary visual information when listening to either their L1 Cantonese or their L3 Mandarin. These findings are consistent with the "tonal hypothesis", which proposes that speakers of tonal languages are less influenced by the visual information compared to auditory information (Sekiyama et al., 2003). This hypothesis has been examined in studies on Japanese, where Japanese listeners showed weaker McGurk effects than English listeners. Cantonese and Mandarin have more complex tonal systems than Japanese (a pitch-accent language), thus the "tonal hypothesis" suggests a similar or less pronounced McGurk effect in these groups. Sekiyama's (1997) study found that the similarities in cultural between Chinese and Japanese, as opposed to Western cultural, may explain why listeners of these languages are less dependent on visual cues (e.g., facial expressions). The "face-avoidance" hypothesis proposes that Japanese listeners avoid staring at a speaker's face during communication, as it is considered impolite. This cultural factor may offer a potential explanation as to why Chinese listeners,

both of Cantonese and Mandarin, may also engage in this practice during interpersonal communication. However, further research on multimodal processing is necessary before a definitive conclusion can be drawn, as the current study requires a higher demand on cognitive resources for focus processing, rather than on simple identification of segments as previous hypotheses were derived.

Furthermore, the present study's findings further suggest that the "non-native speaker effect" hypothesis, which posits that listeners of non-native languages exhibit a stronger McGurk effect than native listeners, may by inadequate to account for multimodal perception in Chinese. This hypothesis was originally proposed from studies that found non-native English listeners perceived visual influence more strongly than native Japanese listeners but is contradicted by comparison between Japanese and Chinese (including Taiwan Mandarin and other Chinese dialects). For instance, in previous studies (Yasuko et al., 1998; Sekiyama, 1997), Chinese listeners were found to be insusceptible to visual information from both Japanese and English stimuli. In the present study, Cantonese listeners did not demonstrate any differences in terms of susceptibility to visual cues when compared to native Mandarin listeners during Mandarin focus processing. These results suggest that tonal language listeners may be less susceptible to visual cues during their decoding of both native or non-native prosodic focus, given that the acoustic input is sufficiently informative and that the cognitive processing demands were high.

However, two additional possibilities that may account for the lack of advantage of additional visual cues observed across groups of participants should be considered in future studies. Firstly, participants may have overlooked the visual information in the current paradigm, as there has been no examination of whether they paid attention to the visual

cues during the rating process. For instance, incorporating a visual-only modality into the current paradigm would be effective in assessing the usability of simultaneous facial cues when processing prosodic focus in trilingual children and adults. Additionally, apart from the study conducted by Fung and Mok (2018), which demonstrates how Cantonese adult speakers mark contrastive focus through gestural prominence in conjunction with prosody, there is a scarcity of literature examining whether Cantonese or Mandarin speakers use facial expressions to mark prosodic focus in their speech communication. Further analysis of real-time facial kinematics data (Zhang et al., 2023) extracted from Cantonese and Mandarin speakers will contribute significantly to our understanding of the role of visual signals in focus processing in tonal languages.

7.2 General discussion

7.2.1 Cross-linguistic approach to focus processing in trilingual adults

The findings of this present study suggest that trilingual adults display a preference for narrow focus over broad focus when mapping prosodic realizations to information structures in their L1 Cantonese. This discovery aligns with observations made by Roettger et al. (2019) regarding the mapping of English focus in native English-speaking adults. The researchers postulated that the prosodic focus processing occurs as a probabilistic inference under conditions of uncertainty, where listeners utilize their prior knowledge of certain focus categories conveyed through various acoustic cues. In particular, the bias against broad focus is thought to stem from the listener's prior knowledge of the uncertainty of the category-cue mapping. One possible explanation proposed by Roettger et al. (2019) is that precursor questions like "What happened?", eliciting broad focus, do not provide any common information to interlocutors and are infrequent in conversational contexts. As a

result, listeners may find it challenging to infer the prosodic patterns of broad focus in incoming utterances. Conversely, English listeners are presumed to possess knowledge that the likelihood of a speaker producing narrow focus is high, given the common occurrence of narrowly focusing on a constituent of utterances. This asymmetric knowledge of focus categories is believed to lead to the tendency among English listeners to prefer mapping ambiguous prosodic forms to the information structure of narrow focus.

Moreover, the biased preference for narrow focus was consistently observed in the subsequent 5-Likert scale task conducted by Roettger et al. (2019), whereby participants previously involved in the two-forced choice task were recruited and presented with the same stimuli, but with the inclusion of more nuanced responses to discern between two focus conditions. In contrast, the bias against broad focus, which was initially observed in the discrimination between two dialogues using a two-forced choice task, disappeared in the subsequent 5-Likert scale task conducted in their English study. However, it should be noted that Roettger et al.'s (2019) experiments were unable to rule out the possible influence of task effects on the same group of participants, as well as the impact of competing focus conditions on discriminating between congruent and incongruent focus mappings. Therefore, the present study evaluates the perceived prosodic naturalness of each target utterance using a 5-Likert scale, without involving discrimination against other matched or mismatched pairs. Additionally, each participant was exposed to the stimuli only once. The results of the present study suggest that trilingual adults consistently exhibited biases against broad focus and in favor of narrow focus, as evidenced by comparative analyses of different pairs. Furthermore, they demonstrate an inherent disfavor for mapping prosodic forms onto the category of broad focus, as demonstrated by

their evaluation of congruous prosody of broad focus in their native language, Cantonese. These findings not only validate the consistent presence of perceptual biases in Cantonese, but also expand our understanding of focus mapping in English (Roettger et al., 2019) by demonstrating that perceptual bias is not solely a result of discrimination between two distinct focus conditions.

The probabilistic approach to focus processing proposed by Roettger et al. (2019) rests on the assumption that narrow focus is more frequently elicited in conversations compared to broad focus in English. However, it is important to note that there seems to be a lack of corpus evidence supporting differential frequency of these two focus categories in English. Additionally, it is challenging to make generalizations about the frequency of focus categories across different languages. Drawing from the findings of the present study on Cantonese focus decoding, it can be inferred that adult listeners of Cantonese, similar to English listeners, likely possess asymmetrical prior knowledge regarding narrow focus and broad focus in the process of focus mapping. The present findings contribute to our understanding that adult listeners in their native languages, as supported by evidence in English and Cantonese, can employ a probabilistic model for decoding prosodic focus.

In general, the probabilistic approach to mapping between acoustic categories and linguistic units has been tested speech perception in general to overcome the pervasive invariance of speech (Clayards et al., 2008; Kleinschmidt et al., 2018; Norris et al., 2016; Chen et al., 2022; see a review by Kleinschmidt & Jaeger, 2015). For example, on the segmental level, listeners store distributions like vowel formants and VOT to make social inferences, such as the speaker gender, age, and dialect (Kleinschmidt et al., 2018). Additionally, a recent study by Chen et al. (2022) proposed that native tonal language

speakers build up and store probabilistic parametric representations of the acoustic cues (F_0) of Cantonese tones. Even when contextual information is added, these stored parameter distributions remain accessible. Chen et al. (2022) revealed the role of prior knowledge in speech perception at the suprasegmental level, particularly among Cantonese speakers. In addition to tonal perception, the present study contributes to the understanding of speech perception that Cantonese adults may accumulate their knowledge of prosodic focus categories and make probabilistic inferences during focus processing.

Furthermore, this study investigates the generalization of the probabilistic approach to focus processing in the context of third language acquisition. The findings provide evidence that the perceptual biases associated with focus comprehension among native Cantonese-speaking adults are not limited to their L1, Cantonese, but are also evident in their L3, Mandarin. Similar to their processing of focus in Cantonese, trilingual adults tend to make probabilistic inferences when interpreting prosodic cues in their L3 Mandarin, based on their prior knowledge of the probability that a speaker intends to convey either narrow focus or broad focus. However, the biases in perceptual patterns pertaining to narrow focus and broad focus, observed in trilingual adults, do not manifest in the focus mapping of native Mandarin-speaking adults. Recall that listeners possess prior knowledge favoring narrow focus over broad focus, which serves as a premise for the preference towards narrow focus and bias against broad focus in the probabilistic model (Roettger et al., 2019). Nevertheless, the findings regarding Mandarin focus of the present study challenge this assumption based on evidence that native Mandarin listeners do not exhibit these biases during their mapping process. As a result, it becomes necessary to reconsider the significance of listeners' asymmetric prior knowledge of narrow focus and broad focus.

While the frequencies of focus categories may indeed play a role in focus processing, it is imperative to recognize that there might be variations in frequency of different focus categories in the pragmatic use across different languages, such as Mandarin and English.

Consequently, the present study suggests that trilingual adults may extend the probabilistic inference approach to focus processing from their L1 Cantonese to their L3 Mandarin. By adopting the probabilistic model used in their L1, they gradually accumulate and update their knowledge of the likelihood of each focus category expressed by a speaker in their exposure to Mandarin speech. The observation from this study suggests that, for trilingual adults, the probability of the category-cue mapping for broad focus is lower than that for narrow focus in their L3 Mandarin, which aligns with the findings in their native Cantonese. The discrepancies between Mandarin focus mapping by native Cantonese-speaking and native Mandarin-speaking adults imply that non-native listeners may possess different prior knowledge of prosodic forms and the corresponding category-cue mappings compared to native speakers. In the context of this study, variations in prior knowledge across languages regarding the mapping between information structures and their prosodic forms may, to some extent, account for the non-native-like decoding of focus exhibited by trilingual adults.

Last but not least, the current study highlights the existence of biased expectations between the information structures of narrow focus and contrastive focus among adult Cantonese and Mandarin listeners when processing incongruous prosody in their respective native languages. Specifically, when listeners are presented with mismatched pairs of broad focus forms, i.e., NB and CB focus conditions, they tend to exhibit greater accuracy in focus comprehension for the preceding discourse of contrastive focus than for narrow focus.

This is due to the expectation among native adults for a strong prominence on the incoming utterance that responds to the prior discourse of contrastive focus, making them sensitive to the prosodic incongruity when encountering the realization of relatively non-prominent broad focus (CB focus condition). Although Roetteger et al. (2019) demonstrated higher accuracy in mapping prosodic forms to the contrastive focus category in English, their study did not provide evidence on the comparison between contrastive focus and narrow focus when they incongruously elicit the same broad focus forms. The current study, however, discovered that native adult listeners do anticipate different prosodic prominence between contrastive and narrow focus when encountering incongruous prosody. The incorporation of the listener's anticipation of prosodic salience within focus mapping contributes to the identified findings that native adult listeners possess prior knowledge regarding the heightened prosodic prominence associated with contrastive focus, thereby enabling them to integrate a top-down probabilistic approach with their interpretation of perceived acoustic cues. Therefore, the present study suggests that native Cantonese- and Mandarin-speaking adults may both apply a probabilistic approach to focus processing, suggesting a cross-language mechanism in prosodic comprehension (German: Roettger & Franke, 2019; Kurumada & Roettger, 2021; English: Roetteger et al., 2019; Roetteger et al., 2021).

7.2.2 Cross-linguistic approach to focus processing in trilingual children

Trilingual children in this study demonstrated differences in their utilization of prior knowledge to decode Cantonese prosodic focus compared to trilingual adults, although they achieved a comparable level of accuracy in evaluating the tested focus conditions. Unlike trilingual adults who exhibited biases towards certain information structures, native

Cantonese-speaking children processed the examined information structures without a bias. As discussed in Chapters 7.2 and 3.3, trilingual adults' probabilistic inference to prosodic focus can be inferred by the assumption that trilingual adults possess greater *prior probabilities* of narrow focus compared to broad focus (i.e., $P(Narrow) > P(Broad)$). This assumption is grounded on the premise that adult listeners accumulate a greater level of experience in perceiving and interpreting the prosodic realizations associated with narrow focus compared to broad focus. Considering the probabilistic method to cognitive development (Bonawitz et al., 2014), it is plausible that trilingual school-aged children have not yet developed an adult-like ability to use prior knowledge during focus decoding.

Regarding developmental aspects, children gradually accumulate linguistic knowledge through their language experiences and demonstrate substantial growth in knowledge acquisition through the childhood (Baltes et al., 2006; Craik & Bialystok, 2006). Nevertheless, it remains uncertain whether they can use their prior knowledge of focus categories as efficiently as adults. Research on cognitive neuroscience perspectives may provide evidence of children's inability to process prior knowledge like adults, specifically in terms of the influence of prior knowledge on memory (see a review by Brod et al., 2013). The difficulties that children face when processing prior knowledge could stem from their underdeveloped prefrontal cortex and medial temporal lobe in the brain, responsible for coordinating the formation and application of knowledge. For instance, children aged 8 to 11 exhibited differences in hippocampus development, an essential structure of the medial temporal lobe, compared to adolescents (aged 14) and young adults in an incidental encoding task (Ghetti et al., 2010). Therefore, the present behavioral findings merely provide a preliminary exploration that the method employed by trilingual children to

comprehend focus is not yet fully adult-like, potentially due to their underdeveloped cognitive processes. Further investigations of neuroscience evidence are necessary before any conclusions can be drawn regarding the development of trilingual children's prior knowledge pertaining to focus processing.

In L3 Mandarin, trilingual children have demonstrated perceptual judgment abilities comparable to those of native Mandarin-speaking peers for each focus condition that was examined. However, minor differences have been observed in their expectations of certain information structures, particularly in terms of their anticipation for the use of prosodic prominence for narrow focus and contrastive focus. Specifically, native Mandarin-speaking children performed similarly well in both the narrow and contrastive focus in the two focus conditions (CB and NB) when each condition was compared separately with the congruous broad focus condition (BF). However, trilingual children were unable to discriminate between the incongruous NB condition and the congruous BF condition, while they were able to differentiate between the incongruous CB condition from the congruous BF condition. One possible explanation for this phenomenon is that trilingual children have an expectation of heightened prominence in the target utterance when it is used to indicate contrastive focus rather than narrow focus, enabling them to accurately detect the prosodic incongruity caused by the use of broad focus form within the context of contrastive focus discourse (CB focus condition). Consequently, trilingual children exhibit partial proficiency in discriminating between congruous and incongruous focus conditions (BF vs. NB), as well as demonstrate superior performance in processing the information structure of contrastive focus compared to narrow focus. Overall, although trilingual children demonstrate comparable performance to their Mandarin-speaking peers

in evaluating each focus condition, they have not completely attained native-like competence in prosodic focus processing when contrasting congruous and incongruous focus conditions.

Moreover, trilingual children are shown to display a developmental advantage in their ability to comprehend contrastive focus over narrow focus during the process of developing focus decoding in their L3. Specifically, trilingual children's developmental advantage in contrastive focus over narrow focus in L3 Mandarin may be partially explained by the idea of cognitive resource consumption supported by evidence like pupillary dilation and ERP responses (Pannekamp et al., 2011; Zellin et al., 2011; Toepel et al., 2009). That is, listeners expend less cognitive resource when comprehending the contrastive information than the new information (narrow focus). The contrastive information clearly contrasts with the previously stated alternative, the corrected information in the precursor question, whereas the number of entities capable of answering *Wh*-questions are infinite, making the focus more challenging to be detected. The study carried out by Pannekamp and colleagues (2011) demonstrated that the developmental advantage in contrastive focus recognition, as opposed to narrow focus recognition, was only apparent in 8-year-old German children, and not in the younger group of children aged five. Thus far, the present study provides the initial evidence of asymmetrical expectations in the comprehension of prosodic focus between narrow focus and contrastive focus among children acquiring a L3 Mandarin. However, in the present study, the developmental advantage in processing information structure of contrastive focus over narrow focus was observed in children's non-native Mandarin language skills rather than in their native Cantonese.

