FIRST LANGUAGE ATTRITION AND SECOND LANGUAGE ATTAINMENT OF MANDARIN-SPEAKING IMMIGRANTS IN HONG KONG: EVIDENCE FROM PROSODIC FOCUS

YIKE YANG

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University
Department of Chinese and Bilingual Studies

First Language Attrition and Second Language Attainment of Mandarin-speaking Immigrants in Hong Kong: Evidence from Prosodic Focus

Yike Yang

A thesis submitted
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

May 2021

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)

_________Yike Yang_________________ (Name of student)

Abstract

In the field of bilingual language development, studies on first language (L1) attrition and second language (L2) attainment have been two separate streams of research. The present thesis represents an attempt to link L1 attrition and L2 attainment and aims to investigate the language development of new immigrants in Hong Kong in depth from the perspective of phonetics, which will increase our understanding of the issues involved in L1 and L2 speech interaction.

Two experiments were conducted with native speakers of Mandarin and Cantonese and with Mandarin-speaking immigrants who had learned Cantonese after their arrival in Hong Kong. In the production experiment, the participants answered questions that elicited various types of focus in Mandarin and Cantonese. The speakers of Mandarin and Cantonese exhibited differences in the use of acoustic cues in marking focus in their native languages. The immigrants' data suggested the bidirectional influences of their L1 and L2, which were reflected in the acoustic measurements of F0, duration and intensity compared to the same metrics for the native speakers. In the perception experiment, the participants were instructed to map between prosody and focus. The results did not show any attrition of the immigrants' L1 Mandarin, and the immigrants showed even higher accuracy rates than native speakers in the Cantonese tasks, which

was attributed to a potential bilingual advantage in speech perception, particularly for the perception of prosody.

Examining the combined results, there was evidence of L1 Mandarin attrition in production but not in perception, and the immigrants were more attuned to acoustic cues than the native Cantonese speakers. As the existing speech learning models cannot explain our data adequately, we proposed a working model (the Bilingual Prosody Transfer Model, or BPTM) to account for the findings of this thesis and to provide a reference for future work on the prosody of bilingual speakers, which requires further testing and refinement. Moreover, to understand the issues surrounding language attrition and acquisition in more detail, future research should explore the developmental sequences of immigrants' L1 and L2, include more language pairs, measure other aspects of speech and language and consider individual variations in speech production and perception.

Acknowledgements

First and foremost, I would like to show my highest gratitude and appreciation to my chief supervisor, Dr Si Chen, for her unlimited guidance, patience and encouragement throughout my graduate study at The Hong Kong Polytechnic University. I am also thankful to my co-supervisor, Dr Roxana Fung, for her valuable feedback on the experimental design and data analysis of the thesis. Moreover, I want to thank the committee members for their constructive suggestions and insightful discussion before and during my oral examination, which really were a tremendous help for a junior researcher like me: Dr Yao Yao (BoE Chair), Professor Ying Chen (External Examiner) and Professor Yiya Chen (External Examiner).

My special thanks go to Professor Chu-Ren Huang for not only sharing various fascinating topics of and approaches to language studies but also providing informative comments on an earlier version of this thesis. I would also like to express my sincere gratitude to Dr Angel Chan, Dr Xiaocong Chen, Dr Yu-Yin Hsu, Dr Steven Lai, Dr Helena Lau, Professor David C. S. Li, Dr Stephen Politzer-Ahles, Dr Jing Shao, Dr Gerald Stell, Dr Doreen Wu and Dr Caicai Zhang for the invaluable discussion on research design, methodology and data analysis.

I am very grateful to all my teachers from Southwest University, The Chinese University of Hong Kong as well as The Hong Kong Polytechnic University, who have helped to shape my thoughts. I appreciate the help from the administration colleagues in Department of Chinese and Bilingual Studies, Faculty of Humanities, Research Office and Graduate School of The Hong Kong Polytechnic University. For the experiments reported in this thesis, I would like to thank all the informants for their participation in one or more tasks, without which this thesis would not have been possible.

Part of this thesis was supported by an ASA International Student Grant from the Acoustical Society of America, to which I would like to express my deep gratitude. I also thank the following academic associations for providing a travel grant or subsidy to me over these years: The International Speech Communication Association, The Acoustical Society of America and The International Society of Experimental Linguistics.

Last but not least, I would like to thank my family and friends for their unconditional love and support throughout these years. This thesis is dedicated to my parents and grandparents.

Table of contents

List of figures

List of tables

List of abbreviations

ACC	Accusative case
BPTM	Bilingual Prosody Transfer Model
CAH	Category assimilation hypothesis
CDH	Category assimilation hypothesis
CL	Classifier
F0	Fundamental frequency
L1	First language
L2	Second language
LDA	Linear discriminant analysis
LILt	L2 Intonation Learning Theory
NOM	Nominative case
NP	Nominal phrase
PAM	Perceptual Assimilation Model
PAM-L2	Perceptual Assimilation Model of Second Language Speech Learning
PFC	Post-focus compression
SFP	Sentence-final particle
SLM	Speech Learning Model
SLM-r	Revised Speech Learning Model
SVO	Subject–verb–object
T1	Tone 1
VOT	Voice onset time
VP	Verb phrase

Chapter 1: Introduction

Late second language (L2) learners frequently retain a foreign accent even after years of extensive exposure to the target language (Flege et al., 1999); during L2 acquisition, their first language (L1) may also undergo some type of alteration due to the influence of the L2, which is defined as the process of L1 attrition (Schmid, 2002). However, examinations on L1 attrition and L2 acquisition have been two separate streams of research on bilingual language development. This study is an attempt to link L1 attrition and L2 attainment and aims to investigate the language development of new immigrants in Hong Kong in depth from the perspective of phonetics, which will increase our understanding of the issues involved in L1 and L2 speech interaction.

This chapter first presents a comprehensive review of topics that are relevant to the scope of this thesis and introduces the significance of and research questions for the current study. Chapter 2 describes the focus production experiments performed by native speakers of Mandarin and Cantonese, while Chapter 3 reports on the focus production experiments performed by immigrants who were invited to produce the utterances in both their L1 (Mandarin) and in their L2 (Cantonese). Chapter 4 shows how native speakers of Mandarin and Cantonese map prosody and focus in their perception, while Chapter 5 presents the mapping tasks performed by the immigrants. Lastly, concluding remarks are presented in Chapter 6.

1.1 L1 attrition and L2 attainment

1.1.1 L1 attrition

'Sequential bilinguals' refers to bilingual speakers who begin to learn their L1 before their L2. Most investigations of sequential bilinguals' speech have focused on the characteristics of their L2s, and have assumed that their L1 will remain intact during the L2 acquisition process. However, recent studies have revealed that an L2 may also interfere with a full-fledged L1, causing the L1 of the learners to diverge from that of monolingual speakers (Ulbrich & Ordin, 2014). These findings are in line with the multicompetence model (Cook, 1991, 2016), which suggests that bilinguals have an 'overall system of a mind or a community that uses more than one language' (Cook, 2016: 2), and thus are not equivalent to two monolinguals. L1 attrition refers to 'a gradual loss' (Schmid, 2002: 24) of one's L1 due to ageing, language disorders or the decreased use of the language. The focus of the present study is the non-age related, non-pathological L1 attrition of late bilinguals (that is, those who begin to learn an L2 after having fully acquired their L1); these bilinguals are usually immigrants who have relocated to a new environment in which their L2 is the dominant language, and their L1 is no longer used or is used less frequently (Köpke & Schmid, 2004). Beginning with the same level of L1 competence, the late bilinguals eliminate the potential maturational constraints in child and adolescent language development (Flege et al., 2006), and are thus ideal population for the investigation of language attrition. Note that L1 attrition does not necessarily imply that the language users have partially or completely lost their abilities in the L1; instead, any L2-induced changes in the L1 are regarded as L1 attrition (e.g., Chamorro, Sturt, & Sorace, 2016).

Mixed results have been found in the attrition of various linguistic domains, such as phonetics and phonology (Oh et al., 2011), lexicon (Baus et al., 2013), morphology (Keijzer, 2010), syntax (Gürel, 2004) and semantics (Chamorro et al., 2016). In the phonetic domain, while some participants manage to maintain their L1 extremely well and do not show significant differences from monolingual speakers after years of exposure to the L2 (Hopp & Schmid, 2013), more studies suggest an L2 influence on L1 production. For example, voice onset time (VOT) in the L1 has been shown to be affected by the L2 in several language pairs, such as English-Spanish (Flege & Eefting, 1987), Japanese-English (Harada, 2003) and English-Portuguese (Major, 1992). Japanese has shorter VOT values than English, and Japanese-English bilinguals produce longer Japanese VOTs than monolingual Japanese speakers, which has been attributed to an L2 influence (Harada, 2003). At the suprasegmental level, de Leeuw, Mennen and Scobbie (2012) provided evidence of L1 attrition in the intonational alignment of pre-nuclear rise, as the German-English bilingual group consistently showed earlier starting points for pre-nuclear rise in German compared to the native control group, although there were individual variations among the bilinguals.

1.1.2 L2 attainment

The notion of L2 ultimate attainment refers to the end state that one can eventually achieve in L2 acquisition, regardless of whether the L2 becomes convergent with or remains divergent from the target language (Birdsong, 2004). Although some domains of linguistic knowledge have been proved to be acquirable in the end state, such as syntax (Sorace & Filiaci, 2006) and lexicon (Saito, 2015), it has been argued that late L2 learners are unlikely to attain native-like pronunciation (Singleton, 2005). The retention of a foreign accent in L2 ultimate attainment has been attributed to the effects of age according to the critical period hypothesis (Singleton, 2005), and has been supported by several studies in which early learners significantly outperformed late learners (Abrahamsson & Hyltenstam, 2009). Baker et al. (2008) examined the production of English vowels by Korean adults and children; the children's pronunciation was more accurate than that of the adults, at least for English vowels that did not have close counterparts in Korean. However, it is not yet clear whether the segmental and prosodic features of the L2 become more native-like after years of immersion in an L2, or whether a foreign accent is still perceivable.

Some earlier studies have suggested a relatively straightforward relationship between L1 attrition and L2 attainment (e.g., Major, 1992), and have assumed that the

L1 and L2 are competing for limited resources, which results in a 'trade-off' in language proficiency; in other words, the greater the attainment in the L2, the greater the attrition in the L1. However, recent evidence has revealed that such a relationship may be domain-dependent and might be modulated by other factors, such as language attitude (e.g., Cherciov, 2013). Schmid and Yilmaz (2018) conducted a comprehensive investigation of the predictor variables that may contribute to L1 attrition and L2 attainment, including language proficiency, language exposure and use, and attitudes. Their findings suggested the key role of L1 and L2 use in daily life in determining whether the L1 can be well maintained; that is, whether the L1 remains active and unchanged in the presence of a more powerful L2. The authors also found a correlation between educational levels and language success/failure.

1.2 Models of speech learning and acquisition

1.2.1 The Speech Learning Model and the Perceptual Assimilation Model

Three models of speech learning have been proposed in the literature to explain issues in phonetic attrition and acquisition. The Speech Learning Model (SLM) was originally proposed by Flege (1995, 2002) to account for differences in the learnability of L2 phonetic segments, and has recently been updated as the Revised Speech Learning Model (SLM-r; Flege & Bohn, 2021). According to the SLM and the SLM-r, the processes and mechanisms that guide L1 speech acquisition (including the ability to form new phonetic categories) remain intact and accessible for learning L2 speech across an individual's lifetime, and should thus also be at work during the process of L1 speech attrition. Another influential model, the Perceptual Assimilation Model (PAM), suggests that, during the early stage of language acquisition, infants attune their perception of speech to the properties of sounds in the ambient language environment; consequently, they become less sensitive to the sounds of a new language and may have difficulty perceiving the unfamiliar sound contrasts (Best, 1991). Similar to the SLM, the extended version of PAM, the Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2), also assumes that the ability to perceive speech continues to be refined throughout the lifespan (Best & Tyler, 2007). Both SLM and PAM-L2 postulate a common phonetic space, which was originally proposed to be a common phonological space in the speaker's mind in SLM and was revised as a phonetic one by Flege and Bohn (2021); this space stores the phonetic categories of both the L1 and the L2. The L1 and L2 categories mutually influence one another, as explained in detail below.

One of the SLM hypotheses is the category assimilation hypothesis (CAH), which claims that, in the common space, an L2 sound that is perceived as being similar to an L1 sound does not form a new category, and is understood as a variant of the L1 sound at an allophonic level; in other words, the cross-linguistic equivalence between the two sounds has been established. In this case, the phonemic variants for interdialectal contact are called diaphones (Weinreich, 1957), and the CAH posits that only one single phonetic category is used to process the two linked diaphones. This mapping of diaphones will eventually give rise to a new merged category in bilingual speakers' mental representations, and will be realised differently from either the L1 sound or the L2 sound in production; this phenomenon has been documented in several studies (Chang, 2012; Major, 1992). A recent study (Ulbrich and Ordin, 2014) examined the post-vocalic /r/ of German-English bilinguals and discovered an influence from the L2 that leads to the assimilation of the consonant pair, which lends support to the CAH.

The SLM (Flege, 1995, 2002) also postulated the category dissimilation hypothesis (CDH), whereby a new category will be established if an L2 sound is absent in the L1 system, which will make the combined phonetic space more crowded; as a result, the phonemes tend to disperse in compensation to maintain the phonetic contrast. When category dissimilation takes place, neither the newly established L2 category nor the closest L1 category will be identical to the categories in monolinguals; consequently, both categories may shift away from their original phonetic spaces. Simonet (2011) provided recent support for CDH, and found that Spanish-Catalan bilinguals had developed two categories to accommodate the mid-back vowels in the two languages.

As discussed above, SLM/SLM-r and PAM-L2 assume a common space for L1 and L2 categories, and can thus be used to account for both L2 speech acquisition and for the possible changes in L1 speech induced by the L2.

1.2.2 L2 Intonation Learning Theory

While SLM and PAM were proposed to account for, and have generally been applied in the research on, the learning of segments, the L2 Intonation Learning Theory (LILt) aims to address issues related to L2 learners' learning of intonation (Mennen, 2015). Mennen (2015) differentiated clearly between phonological representation and its phonetic realisation in LILt, and proposed four dimensions to capture the similarities and differences between L1 and L2 intonation based on Ladd (1996), namely the systemic dimension, the realisational dimension, the semantic dimension and the frequency dimension. Each dimension will be explained separately in the following.

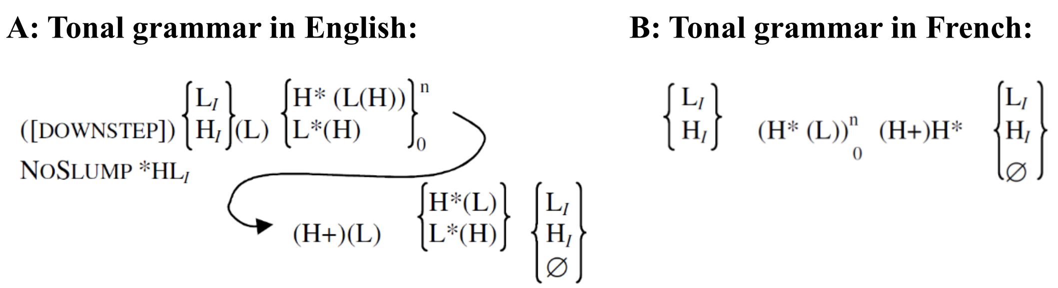

Figure 1-1: Tonal grammar in French and in English; adapted from Post et al. (2007: 192).

The first dimension is the 'systemic dimension', which concerns the inventory and distribution of phonological categories. For example, English and French have been shown to differ with regard to possible tonal sequences within an intonation phrase

(Gussenhoven, 2004; Post, 2000). As illustrated in Figure 1-1, English has a greater number of elements that may occupy more positions, as well as more possible combinations of elements and positions than French; thus, English can make use of more grammatical distinctions to express meaning (Post, D'Imperio, & Gussenhoven, 2007). Such typological differences between the L1 and the L2 make it difficult for L2 learners to acquire the target phonological category. Using longitudinal corpus data, Mennen, Chen and Karlsson (2010) examined the development of English pitch accents in Italian and Punjabi learners, and found that the learners did not use any complex English pitch accents (H*LH or L*HL) over time, thus suggesting that they had not formed these phonological categories in their L2.

Secondly, the phonetic implementation and realisation of the phonological categories are part of the 'realisational dimension'. Examples include how pitch accents are scaled in different languages and how pitch accents are aligned within an utterance. It has been suggested that both nuclear and pre-nuclear peaks are aligned later in Scottish Standard English than they are in Southern British English (Ladd, Schepman, White, Quarmby, & Stackhouse, 2009). If the realisation of one phonological category differs among languages or dialects, it is logical to assume an interaction of these languages or dialects. For example, Mennen (2004) investigated the pitch alignment in bilingual speakers of Dutch and Greek, and found bidirectional influences of the bilinguals' Dutch and Greek languages for the majority of the speakers, resulting in both the L1 and the L2 diverging from those of native speakers of both languages.

The third dimension pertains to the function of the phonological categories, and is called the 'semantic dimension'. For example, the function of the rising intonation varies within English dialects; it signals questions in most varieties of English, but is used to mark statements in Belfast English (Grabe, 2004).

The last is the 'frequency dimension', which concerns the frequency of use of the phonological categories. Even when languages share some identical phonological categories, the frequency of occurrence may differ. For example, the rising pitch accent appears more often among female German speakers than it does among female English speakers, although it exists in both German and in English (Mennen, Schaeffler, & Docherty, 2012). This difference in frequency of occurrence might also have some influences on L2 learning.

However, as Mennen (2015) stated, determining the dimension of the LILt according to which an observed deviance or phenomenon should be classified is not always straightforward. Moreover, there might be interactions among the dimensions. For example, Mandarin- and Korean-speaking learners of English were compared to native speakers of English with regard to how they marked prominence in English (McGory, 1997). Unlike native speakers, who consistently produced congruent pitch accents in each condition, the Mandarin and Korean learners mainly produced LHL tonal patterns. The difference between L1 and L2 production reveals a clear deviance in the realisational dimension; because of this, the utterances produced by the L2 learners signal different linguistic meanings from those produced by L1 speakers (semantic dimension). Apart from the realisational and semantic dimensions, the cause of the deviance might be the L2 learners' underlying difficulty in retrieving or accessing the appropriate phonological categories from the systemic dimension.

As in the case of SLM and PAM, which posit that the L1 and L2 segmental categories are stored in a common space, the LILt also advances a similar claim and provides evidence for L1 and L2 interaction at the intonation level. For example, in a study of the L1 attrition of German intonation, de Leeuw et al. (2012) found that German-English bilinguals showed merged values in the alignment of pre-nuclear rising pitch accents in their L1 German and L2 English (category assimilation). In addition, two bilinguals in the study had later alignment of pre-nuclear peaks in L1 German than the monolingual controls, which was similar to monolingual English speakers but different from monolingual German speakers. This category dissimilation of pre-nuclear peak alignment in the bilinguals' L1 contributed to maintain an even larger difference between their L1 and their L2.

1.3 Focus

During speech communication, speakers structure utterances to convey information based on shared knowledge or discourse; focus is an essential concept of information structure, and has been studied extensively. According to Krifka (2008: 247), as a linguistic device, 'focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions'.

1.3.1 Classification of focus

Different types of focuses have been identified by researchers in the field (Dik, 1980; Gussenhoven, 2007).

The term 'focus' is usually understood as presentational/informational focus or new information focus. As the most commonly used type of focus, presentational focus brings new information to the dialogue and the component containing the new information is the focused part of the sentence, which is usually prompted by a *wh*-question (Büring, 2012). As shown in Example (1), 'an apple' is the new information, and is thus the focus of the utterance. Contrastive focus, which is also known as the identificational focus or the alternative focus, refers to cases in which there are alternatives that fall within the same set as the contrastively focused component (Selkirk, 2008). In Example (2), a banana, a peach and an apple are from the same set, and the speaker has established a contrast with the first two. Another type of focus is corrective focus, which is regarded as a subtype of contrastive focus, and is used to directly reject an alternative (Gussenhoven, 2007). In Example (3), the answerer rejected ‘a banana’ by marking the focus as ‘an apple’.

(1) Q: What did you have just now?

A: I had [**an apple**]*F*.

(2) I didn’t have a banana or a peach. I had [**an apple**]*F*.

(3) Q: Did you have a banana just now?

A: No, I had [**an apple**]*F*.

In addition, focus can be classified according to the size of the focus (Büring, 2012). As shown below, the answers in Examples (4-6) are all instances of presentational focus, but they vary in terms of the size of focus. Example (4) is called sentential or broad focus, while Examples (4) and (5) are examples of narrow focus due to the relatively narrow scope that they cover.

(4) Q: What happened?/What did you say?

A: [**John got a book**]*F*.

(5) Q: What did John do?

A: John [**got a book**]*F*.

(6) Q: What did John get?

A: John got [**a book**]*F*.

It has been shown that the narrow presentational focus and the contrastive focus are realised similarly in Mandarin and cannot be differentiate clearly via the perception of native listeners (Yang, in preparation); therefore, we will only consider presentational focus in this thesis, and will use ‘focus’ to refer to ‘presentational focus’ hereafter.

1.3.2 Realisation of focus

To differentiate the focused component from the alternatives in the set and other components in the sentence, the focused component receives more prominence and is usually marked by certain linguistic means (Bolinger, 1972; Gussenhoven, 1983). For example, as shown in Example (7), focus can be marked by dislocation in Cantonese, a syntactic means that involves some movement of the elements in the sentence (Lee, 2017). Sentence (7a) follows the canonical subject-verb-object (SVO) word order in Cantonese, and it does not mark focus. In Sentence (7b), the subject *keoi* ‘he’, the auxiliary verb *wui* ‘will’ and the main verb *maai* ‘buy’ are moved to the right of the sentence-final particle (SFP) *aa3*, and the remaining object *jat bou dinnou* ‘one computer’ is thus emphasised and receives the focus (Cheung, 2009). This is an example of how focus can be marked via syntax.

(7) a. Keoi wui maai jat bou dinnou aa3.

he will buy one CL computer SFP

'He will buy a computer.'

b. [**Jat bou dinnou**]*F* aa3, keoi wui maai.

one CL computer SFP he will buy

'He will buy a computer.' (Cheung, 2009: 199-200)

It is also possible to mark focus via morphological means. In Japanese, focus particles are attached to the right of the focused component. As is evidenced in (8a) and (8b), although the two sentences generally share the same syntactic structure and lexical elements, the interpretations are quite different due to the different locations of the focus particle *mo*, which marks the focused component of each sentence (Matsuoka et al., 2006).

(8) a. [**Yusuke**]*F*-***mo*** jitensha-o kat-ta

Yusuke-also bicycle-ACC buy-PAST

'Yusuke also bought a bicycle (in addition to other people)'.

b. Yusuke-ga [**jitensha**]*F*-***mo*** kat-ta

Yusuke-NOM bicycle-also buy-PAST

'Yusuke bought a bicycle, too (in addition to other belongings)'.

(Matsuoka et al., 2006: 1)

Furthermore, focus can be realised via prosody in many languages. The realisation of prosodic focus has been documented extensively in various languages, and the focused components generally show expanded pitch range, increased intensity and lengthened duration (Yiya Chen, 2006; Cooper, Eady, & Mueller, 1985; Xu, Chen, & Wang, 2012). More recently, evidence has suggested that a post-focus component may be associated with a reduced or compressed pitch range, and this phenomenon has been termed post-focus compression (PFC) in Xu (2011). However, the presence or absence of PFC seems to diverge even within the same language family; for example, PFC has been found in Beijing Mandarin but not in Taiwan Mandarin, and its absence may be due to close contact with Taiwanese, which also lacks PFC (Xu et al., 2012). Chen, Xu and Guion-Anderson (2014) compared the production of focus by bilingual speakers of Southern Min (L1) and Mandarin (L2) in both languages. Although PFC is absent in Southern Min, the bilinguals produced PFC in Mandarin, and the younger group even showed clear PFC, thus suggesting that PFC can be acquired by L2 learners.

Another issue worth examining is the potential interaction between prosodic focus and lexical tones in tone languages. Tone languages use pitch to distinguish lexical items, which may interact with or even prevent the realisation of prosodic focus. Kügler and Skopeteas (2007) showed that focus was not expressed via prosodic means (particularly pitch accents) in Yucatec Maya, and the authors attributed this to the fact that pitch was already used for marking tones. On the other hand, Xu (1999) reported the effect of focus on the pitch curves of local tone-bearing units in Mandarin Chinese, thus revealing the influence of both tone and focus on pitch realisation. More recently, Yang, Chen and Li (2018) investigated how prosodic focus interacted with lexical tones in Chongming Chinese, a Chinese dialect that has eight tones. Their results revealed significant effects of focus and tone on the realisation of the F0 contour and the F0 range and suggested the presence of PFC in some tones but not in others. The authors proposed that, instead of being an all-or-nothing phenomenon, PFC might be tone dependent and should be examined according to different tonal combinations.

Following this line of research, the current project aims to investigate the realisation of prosodic focus in immigrants' Mandarin and Cantonese. With regard to the realisation of prosodic focus, Mandarin and Cantonese are reported to exhibit different features; Cantonese relies largely on duration and intensity for marking focus (Wu & Xu, 2010), while Mandarin makes extensive use of F0 (Xu et al., 2012). Moreover, PFC is reported to be absent in Cantonese and is well documented in Mandarin (Chen & Gussenhoven, 2008; Gu & Lee, 2007; Wu & Xu, 2010; Xu, 2011). Thus, it is proposed that PFC is a complex phenomenon, and that its absence or presence may be dependent on other constraints, such as the tonal contexts suggested by Chen (2010).

Very few studies have addressed the realisation of L2 prosody, and these studies have attempted to investigate whether prosodic features are transferrable from the L1 to the L2. Wu (2013) showed that PFC was absent in the Cantonese data of Cantonese-English bilinguals but, for some bilinguals, PFC could be found in their English production, which is consistent with monolingual English speakers. Thus, Wu (2013) concluded that PFC was a feature that could be lost but not gained. Chen (2014) tested the realisation of prosodic focus in the L2 Mandarin of English speakers, although PFC was present in their L1 English, the English speakers had difficulty transferring PFC to their L2 Mandarin. Similarly, Chen (2014) noted the absence of PFC in the L2 English of L1 Mandarin speakers. Both studies suggested that it is difficult to transfer the feature of PFC to an L2. In this study, we aim to explore whether the PFC of some acoustics cues will disappear in an L1 due to extensive exposure to an L2 without PFC in more depth.

1.4 The current study

As stated above, controversial results have been reported in studies of L1 phonetic attrition. Some participants retained native-like after years of exposure to an L2 (Hopp & Schmid, 2013), while others exhibited systematic changes in their pronunciation of the L1 even in the first weeks of L2 immersion (Chang, 2012). With regard to L2 speech acquisition, whether L2 prosody can be fully acquired in the immigration setting also remains to be explored. The current study will carefully design new experiments to address the issues in L1 phonetic attrition in more depth while focusing on the prosodic level of phonetic attrition. Moreover, the phonetic aspect of the L2 is also worth examining via long-term interactions between the L1 and the L2, which will not only test the postulates of the speech learning models but will also answer questions about the ultimate attainment of L2 phonology.