Why was this phenomenon not observed in their native Cantonese? In the case of Mandarin, the superior performance of trilingual children in distinguishing contrastive focus from narrow focus was determined through a comparative analysis of different focus conditions, namely BF versus CB and BF versus NB focus conditions. It is speculated that this advantage is specific to the acquisition of focus mapping in a non-native language, at least for children around the age of nine. This is supported by the fact that native Mandarin-speaking children exhibit a similar ability to differentiate between contrastive and narrow focus on the same syntactic position in Mandarin. Then native Mandarin-speaking children undergo a period of development to acquire an adult-like approach to processing prosodic focus in their native language, leading to an enhanced ability to accurately identify contrastive focus compared to narrow focus (i.e., NB versus CB focus conditions). Furthermore, this advantage in focus processing of L3 Mandarin may be absent in the acquisition of Cantonese focus processing, as observed in trilingual adults. These adults tended to mistakenly rate congruous BF as having less natural prosody than incongruous CB due to their strong bias against broad focus. In contrast, trilingual children did not exhibit similar perceptual biases in Cantonese focus mapping, indicating a deviation in their processing approaches compared to trilingual adults.

Finally, the impact of L2 acquisition on L3 prosodic development likely contributes to the disparity in L3 focus processing between trilingual adults and children. Trilingual acquisition is inherently complex and dynamic (e.g., Herdina & Jessner, 2002; Wrembel, 2015), given that L3 prosody is potentially influenced by various sources of transfer, not only from L1 or L2, but also from cross-linguistic influences among multiple languages (Wrembel, 2015; Zhu et al., 2019). Ge et al. (2018; 2021) investigated the focus

comprehension of L2 English among native Cantonese adult speakers, demonstrating accurate identification of congruent and incongruent prosody in focus mapping. Additionally, native Cantonese adults can use on-focus expansion of F_0 ranges to mark focus in their L2 English, which are absent in their L1 Cantonese (Fung & Mok, 2014). As L3 learners can draw on conscious linguistic knowledge and language-learning strategies acquired during L2 acquisition (e.g., De Angelis, 2007), it is plausible that the acquisition of L3 focus processing by trilingual adults is influenced by the decoding methods applied in L2 English. However, school-aged children have limited learning experience and exposure to L2 English compared to adults in the same Hong Kong community. According to the Dynamic Model of Multilingualism (Jessner, 2006), the acquisition of L3 is the result of the combined influence of the structures of L1, L2, and L3, with their respective weight being determined by a variety of factors, along with their language awareness and learning strategies. In the present study, the influence of L2 on the acquisition of L3 focus processing played a limited role in trilingual children compared to adults. Therefore, the differences in L2 transfer and/or L1-L2 cross-linguistic influence between trilingual children and adults may account for the significantly distinct focus decoding in L3 Mandarin in comparison with their respective counterparts.

7.3 The Prosody Processing Model in L1/L3 (PPM-L1/L3): A working model

In the domain of speech learning, the existing Speech Learning Model (SLM, Flege, 1995, 2002) and its revised version (SLM-r, Flege & Bohn, 2021) as well as the Perceptual Assimilation Model of L2 (PAM-L2, Best & Tyler, 2007) formalize the theoretical hypotheses regarding the speech perception and production of segments in second language learning. Furthermore, the L2 Intonation Learning Theory (LILt, Mennen, 2015) offers a

comprehensive model for acquiring suprasegmental features (i.e., intonation) in L2 production, building upon the underlying assumptions of the SLM. The LILt posits four dimensions (systemic, realizational, semantic, and frequency dimensions) to predict the intonational deviations encountered by language learners during their L2 phonetic implementation. The realizational and semantic dimensions are particularly relevant to the thesis. Specifically, the realizational dimension pertains to the phonetic realization of categorical elements, while the semantic dimension characterizes the functionality of these categorical elements. The present study aims to investigate how listeners identify the focus category (semantic dimension) when presented with various prosodic forms (realizational dimension).

The LILt posits that L2 learners can apply the learning mechanisms utilized in their L1 to acquire L2 intonation and approximate the norms of L2 production, particularly in certain intonation dimensions. This thesis demonstrates that trilingual children are capable of evaluating the prosody of matched and mismatched focus conditions as accurately as native L3 Mandarin speakers, indicating that L3 learners can process intonational categories of prosodic focus in L3 and achieve an approximate accuracy comparable to that of native speakers. These findings contribute to the LILt by providing evidence of perception and comprehension of speech intonation, suggesting that L2/L3 learners are likely to succeed in perceiving L2 intonation in both the realizational and semantic dimensions.

Additionally, the LILt suggests a potential bidirectional interaction between L1 and L2 at the intonation level. This interaction could involve the assimilation or merging of L1 and L2 characteristics, resulting in L2 learners producing values that are intermediate

between those of their L1 and L2. The findings of the current study provide evidence that the bidirectional influences can also be observed in the perception and comprehension of L3 intonation. Specifically, trilingual children rated matched and mismatched broad focus as equally natural in L1 Cantonese, in contrast to trilingual adults who erroneously rated the matched broad focus as less natural than the mismatched prosody, indicating biases. Conversely, they were able to accurately process broad focus in L3 Mandarin, rating the matched broad focus as more natural prosody than the mismatched ones. On one hand, trilingual children, influenced by their L1 prosody, did not accurately rate the incongruous broad focus when preceded by narrow focus context in their L3 Mandarin, as compared to their native Mandarin-speaking peers. On the other hand, their processing of the same focus conditions in L1 was not as inaccurate as that of native Cantonese-speaking adults, possibly due to the influence of their L3. The focus mapping of L1 Cantonese in school-aged trilingual children was found to be influenced by the bidirectional interaction between L1 and L3. The bidirectional influences between L1 and L3 may result in three scenarios of trilingual children's language focus mapping in L1 as they transition into adulthood. It can be hypothesized that trilingual children may accurately decode broad focus in their L1 without bias in their adulthood, considering the influences of L3. However, it is also possible that they may erroneously reject broad focus as unnatural prosody, similar to trilingual adults of native Cantonese in the present study, if the effect of L3 prosody plays a minimal role in their prosody development of L1. Furthermore, it is challenging to exclude the possibility that the interaction between L1 and L3 in focus processing is not temporary, as observed in the intermediate values in production in several studies (e.g., Mennen, 2014; De Leeuw et al., 2012). Further investigations are necessary to comprehend

the developmental trajectory of the focus decoding approach employed by trilingual children in certain intonation dimensions.

However, the current models of speech learning, which primarily focus on comparing phonological and phonetic categories across languages, appear to be insufficient in explaining the findings that involve multiple levels of linguistic knowledge (i.e., pragmatic and prosodic knowledge) beyond mere phonetic implementation. Specifically, the explanation of comparing phonological categories across languages falls short in accounting for the perceptual mapping between information structure and prosodic realizations in spontaneous dialogues. This inadequacy stems not only from the absence of a stationary one-to-one mapping between linguistic unities and acoustic forms (e.g., Cruttenden, 1997; Cangemi et al., 2015; Grice et al., 2017; Turnbull et al., 2017), but also from the influences of factors beyond bottom-up acoustic cues, such as the linguistic context and listeners' expectations (e.g., Kleinschmidt & Jaeger, 2015; Kuperberg & Jaeger, 2016; Roettger & Franke, 2018; Roettger et al., 2019).

For a deeper comprehension of how listeners interpret prosody associated with information structure, the study conducted by Roettger et al. (2019) proposes that English listeners engage in a probabilistic inference of the acoustic cues they perceive, thereby making judgments about the speakers' intentions, such as the focus of their speech. This inference process entails the integration of prior knowledge regarding the likelihood of a speaker utilizing specific prosodic patterns within a given context. Figure 7.1 offers a visual representation of the mapping process for English prosodic focus based on the research by Roettger et al. (2019). As depicted in the chart, English listeners employ their top-down probabilistic expectations in conjunction with acoustic signals when processing prosody

within a particular context, such as Q-A dialogues that elicit different information structures within the responses. During this process, listeners activate their stored prior knowledge of linguistic relevance by calculating the likelihood of the specific speaker producing the perceived prosodic forms, as well as the likelihood of these forms being used within the given context. These sequential steps ultimately contribute to the listener's probabilistic inference and may result in biases towards or against specific information structures. For instance, as indicated by Roettger et al. (2019), English listeners tend to erroneously reject broad focus, possibly due to their prior knowledge suggesting a lower likelihood of encountering this prosodic form in relation to its discourse function. Particularly in experimental tasks, the information structure of broad focus often fails to pre-activate sufficient prior knowledge, as questions like "What happened?" induce unexpected interpretation of subsequent responses, resulting in increased uncertainty during the process of probabilistic inference. The investigation by Roettger and colleagues provides a significant contribution to our understanding of how listeners interpret prosody in relation to information structure, underscoring the essential nature of integrating top-down approaches with acoustic signals in the analysis of human speech.

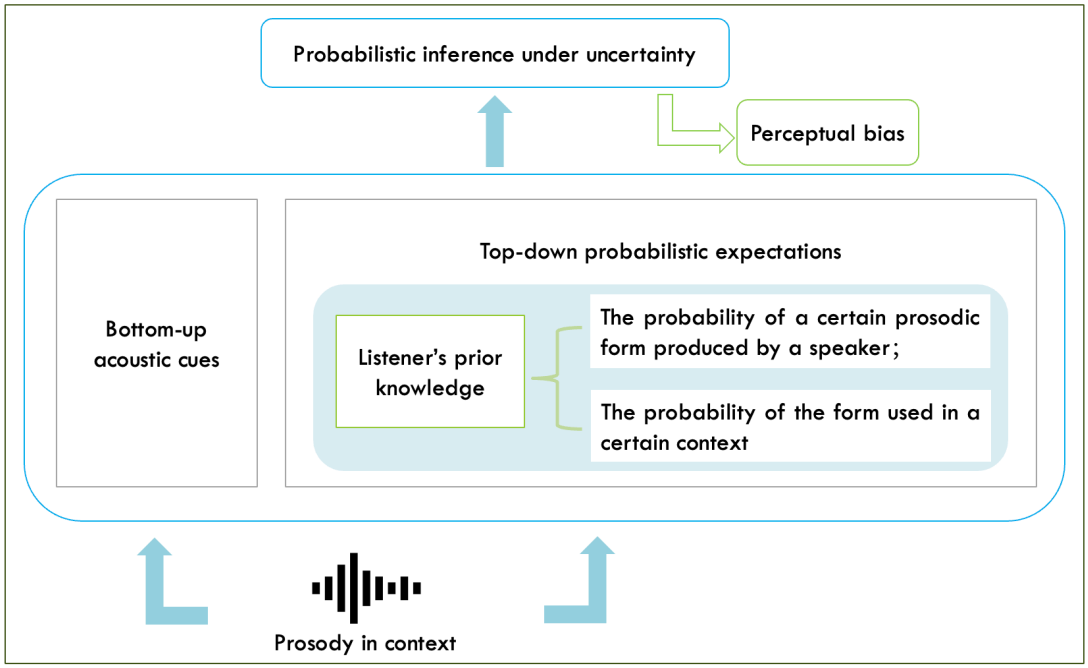


Figure 7.1 The process of focus mapping by English listeners (Based on Roettger et al., 2019)

However, the probabilistic method employed by Roettger et al. (2019) assumes a presupposition regarding the asymmetry in prior beliefs held by listeners concerning the likelihood of different information structures. Specifically, they presumed that listeners possess a prior belief that speakers are more inclined to use narrow focus over broad focus in speech communication, which results in a higher probability of mapping uncertain prosodic cues to narrow focus instead of broad focus. Considering the role of belief updating in speech comprehension (Kuperberg & Jaeger, 2016), listeners can incorporate new evidence from prosody to revise and refine their initial beliefs or probabilities. Consequently, it was assumed that listeners lacked the knowledge about the consistent mapping between prosodic forms and information structures for certain focus categories (e.g., broad focus in this context). Nevertheless, there is no evidence from any corpus study

suggesting that questions eliciting broad focus are asked less frequently than those eliciting narrow focus. For instance, there is no indication that questions like “What happened?” are asked less frequently than questions like “What did someone do to something?”. Moreover, even if there is evidence indicating a frequency discrepancy in the usage of these question types in English, the variations should be carefully considered in cross-language studies. The outcomes from the processing of prosodic focus in Mandarin by trilingual adults challenge the aforementioned assumption, as the perceptual biases observed in trilingual adults were absent in native Mandarin-speaking adults. This discovery signifies that the notion of asymmetry in prior knowledge between narrow focus and broad focus in focus processing may be susceptible, necessitating further investigations into the factors contributing to perceptual biases in English and Cantonese. Alternatively, although listeners of specific languages may possess knowledge of such asymmetry in prosodic focus, the frequency of using distinct focus categories remains language-specific. Specifically, it is questionable whether narrow focus is more frequently employed than broad focus in Mandarin speech communication.

Furthermore, while Roettger and colleagues provide a hypothesis that elucidates the mapping mechanism of focus in English among adult listeners, there is a lack of evidence regarding whether probabilistic processing of prosodic focus extends to children’s perception, let alone whether this perception mechanism remains intact when processing non-native languages. Given that the inadequate coverage of existing literature in explaining the findings of this dissertation, the current study posits a working model: the Prosody Processing Model in L1/L3 (PPM-L1/L3). The PPM-L1/L3 puts forward three hypotheses to enhance our understanding of the decoding process of prosodic focus for

both adults and children in their native language as well as in a third language.

Hypothesis 1 presumes that the application of a probabilistic approach in processing prosodic focus by native adult listeners in their respective native languages is a cross-linguistic phenomenon, wherein their prior knowledge of prosodic forms and the co-occurrence information structure are utilized to discern the speaker's intention. The current study demonstrates that native adult listeners of Cantonese exhibit a stronger reliance on their prior knowledge of category-cue mapping for broad focus, compared to bottom-up perceptual cues. Similarly, the adult Mandarin-speaking listeners in the control group also demonstrate increased uncertainty in identifying broad focus compared to other focus categories in Mandarin. Note that the present study does not provide a thorough investigation into whether listeners possess prior knowledge of prosodic cues for each focus category. Nevertheless, it does provide empirical support for the notion that listeners possess a knowledge of the probability associated with the occurrence of prosodic cues within distinct contextual settings. In addition, consistent with findings in English listeners (Roettger et al., 2019), Kurumada and Roettger (2021) have developed comprehensive probabilistic frameworks for intonational processing, taking into account talker variability in German. The evidence from the aforementioned languages (Cantonese, Mandarin, English and German) suggests that this probabilistic mechanism of focus mapping may be a general one in adult listeners' native language comprehension. The present findings provide a cross-linguistic perspective on the probabilistic method of prosody perception, encompassing both Indo-European and Sino-Tibetan languages, although further investigations on additional languages are necessary for a comprehensive understanding of focus mapping.

Hypothesis 2 proposes that the retention of the probabilistic approach to focus processing, employed by listeners in their first language, can extend to their second or third languages. This study aims to explore the processing of prosodic focus in both the participants' native language (Cantonese) and a non-native language (Mandarin), thereby providing evidence on whether trilingual listeners utilize the probabilistic approach to the category-cue mapping in a non-native language. Figure 7.2 illustrates the process through which trilingual listeners make judgments of prosody in context in both their native language (L1) and their third language (L3). Listeners employ acoustic cues and their expectations of the speaker's intentions to make probabilistic judgments regarding focus categories. The present study's findings indicate that trilingual adult listeners have not attained a native-like proficiency in L3 Mandarin focus decoding, likely influenced by negative transfer from their L1 and L2. Meanwhile, they extensively utilize their probabilistic knowledge during the processing of Cantonese, also demonstrate a comparable perceptual mechanism when processing Mandarin. Specifically, they demonstrate comparable patterns of perceptual biases in both their native language and their third language, commonly rejecting broad focus and favoring narrow focus incorrectly. Figure 7.2 symbolizes the comparable results between L1 and L3 using the shape of a balance scale, highlighting the observed similar perceptual biases in both languages for trilingual adult listeners.

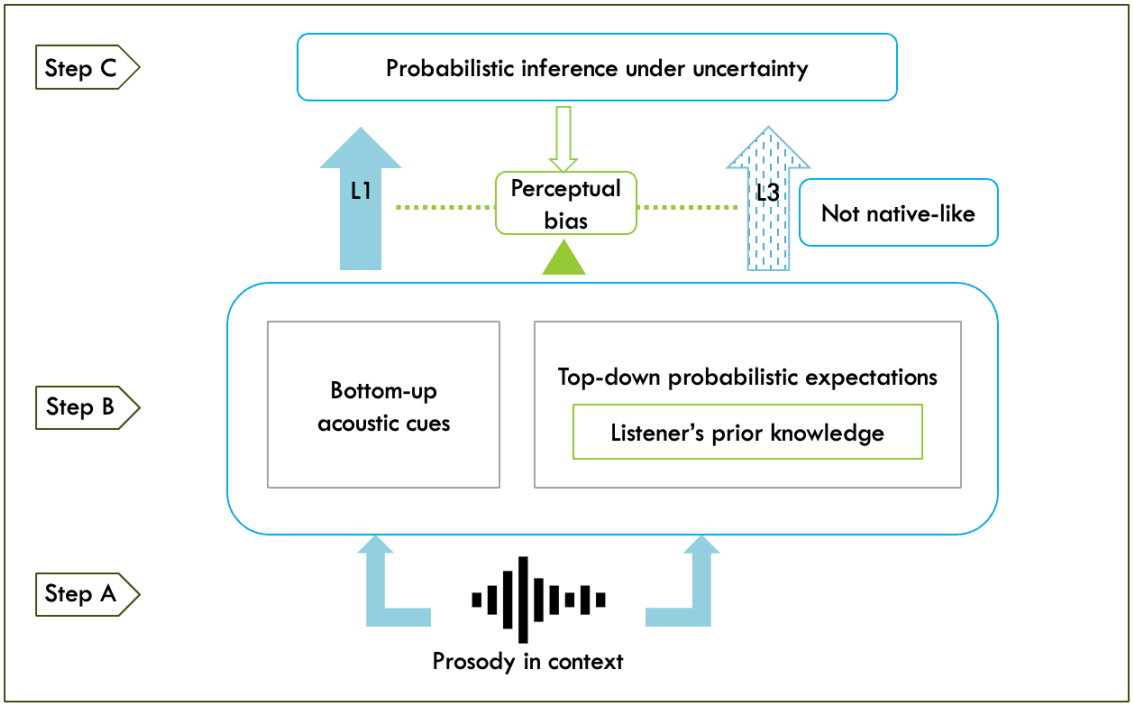


Figure 7.2 The process of focus mapping by adult listeners in L1 and L3

Hypothesis 3 posits that the effective utilization of a probabilistic approach to focus mapping may be more commonly observed in adult listeners rather than school-aged child listeners in their native language. The rational process of making inferences under uncertainty occurs incrementally, as listeners adjust their initial probabilities to updated probabilities upon encountering prosodic forms (Kuperberg & Jaeger, 2016; Roettger et al., 2019). This updated *posterior probability* then serves as the new *prior probability* in the subsequent cycle. It is hypothesized that children, with less language experience than adults, have not accumulated sufficient probabilistic knowledge to effectively process prosodic focus. The findings of the current study support this hypothesis indicating that trilingual children who are native speakers of Cantonese do not employ an adult-like probabilistic approach when decoding Cantonese focus. In contrast to their trilingual adult counterparts, children depend more on the acoustic input rather than their expectations of information

structure pertaining to broad focus and narrow focus. Figure 7.3 draws the processes that trilingual children and adults decode the prosodic focus in their L1. From step B to step C, the yellow-colored lines and shapes represent the processing procedure employed by trilingual children, while the blue-colored lines and shapes indicate the procedure adopted by trilingual adults. The disparities in the utilization of probabilistic expectations between children and adults result in discernible differences in perceptual biases concerning focus processing. It is a valid proposition to hypothesize that children require a prolonged period of time to acquire a fully developed probabilistic approach to efficiently process prosodic focus, similar to that exhibited by adults.

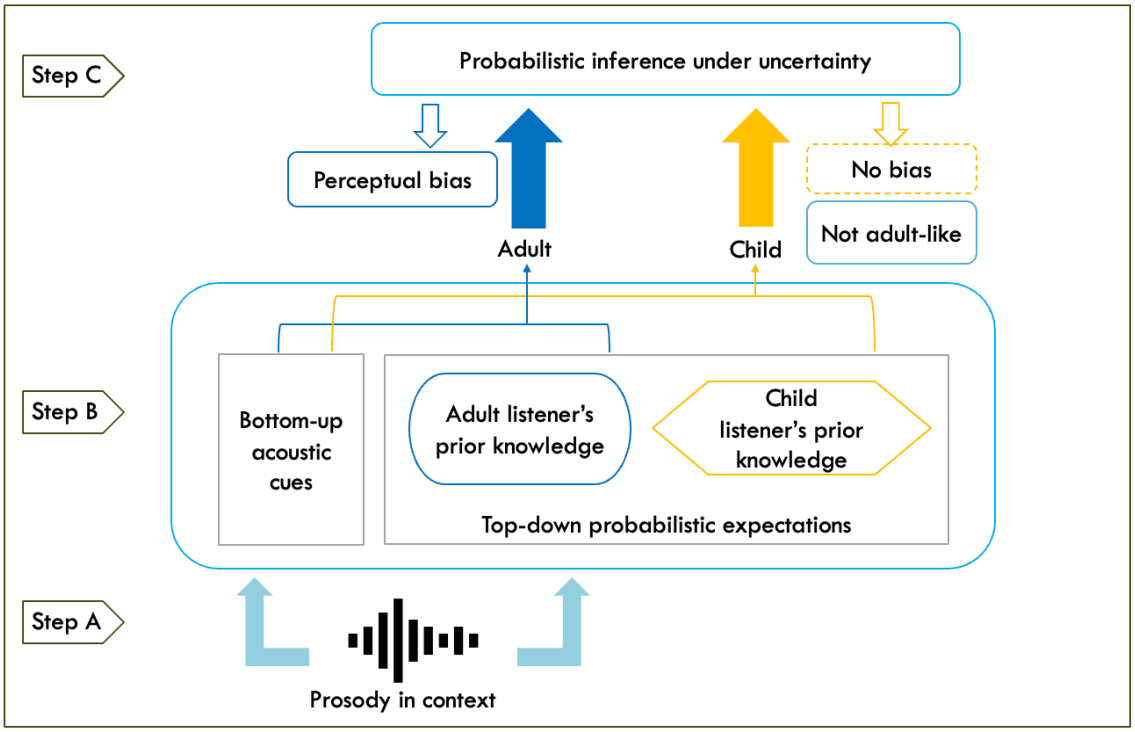


Figure 7.3 The process of focus mapping by child and adult listeners in L1

Hypothesis 4 proposes that school-aged trilingual children have not attained the native-like proficiency in processing prosodic focus in their L3, and the approach used in L1 has not been fully transferred to their focus processing in L3. The discrepancy in

decoding Mandarin focus mainly manifests in their comprehension of the natural prosody of broad focus; however, the inaccuracy of identifying broad focus does not entirely align with the perceptual pattern observed in their L1 focus processing. Additionally, trilingual children demonstrate differences in their ability to anticipate the prosodic prominence related to specific focus categories in their L3 Mandarin compared to their native Mandarin-speaking peers, despite both groups showing no perceptual bias. Figure 7.4 includes the processes of focus decoding in L3 by trilingual children alongside the process of L1 focus decoding tested in Hypothesis 3. It is a reasonable proposal to suggest that children at school ages need substantial exposure to the L3 prosody in order to develop a fully native-like ability to effectively process prosodic focus. Simultaneously, due to limited exposure to L2, their acquisition of L3 appears to be subject to less transfer from L2 compared to trilingual adults.

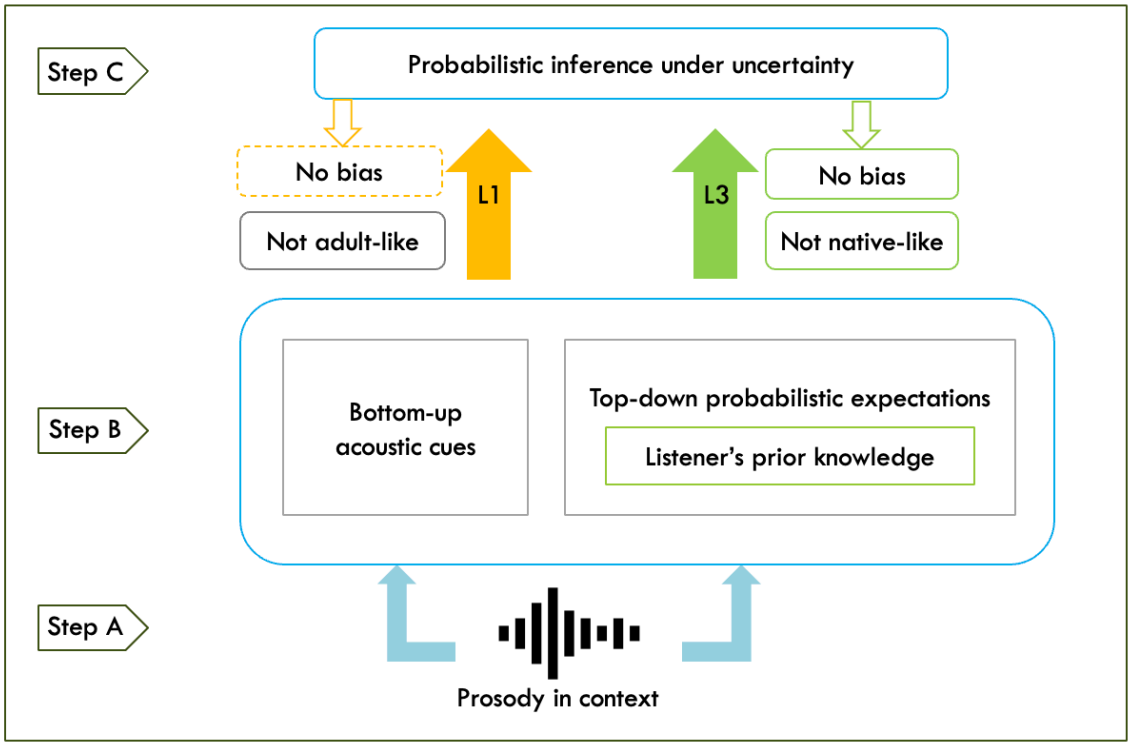


Figure 7.4 The process of focus mapping by child listeners in L1 and L3

7.4 Limitations and future directions

This dissertation has several limitations, which implies that there are several directions for future research to explore. Firstly, the recruitment of only one age group of trilingual children, with an average age of nine, limits our understanding of the developmental path to focus processing experienced by trilingual children. The trilingual children in this study had already attained an adult-like proficiency level in Q-A congruence tasks in their native Cantonese. To more fully understand the age-related differences in focus decoding accuracy, it will be necessary to include younger trilingual children in future studies. Furthermore, it is essential to explore the cognitive processing development of focus decoding in trilingual adolescents, specifically to examine their ability to process paralinguistic cues like prior knowledge of information structures. Such investigations would be best conducted using cutting-edge methodologies such as EEG or eye-tracking, which would allow us to more accurately track the developmental changes in focus decoding accuracy across different age groups.

Also, it should be noted that the current study had limited focus conditions of mismatched pairs, which may have restricted our exploration of trilingual speakers' complete mapping patterns. Specifically, the mismatched pairs in this study only involved broad focus forms and contrastive focus forms preceded by certain discourse. Therefore, it remains unclear how listeners map incongruous prosodic realizations of narrow/contrastive focus when they are preceded by a question that elicits broad focus. Furthermore, the present study did not assess listeners' sensitivity to prosodic modulations with incongruous focus positions. To address these limitations, future studies should aim to recruit a larger

number of participants to examine the entire focus processing mechanism across various focus positions and conditions of mismatched pairs.

Additionally, the current study solely examines the comprehension of prosodic focus in L1 Cantonese and L3 Mandarin and does not investigate the data pertaining to focus production of trilingual speakers. While the findings are significant and contribute to our understanding of cross-linguistic focus decoding, further research is necessary to explore the methods employed for focus encoding and the comprehension-production relationship in trilingual speakers. For the development of prosodic focus, Chen and Bergh (2020) have found that the comprehension-production link develops at various rates among Dutch children, depending on the position of focus, rather than following a simple relationship of comprehension preceding production or vice versa in language acquisition (as reviewed by Hendriks & Koster, 2010). Moreover, the probabilistic approach to focus comprehension underscores the significance of talker variability in listeners' probabilistic inference of focus categories under uncertainty (Kurumada & Roettger, 2021). Further investigations into focus production may provide a corpus with which to more comprehensively examine probabilistic methods for decoding focus in both children and adults.

Finally, this dissertation exclusively examines trilingual speakers' prosodic focus decoding mechanisms in their L1 Cantonese and L3 Mandarin, and does not investigate their L2, English. Cross-linguistic influence (CLI) suggests that multilingual speakers may experience simultaneous influence from more than one language they have acquired (Wrembel, 2015). To expand current knowledge of trilingual prosodic processing, future studies should examine focus processing mechanisms across languages and consider the

typological distance between Cantonese, English, and Mandarin. This research direction has the potential to contribute to theoretical models of multilingual acquisition, as well as provide pedagogic insights for both child and adult learners in multilingual communities.

7.5 Conclusions

In summary, this dissertation aimed to investigate the prosodic focus processing abilities of trilingual adults and school-aged children in both their L1 Cantonese and L3 Mandarin. The perceptual experiments revealed that trilingual adults and children differed in their usage of probabilistic inference for focus categories. Trilingual adults exhibited inference biases for specific information structures under both congruous and incongruous prosody in their native Cantonese. The observed mapping biases in L1 were also present in L3 Mandarin, resulting in the trilingual adults' non-native-like performance in Mandarin focus processing. Trilingual children at school age demonstrated proficiency in evaluating prosodic naturalness comparable to native Cantonese-speaking adults, while still showed aspects of not fully adult-like probabilistic inference during focus processing. In L3 Mandarin, trilingual children attained similar levels of accuracy in specific focus categories when compared with their native Mandarin-speaking peers, despite variations in expectations for the prosodic prominence of specific information structures. This dissertation makes a significant contribution to the probabilistic model applied to the comprehension of prosodic focus, highlighting that while adults can employ their prior knowledge of category-cue mapping to process focus in both their L1 and L3, school-aged trilingual children are still unable to probabilistically infer focus during speech processing in an adult-like manner. The PPM-L1/L3 was proposed as a theoretical framework to comprehensively elucidate the probabilistic approach to focus decoding in both child and

adult populations across different languages. In future research, investigating the developmental trajectory of focus decoding and exploring the link between focus comprehension and production in trilingual speakers are important directions to further understand the development of probabilistic approach to focus processing. Additionally, including their L2 English in future research will aid in deepening our knowledge of cross-linguistic influence.

Appendices

Appendix 1. Target sentences in Cantonese for Experiments 1 and 3

1. 張生揸飛機

   /tsœŋ55 saŋ55 tsa55 fei55 kei55/

   Zoeng55 saang55 zaa55 fei55 gei55

   Mr. Cheung operated an airplane.

2. 婉婉摸狗仔

   /jyn25 jyn25 mɔ25 kɐu25 tsɐi25/

   jyun25 jyun25 mo25 gau25 zai25

   Jyunjyun touched a dog.

3. 秀秀讚燕燕

   /sɐu33 sɐu33 tsan33 jin33 jin33/

   Sau33 sau33 zaan33 jin33 jin33

   Sausau praised jinjin.

4. 雅雅買泡泡

   /ŋa23 ŋa23 mai23 p^hou23 p^hou23/

Ngaa23 ngaa23 maai23 pou23 pou23

Ngaangaa bought bubbles.

5. 佩佩開香檳

/pʰui33 pʰui33 hɔi55 hœŋ55 pɐn55/

pui33 pui33 hoi55 hoeng55 ban55

Puipui opened a bottle of champagne.

6. 丙仔吹風車

/pɪŋ25 tsɐi25 tsʰɵy55 fʊŋ55 tsʰɛ55/

bing25 zai25 ceoi55 fung55 ce55

Bingzai blew a windmill.

7. 偉偉搬傢私

/wɐi23 wɐi23 pun55 ka55 si55/

Wai23 wai23 bun55 gaa55 si55

Waiwai moved the furniture.