1.4.1 Significance of the study

This study is among the first attempts to associate L1 attrition with ultimate L2 attainment in the field, which will increase our understanding of the issues involved in L1 and L2 interaction in speech. In addition, this is the first study to systematically investigate the phonetic development of new immigrants in Hong Kong; thus, it will provide more insights into Chinese linguistics, phonetics and L2 acquisition. Moreover, as prosody is a relatively under-studied topic in language acquisition and attrition, this study aims to fill this research gap.

Studies of L1 attrition and L2 ultimate attainment have emerged in recent decades, and have formed two separate directions in research on bilingual language development. With regard to phonetics, a general conclusion drawn from these two branches is that exposure to an L2 might influence L1 speech in some way, while extensive immersion may fail to result in native-like pronunciation in L2 ultimate attainment. The speech learning models reviewed above suggest a common phonological space in bilinguals' mental representations in which a bidirectional influence between the sounds of the two languages is likely to occur. This may cause the assimilation or dissimilation of the sounds in both languages. The current study examines this bidirectional influence on prosodic realisation between the immigrants' L1 Mandarin and L2 Cantonese, the results of which will provide evidence for or against the postulates of these models.

Immigrants usually have sufficient exposure to the language in the new environment and may not maintain frequent use of their native language. Consequently, immigrants serve as an optimal group for testing L1 attrition and L2 ultimate attainment. However, the language development of new immigrants in Hong Kong has received little attention. This study aims to investigate the phonetic development of the immigrants' similar language pairs, namely Mandarin and Cantonese. Although the two languages (dialects) are structurally similar, the different phonetic systems may require the immigrants to establish new categories or to modify existing categories in the phonetic or phonological space, which will result in interactions within this space. This pioneering study will be the first step in the investigation of immigrants' language development. Based on our results, subsequent studies may proceed towards other aspects, as well as interfaces between different linguistic domains (such as phonetics-phonology and syntax–phonology), which will be of benefit to the study of Chinese linguistics, phonetics and L2 acquisition by providing more empirical data.

Moreover, previous works on L1 attrition and L2 acquisition have mainly concerned the segmental level (for example, vowel formants), while whether L1 and L2 prosody will interact with each other remains unclear. This study is designed to investigate the issue of L1 and L2 interaction from the suprasegmental level; thus, the results of this study will complement previous literature in this field and will provide a more complete picture of phonetic attrition and acquisition in the immigration setting.

1.4.2 Research questions

Research question 1: How do native speakers of Mandarin and Cantonese mark prosodic focus in their native languages when complex subject NPs are involved? How are Mandarin and Cantonese different in terms of focus marking? Unlike previous studies, we used complex NPs as the subjects in the test sentences to make the stimuli more natural. It is predicted that focus production in Mandarin and Cantonese will be unaffected by the complex NPs when the entire subject phrase receives prominence (namely, under focus), and that the previously reported differences in Mandarin and Cantonese focus marking (Mandarin relies largely on F0 while Cantonese makes use of duration) can be reduplicated.

Research question 2: How do immigrants mark prosodic focus in their L1 Mandarin and L2 Cantonese? Does their focus production resemble that of native speakers? Due to long-term exposure to the L2, which we think will lead to some cross-linguistic influences, it is expected that the immigrants' L1 Mandarin and L2 Cantonese will show some divergence from that of native speakers of the respective languages, and that the immigrants' L1 and L2 may exhibit some assimilation.

Research question 3: Can native speakers of Mandarin and Cantonese map prosody and focus correctly in their native languages? Do these participants show different accuracy rates when the direction of the mapping is reversed? Two types of tasks will be performed, one involving mapping prosody onto focus and the other mapping focus onto prosody. Compared to deciding on the acoustic cues for a specific focus type, it should be easier to notice the differences in the acoustic cues for various focus types; it is thus predicted that the participants will perform more accurately in the mapping from prosody onto focus.

Research question 4: Can immigrants map prosody and focus correctly in their L1 Mandarin and L2 Cantonese? Due to reduced contact in the L1, the immigrants are expected to show attrition in focus perception L1 Mandarin. In addition, they should be less accurate than native speakers with regard to the perception of focus in Cantonese.

Chapter 2: Production of Mandarin and Cantonese Focus by Native Speakers

2.1 Introduction

This chapter reports the production experiment of Mandarin and Cantonese focus marking by native speakers.

Mandarin and Cantonese are two varieties/dialects of Chinese, but they have different phonological systems and are not mutually intelligible (Zhang, 1998). Like other Chinese language, Mandarin and Cantonese are tone languages. There are four lexical tones in Mandarin and six in Cantonese (Bauer & Benedict, 1997; Chao, 1948), as shown in Table 2-1.

Language	**Tone name**	**Tone category**	**Tone letter**	**Example**
Mandarin	Tone 1	High Level	55	*ma55* ‘mother’
	Tone 2	High Rising	35	*ma35* ‘hemp’
	Tone 3	Low Dipping	214	*ma214* ‘horse’
	Tone 4	High Falling	51	*ma51* ‘to scold’
Cantonese	Tone 1	High Level	55	*si55* ‘teacher’
	Tone 2	High Rising	25	*si25* ‘history
	Tone 3	Mid Level	33	*si33* ‘test’

Tone 4	Mid-low Falling	21	*si21* ‘time’
Tone 5	Mid-low Rising	23	*si23* ‘market’
Tone 6	Mid-low Level	22	*si22* ‘matter’

Table 2-1: Tonal systems in Mandarin and Cantonese.

Previous studies on Mandarin focus production suggest the role of F0 in focus marking and such manipulation of F0 can be found in both on-focus and post-focus components (Yiya Chen & Gussenhoven, 2008; Jin, 1996; Xu, 1999). The phenomenon of a post-focus component having reduced or compressed F0 is called post-focus compression (PFC) (Xu, 2011), but the presence or absence of PFC seems to diverge even within the same language family. For example, PFC has been found in Beijing Mandarin but not in Taiwan Mandarin, and its absence may possibly be due to the close contact with Taiwanese, which also lacks PFC (Xu et al., 2012). Cantonese is a relatively less-studied language and there are only a couple of studies that worked on focus prosody in Cantonese (Gu & Lee, 2007; Hsu, Xu, & Ngai, 2018; Wu & Xu, 2010). It is shown that Cantonese does not make use of F0 or PFC to mark focus; rather, duration is an essential acoustic correlate of prosodic focus in Cantonese. The role of intensity in focus production is also documented (Wu & Xu, 2010).

Besides focus location, the issue of focus domain has rarely been studied in speech production. It is thus necessary to examine whether the realisation of relatively broad focus differs from that of narrow focus. This study takes this issue into account and investigates the acoustic correlates of Mandarin and Cantonese prosodic focus when the focused components differ in size and position.

The specific questions to be addressed in this chapter include the following:

1) How do native speakers of Mandarin and Cantonese mark prosodic focus in their native languages?

2) Does focus domain/size influence the prosodic marking of focus in Mandarin and Cantonese?

2.2 Methodology

2.2.1 Participants

Twenty-one native speakers of Mandarin (11 females, 10 males; aged: 24.95 ± 3.75) and 21 native speakers of Cantonese (10 females, 11 males; aged: 20.78 ± 2.56) were recruited from the Hong Kong Polytechnic University and participated in a production experiment at the Speech and Language Sciences Laboratory of the Hong Kong Polytechnic University. The Mandarin-speaking participants were born in northern China, where Mandarin is the native and dominant language, and they all reported having spent most of their life in Mandarin-speaking regions. The participants all speak English as an L2 and none of them speak Cantonese. The Cantonese-speaking participants were born and brought up in Hong Kong, where Cantonese is the native and dominant language. The Cantonese participants also speak English and Mandarin, but they are native and dominant in Cantonese only. No participants had any history of speaking, hearing or language difficulties

2.2.2 Experiment design

To answer our research questions, simple declarative subject-verb-object (SVO) sentences were selected. Because lexical tone was irrelevant to this study, the high-level first tone (T1) was used in the stimuli as much as possible. All the content words were in T1, with the exceptions being the demonstrative and classifiers, the tone of which is very difficult to control.

In the SVO sentences, complex subject nominal phrases (NPs) were adopted to make the sentences more natural and also to test the effect of focus size. The SVO sentences shared the same syntactic structure, with a tetrasyllabic subject, a monosyllabic verb and a disyllabic object, as exemplified in Tables 2-2 and 2-3. There were five focus types, which were all elicited by precursor questions asked by the experimenter. In the broad focus condition, the interlocutor needs information about the whole event, and the whole utterance is under focus. In the subject focus condition, the question concerns the agent of the event, and the subject NP is under focus. In the verb focus condition, the question asks about the action of the event, and the verb is under focus. In the object focus condition, the information of the theme or patient is unknown, and the object is under focus. In the VP (verb phrase) focus condition, the given information is the agent and the question asks about the event.

Focus types	Precursor questions	Target sentences
Broad focus	*ni3 shuo1 shen2me0* you say what 'What did you say?'	*na4 wei4 yi1sheng1 he1 ka1fei1* that CL doctor drink coffee 'The doctor drank coffee.'
Subject focus	*shei2 he1 ka1fei1* who drink coffee 'Who drank coffee?'	*na4 wei4 yi1sheng1 he1 ka1fei1* that CL doctor drink coffee 'The doctor drank coffee.'
Verb focus	*na4 wei4 yi1sheng1 zen3me0 ka1fei1* that CL doctor how coffee 'What did the doctor do to coffee?'	*na4 wei4 yi1sheng1 he1 ka1fei1* that CL doctor drink coffee 'The doctor drank coffee.'
Object focus	*na4 wei4 yi1sheng1 he1 shen2me0* that CL doctor drink what 'What did the doctor drink?'	*na4 wei4 yi1sheng1 he1 ka1fei1* that CL doctor drink coffee 'The doctor drank coffee.'
VP focus	*na4 wei4 yi1sheng1 zuo4 shen2me0* that CL doctor do what 'What did the doctor do?'	*na4 wei4 yi1sheng1 he1 ka1fei1* that CL doctor drink coffee 'The doctor drank coffee.'

Table 2-2: Focus conditions of the Mandarin SVO sentences.

Various focus types were elicited by precursor questions. The focused regions of the target sentences have been underlined.

Focus types	Precursor questions	Target sentences
Broad focus	*nei5 waa6 mat1je5* you say what 'What did you say?'	*go2 go3 gaa1ban1 ling1 gaau1zeon1* that CL guest carry plastic_bottle 'The guest carried a plastic bottle.'
Subject focus	*bin1go3 ling1 gaau1zeon1* who carry plastic_bottle 'Who carried a plastic bottle?'	*go2 go3 gaa1ban1 ling1 gaau1zeon1* that CL guest carry plastic_bottle 'The guest carried a plastic bottle.'
Verb focus	*go2 go3 gaa1ban1 mat1je5 gaau1zeon1* that CL guest how plastic_bottle 'What did the guest do to a plastic bottle?'	*go2 go3 gaa1ban1 ling1 gaau1zeon1* that CL guest carry plastic_bottle 'The guest carried a plastic bottle.'
Object focus	*go2 go3 gaa1ban1 ling1 mat1je5* that CL guest carry what 'What did the guest carry?'	*go2 go3 gaa1ban1 ling1 gaau1zeon1* that CL guest carry plastic_bottle 'The guest carried a plastic bottle.'

VP focus	*go2 go3 gaa1ban1 zou6 mat1je5*	*go2 go3 gaa1ban1 ling1 gaau1zeon1*
	that CL guest do what	that CL guest carry plastic_bottle
	'What did the guest do?'	'The guest carried a plastic bottle.'

Table 2-3: Focus conditions of the Cantonese SVO sentences.

Three Mandarin target sentences and four Cantonese target sentences were prepared for the native speakers, which were repeated twice in the recording sessions. In total, there were 2,205 target sentences (3 sentences * 5 focus types * 3 repetitions * 21 Mandarin natives + 4 sentences * 5 focus types * 3 repetitions * 21 Cantonese natives).

2.2.3 Procedures

As mentioned in the previous section, the target stimuli were elicited as answers to precursor questions, and the question and answer pairs were prepared in written form (namely, in Chinese characters) in this study. During the experiment, the stimuli were randomly presented on a computer screen in E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2012). The author, a native speaker of Mandarin and a fluent speaker of Cantonese, asked the questions to the participants, and the participants were instructed to answer them as naturally as possible. This approach helped us to collect data that resemble naturalistic speech. The question and answer pairs were recorded at a sampling rate of 44,100 Hz in Audacity (Audacity Team, 2019) on another computer. Only the answers were processed for further analysis.

This project has been approved by the Human Subjects Ethics Sub-committee (HSESC) of the Hong Kong Polytechnic University (Reference #: HSEARS20190102001). All participants gave their written informed consent prior to the recording sessions.

2.2.4 Data analysis

The syllables in the target sentences were manually segmented by trained phoneticians in Praat (Boersma & Weenink, 2015). Relevant acoustic values, including 20 time-normalised F0 points, mean F0, duration and mean values for each syllable were extracted using the ProsodyPro Praat script (Xu, 2013). The F0 values, originally measured in Hz, were then converted to semitones (st) individually, with mean F0 of each speaker as reference (Nolan, 2003).

The time-normalised F0 points were used to plot the F0 contours, which were smoothed with the 'geom_smooth' function. The extracted mean values, including mean F0, duration and intensity, were analysed with linear mixed-effects modelling using the 'lme4' package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2018; RStudio Team, 2016). Although some studies also tested variables such F0 range, F0 maxima and F0 minima, we have shown in a pilot study that, mean F0 is a better indicator than other F0-related values (Yang & Chen, 2020b). So only mean F0 was included in our analysis here. Because we considered several focus types and also all the syntactic positions, it would be overwhelming to put all the focus types together. Pairwise comparisons were made between broad focus with the remaining focus types. In the models, the mean values were used as the dependent variables, Focus Type (hereinafter Focus) was included as one fixed effect, and Speaker, Sentence and Repetition as the random effects. The figures were plotted with the ‘ggplot2’ package (Wickham, 2016). Besides, to determine the contribution of the acoustic cues on the differentiation of each focus type from broad focus, we also carried out linear discriminant analyses (LDAs) on the acoustic cues across different locations following Breen et al. (2010). The LDAs were conducted for each focus pair for each language and the models were used to predict the accuracy of the differentiation within each focus pair.