8. 坤坤洗水果

/kʰwɐn55 kʰwɐn55 sɐi25 sɵy25 kwɔ25/

Kwan55 kwan55 sai25 seoi25 gwo25

Kwankwan washed the fruit.

9. 駿駿整餃子

[tsɵn33 tsɵn33 tsɪŋ25 kau25 tsi25]

zeon33 zeon33 zing25 gaau25 zi25

Zeonzeon made dumplings.

10. 敏敏剪海藻

[mɐn23 mɐn23 tsin25 hɔi25 tsou25]

Man23 man23 zin25 hoi25 zou25

Manman cut seaweeds.

11. 欣欣餵兔兔

[jɐn55 jɐn55 wɐi33 tʰou33 tʰou33]

Jan55 jan55 wai33 tou33 tou33

Janjan fed the rabbit.

12. 表姐戴鑽戒

[piu25 tsɛ25 tai33 tsyn33 kai33]

biu25 ze25 daai33 zyun33 gaai33

The cousin wore a diamond ring.

13. 允允笑杉杉

[wɐn23 wɐn23 siu33 tsʰam33 tsʰam33]

wan23 wan23 siu33 caam33 caam33

Wanwan laughed at Caamcaam.

14. 姜生養螞蟻

[kœŋ55 saŋ55 jœŋ23 ma23 ŋɐi23]

goeng55 saang55 joeng23 maa23 ngai23

Mr. Goeng kept ants.

15. 廠長買老馬

[tsʰɔŋ25 tsœŋ25 mai23 lou23 ma23]

cong25 zoeng25 maai23 lou23 maa23

The factory director bought old horses.

16. 貝貝抱美美

[pui33 pui33 pʰou23 mei23 mei23]

bui33 bui33 pou23 mei23 mei23

Buibui bugged Meimei.

17. 詠詠燒青瓜

[wɪŋ22 wɪŋ22 siu55 tsʰɛŋ55 kwa55]

wing22 wing22 siu55 cing55 gwaa55

Wingwing cooked cucumbers.

18. 護士剪紙板

[wu22 si22 tsin25 tsi25 pan25]

wu22 si22 zin25 zi25 baan25

The nurse cut the paperboard.

19. 侍衛怕暗器

[si22 wɐi22 pʰa33 ɐm33 hei33]

Si22 wai22 paa33 am33 hei33

The guard scared of hidden weapons.

20. 耀華咬蟹柳

[jiu22 wa22 ŋau23 hai23 lɐu23]

Jiu22 waa22 ngaau23 haai23 lau23

Jiuwaa bit the crab stick.

Appendix 2. Filler sentences in Cantonese for Experiments 1 and 3

1. 嫲嫲扶盲人

[ma21 ma21 fu21 maŋ21 jɐn21]

maa21 maa21 fu21 maang21 jan21

Grandmother supported the blind man with a hand.

2. 耀耀畫大象

[jiu22 jiu22 wak22 tai22 tsœŋ22]

jiu22 jiu22 waak22 daai22 zoeng22

Jiujiu drew an elephant.

3. 西西做拌麵

[sɐi55 sɐi55 tsou22 pun22 min22]

sai55 sai55 zou22 bun22 min22

Saisai made Lo mein (noodles).

4. 小凱運豆腐

[siu25 hɔi25 wɐn22 tɐu22 fu22]

siu25 hoi25 wan22 dau22 fu22

Siuhoi delivered the tofu.

5. 壯壯賣電視

   [tsɔŋ33 tsɔŋ33 mai22 tin22 si22]

   Zong33 zong33 maai22 din22 si22

   Zongzong sold televisions.

6. 朗朗望大雁

   [lɔŋ23 lɔŋ23 mɔŋ22 tai22 ŋaŋ22]

   Long23 long23 mong22 daai22 ngaan22

   Longlong looked at wild geese.

7. 姑媽搽牛油

   [ku55 ma55 tsʰa21 ŋɐu21 jɐu21]

   Gu55 maa55 caa21 ngau21 jau21

   The aunt spread the butter.

8. 小廣傳籃球

   [siu25 kwɔŋ25 tsʰyn21 lam21 kʰɐu21]

   siu25 gwong25 cyun21 laam21 kau21

   Siugwong passed the basketball.

9. 翠翠縫長裙

   [tsʰɵy33 tsʰɵy33 fʊŋ21 tsʰœŋ21 kʰwɐn21]

   ceoi33 ceoi33 fung21 coeng21 kwan21

   Ceoiceoi sewed a dress.

10. 永偉聞麻油

    [wɪŋ23 wɐi23 mɐn21 ma21 jɐu21]

    wing23 wai23 man21 maa21 jau21

    Wingwai smelled the sesame oil.

11. 佑佑彈提琴

    [jɐu22 jɐu22 tʰan21 tʰɐi21 kʰɐm21]

    Jau22 jau22 taan21 tai21 kam21

    Jaujau played the violin.

12. 豺狼偷西瓜

    [tsʰai21 lɔŋ21 tʰɐu55 sɐi55 kwa55]

caai21 long21 tau55 sai55 gwaa55

The wolf stole the watermelon.

13. 農民飲井水

[nʊŋ21 mɐn21 jɐm25 tsɪŋ25 sɵy25]

nung21 man21 jam25 zing25 seoi25

The farmer drank the well water.

14. 王明挖貝殼

[wɔŋ21 mɪŋ21 wat33 pui33 hɔk33]

wong21 ming21 waat33 bui33 hok33

Wongming dug the shells.

15. 黃晴買淡奶

[wɔŋ21 tsʰɪŋ21 mai23 tʰam23 nai23]

wong21 cing21 maai23 daam23 naai23

Wongcing bought the condensed milk.

16. 楊洋畫月亮

[jœŋ21 jœŋ21 wa22 jyt22 lœŋ22]

Joeng21 joeng21 waak22 jyut22 loeng22

Joengjoeng drew the moon.

Appendix 3. Target sentences in Mandarin for Experiments 2 and 4

1. 軍官背書包

   /tɕyn55 kwan55 peɪ55 ʂu55 pɑʊ55/

   Jun55 guan55 bei55 shu55 bao55

   The officer carried a schoolbag.

2. 汪叔提茶壺

   /wɑŋ55 ʂu55 tʰi35 ʈʂʰa35 xu35/

   Wang55 shu55 ti35 cha35 hu35

   Uncle Wang held a teapot.

3. 周歡買木凳

   /ʈʂoʊ55 xwan55 maɪ51 mu51 tɤŋ51/

   Zhou55 huan55 mai51 mu51 deng51

   Zhouhuan sold a wooden stool.

4. 劉寧摸貓咪

   /ljoʊ35 nɪŋ35 mɔ55 mɑʊ55 mi55/

   Liu35 ning35 mo55 mao55 mi55

   Liuning touched the cat.

5. 王姨縫棉服

   /wɑŋ35 i35 fɤŋ35 mjɛn35 fu35/

   Wang35 yi35 feng35 mian35 fu35

Aunt Wang sewed the cloths.

6. 于洋畫樹葉

/y35 jɑŋ35 xwa51 ʂu51 jɛ51/

Yu35 yang35 hua51 shu51 ye51

Yuyang drew a left.

7. 大舅抓青蛙

/ta51 tɕjoʊ51 ʈʂwa55 tɕʰɪŋ55 wa55/

Da51 jiu51 zhua55 qing55 wa55.

The uncle caught the frog.

8. 大壯投籃球

/ta51 ʈʂwɑŋ51 tʰoʊ35 lan35 tɕʰjoʊ35/

Da51 zhuang51 tou35 lan35 qiu35

Dazhuang threw the basketball.

9. 魏麗看電視

/weɪ51 li51 kʰan51 tjɛn51 ʂɚ51/

Wei51 li51 kan51 dian51 shi51

Weili watched the TV.

10. 方英開冰箱

/fɑŋ55 ɪŋ55 kʰaɪ55 pɪŋ55 ɕjɑŋ55/

Fang55 ying55 kai55 bing55 xiang55

Fangying opened the refrigerator.

11. 張音扶盲人

/ʈʂɑŋ55 ɪn55 fu35 mɑŋ35 ʐən35/

Zhang55 yin55 fu35 mang35 ren35

Zhangyin supported the blind.

12. 高冰望日落

/kɑʊ55 pɪŋ55 wɑŋ51 ʐɚ51 luɔ51/

Gao55 bing55 wang51 ri51 luo51

Gaobing looked at the sunset.

13. 王元吹風車

/wɑŋ35 ɥœn35 ʈʂʰweɪ55 fɤŋ55 /ʈʂʰɤ55/

Wang35 yuan35 chui55 feng55 che55

Wangyuan blew the windmill.

14. 農民嘗檸檬

/nʊŋ35 mɪn35 ʈʂʰɑŋ 35nɪŋ35 mɤŋ35/

Nong35 min35 chang35 ning35 meng35

The farmer tasted the lemon.

15. 吳林餵四妹

/u35 lɪn35 weɪ51 sɯ51 meɪ51/

Wu35 lin35 wei51 si51 mei51

Wulin fed the young sister.

16. 孟亮修單車

/mɤŋ51 ljɑŋ51 ɕjoʊ55 tan55 ʈʂʰɤ55/

Meng51 liang51 xiu55 dan55 che55

Mengliang fixed the bike.

17. 曼麗傳排球

/man51 li51 ʈʂʰwan35 pʰaɪ35 tɕʰjoʊ35/

Man51 li51 chuan35 pai35 qiu35

Manli passed the volleyball.

18. 趙娜賣蜜柚

/ʈʂɑʊ51 na51 maɪ51 mi51 joʊ51/

Zhao51 na51 mai51 mi51 you51

Zhaona sold honey pomelos.

Appendix 4. Filler sentences in Mandarin for Experiments 2 and 4

1. 高叔挖珍珠

   /kɑʊ55 ʂu55 wa55 ʈʂən55 ʈʂu55/

   Gao55 shu55 wa55 zhen55 zhu55

   Uncle Gao dug pearls.

2. 佳欣摘樱花

   /tɕja55 ɕɪn55 ʈʂaɪ55 ɪŋ55 xwa55/

   Jia55 xin55 zhai55 ying55 hua55

   Jiaxin picked cherry blossoms

3. 表姐喝椰汁

   /pjɑʊ214 tɕjɛ214 xɤ55 jɛ55 ʈʂɚ55/

   Biao214 jie214 he55 ye55 zhi55

   The cousin drunk the coconut.

4. 保姆拿陀螺

   /pɑʊ214 mu214 na35 tʰuɔ35 luɔ35/

   Bao214 mu214 na35 tuo35 luo35

   The babysitter held the top.

5. 李伟剷野草

   /li214 weɪ214 ʈʂʰan214 jɛ214 tsʰɑʊ214/

   Li214 wei214 chan214 ye214 cao214

   Liwei shovelled the weed.

6. 小美賣電話

   /ɕjɑʊ214 meɪ214 maɪ51 tjɛn51 xwa51/

   Xiao214 mei214 mai51 dain51 hua51

   Xiaomei sold phones.

7. 周斌打老虎

   / ʈʂoʊ55 pɪn55 ta214 lɑʊ214 xu214/

   Zhou55 bin55 da214 lao214 hu214.

   Zhoubin fought the tiger.

8. 王龍找海藻

   /wɑŋ35 lʊŋ35 ʈʂɑʊ214 xaɪ214 tsɑʊ214/

   Wang35 long35 zhao214 hai214 zao214

   Wanglong was finding seaweeds.

9. 老馬掃廣場

   /lɑʊ214 ma214 sɑʊ214 kwɑŋ214 ʈʂʰɑŋ214/

   Lao214 ma214 sao214 guang214 chang214.

   Laoma cleaned the square.

10. 夏夢撿海膽

    /ɕja51 mɤŋ51 tɕjɛn214 xaɪ214 tan214/

    Xia51 meng51 jian214 hai214 dan214

    Xiameng picked up sea urchins.

Appendix 5. Statistical results of Experiment 1

Table 3.7 Post-hoc analyses of mismatched focus conditions in Experiment 1

Focus	Tone	Modality	Focus	Tone	Modality	Estimate	SE	Z.ratio	P.value
Focus contrasts: NB vs. NC									
NB	LL	AO	NC	LL	AO	-0.073	0.324	-0.226	1.000
NB	RL	AO	NC	RL	AO	-1.116	0.387	-2.884	0.709
NB	LR	AO	NC	LR	AO	-0.775	0.322	-2.409	0.963
NB	RR	AO	NC	RR	AO	-0.854	0.389	-2.195	0.992
NB	LL	AV	NC	LL	AV	-0.825	0.317	-2.604	0.894
NB	RL	AV	NC	RL	AV	-1.279	0.446	-2.870	0.720
NB	LR	AV	NC	LR	AV	-1.054	0.331	-3.182	0.450
NB	RR	AV	NC	RR	AV	-1.081	0.372	-2.905	0.691
Focus contrasts: CB vs. NC									
CB	LL	AO	NC	LL	AO	-0.622	0.319	-1.948	0.999
CB	RL	AO	NC	RL	AO	-2.249	0.402	-5.588	0.000
CB	LR	AO	NC	LR	AO	-1.110	0.319	-3.476	0.233
CB	RR	AO	NC	RR	AO	-1.454	0.384	-3.784	0.095
CB	LL	AV	NC	LL	AV	-1.369	0.317	-4.324	0.013
CB	RL	AV	NC	RL	AV	-2.028	0.435	-4.664	0.003
CB	LR	AV	NC	LR	AV	-1.352	0.332	-4.069	0.035
CB	RR	AV	NC	RR	AV	-1.370	0.379	-3.610	0.162

Table 3.8 Post-hoc analyses of focus contrasts (NV vs. NB) in Experiment 1

Focus	Tone	Modality	Focus	Tone	Modality	Estimate	SE	Z.ratio	P.value
Focus contrasts: NV vs. NB									
NV	LL	AO	NB	LL	AO	0.131	0.328	0.399	1.000

NV	RL	AO	NB	RL	AO	0.898	0.397	2.261	0.987
NV	LR	AO	NB	LR	AO	0.726	0.326	2.230	0.989
NV	RR	AO	NB	RR	AO	1.110	0.392	2.831	0.750
NV	LL	AV	NB	LL	AV	1.297	0.334	3.883	0.069
NV	RL	AV	NB	RL	AV	0.092	0.383	0.241	1.000
NV	LR	AV	NB	LR	AV	0.908	0.320	2.842	0.742
NV	RR	AV	NB	RR	AV	1.863	0.417	4.466	0.007

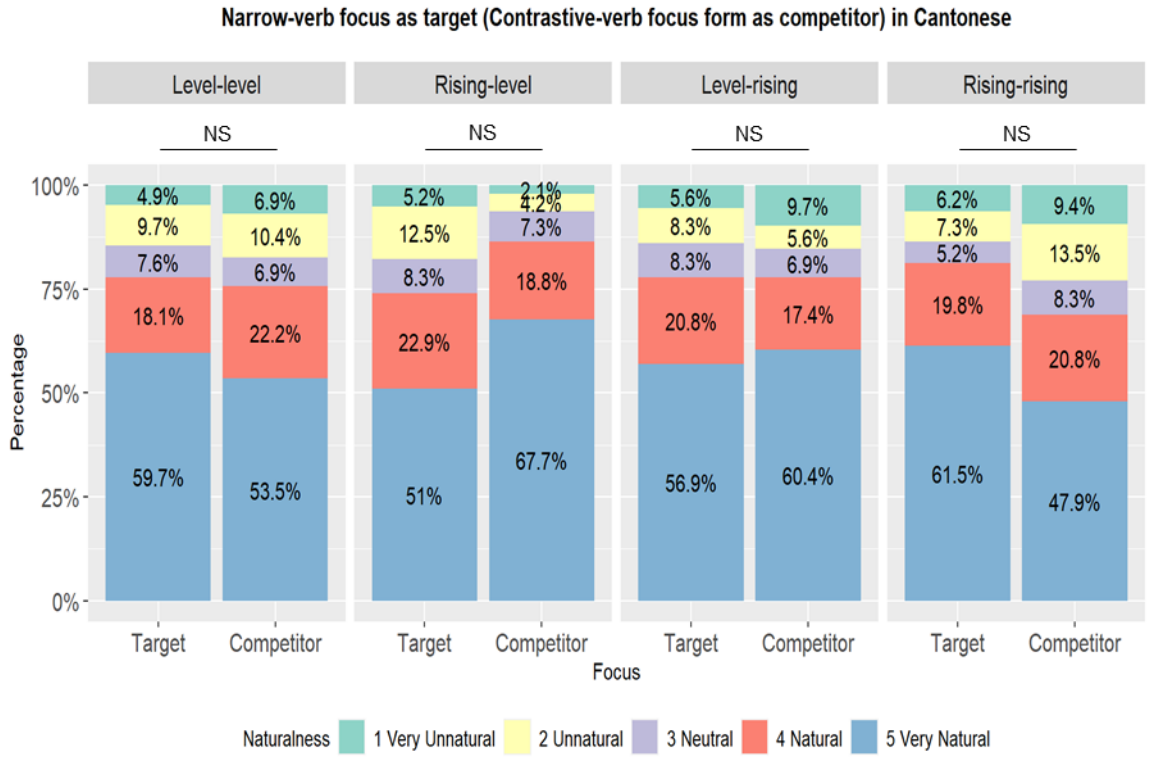


Figure 3.9 Adults' performance in Cantonese focus contrasts with different prosodic forms (Narrow-Contrastive).

Targets refer to pairs with narrow-verb focus context and matched focus forms; Competitors refer to pairs with narrow-verb focus context and mismatched contrastive-verb focus forms.

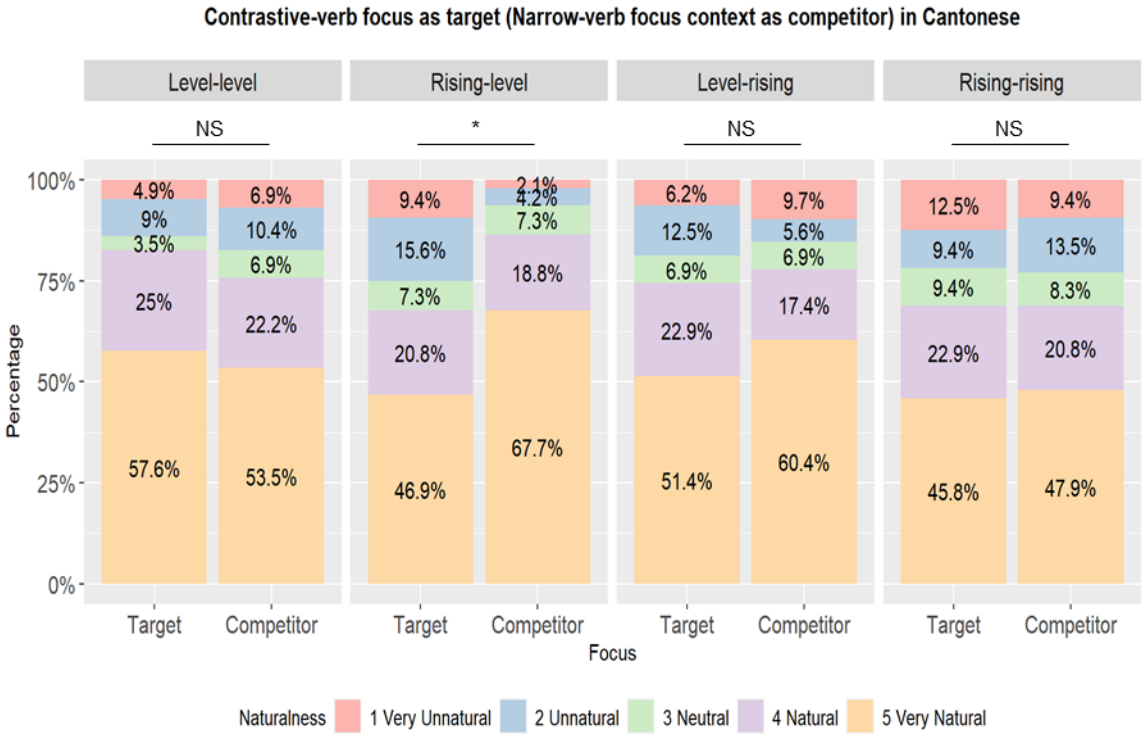


Figure 3.12 Adults' performance in Cantonese focus contrasts with different information structures (Narrow-Contrastive).

Targets refer to pairs with matched contrastive-verb focus contexts and forms; Competitors refer to pairs with mismatched narrow-verb focus contexts and contrastive-verb focus forms.

Appendix 6. Statistical results of Experiment 2

Table 3.9 Post-hoc analyses of mismatched focus conditions in Experiment 2

Group	Focus	Tone	Group	Focus	Tone	Estimate	SE	Z.ratio	P.value
Group: native Cantonese adults (CA)									
Focus contrasts: NB vs. NC									
CA	NB	TL	CA	NC	TL	-0.410	0.297	-1.381	1.000
CA	NB	TR	CA	NC	TR	-1.016	0.292	-3.475	0.156
CA	NB	TF	CA	NC	TF	-1.476	0.301	-4.907	0.001
Focus contrasts: CB vs. NC									
CA	CB	TL	CA	NC	TL	-1.395	0.292	-4.781	0.001
CA	CB	TR	CA	NC	TR	-0.603	0.299	-2.015	0.991
CA	CB	TF	CA	NC	TF	-1.822	0.296	-6.153	0.000
Group: native Mandarin adults (MA)									
Focus contrasts: NB vs. NC									
MA	NB	TL	MA	NC	TL	-2.132	0.305	-6.999	0.000
MA	NB	TR	MA	NC	TR	-3.431	0.328	-10.452	0.000
MA	NB	TF	MA	NC	TF	-2.724	0.309	-8.800	0.000
Focus contrasts: NB vs. NC									
MA	CB	TL	MA	NC	TL	-4.874	0.344	-14.171	0.000
MA	CB	TR	MA	NC	TR	-2.356	0.337	-6.988	0.000
MA	CB	TF	MA	NC	TF	-2.716	0.254	-10.686	0.000

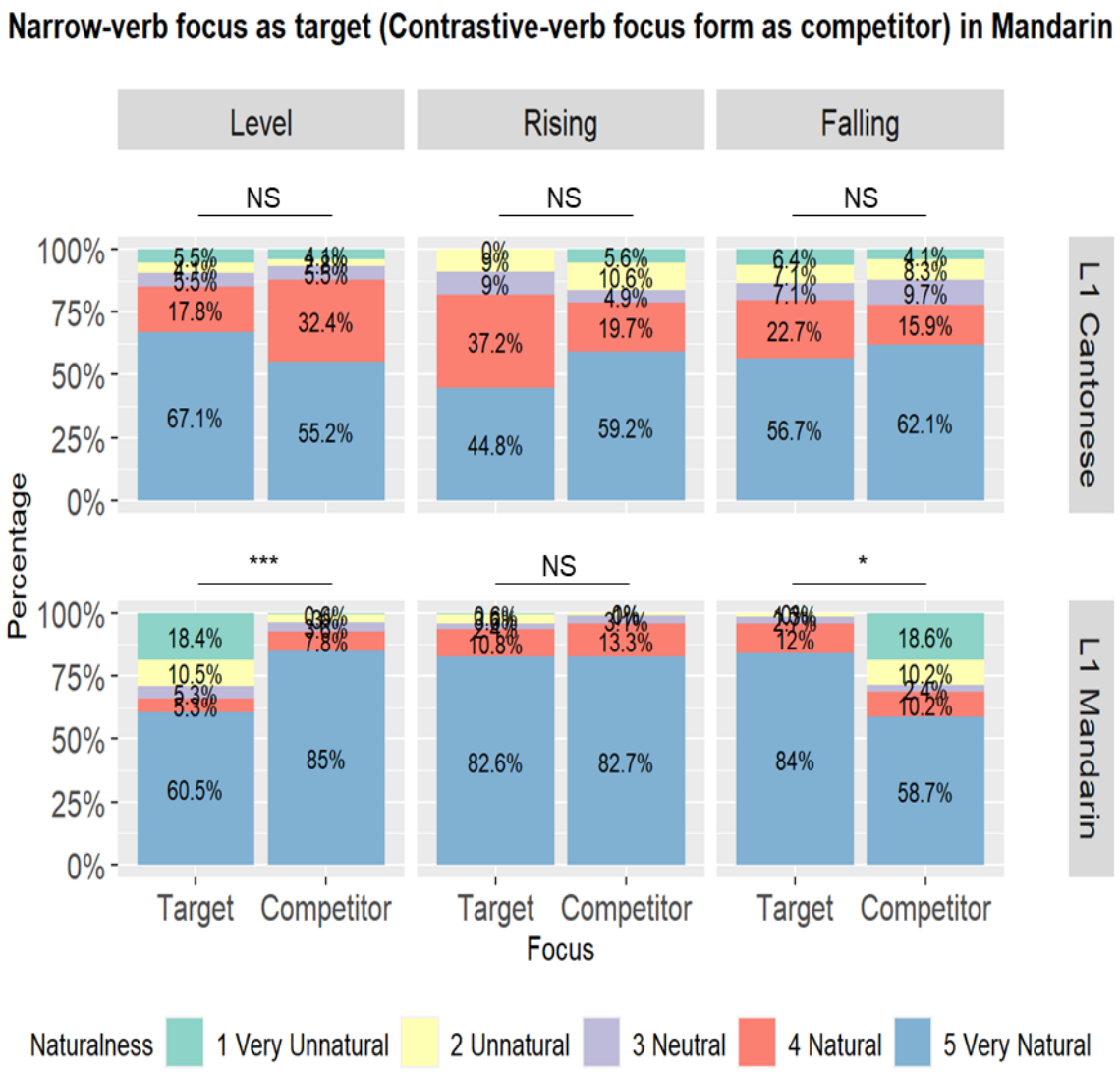


Figure 4.9 Adults' performance in Mandarin focus contrasts with different prosodic forms (Narrow-Contrastive).

Targets refer to pairs with narrow-verb focus context and matched focus forms; Competitors refer to pairs with narrow-verb focus context and mismatched contrastive-verb focus forms.

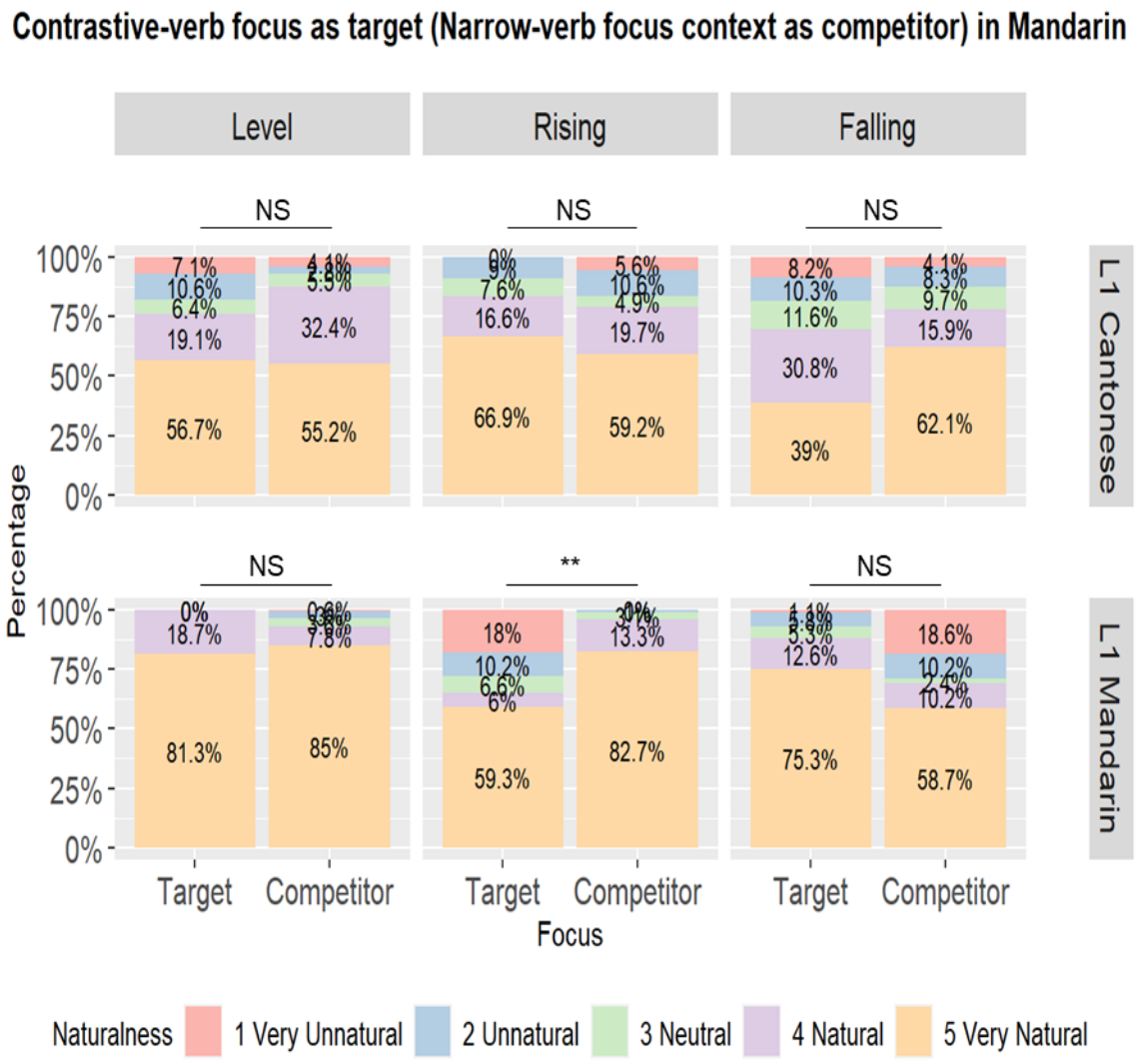


Figure 4.12 Adults' performance in Mandarin focus contrasts with different information structures (Narrow-Contrastive).

Targets refer to pairs with matched contrastive-verb focus contexts and forms; Competitors refer to pairs with mismatched narrow-verb focus contexts and contrastive-verb focus forms.