After presenting the results from native speakers of Mandarin and Cantonese separately, we also compared data from these two speaker groups in another subsection. Following previous studies (e.g., Chen, 2014), we compared the differences of the tested acoustic cues between the focus types in the verb position. Specifically, we calculated the changes in mean F0, duration and intensity between broad focus and other focus types (subject, verb and object) by token and then used linear mixed-effects models to analyse the data.

2.3 Results

This section presents the production data from native speakers of Mandarin and Cantonese. Descriptions of the F0 contours are provided before the statistical analyses of the mean values are reported.

2.3.1 Mandarin focus marking by native speakers

2.3.1.1 Broad focus vs subject focus

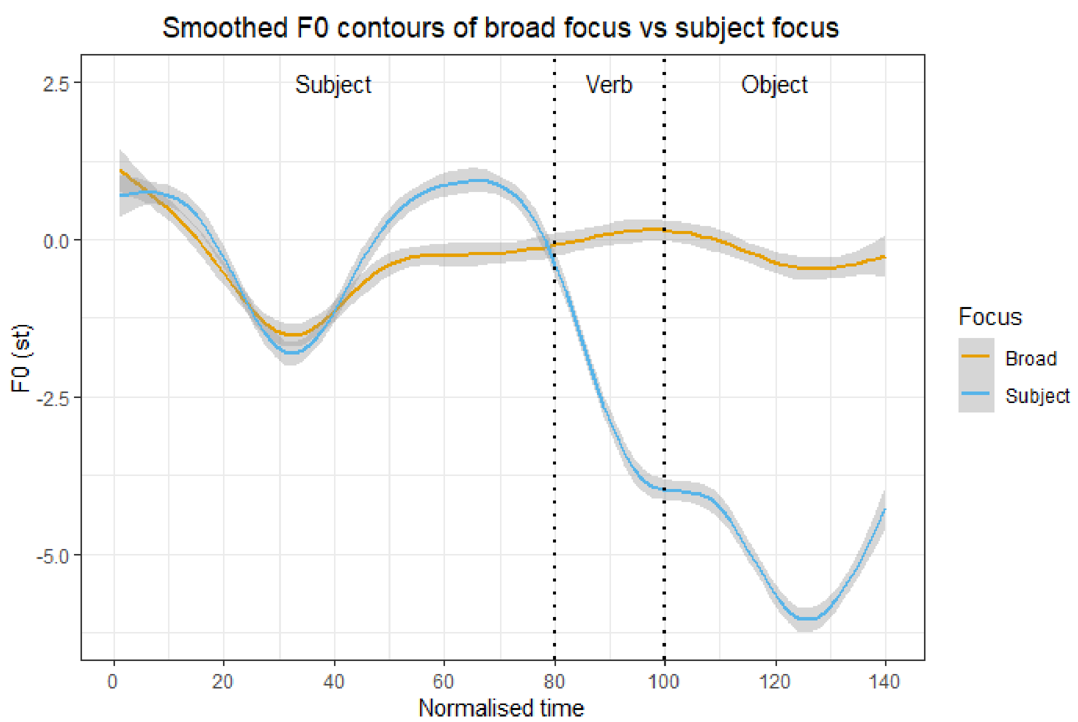

Figure 2-1: F0 contours of broad focus vs subject focus by Mandarin speakers.

F0 contour: Figure 2-1 presents the smoothed F0 contours of Mandarin SVO sentences under broad focus and subject focus produced by native speakers of Mandarin. The F0 contour of the broad focus condition was in general a flat one. For the subject focus condition, a clear focus-induced raising of F0 contours can be found in the subject nouns (Points 41 to 80) but not in the demonstratives (Points 1 to 20) or the classifiers (Points 21 to 40), indicating that Mandarin speakers are sensitive to the lexical categories and have placed their focus on content words only. The F0 contour had a sharp decrease in the verb position and continued to drop in the object position.

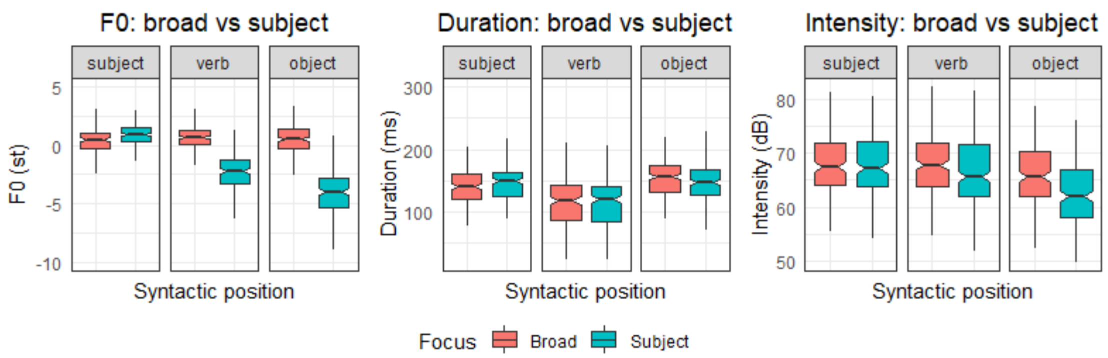

Figure 2-2: Mean values of broad focus vs subject focus by Mandarin speakers.

Mean F0: The main effect of focus on mean F0 reached significance in the subject ($\chi^2(1) = 24.001$, $p < .001$), verb ($\chi^2(1) = 324.650$, $p < .001$) and object ($\chi^2(1) = 343.650$, $p < .001$) positions. More specifically, focus increased the mean F0 by 0.514 ± 0.103 st in the subject position and decreased the mean F0 by 3.069 ± 0.131 st and 4.835 ± 0.199 st in the verb and object positions, respectively. There was an increase of mean F0 when the subject was under focus, which was followed a sharp decrease of mean F0 on the post-focus regions.

Duration: There was a main effect of focus on duration in the subject position ($\chi^2(1) = 15.963$, $p < .001$), lengthening one syllable by 5.215 ± 1.292 ms. There was also an effect on duration in the object position ($\chi^2(1) = 8.432$, $p = .004$), shortening one syllable by 5.368 ± 1.840 ms. Although there was a slight shortening of duration (1.106 ± 1.575) in the verb position, such effect was non-significant ($\chi^2(1) = .495$, $p = .482$).

Intensity: There was no on-focus increase of intensity in the subject position ($\chi^2(1) = .146$, $p = .706$), but the effect of focus was found in the verb ($\chi^2(1) = 44.581$, $p < .001$) and object ($\chi^2(1) = 178.270$, $p < .001$) positions, which lowered the intensity by 1.630 ± 0.236 dB and 3.621 ± 0.267 dB, respectively.

2.3.1.2 Broad focus vs verb focus

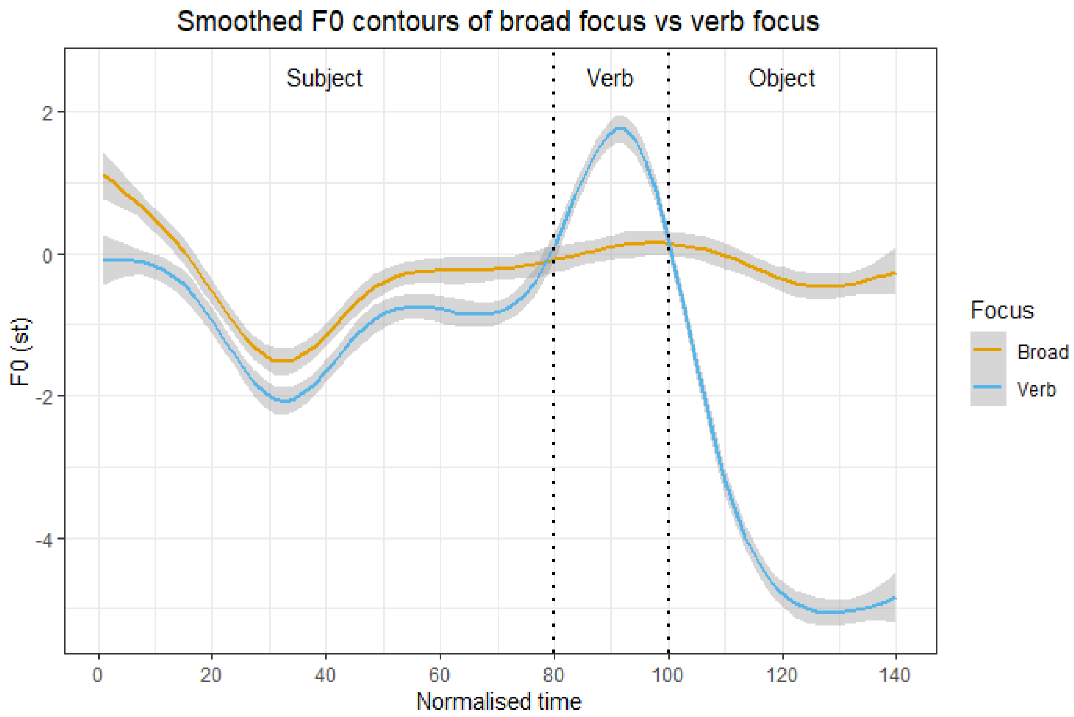

Figure 2-3: F0 contours of broad focus vs verb focus by Mandarin speakers.

F0 contour: As can be seen from Figure 2-3, the verb focus significantly raised the F0 contour in the verb position, which resulted in an F0 peak. There was also an effect of focus in the pre- and post-focus regions, which lowered the F0 contour, and the degree of post-focus lowering was much greater than that of pre-focus lowering.

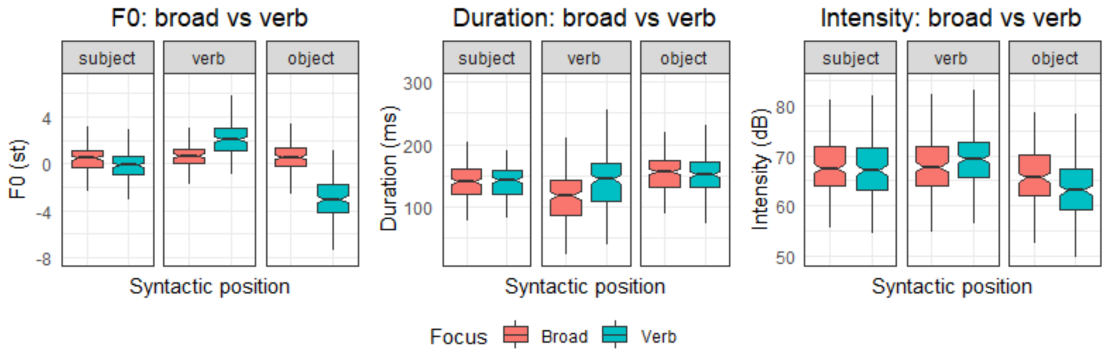

Figure 2-4: Mean values of broad focus vs verb focus by Mandarin speakers.

Mean F0: The main effect of focus on mean F0 reached significance in the subject ($\chi^2(1) = 21.631$, $p < .001$), verb ($\chi^2(1) = 127.900$, $p < .001$) and object ($\chi^2(1) = 251.460$, $p < .001$) positions. More specifically, focus increased the mean F0 by 1.456 ± 0.117 st in the verb position and decreased the mean F0 by 0.470 ± 0.099 st and 3.767 ± 0.196 st in the subject and object positions, respectively. The compression of mean F0 was much stronger in the post-focus region than the pre-focus region.

Duration: In the verb position, there was a lengthening of one syllable by 27.459 ± 1.949 ms ($\chi^2(1) = 156.550$, $p < .001$). The main effect of focus on the duration did not reach significance in the subject ($\chi^2(1) = .003$, $p = .986$) or object ($\chi^2(1) = .328$, $p = .567$) positions. The results suggested that only on-focus duration was manipulated to mark the verb focus in Mandarin.

Intensity: There was an on-focus increase of intensity by 1.262 ± 0.254 dB in the verb position ($\chi^2(1) = 23.990$, $p < .001$). The effect of focus was also found in the subject

($\chi^2(1)$ = 10.3224, p = .001) and object ($\chi^2(1)$ = 85.031, p < .001) positions, which lowered the intensity by 0.683 ± 0.212 dB and 2.386 ± 0.243 dB, respectively. Similarly, the lowering of intensity was more robust in the object position than in the subject position.