Appendix 7. Statistical results of Experiment 3

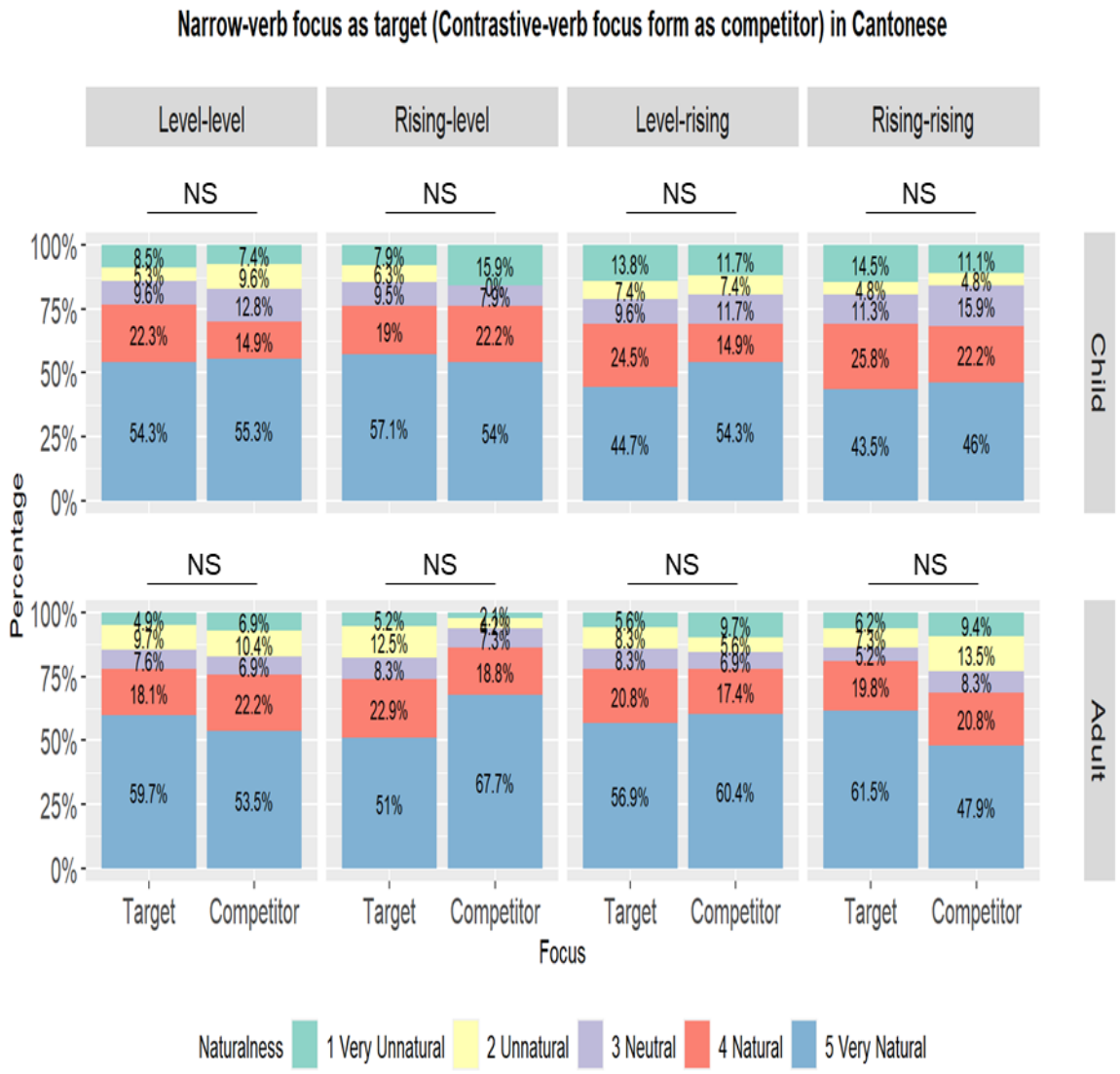


Figure 5.5 Children and adults' performance in Cantonese focus contrasts with different prosodic forms (Narrow-Contrastive).

Targets refer to pairs with narrow-verb focus context and matched focus forms; Competitors refer to pairs with narrow-verb focus context and mismatched contrastive-verb focus forms

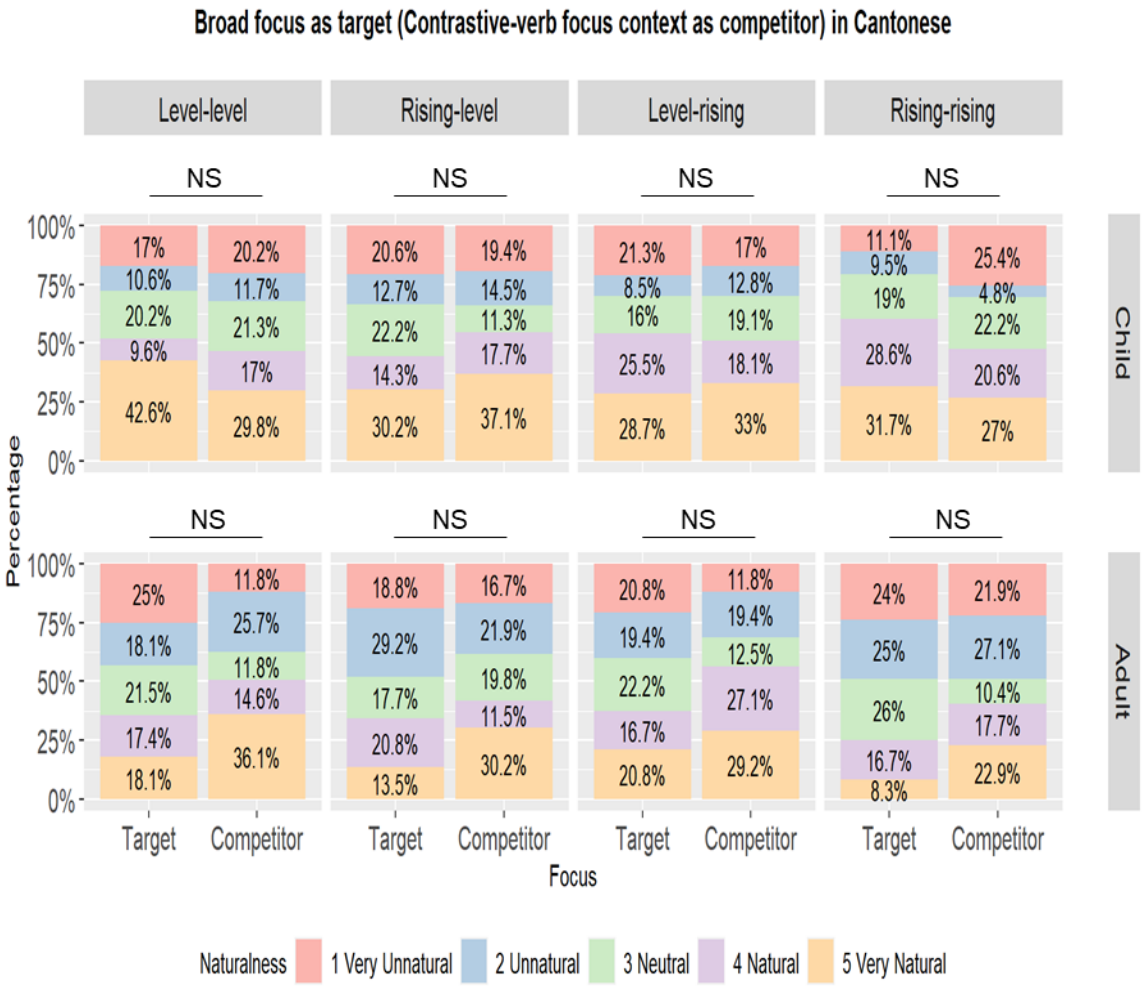


Figure 5.7 Children's performance in Cantonese focus contrasts with different information structures (Contrastive-Broad).

Targets refer to pairs with matched broad focus contexts and forms; Competitors refer to pairs with mismatched contrastive-verb focus contexts and broad focus forms.

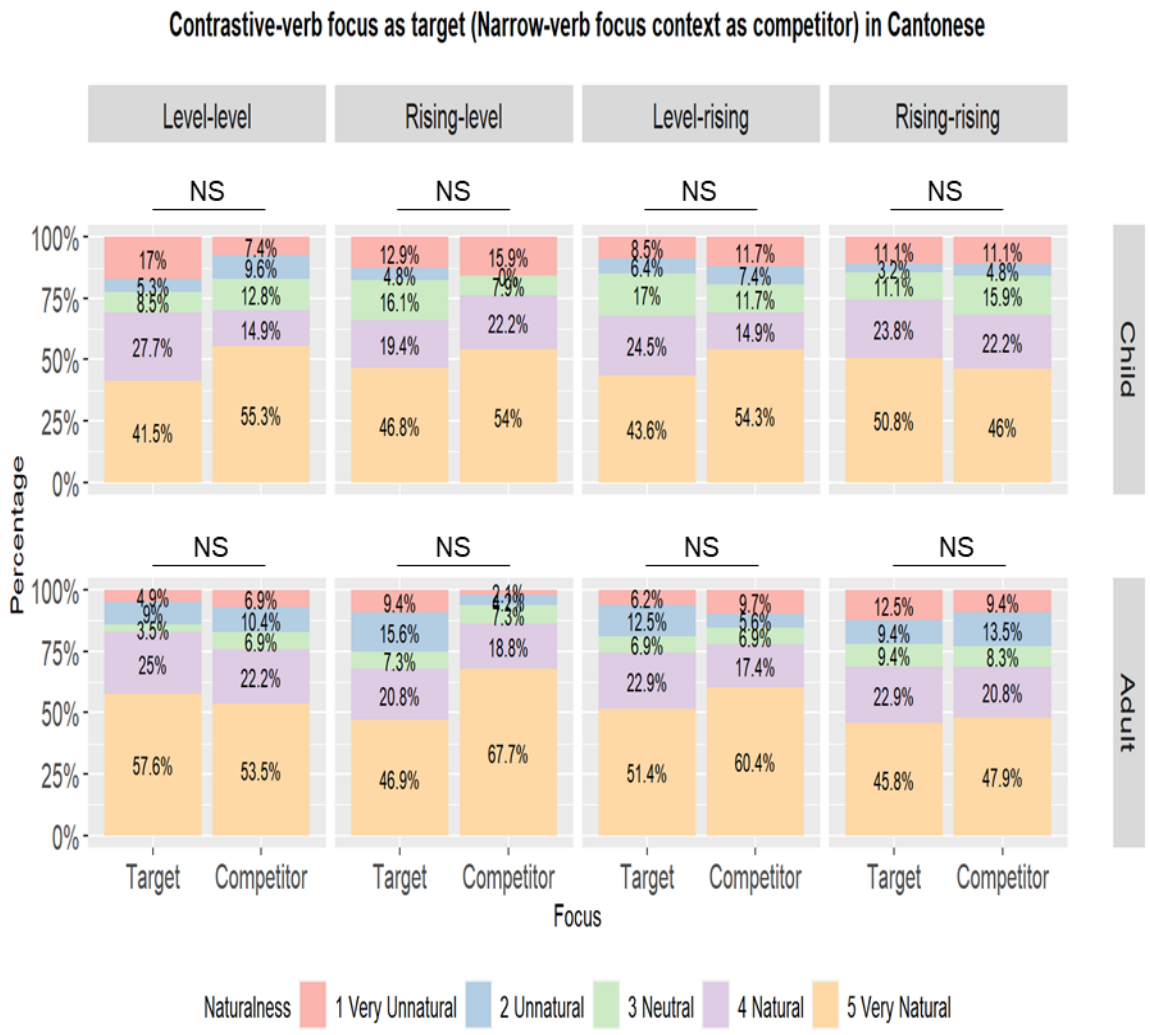


Figure 5.8 Children's performance in Cantonese focus contrasts with different information structures (Narrow-Contrastive).

Targets refer to pairs with matched contrastive-verb focus contexts and forms; Competitors refer to pairs with mismatched narrow-verb focus contexts and contrastive-verb focus forms.

Appendix 8. Statistical results of Experiment 4

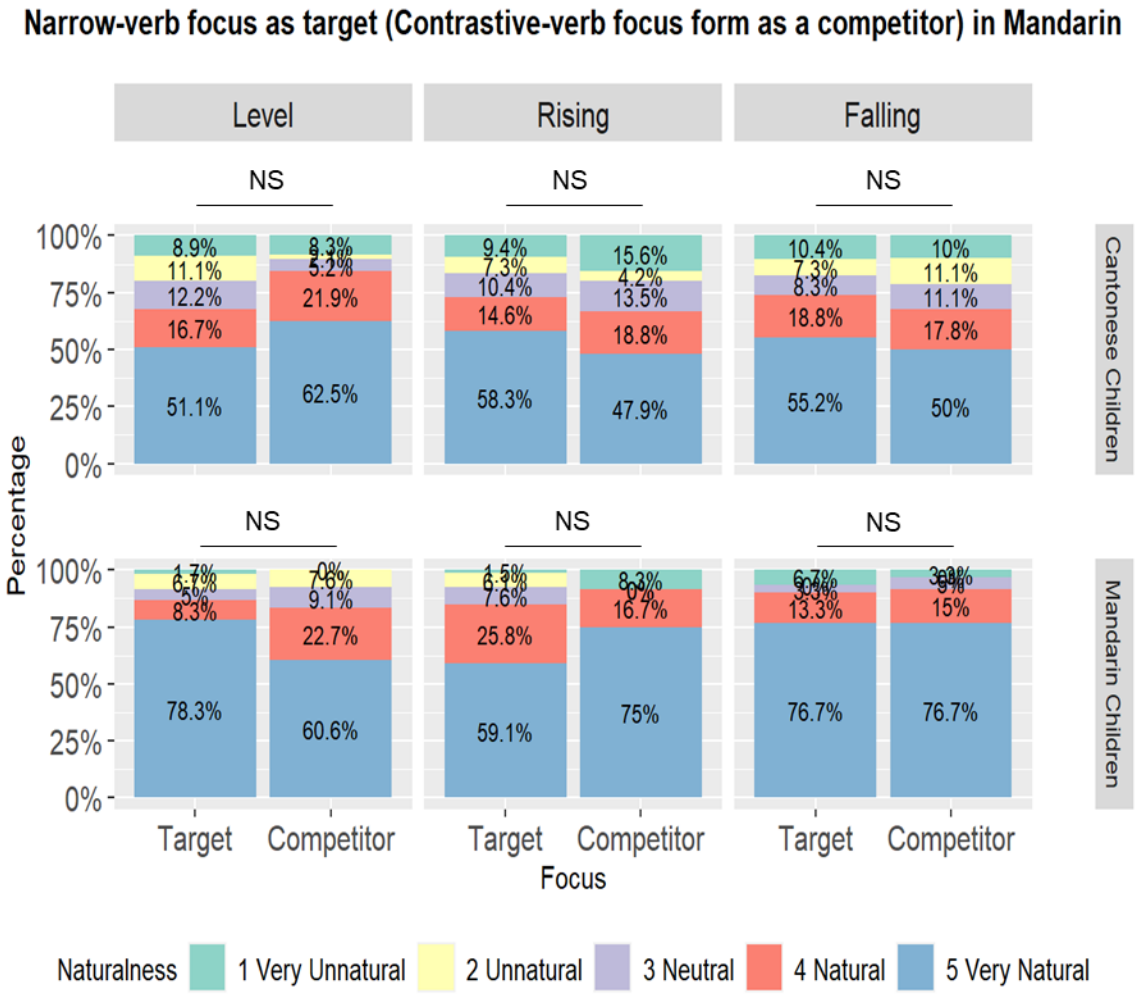


Figure 6.5 Children's performance in Mandarin focus contrasts with different prosodic forms (Narrow-Contrastive).

Targets refer to pairs with narrow-verb focus context and matched focus forms; Competitors refer to pairs with narrow-verb focus context and mismatched contrastive-verb focus forms.

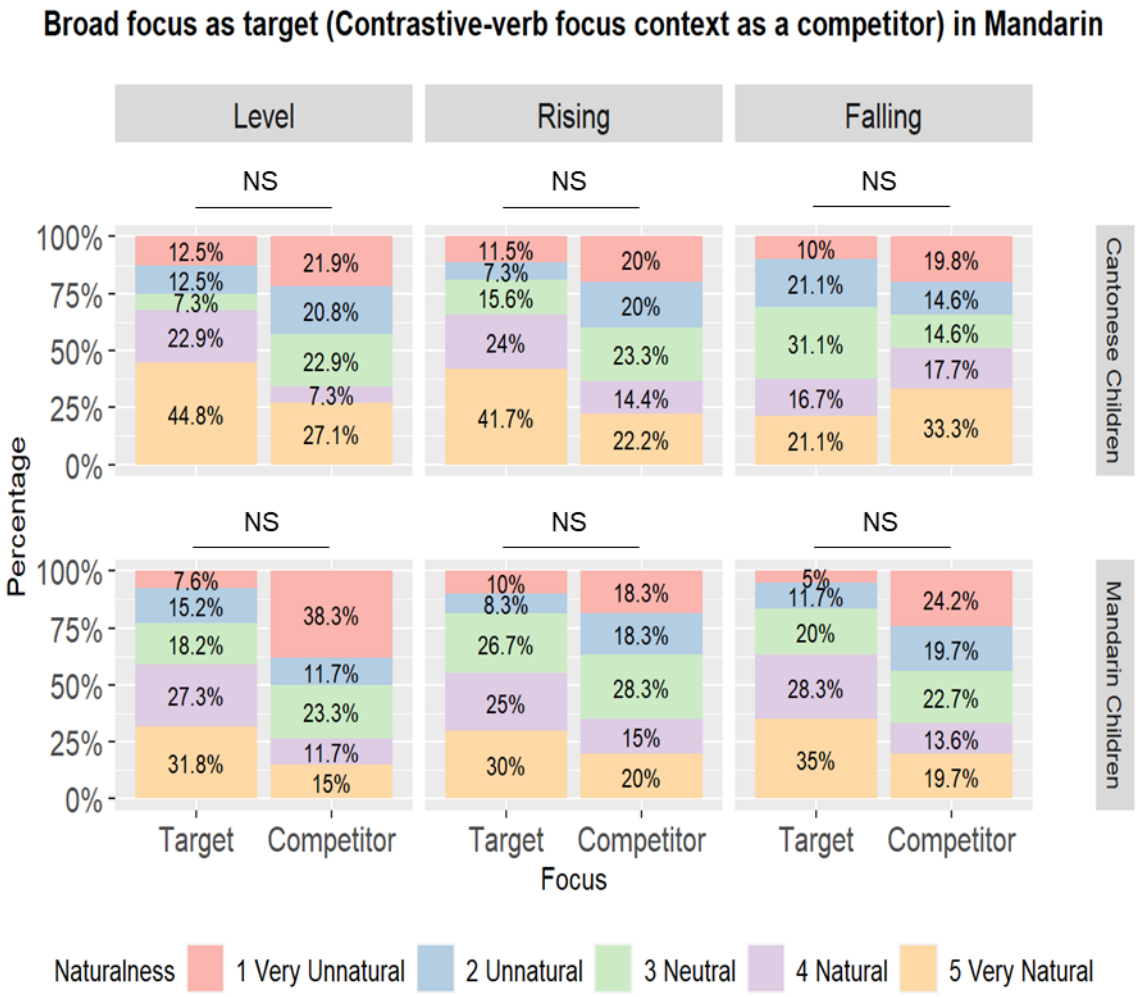


Figure 6.7 Children's performance in Mandarin focus contrasts with different information structures (Contrastive-Broad).

Targets refer to pairs with matched broad focus contexts and forms; Competitors refer to pairs with mismatched contrastive-verb focus contexts and broad focus forms.

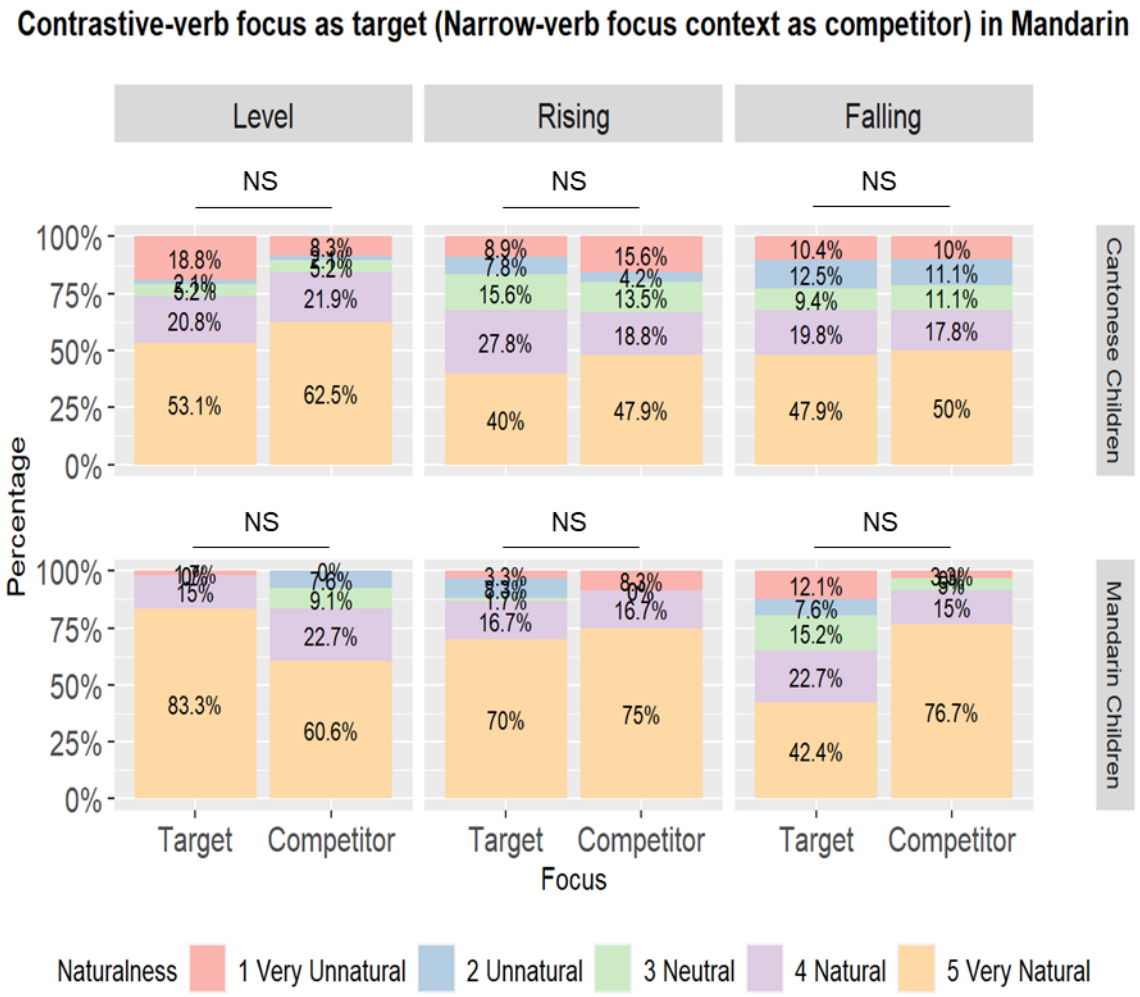


Figure 6.8 Children's performance in Mandarin focus contrasts with different information structures (Narrow-Contrastive).

Targets refer to pairs with matched contrastive-verb focus contexts and forms; Competitors refer to pairs with mismatched narrow-verb focus contexts and contrastive-verb focus forms.

References

Arnhold, A., Chen, A., & Järvikivi, J. (2016). Acquiring complex focus-marking: Finnish 4- to 5-year-olds use prosody and word order in interaction. *Frontiers in Psychology*, *7*(DEC), 1–19. https://doi.org/10.3389/fpsyg.2016.01886

Baltes, P. B., Lindenberger, U., and Staudinger, U. M. (2006). "Life-span theory in developmental psychology," in *Handbook of Child Psychology: Vol. 1. Theoretical Models of Human Development*, Series edited by W. Damon and Vol. edited by R. M. Lerner, 6th ed., (New Jersey: Wiley), 569–664.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Berlin: Mouton de Gruyter.

Baumann, S., Mertens, J., & Kalbertodt, J. (2021). The influence of informativeness on the prosody of sentence topics. *Glossa: A Journal of General Linguistics 6(1)*, *35*(1), 1–28. https://doi.org/10.16995/GLOSSA.5871

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, *70*, 20–38. https://doi.org/10.1016/j.wocn.2018.05.004

Best, C. T., Tyler, M., Bohn, O., & Munro, M. (2007). Nonnative and second-language speech perception. *Language experience in second language speech learning*, 13-34.

Beskow, J., Granström, B., & House, D. (2006). Visual correlates to prominence in several expressive modes. In *Ninth International Conference on Spoken Language Processing*.

Bergsma, Wenda (2006). (Un) stressed ook in Dutch. in Veerle van Geenhoven (ed.), *Semantics in acquisition*. Springer, Dordrecht, 329–348.

Bishop, J. (2011). English Listeners' knowledge of the broad versus narrow focus contrast. *ICPHS2011*. HK, August, 312–315.

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. *Prosody and meaning*, 25, 239.

Bishop, J. (2016). Individual differences in top-down and bottom-up prominence perception. *Proceedings of the International Conference on Speech Prosody*, *2016-Janua*(May 2016), 668–672. https://doi.org/10.21437/SpeechProsody.2016-137

Bishop, J., Kuo, G., & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors

in the perception of prosodic prominence: evidence from rapid prosody transcription. *Journal of Phonetics*, *82*(July).

Bod, R. et al., eds. (2003) *Probabilistic Linguistics,* MIT Press.

Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.02, retrieved 29 November 2022 from http://www.praat.org/

Bonawitz, E., Denison, S., Griffiths, T. L., & Gopnik, A. (2014). Probabilistic models, learning algorithms, and response variability: Sampling in cognitive development. *Trends in Cognitive Sciences*, *18*(10), 497–500.

Botinis, A., Fourakis, M., & Gawronska, B. (1999). Focus Identification in English, Greek and Swedish. *Proceedings of the XIV the International Congress of Phonetic Sciences*, *July 2014*, 1557–1560.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Brod, G., Werkle-Bergner, M., & Lee Shing, Y. (2013). The influence of prior knowledge on memory: A developmental cognitive neuroscience perspective. *Frontiers in Behavioral Neuroscience*, *7*(OCT), 1–13. https://doi.org/10.3389/fnbeh.2013.00139

Burnham, D., Attina, V., & Kasisopa, B. (2011). Auditory-Visual Discrimination and Identification of Lexical Tone Within and Across Tone Languages. *Auditory-Visual Speech Processing 2011, AVSP 2011*, 37–42.

Burnham, D., Ciocca, V., & Stokes, S. (2001). Auditory-Visual perception of lexical tone. *Eurospeech 2001 - scandinavia - 7th European Conference on Speech Communication and Technology*, 395–398.

Burnham, D., Vatikiotis-Bateson, E., Vilela Barbosa, A., Menezes, J. V., Yehia, H. C., Morris, R. H., Vignali, G., & Reynolds, J. (2022). Seeing lexical tone: Head and face motion in production and perception of Cantonese lexical tones. *Speech Communication*, *141*(March), 40–55. https://doi.org/10.1016/j.specom.2022.03.011

Büring, D. (2012). Focus and intonation. In G. Russell & D. Graff Fara (Eds.), *The Routledge companion to the philosophy of language* (pp. 103-115). London: Routledge.

Calhoun, S. (2010). The centrality of metrical structure in signaling information structure :

a probabilistic perspective. *Language*, *86*(1), 1–42.

Calhoun, S., Wollum, E., & Kruse Va'ai, E. (2021). Prosodic Prominence and Focus: Expectation Affects Interpretation in Samoan and English. *Language and Speech*, *64*(2), 346–380. https://doi.org/10.1177/0023830919890362

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific productions in intonation. *Individual Differences in Speech Production and Perception*, *3*, 123–145. https://doi.org/10.3726/978-3-653-05777-5

Carignan, C., Esteve-gibert, N., Lœvenbruck, H., Dohen, M., & Imperio, M. D. (2021). Strategies of head nod alignment with pitch prominence in French focus. In *https://issp2020. yale. edu/ProcISSP2020. pdf* (pp. 9-12). Haskins Press.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. *Subject and topic*.

Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken Chinese*. Harvard University Press.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335–344. https://doi.org/10.1016/j.tics.2006.05.006

Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua*, *120*(8), 1926–1939.

Chen, A. (2011). Tuning information packaging: intonational realization of topic and focus in child Dutch. *Journal of Child Language*, *38*(5), 1055–1083. https://doi.org/10.1017/S0305000910000541

Chen, A. (2018). Get the focus right across languages Acquisition Acquisition of prosodic focus-marking in production. In *The Development of Prosody in First Language Acquisition* (Vol. 8, Issue 3, pp. 285–301). https://doi.org/10.1075/gest.8.3.02str

Chen, A., & Bergh, H. Van Den. (2020). *The production - comprehension link in prosodic development and individual differences*.

Chen, A., & Höhle, B. (2018). Four- to five-year-olds' use of word order and prosody in focus marking in Dutch. *Linguistics Vanguard*, *4*(March).

Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. *Phonetics and phonology: Interactions and interrelations. Amsterdam:*

Benjamins.

Chen, H. (2022). *Acquisition of focus-in a cross-linguistic perspective*. PhD Dissertation.

Chen, H. C., Szendrői, K., Crain, S., & Höhle, B. (2019). Understanding Prosodic Focus Marking in Mandarin Chinese: Data from Children and Adults. *Journal of Psycholinguistic Research*, *48*(1), 19–32. https://doi.org/10.1007/s10936-018-9580-9

Chen, S. H. E. (1998). Surface cues and the development of given/new interpretation. *Applied Psycholinguistics*, *19*(4), 553–582.

Chen, S. W., Wang, B., & Xu, Y. (2009). Closely related languages, different ways of realizing focus. *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech*, 1007–1010.

Chen, S., Zhang, C., Lau, P., Yang, Y., & Li, B. (2022). Modelling representations in speech normalization of prosodic cues. *Scientific Reports*, *12*(1), 1–21. https://doi.org/10.1038/s41598-022-18838-w

Chen, Y. (2014). *Prosodic realization of focus in second language speech: Effects of language experience* [Doctoral dissertation, University of Oregon].

Chen, Y., Guion-Anderson, S., & Xu, Y. (2012). Post-focus compression in second language mandarin. *Proceedings of the 6th International Conference on Speech Prosody, SP 2012*, *1*(April 2015), 410–413.

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.

Christensen, R.H.B. (2015). Package 'ordinal'. cran.rproject.org/web/packages/ordinal/ordinal.pdf.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *Journal of the Acoustical Society of America*, 77, pp. 2142-2156.

Craik, F. I. M., and Bialystok, E. (2006). Cognition through the lifes- pan: mechanisms of change. Trends Cogn. Sci. 10, 131–138. doi:10.1016/ j.tics.2006.01.007

Cruttenden, A. (1997). Intonation (2nd ed.). Cambridge, England: Cambridge University Press.

Cvejic, E., Kim, J., & Davis, C. (2010). Prosody off the top of the head: Prosodic contrasts can be discriminated by head motion. *Speech Communication*, *52*(6), 555–564. https://doi.org/10.1016/j.specom.2010.02.006

Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(5), 441–452. https://doi.org/10.1002/wcs.1355

De Leeuw, E., I. Mennen, and J. M. Scobbie. 2012. Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism* 16:101–116.

Dohen, M., & LAœvenbruck, H. (2009). Interaction of Audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, *52*(2–3), 177–206. https://doi.org/10.1177/0023830909103166

Dohen, M., & Lœvenbruck, H. (2005). Audiovisual production and perception of contrastive focus in French: A multispeaker study. *9th European Conference on Speech Communication and Technology*, 2413–2416. https://doi.org/10.21437/interspeech.2005-49

Dohen, M., Loevenbruck, H., & Hill, H. (2006). Visual correlates of prosodic contrastive focus in french: Description and inter-speaker variability. *Proceedings of the International Conference on Speech Prosody*, 2–5.

Dupoux, E., Kaheki, K., Hirose, Y., Pallier, C. & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performanc*e 25, 1568-1578.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, 80, pp. 402-415.

Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017). The timing of head movements: The role of prosodic heads and edges. *The Journal of the Acoustical Society of America*, *141*(6), 4727–4739.

https://doi.org/10.1121/1.4986649

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, *June*, 233–277. https://doi.org/10.1111/j.1600-0404.1995.tb01710.x

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, 3-83.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language Games. Science, 336(6084), 998–998. https://doi.org/10.1126/science.1218633

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765–768.

Fung, H. S. H., & Mok, P. P. K. (2014). Realization of narrow focus in Hong Kong English declaratives—a pilot study. *Proceedings of Speech Prosody 2014*, 964–968.

Ge, H., Liu, F., Kwan, H., Aishu, Y., & Virginia, C. (2022). Comprehension of prosodically and syntactically marked focus in Cantonese - speaking children with and without autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *0123456789*. https://doi.org/10.1007/s10803-022-05770-1

Ghetti, S., DeMaster, D. M., Yonelinas, A. P., & Bunge, S. A. (2010). Developmental differences in medial temporal lobe function during memory encoding. *Journal of Neuroscience*, *30*(28), 9548–9556. https://doi.org/10.1523/JNEUROSCI.3500-09.2010

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818–829. https://doi.org/10.1016/j.tics. 2016.08.005

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90–107. https://doi.org/10.1016/j.wocn.2017.03.003

Gu, W., & Lee, T. (2007). Effects of tonal context and focus on Cantonese F_0. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007) (pp. 1033-1036).*

Gu, W., & Lee, T. (2008). Effects of Tone and Emphatic Focus on Speech Prosody - A Comparison between Standard Chinese and Cantonese. *Chinese Journal of*

Phonetics, 2, 133-147.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Foris, Dordrecht.