2.3.1.3 Broad focus vs object focus

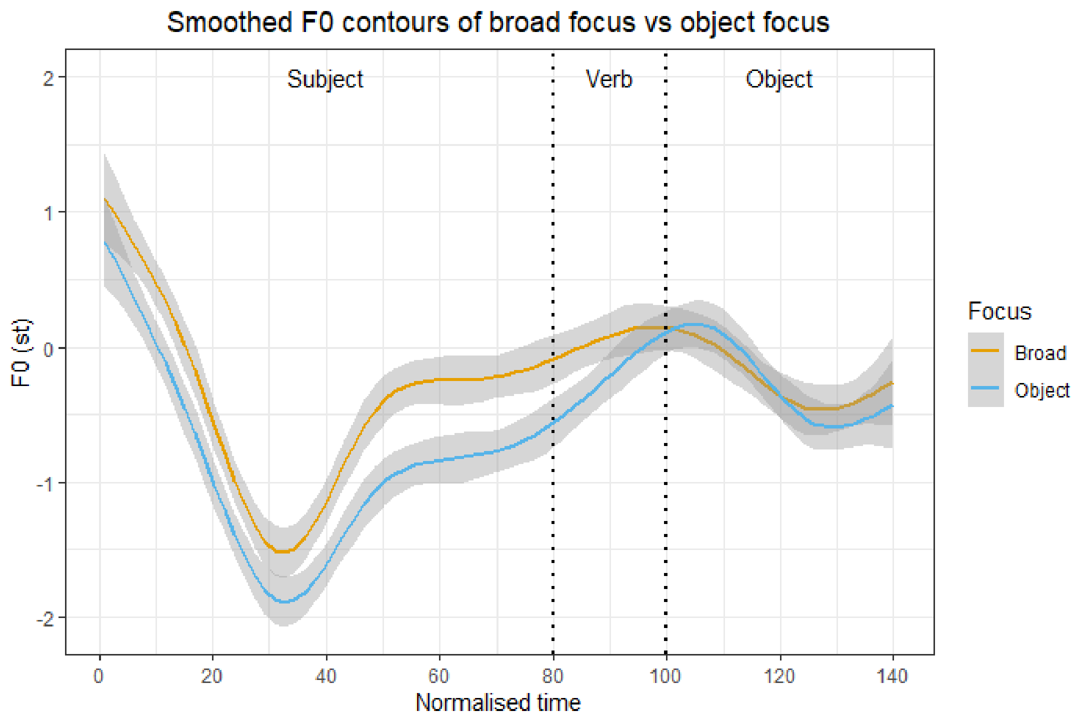

Figure 2-5: F0 contours of broad focus vs object focus by Mandarin speakers.

F0 contour: When the object was under focus, it was difficult to distinguish between broad focus and object focus in the object position because there was no focus-induced increase of F0. In fact, the contours of the two focus types overlapped in the object position. However, a pre-focus decrease of F0 was evident in both the subject and verb positions, as the curve representing the object focus condition was always below the one for the broad focus condition.

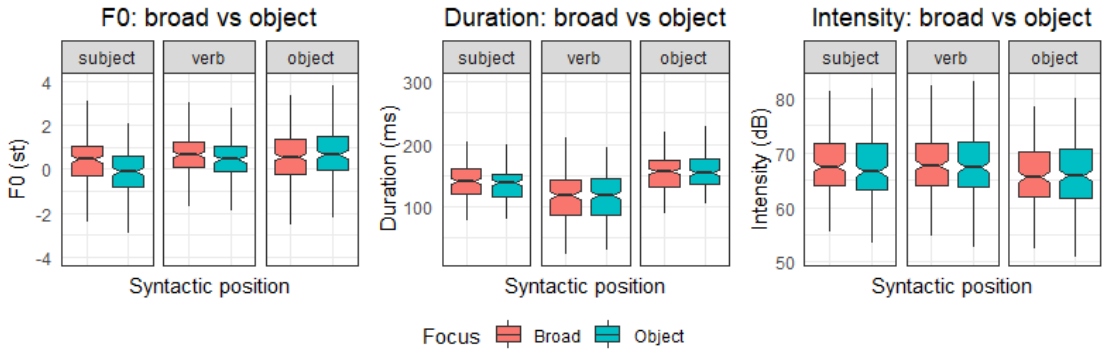

Figure 2-6: Mean values of broad focus vs object focus by Mandarin speakers.

Mean F0: The main effect of focus on mean F0 reached significance in the subject ($\chi^2(1) = 21.962$, $p < .001$) and verb ($\chi^2(1) = 6.037$, $p = .014$) positions, which decreased the mean F0 by 0.455 ± 0.096 st and 0.232 ± 0.094 st in the subject and verb positions, respectively. There was no effect of focus on mean F0 in the object position($\chi^2(1) = .046$, $p = .831$), indicating no on-focus increase of mean F0 for object focus.

Duration: There was a main effect of focus on duration in the subject position ($\chi^2(1) = 7.167$, $p = .007$), shortening one syllable by 3.525 ± 1.312 ms. There was also an effect on duration in the verb position ($\chi^2(1) = 5.690$, $p = .017$), lengthening one syllable by 4.373 ± 1.828 ms. No effect was found on duration in the object position ($\chi^2(1) = 2.628$, $p = .105$)

Intensity: The effect of focus on intensity was significant in the subject position ($\chi^2(1) = 4.113$, $p = .043$), lowering the intensity by 0.414 ± 0.204 dB. No effect of focus was found in the verb ($\chi^2(1) = .091$, $p = .764$) or object positions ($\chi^2(1) = 1.872$, $p = .171$).

2.3.1.4 Broad focus vs VP focus

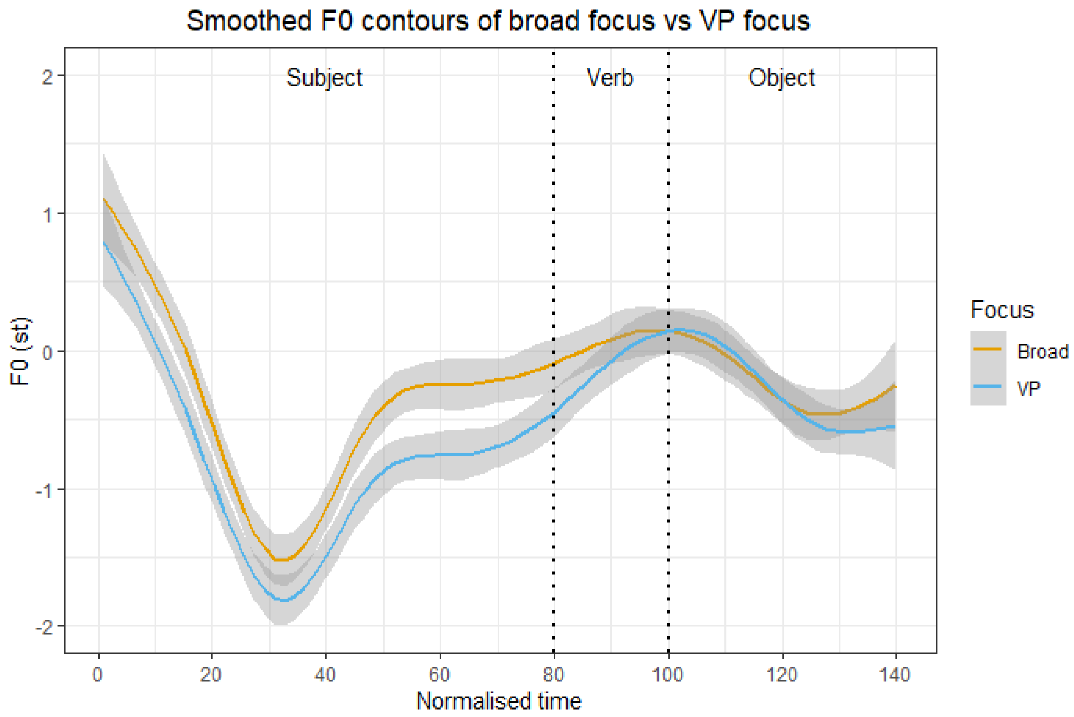

Figure 2-7: F0 contours of broad focus vs VP focus by Mandarin speakers.

F0 contour: In the on-focus verb and object regions, the F0 contour of the VP focus did not surpass the F0 contour of the broad focus. Rather, there was a slight lowering of F0 in the initial and final portions of the VP phrase. Meanwhile, in the pre-focus region, the F0 contour of the VP focus was lowered. As a result, there was an F0 peak in the on-focus VP region.

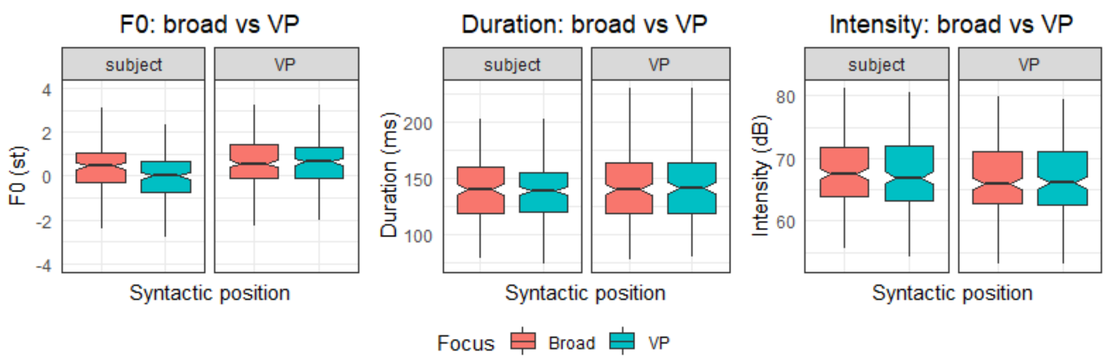

Figure 2-8: Mean values of broad focus vs VP focus by Mandarin speakers.

Mean F0: There was a main effect of focus on mean F0 in the subject position ($\chi^2(1)$ = 17.878, p < .001), which lowered the mean F0 by 0.417 ± 0.097 st. There was no effect of focus on mean F0 in the VP position ($\chi^2(1)$ = .291, p = .590).

Duration: The main effect of focus on duration was significant in the subject position ($\chi^2(1)$ = 5.804, p = .016), shortening one syllable by 3.063 ± 1.268 ms. There was a minor lengthening of duration by 1.050 ± 1.417 ms in the VP position, but the effect was non-significant ($\chi^2(1)$ = .550, p = .458).

Intensity: In the subject position, there was an effect of focus on intensity($\chi^2(1)$ = 5.079, p = .024), lowering it by 0.478 ± 0.212 dB. The effect of focus on intensity did not reach significance in the VP position ($\chi^2(1)$ = .204, p = .651).

2.3.1.5 Linear discriminant analysis

We also conducted linear discriminant analyses to each focus pair in different sentence positions to test the prediction accuracy of the acoustic cues in the Mandarin data. The results presented in Table 2-4 are the average accuracy rates of the three sentence positions for each focus pair.

Focus pair	Accuracy of prediction		
	Mean F0	**Duration**	**Intensity**
Broad vs subject	94.3%	56.5%	80.0%
Broad vs verb	88.9%	61.8%	75.8%
Broad vs object	61.1%	54.1%	56.0%
Broad vs VP	57.1%	51.7%	55.0%

Table 2-4: Predicted accuracy from the LDA (Mandarin natives).

The results in the table were generally in line with the results from the linear mixed-effects models. The Mandarin speakers clearly distinguished broad focus and subject focus as well as broad focus and verb focus, and the most useful acoustic cue was mean F0. For the other two focus pairs, the prediction accuracies were just above chance level, suggesting that these cues may not help to distinguish these two pairs.

2.3.1.6 An interim summary

Here we presented the focus production data of Mandarin native speakers, with broad focus as the reference for pairwise comparisons. We have shown that, the Mandarin speakers clearly distinguished broad focus with subject focus and verb focus by manipulating F0, duration and intensity, but when comparing broad focus with object focus and VP focus, they only marked pre-focus F0 and duration and did not show on-focus changes of the acoustic cues.

2.3.2 Cantonese focus marking by native speakers

2.3.2.1 Broad focus vs subject focus

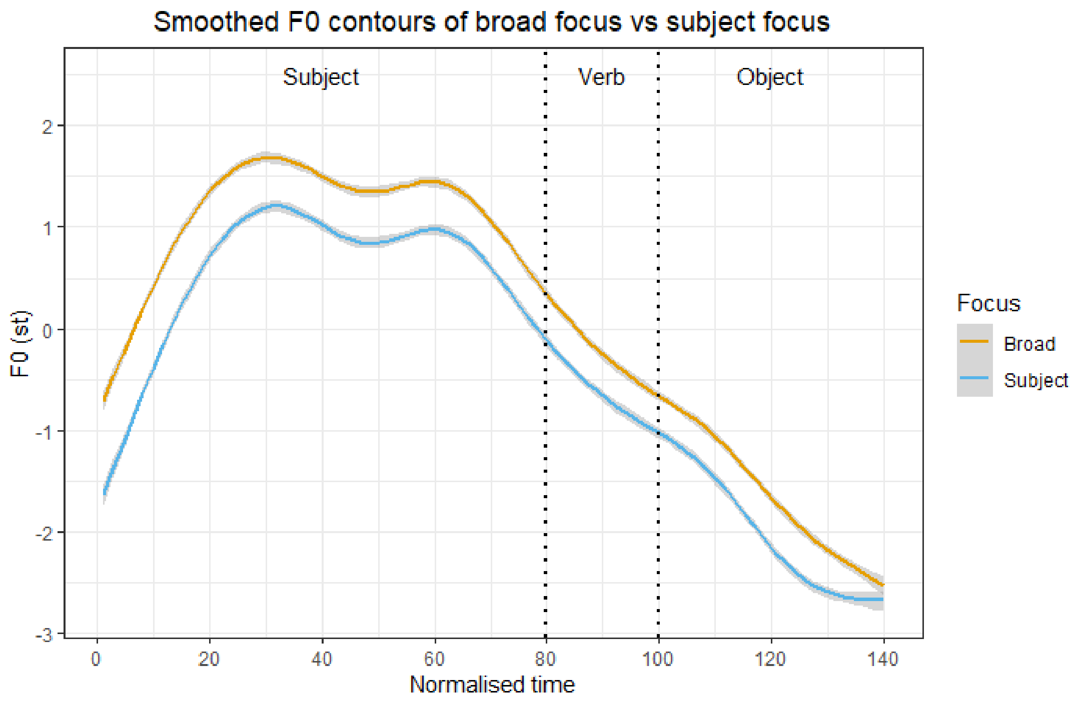

Figure 2-9: F0 contours of broad focus vs subject focus by Cantonese speakers.

F0 contour: Figure 2-9 shows the F0 contours of broad focus and subject focus produced by Cantonese speakers. These two figures share a very similar shape, but the one for broad focus consistently had higher F0 than the one for subject focus.

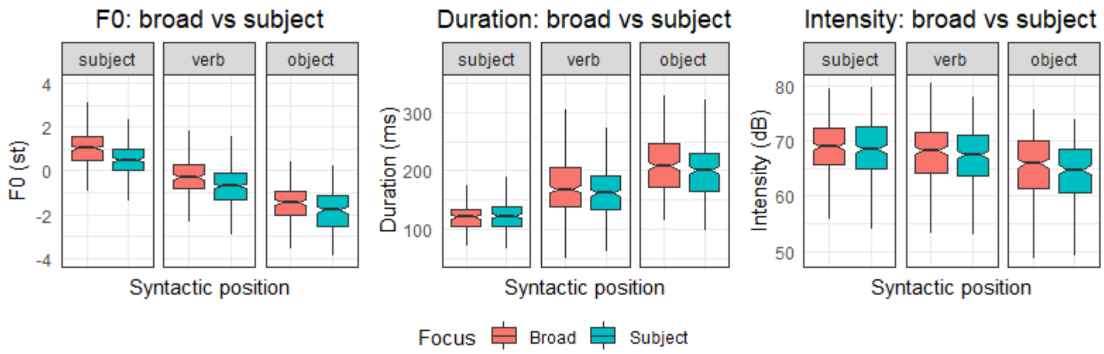

Figure 2-10: Mean values of broad focus vs subject focus by Cantonese speakers.

Mean F0: The main effect of focus on mean F0 reached significance in the subject ($\chi^2(1) = 40.058$, $p < .001$), verb ($\chi^2(1) = 42.895$, $p < .001$) and object ($\chi^2(1) = 20.945$, $p < .001$) positions. Broad focus was always higher in mean F0 than subject focus, and the difference was 0.542 ± 0.077 st in the subject position, 0.435 ± 0.065 st in the verb position, and 0.408 ± 0.089 st in the object position.

Duration: There was a main effect of focus on duration in the subject position ($\chi^2(1) = 6.194$, $p = .013$), lengthening one syllable by 3.951 ± 1.584 ms. There was also an effect of focus on duration in the verb ($\chi^2(1) = 7.096$, $p = .008$) and object ($\chi^2(1) = 18.922$, $p < .001$) positions, shortening one syllable by 8.000 ± 2.995 ms and 10.097 ± 2.060 ms, respectively.

Intensity: There was no effect of focus on intensity in the subject position ($\chi^2(1) = 3.726$, $p = .054$), but the effect of focus was found in the verb ($\chi^2(1) = 17.310$, $p < .001$) and object ($\chi^2(1) = 34.590$, $p < .001$) positions, which lowered the intensity by 0.866 ± 0.207 dB and 1.138 ± 0.190 dB, respectively.

2.3.2.2 Broad focus vs verb focus

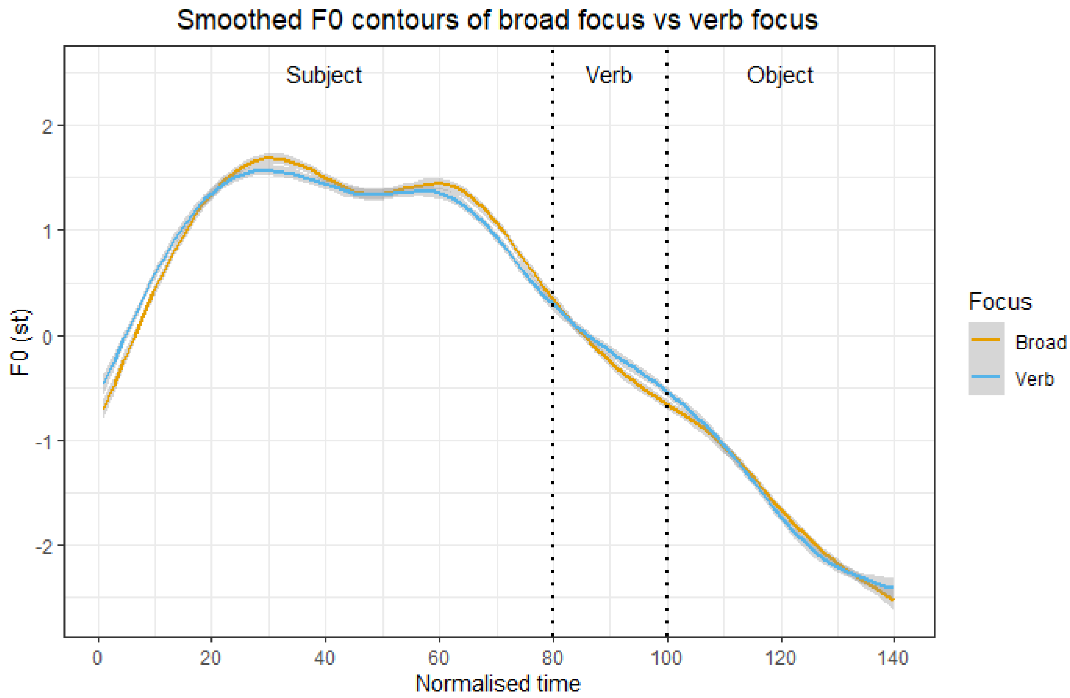

Figure 2-11: F0 contours of broad focus vs verb focus by Cantonese speakers.

F0 contour: The F0 contours of broad focus and verb focus is presented in Figure 2-11, which suggests that, although the contours overlapped most of the time, there was a slight raising in the verb position for verb focus.

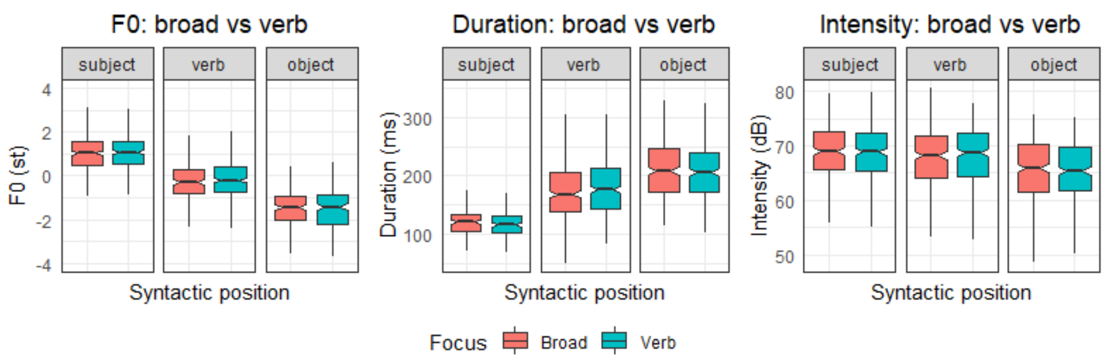

Figure 2-12: Mean values of broad focus vs verb focus by Cantonese speakers.

Mean F0: No effect of focus on mean F0 was found in the subject ($\chi^2(1) = = .117$, *p*

= .732), verb ($\chi^2(1) = .991$, $p = .320$) or object ($\chi^2(1) = .018$, $p = .893$) positions.

Duration: In the verb position, there was a lengthening of one syllable by 18.744 ± 3.808 ms ($\chi^2(1) = 23.689$, $p < .001$). The main effect of focus on the duration did not reach significance in the subject ($\chi^2(1) = 2.561$, $p = .110$) or object ($\chi^2(1) = .890$, $p = .346$) positions.

Intensity: There was no effect of focus on intensity in the subject ($\chi^2(1) = 1.702$, $p = .192$), verb ($\chi^2(1) = .799$, $p = .371$) or object ($\chi^2(1) = 1.247$, $p = .264$) positions.

2.3.2.3 Broad focus vs object focus

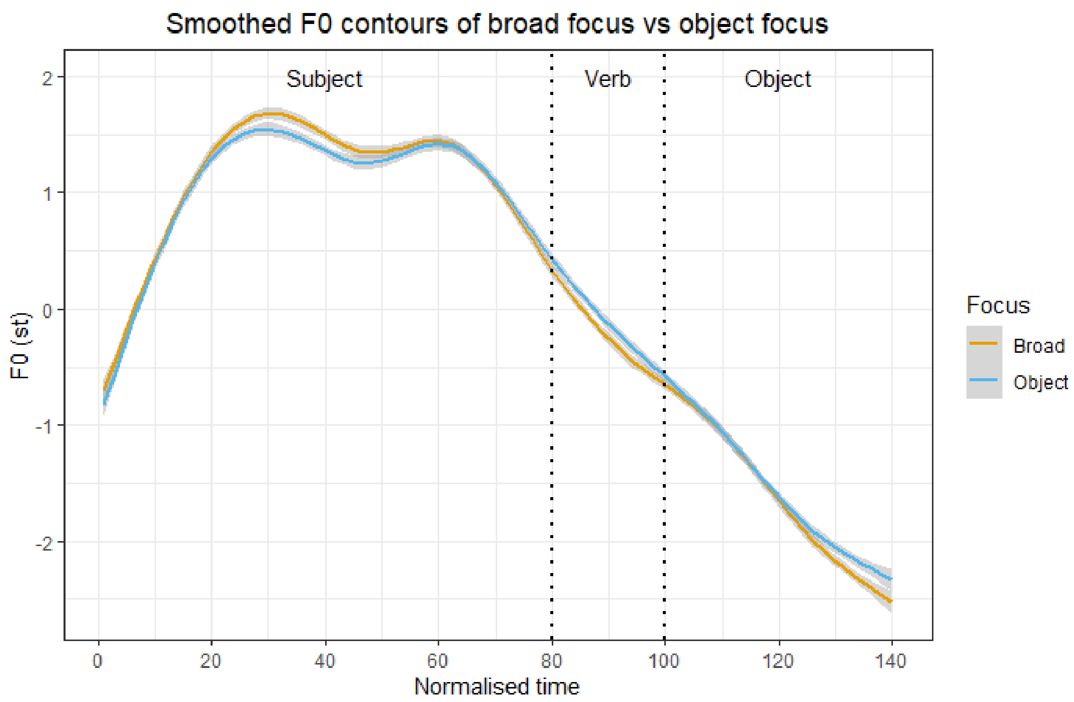

Figure 2-13: F0 contours of broad focus vs object focus by Cantonese speakers.

F0 contour: Figure 2-13 is the F0 contours of broad focus and object focus. There was a minor lowering of F0 in the subject and verb positions and a minor raising of F0 towards the end of the object.

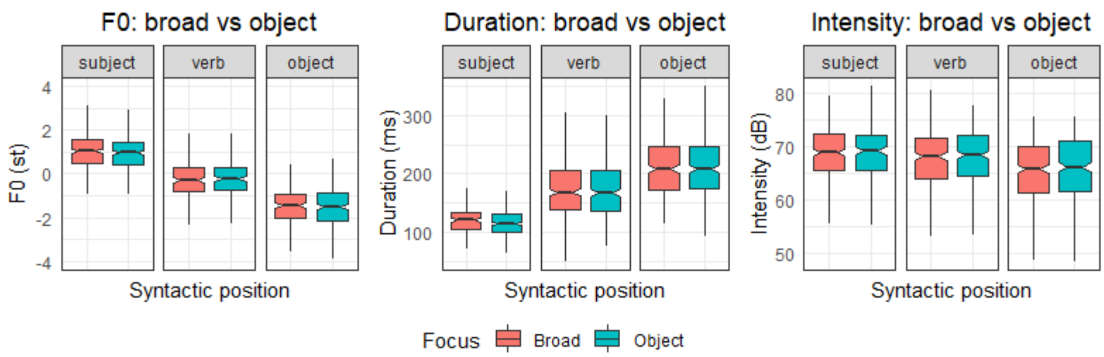

Figure 2-14: Mean values of broad focus vs object focus by Cantonese speakers.

Mean F0: No effect of focus on mean F0 was found in the subject ($\chi^2(1)$ =.433, p = .510), verb ($\chi^2(1)$ = .513, p = .474) or object ($\chi^2(1)$ = 1.039, p = .308) positions.

Duration: There was a main effect of focus on duration in the subject position ($\chi^2(1)$ = 7.126, p = .008), shortening one syllable by 3.875 ± 1.441 ms. There was also a marginal effect of focus on duration in the object position ($\chi^2(1)$ = 3.308, p = .069), lengthening one syllable by 4.091 ± 2.248 ms. No effect of focus was found in the verb position ($\chi^2(1)$ = .084, p = .772).

Intensity: There was no effect of focus on intensity in the subject ($\chi^2(1)$ = .874, p = .360), verb ($\chi^2(1)$ = .495, p = .482) or object ($\chi^2(1)$ = 1.599, p = .206) positions.