Gussenhoven, C. (1999). Discreteness and gradience in intonational contrast. *Language and Speech*, 42, 281 – 305.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511616983

Gussenhoven, C., & Rietveld, T. (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech*, 43(2), 183-203.

Hannah, B., Wang, Y., Jongman, A., Sereno, J. A., Cao, J., & Nie, Y. (2017). Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers. *Frontiers in Psychology*, *8*(DEC), 1–15. https://doi.org/10.3389/fpsyg.2017.02051

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279. https://doi.org/10.1016/J.WOCN.2011.11.001

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, *120*(8), 1887–1897. https://doi.org/10.1016/j.lingua.2010.02.002

Herdina, P., & Jessner, U. (2002). A dynamic model of multilingualism: Changing the psycholinguistic perspective.

Ishihara, S. (2002). Invisible but audible *Wh*-scope marking: *Wh*-constructions and deaccenting in Japanese. In *Proceedings of WCCFL* (Vol. 21, pp. 180-193).

Jia, Y., Xiong, Z., & Li, A. (2006). Phonetic and phonological analysis of focal accents of disyllabic words in Standard Chinese. In *Chinese Spoken Language Processing: 5th International Symposium*, ISCSLP 2006, Singapore, December 13-16, 2006. Proceedings (pp. 55-66). Springer Berlin Heidelberg.

Kim, J., Cvejic, E., & Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Communication*, *57*, 317–330. https://doi.org/10.1016/j.specom.2013.06.003

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust Speech Perception: Recognize the Familiar, Generalize to the Similar, and Adapt to the Novel. *Psychological Review*,

305–333. https://doi.org/10.1002/9781119203063.ch12

Kleinschmidt, D. F., Weatherholtz, K., & Florian Jaeger, T. (2018). Sociolinguistic Perception as Inference Under Uncertainty. *Topics in Cognitive Science*, *10*(4), 818–834. https://doi.org/10.1111/tops.12331

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, 118(2), 1038-1054.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*(1), 29–53. https://doi.org/10.1177/00238309050480010201

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

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust Speech Perception: Recognize the Familiar, Generalize to the Similar, and Adapt to the Novel. *Psychological Review*, 305–333. https://doi.org/10.1002/9781119203063.ch12

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, cognition and neuroscience*, 31(1), 32-59.

Kurumada, C., & Roettger, T. B. (2021). Thinking probabilistically in the study of intonational speech prosody. In *Wiley Interdisciplinary Reviews: Cognitive Science* (Vol. 13, Issue 1). https://doi.org/10.1002/wcs.1579

Kurumada, C., & Roettger, T. B. (2022). Thinking probabilistically in the study of intonational speech prosody. In *Wiley Interdisciplinary Reviews: Cognitive Science* (Vol. 13, Issue 1). John Wiley and Sons Inc. https://doi.org/10.1002/wcs.1579

Hüttner, T., Drenhaus, H., Van de Vijver, R., & Weissenborn, J. (2004, June). The acquisition of the German focus particle auch ‘too’: Comprehension does not always precede production. In *Proceedings of the 28th Annual Boston University Conference on Language Development*.

Jessner, U. (2006). *Linguistic awareness in multilinguals*. Edinburgh, UK: Edinburgh University Press.

Ladd, DR. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Ind. Univ. Press.

Ladd, DR. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Ladd, DR. (2008). *Intonational Phonology*. Cambridge, UK: Cambridge Univ. Press. 2nd ed.

Lambrecht, K. (1994). Information Structure and Sentence Form. *Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Lee, A., Chiu, F., & Xu, Y. (2022). Focus perception in Japanese: Effects of lexical accent and focus location. *PLoS ONE*, *17*(9 September), 1–18. https://doi.org/10.1371/journal.pone.0274176

Lee, Y., & Xu, Y. (2010). Phonetic realization of contrastive focus in Korean. *Proceedings of Speech Prosody 2010*, 100033:1-4.

Lenth R (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.4-1, <https://CRAN.R-project.org/package=emmeans>.

Li, B., Guan, Y., & Chen, S. (2020). Carryover effects on tones in Hong Kong Cantonese. *Proceedings of the International Conference on Speech Prosody*, *2020-May*(May), 489–493. https://doi.org/10.21437/SpeechProsody.2020-100

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, Fang & Xu, Y. 2007. Question intonation as affected by word stress and focus in English. *The 16th International Congress of Phonetic Sciences*, 1189–1192. Saarbrücken.

Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition*, *174*(February), 55–70. https://doi.org/10.1016/j.cognition.2018.01.003

Liu, Z., Chen, A., & Velde, H. Van De. (2016). Prosodic Focus Marking in Minority L1 Bai-Children Learning Mandarin Chinese as L2. *Proceedings of the 40th annual Boston University Conference on Language Development*.

Magne, C., Astesano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of "pop-out'' words in spoken French dialogues. *Journal of Cognitive Neuroscience*, 17(5), 740–756. https://doi.org/10.1162/0898929053747667.

Man, C. H. V. (1999). *An acoustic study of the effects of sentential focus on Cantonese*

tones.

Manning, C. (2003) Probabilistic Syntax. In *Probabilistic Linguistics* (Bod, R. et al., eds), pp. 289–341, MIT Press.

Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar*. Routledge.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748.

Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations* (pp. 171-188). Berlin, Heidelberg: Springer Berlin Heidelberg.

Mennen, I., F. Schaeffler, and C. Dickie. 2014. Second language acquisition of pitch range in Ger man learners of English. *Studies in Second Language Acquisition* 36:303–329.

Ménard, L., Loevenbruck, H., & Savariaux, C. (2006). Articulatory and acoustic correlates of contrastive focus in French children and adults. *Speech Production: Models, Phonetic Processes and Techniques*, 227–251.

Mok, P. P. K., Fung, H. S. H., & Li, V. G. (2019). Assessing the link between perception and production in Cantonese tone acquisition. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1243-1257.

Mok, P. P. K., Li, V. G., & Fung, H. S. H. (2020). Development of phonetic contrasts in cantonese tone acquisition. *Journal of Speech, Language, and Hearing Research*, *63*(1), 95–108. https://doi.org/10.1044/2019_JSLHR-19-00152

Nava, E., & Zubizarreta, M. L. (2008). Prosodic Transfer in L2 Speech: Evidence from Phrasal Prominence and Rhythm. *Proceedings of Speech Prosody 2008*.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18. https://doi.org/10.1080/23273798.2015.1081703

Pannekamp, A., Van Der Meer, E., & Toepel, U. (2011). Context- and prosody-driven ERP markers for dialog focus perception in children. *Brain Topography*, *24*(3–4), 229–242. https://doi.org/10.1007/s10548-011-0194-x

Paterson, K. B., Liversedge, S. P., Rowland, C., & Filik, R. (2003). Children's

comprehension of sentences with focus particles. *Cognition*, 89(3), 263-294.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Unpublished PhD dissertation. MIT.

Prieto i Vives, P., & Esteve-Gibert, N. (2018). The development of prosody in first language acquisition. *The Development of Prosody in First Language Acquisition*.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/.

Psychology Software Tools, Inc. [E-Prime Go]. (2020). Retrieved from https://support.pstnet.com/.

R Core Team. (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Rapin, L., & Ménard, L. (2019). The Multimodal Perception of Contrastive Focus in French: A Developmental Study. *Frontiers in Communication*, *3*(February), 1–11. https://doi.org/10.3389/fcomm.2018.00060

Roettger, T. B., & Franke, M. (2019). Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation. *Cognitive Science*, *43*(7). https://doi.org/10.1111/cogs.12745

Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning–the case of information structure in American English. *Language, Cognition and Neuroscience*, *34*(7), 841–860. https://doi.org/10.1080/23273798.2019.1587482

Roettger, T. B., & Rimland, K. (2020). Listeners' adaptation to unreliable intonation is speaker-sensitive. *Cognition*, *204*(November 2019), 104372. https://doi.org/10.1016/j.cognition.2020.104372

Romøren, A. S. H., & Chen, A. (2017). The acquisition of focal lengthening in Stockholm Swedish. *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech*, *2017-Augus*, 699–703. https://doi.org/10.21437/Interspeech.2017-1065

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics: An International Journal of Semantics and Its Interfaces in Grammar*, *1*(1), 75–116. https://doi.org/10.1007/BF02342617

Rump, H.H., Collier, R. 1996. Focus conditions and prominence of pitch-accented syllables. *Language and Speech, 29*, 1-17.

Samuel, Arthur (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology*, General 110: 474-494.

Scarborough, R., Keating, P., Mattys, S. L., Cho, T., & Alwan, A. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech*, 52(2-3), 135-175.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception and Psychophysics*, *59*(1), 73–80. https://doi.org/10.3758/BF03206849

Sekiyama, K., Burnham, D., Tam, H., & Erdener, D. (2003). *Auditory-visual speech perception development in Japanese and English speakers*. In *AVSP 2003-International Conference on Audio-Visual Speech Processing*.

Selkirk, Elisabeth. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT press.

Selkirk, Elisabeth. (1996). Sentence prosody: Intonation, stress, and phrasing. *The hand-book of phonological theory*, ed. by John A. Goldsmith, 550–69. Oxford: Blackwell.

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, *58*(2), 574–608. https://doi.org/10.1016/j.jml.2007.08.001

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). *Predictive top-down integration of prior knowledge during speech perception*. *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Steffman, J. (2019). Phrase-final lengthening modulates listeners' perception of vowel duration as a cue to coda stop voicing. *The Journal of the Acoustical Society of America*, *145*(6), EL560–EL566. https://doi.org/10.1121/1.5111772

Surányi, B., & Pintér, L. (2021). Children's comprehension of prosodically marked focus in Hungarian: How mandatory syntactic focus-marking affects the trajectory of acquisition. *Journal of Child Language*, 824–838. https://doi.org/10.1017/S0305000921000313

Swerts, M., & Krahmer, E. (2004). Congruent and incogruent audiovisual cues to

prominence. *Proceedings of the Speech Prosody 2004 Conference, Nara (Japan), March 23-26, 2004*, *1*, 69–72.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94. https://doi.org/10.1016/j.jml.2005.02.003

Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, *36*(2), 219–238. https://doi.org/10.1016/j.wocn.2007.05.001

Szendrői, K., Bernard, C., Berger, F., Gervain, J., & Höhle, B. (2018). Acquisition of prosodic focus marking by English, French, and German three-, four-, five- and six-year- olds. *Journal of Child Language*, *45*(1), 219–241. https://doi.org/10.1017/S0305000917000071

Tang, P., Yuen, I., Demuth, K., & Rattanasone, N. X. (2022). The Acquisition of Contrastive Focus During Online Sentence-Comprehension by Children Learning Mandarin Chinese. *Developmental Psychology*. https://doi.org/10.1037/dev0001498

Toepel, U., Pannekamp, A., & Alter, K. (2007). Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions*, *3*, 1–13. https://doi.org/10.1186/1744-9081-3-53

Toepel, U., Pannekamp, A., & van der Meer, E. (2009). Fishing for information: The interpretation of focus in dialogs. *Brain talk: discourse with and in the brain*, 175-190.

Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America*, 99(6), 3782-3790.

Turnbull, R., Royer, A. J., Ito, K., & Speer, S. R. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience*, *32*(8), 1017–1033. https://doi.org/10.1080/23273798.2017.1279341

Vainio, M., & Järvikivi, J. (2006). Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics*, *34*(3), 319–342. https://doi.org/10.1016/j.wocn.2005.06.004

Visser, M., Krahmer, E., & Swerts, M. (2011). Children's expression of uncertainty in collaborative and competitive contexts. *Auditory-Visual Speech Processing 2011,*

AVSP 2011, 25–30.

Wagner, P. (2005). Great expectations - Introspective vs. perceptual prominence ratings and their acoustic correlates. *9th European Conference on Speech Communication and Technology*, 2381–2384.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science,* 167(3917), 392–393. doi:10.1126/science. 167.3917.392

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, 40(1), 141-151.

Wong, Y. W. (2007). *Production and perception of tones in Cantonese continuous speech*. *January*.

Wonnacott, E., & Watson, D. G. (2008). Acoustic emphasis in four year olds. *Cognition,* 107, 1093–1101.

Wrembel, M. (2015). In search of a new perspective: Cross-linguistic influence in the acquisition of third language phonology. http://hdl.handle.net/10593/14647

Wu, W. L. (2013). Cantonese prosody: Sentence-final particles and prosodic focus. *PhD Thesis*.

Wu, W. L., & Xu, Y. (2010). Prosodic Focus in Hong Kong Cantonese without post-focus Compression. *Proceedings Speech Prosody 2010*, 1–4. https://doi.org/10.1007/s10831-011-9087-y

Xie, X., Buxó-Lugo, A., & Kurumada, C. (2021). Encoding and decoding of meaning through structured variability in intonational speech prosody. *Cognition*, *211*(February). https://doi.org/10.1016/j.cognition.2021.104619

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f 0 contours. *Journal of Phonetics*, *27*(1), 55–105. https://doi.org/10.1006/jpho.1999.0086

Xu, Y. (2011). Speech prosody: A methodological review. *Journal of Speech Sciences*, *1*(1), 85–115. http://discovery.ucl.ac.uk/1322252/

Xu, Y. (2013). ProsodyPro - A tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France. 7-10.

Xu, Y., Chen, S. W., & Wang, B. (2012). Prosodic focus with and without post-focus

compression: A typological divide within the same language family? *Linguistic Review*, *29*(1), 131–147. https://doi.org/10.1515/tlr-2012-0006

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319–337. https://doi.org/10.1016/S0167-6393(00)00063-7

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, A. (2017). *The acquisition of prosodic focus-marking in Mandarin Chinese- and Seoul Korean-speaking children.* PhD Thesis.

Yang, A., & Chen, A. (2014). Prosodic focus-marking in Chinese four- and eight-year-olds. *Speech Prosody*, 1–5.

Yang, A., & Chen, A. (2018). The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children. *First Language*, *38*(1), 26–46. https://doi.org/10.1177/0142723717733920

Yang, A., Cho, T., Kim, S., & Chen, A. (2015). Phonetic focus-marking in Korean-speaking 7- to 8-year-olds and adults. *ICPhS 2015*, *1*, 1–5.

Yang, Y. (2022). *First language attrition and second language attainment of Mandarin-speaking immigrants in Hong Kong Evidence from prosodic focus.* PhD Thesis.

Yasuko, Hayashi & Kaoru, S. (1998). Native-foreign langage effect in the mcgurk effect: a test with Chinese and Japanese. *In AVSP'98 International Conference on Auditory-Visual Speech Processing.*

Zellin, M., Pannekamp, A., Toepel, U., & van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, *81*(3), 133–141. https://doi.org/10.1016/j.ijpsycho.2011.05.009

Zhang, Y., Chen, X., Chen, S., Meng, Y., & Lee, A. (2023). Visual-auditory perception of prosodic focus in Japanese by native and non-native speakers. *Frontiers in Human Neuroscience*, *17*.

Zheng, X., & Pierrehumbert, J. B. (2010). The effects of prosodic prominence and serial position on duration perception. *The Journal of the Acoustical Society of America*, 128(2), 851-859.

Zhou, P., Crain, S., & Zhan, L. (2012a). Sometimes children are as good as adults: The

pragmatic use of prosody in children's on-line sentence processing. *Journal of Memory and Language*, *67*(1), 149–164. https://doi.org/10.1016/j.jml.2012.03.005

Zhou, P., Su, Y., Crain, S., Gao, L., & Zhan, L. (2012b). Children's use of phonological information in ambiguity resolution: A view from Mandarin Chinese. *Journal of Child Language*, *39*(4), 687–730. https://doi.org/10.1017/S0305000911000249

Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children*. Clevedon: Multilingual Matters Ltd.