2.3.2.4 Broad focus vs VP focus

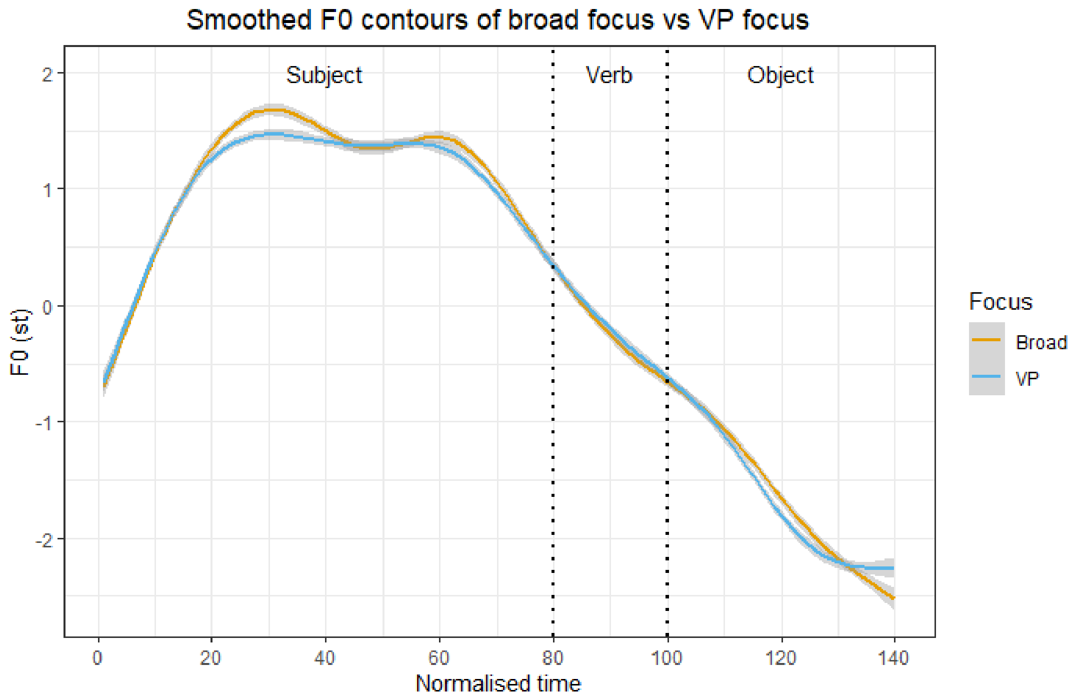

Figure 2-15: F0 contours of broad focus vs VP focus by Cantonese speakers.

F0 contour: Figure 2-15 shows the F0 contours of broad focus and VP focus produced by Cantonese speakers. The contour of VP focus had a lowering in the subject and object positions but then raised lightly towards the final points of the object.

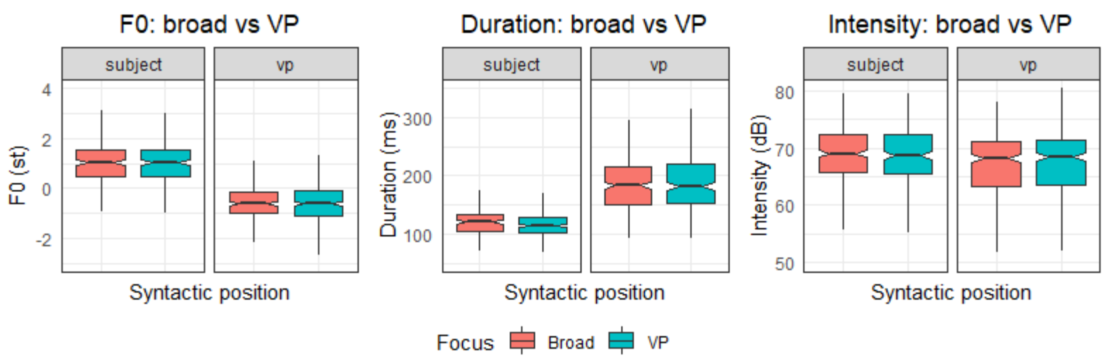

Figure 2-16: Mean values of broad focus vs VP focus by Cantonese speakers.

Mean F0: There was no effect of focus on mean in the subject ($\chi^2(1) = 1.555$, $p = .212$)

or VP ($\chi^2(1) = .119$, $p = .730$) positions.

Duration: The main effect of focus on duration was significant in the subject position ($\chi^2(1) = 4.113$, $p = .043$), shortening one syllable by 2.955 ± 1.455 ms. There was a minor lengthening of duration by 2.396 ± 2.407 ms in the VP position, but the effect was non-significant ($\chi^2(1) = .992$, $p = .319$).

Intensity: The effect of focus on intensity did not reach significance in the subject ($\chi^2(1) = 1.177$, $p = .278$) or VP ($\chi^2(1) = .082$, $p = .775$) positions.

2.3.2.5 Linear discriminant analysis

We also conducted linear discriminant analyses to each focus pair to test the prediction accuracy of the acoustic cues in the Cantonese data. The results are presented in Table 2-5.

Focus pair	Accuracy of prediction		
	Mean F0	Duration	Intensity
Broad vs subject	62.5%	57.3%	56.2%
Broad vs verb	48.6%	56.5%	52.0%
Broad vs object	54.2%	54.8%	52.0%
Broad vs VP	48.0%	54.4%	53.0%

Table 2-5: Predicted accuracy from the LDA (Cantonese natives).

The LDA results suggest that none of the tested acoustic cues contribute to the differentiation between the four focus pairs, as all the prediction accuracies are at chance level. However, it is clear that Cantonese speaker did use duration to some extent in focus marking. The predicted accuracy rates are highest for duration, except for the pair of broad focus and subject focus.

2.3.2.6 An interim summary

We presented the pairwise comparisons of broad focus and other focus types produced by native speakers of Cantonese. It has been shown that the Cantonese speakers did not use F0 or intensity in focus marking for most of the time, but they manipulated duration to mark subject focus and verb focus. The Cantonese speakers also showed pre-focus shortening of the duration for object focus and VP focus.

2.3.3 Comparing focus production by Mandarin and Cantonese speakers

2.3.3.1 F0 contours

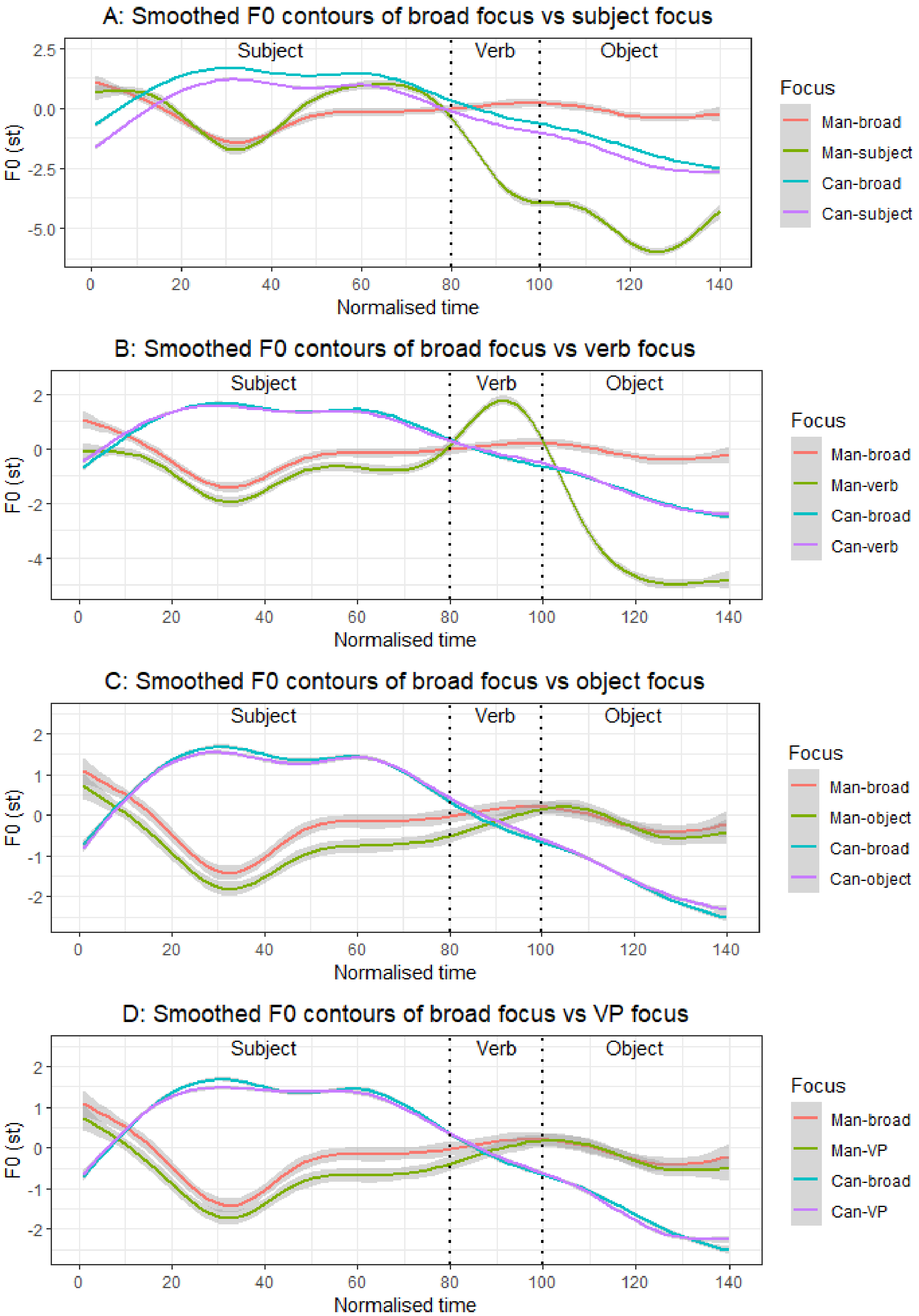

Figure 2-17: F0 contours of focus marking by native speakers.

Man= Mandarin speakers; Can = Cantonese speakers.

Figure 2-17 shows the smoothed F0 contours of Mandarin and Cantonese focus production by native speakers. The red and blue lines are the contours for broad focus by Mandarin and Cantonese native speakers, which have very different patterns. Recall that the first two syllables are determiners and classifiers, and the tones are different in Mandarin and Cantonese, so it is reasonable that the first 40 points do not converge. But from Points 41 on, the contour for Mandarin has a steady raising until a minor declination at around Points 100, and the contour for Cantonese only has a gradual declination towards the end. This suggests the typological differences in the intonation between Mandarin and Cantonese, which needs further investigation.

The green and purple lines from Figure 2-17 represent F0 contours of different focus conditions by Mandarin and Cantonese native speakers. It can be seen from the figures that Mandarin speakers mark different focus types by manipulating pre-, on- or post-focus F0, while the Cantonese speakers hardly make use of F0 to mark focus.

2.3.3.2 Differences in the acoustic cues

In this subsection, we present the results of the differences between broad focus and each of the other focus types (subject, verb and object) by native speakers, which were calculated by subtracting the values of the other focus types from the values of the broad focus.

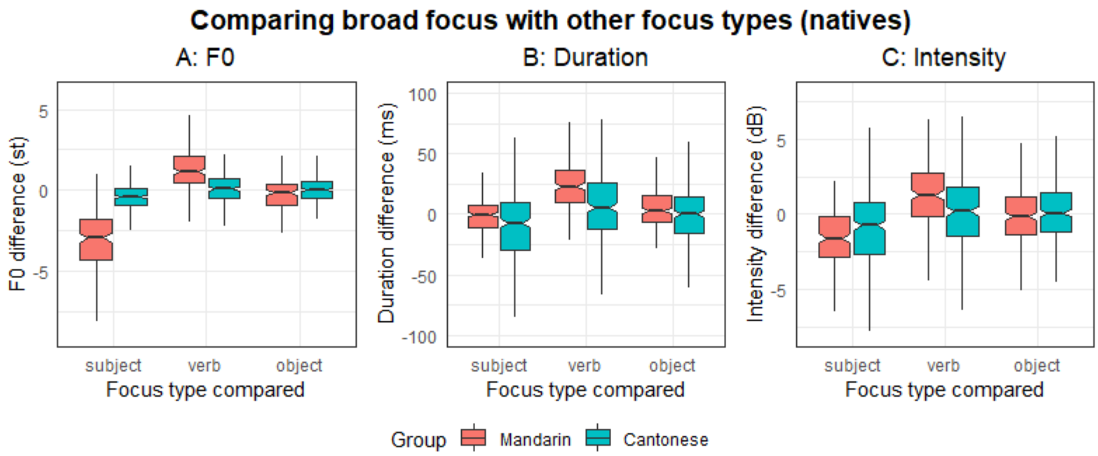

Figure 2-18: Differences between broad and other focus types (natives).

Mean F0 difference: Figure 2-18-a presents the differences in mean F0 between broad and other focus types by native speakers of Mandarin and Cantonese. There were main effects of focus ($\chi^2(1) = 399.992$, $p < .001$) and group ($\chi^2(1) = 6.179$, $p = .013$) on mean F0 difference. The interaction of focus and group also reached significance ($\chi^2(3) = 470.550$, $p < .001$). Post-hoc tests showed that native speakers of Mandarin marked pre-, on- and post-focus components with mean F0, while native speakers of Cantonese did not mark focus with mean F0.

Duration difference: Figure 2-18-b shows the differences in duration between broad and other focus types by native speakers of Mandarin and Cantonese. According to the linear mixed-effects models, there were main effects of focus ($\chi^2(1) = 16.734$, $p < .001$) and group ($\chi^2(1) = 125.183$, $p < .001$) as well as an interaction of focus and group ($\chi^2(3) = 19.418$, $p < .001$) on duration difference. Post-hoc tests suggested the following: 1) the effect of on-focus lengthening was found in both speaker groups but was more robust in Mandarin; 2) the Mandarin speakers hardly marked pre- or post-focus with duration, and the Cantonese speakers showed clear pre- and post-focus shortening of duration.

Intensity difference: Figure 2-18-c represents the differences in intensity between broad and other focus types by native speakers of Mandarin and Cantonese. There were main effects of focus ($\chi^2(1) = 103.534$, $p < .001$) and speaker group ($\chi^2(1) = 7.038$, $p = .008$) on intensity difference. The interaction of these two variables was also found ($\chi^2(4) = 37.368$, $p < .001$). Specifically, there was on-focus increase of intensity and post-focus decrease of intensity in both groups, but the native speakers of Mandarin showed a higher degree of increase and decrease. No pre-focus decrease of intensity was found in either group.

2.4 Discussion

2.4.1 Summary of findings

This chapter examined how native speakers of Mandarin and Cantonese mark prosodic focus in their native languages. Simple SVO sentences with complex subject NPs were used as the stimuli and the production data were elicited with precursor questions. Pairwise comparisons of different focus types were reported in this chapter, and a summary of the acoustic cues that marked prosodic focus by native speakers of Mandarin and Cantonese is presented in Table 2-6.

Focus pair	Group	Subject position	Verb position	Object position
Broad vs subject	Mandarin	F0 ↑, duration ↑	F0 ↓, intensity ↓	F0 ↓, duration ↓, intensity ↓
	Cantonese	F0 ↓, duration ↑	F0 ↓, duration ↓, intensity ↓	F0 ↓, duration ↓, intensity ↓
Broad vs verb	Mandarin	F0 ↓, intensity ↓	F0 ↑, duration ↑, intensity ↑	F0 ↓, intensity ↓
	Cantonese	N/A	duration ↑, intensity ↑	N/A
Broad vs object	Mandarin	F0 ↓, duration ↓, intensity ↓	F0 ↓, duration ↑	N/A
	Cantonese	Duration ↓	N/A	N/A
Broad vs VP	Mandarin	F0 ↓, duration ↓, intensity ↓	N/A	
	Cantonese	Duration ↓	N/A	

Table 2-6: Acoustic cues for focus marking in Mandarin and Cantonese (natives).

Note: The acoustic cues that have successfully distinguished the target focus condition from the broad focus condition for each position are listed in this table. The upward/downward arrows beside the acoustic cues suggest significantly higher/lower values of the target focus condition compared with the broad focus condition for the specific cue.

Both languages mark prosodic focus but with different acoustic cues on various positions. For subject focus and verb focus, Mandarin speakers showed pre-, on- and post-focus changes of F0 and intensity, but Cantonese speakers mostly manipulated on-focus change of duration. Also, both groups of speakers did not mark object focus or VP focus, but there were some pre-focus changes of different acoustic cues.

2.4.2 Focus marking in Mandarin

The Mandarin speakers distinguished broad focus from other focus types in their production: pre-, on- and post-focus changes were identified for marking verb focus; on- and post-focus changes were used for marking subject focus; for object focus and VP focus, when the focused components were in the sentence-final regions, there was no clear on-focus marking, but there was pre-focus lowering of various acoustic cues, which would make the focused region the most prominent in the utterance. Although complex NP subjects were used, our findings were still consistent with data from previous studies on Mandarin focus production in that the focused subjects showed higher F0 and longer duration regardless of the complexity (Yiya Chen & Gussenhoven, 2008; Xu et al., 2012), suggesting that focus can be projected onto a region larger than one single word. However, if we examine further on Figure 2-1, it is obvious that the F0 contours diverge from Points 41 to 80 only, which correspond to the head nouns. Points 1 to 40, which correspond to the determiners and the classifiers, show no difference between broad focus and subject focus. To the best of our knowledge, there is only one study that has worked on focus production of complex NPs in Mandarin (Hsu & Xu, 2017), which included one disyllabic numeral, one monosyllabic classifier and one monosyllabic noun within each NP. When the whole NP was the focused constituent, the noun also received more prominence than the numeral, which is similar to our observations. Yet it still needs to be investigated whether the observed focus marking pattern on nouns results from syntactic position or part of speech, which will advance our understanding of focus marking on complex structures in Mandarin.

Also, when focus domain was taken into account, the Mandarin speakers showed differences in marking broad focus vs VP focus, verb focus vs VP focus and broad focus vs object focus, but they did not distinguish object focus from VP focus, which indicates the possibility of having placed focus on a more complex syntactic unit than the object of the sentence (namely, object focus is placed on the whole VP, while not on the object position only). This result partially supports the hypothesis from focus projection theories that the narrow focus on sentential object may be projected further onto a larger unit such as the whole VP (Bishop, 2017). Of particular importance is the Default Prosody approach within focus projection theories, which suggests that prosodic prominence is determined by default principles of prosody within a focused constituent (Arregi, 2016). By default, the most prominent element of a constituent in English is the rightmost one (Chomsky & Halle, 1968), and consequently, when focus is on the object, on the VP, or on the full sentence, the prominence is always on the object position. With data from Chinese languages, we also showed that the Mandarin and Cantonese speakers did not differentiate object focus and VP focus, but they did exhibit some minor differences in the pre-focused components to distinguish object focus and VP focus from broad focus, suggesting that focus can be projected from object into the whole VP, but it remains to be tested whether focus can be projected to the whole utterance (broad focus). Also, more studies are needed to provide satisfactory explanations as to why such projection is possible or preferred and how is the projection conditioned.

2.4.3 Focus marking in Cantonese

Results from the Cantonese group suggest that, Cantonese speakers distinguished broad focus and other focus types mainly with manipulation of duration in the on-focus region, although there were some pre- and post-focus changes of duration. In some cases, intensity was also used by Cantonese speakers to mark focus. Besides, we did not find any contribution of F0 in focus marking from Cantonese speakers' data (except for the subject focus, but we have not found any satisfactory explanations as to why the F0 of subject focus parallels but is always lower than broad focus), which might be due to the fact that there are six lexical tones in Cantonese and the main role of F0 is to maintain lexical contrasts (Wu, 2013). Recall that we used the high-level tone only in our stimuli, and because it is the highest tone in the phonological space of the Cantonese speakers, it is possible that the native speakers tend not to manipulate the F0 of the T1 syllables as the T1 syllables may have already reached the upper limit of the speakers' F0 range.

Even we used complex NP subjects in the stimuli, the Cantonese speakers still clearly distinguished subject focus from broad focus with the manipulation of on-focus and post-focus duration. But it is surprising that broad focus had higher F0 and intensity than subject focus across all the syntactic positions, especially when the Cantonese speakers could manipulate duration in marking broad focus vs other focus types (but not subject focus). This finding contradicts the results from previous studies. For focus production of Cantonese Tone 1 syllables, Wu (2013) found that broad focus always had the lowest F0, and the other focus types consistently showed higher F0 than broad focus across the whole sentence. Moreover, in another study that examined Cantonese focus production in different tones (Gu & Lee, 2007), focus brings a wide-range increase of F0, namely, F0 is increased not only on the focused components but also on surrounding components. It is not easy to directly compare our results with these two studies given the methodological divergence. Gu and Lee (2007) invited only one participant for the recording and Wu (2013) asked the participants to read the questions themselves. In this study, sentences with different focus conditions were elicited with precursor questions asked by the experimenter. This puzzle, therefore, still requires further investigations.

Besides, our data do not confirm the lack of PFC in Cantonese. For verb focus, it is clear that Cantonese does not have PFC of F0 and intensity, which echoes the findings from Wu (2013). But for subject focus, there is evidence of post-focus lowering of intensity as well as shortening of duration in the verb and object positions. One might argue that the lowering of intensity in our data is due to the decrease of F0 in the verb and object positions and therefore cannot be used to support PFC in Cantonese. Recall that for broad focus and subject focus, F0 is always lower under subject focus than under broad focus, including in the subject position, where there is on-focus increase of intensity. This suggests that the lowering of intensity should not be caused by the association between F0 and intensity (Zee, 1978); otherwise, it is difficult to explain the observed on-focus increase of intensity and decrease of F0 in our data. While no conclusive point could be made for the interaction between F0 and intensity, it is evident that PFC exists in the intensity and duration of subject focus in Cantonese. In fact, studies on focus production of other Chinese languages (all tonal) show that, rather than an all-or-nothing phenomenon, the actual realisation of PFC might be dependent on different factors, such as tonal contexts (Yiya Chen, 2010), segmental contexts (Huang & Hsieh, 2016), and lexical tones (Yang, Chen, & Li, 2019). For the high-level tone in

Chongming Chinese, a variety of Wu Chinese dialect, which is similar to the high-level tone we tested in this study, there is no PFC of F0 or intensity, but a post-focus shortening of syllable duration is found (Yang et al., 2018, 2019). No PFC is observed for other tones in Chongming Chinese. It is proposed that we should not simply classify a language as having PFC or not. Rather, future studies are called for to examine the constraints on PFC in Cantonese and in other languages.

2.4.4 Towards a typology of focus marking

The results above indicate the differences in focus production of Mandarin and Cantonese, although exactly the same syntactic structure and focus conditions were adopted in the test design. One noticeable difference in focus marking is the extensive use of F0 in Mandarin and the exclusive use of duration in Cantonese. This points to the possibility of a typological difference in Mandarin and Cantonese focus marking, despite the fact that these two languages are closely related. While more data are needed for a better understanding of such a typology, we can at least provide some explanations for the observed divergence here.

First, although both Mandarin and Cantonese are tone languages with a simple syllable structure, there is a typological difference from the perspective of prosody: Mandarin has lexical stress in addition to lexical tones while Cantonese does not have stress (Jun, 2005). Syllables in Cantonese generally receive the same weight (Bauer &

Benedict, 1997) and consequently have relatively equal intensity and pitch height (across the same lexical tone). In Mandarin unstressed syllables, the duration, amplitude and pitch height are reduced significantly (Chao, 1968). It is plausible that Mandarin speakers are used to manipulating these acoustic cues (particularly F0 and intensity) and can employ them for communication purposes such as marking focus. Because Cantonese does not have stress at the lexical level, native speakers of Cantonese may not rely much on the acoustic cues to mark focus. This claim can be backed up by data from the acquisition of English lexical stress by Mandarin- and Cantonese-speaking learners, wherein the Mandarin speakers had higher accuracy rates in the perception task and showed a higher degree of contrast in the production task than the Cantonese speakers (Li & Grigos, 2018). In a more recent study, the Mandarin speakers' advantage in the acquisition of English lexical stress is also attributed to the existence of lexical stress in their native Mandarin (Li & Grigos, 2021). If this hypothesis is true, our data then complement previous findings in that the acoustic sensitivity benefited from an L1 (such as from the neutral tone in this case) can also be applied to L2 learning of larger prosodic units like prosodic focus. Also, Cantonese has a more complex tonal system with six lexical tones. To maintain the identity of the tones, the Cantonese speakers may tend not to manipulate F0. However, whether these explanations hold still needs to be tested in future studies.

Besides, the differences between Mandarin and Cantonese focus marking might be caused by the different weighting of means of focus realisation. As shown in Chapter 1, there are various means to mark focus. Native speakers of a particular language may have a preference for one means over another, which has been supported by some recent empirical evidence. For example, although Mandarin and English allow both prosody and cleft structure to mark focus, in a processing study that adopted speeded false alternative rejection tasks, prosodic cues are more effective than syntactic clefting for both Mandarin and English participants, indicating the native speakers' favouring prosodic cues over syntactic cues in focus marking (Yan & Calhoun, 2020). We have provided an example of Cantonese marking focus with dislocation in Example (7). In addition to that, Cantonese also uses cleft structure to mark focus. Given the fact that there are already several syntactic means for focus marking, it is likely that Cantonese speakers may rely more on these syntactic cues than prosodic cues in focus marking, which is contrary to the case for Mandarin speakers. This hypothesis is supported by a recent study on focus perception by Cantonese-speaking children and adults (Ge & Yuen, 2020). In their study, two types of stimuli were manipulated: stimuli with syntactic marking of focus, and stimuli with prosodic marking of focus. The adult participants showed a preference for the syntactic cues over prosodic cues, and the children were only sensitive to prosodic cues. The results of these perception studies further support our claim that Mandarin and Cantonese may be typologically different in terms of focus marking. Mandarin may be a language that favours prosodic marking of focus, while Cantonese may have a preference for syntactic marking of focus.

This chapter compared the production of prosodic focus by native speaker of Mandarin and Cantonese. Both languages mark prosodic focus but with different acoustic cues on various positions. For subject focus and verb focus, Mandarin speakers showed pre-, on- and post-focus changes of F0 and intensity, but Cantonese speakers mostly manipulated on-focus change of duration. Also, both groups of speakers did not mark object focus or VP focus, but there were some pre-focus changes of different acoustic cues. Relevant issues regarding focus production are discussed and a typological difference in Mandarin and Cantonese focus marking is proposed.

Chapter 3: Production of Mandarin and Cantonese Focus by Immigrants

3.1 Introduction

This chapter reports how immigrants mark prosodic focus in their L1 Mandarin and L2 Cantonese and also compares the immigrants' production with the native speakers' production in the respective language.

While prosody in L1 attrition and L2 acquisition is a relatively under-researched area, some studies have shown the cross-linguistic inference between bilingual's two languages. For example, Gut and Pillai (2014) investigated the prosodic marking of English information structure by native speakers of Malaysian. The results suggested that the Malaysian speakers had similar patterns when marking focus in Malay and English, pointing to the L1 influence on speech prosody. Moreover, de Leeuw, Mennen and Scobbie (2012) showed evidence of L1 attrition in the intonational alignment of pre-nuclear rise, where the German-English bilingual group consistently had earlier starting points of pre-nuclear rise in German compared with the native control group.

As has been shown in Chapter 2, Mandarin and Cantonese speakers do not resemble each other in the production of prosodic focus. The Mandarin speakers make extensive use of F0 in focus marking whereas the Cantonese speakers mostly use duration to indicate focus. Also, for sentence-final focus types, namely, object focus and VP focus in our design, only the Cantonese speakers have on-focus change. If the claim from LILt that L1 and L2 share the same phonological space for the contrasts is true, there should be some interactions between the immigrants' L1 and L2 in their actual implementation of focus marking.

The specific questions to be addressed in this chapter include the following:

1) How do immigrants mark prosodic focus in their L1 Mandarin and L2 Cantonese?

2) Is the immigrants' Mandarin influenced by their L2 Cantonese?

3) Is the immigrants' Cantonese influenced by their L1 Mandarin?

3.2 Methodology

3.2.1 Participants

The participants consisted of 22 immigrants (19 females, 3 males; aged: 30.14 ± 4.30) who speak Mandarin as the L1 and have been exposed to Cantonese since their arrival in Hong Kong. Although the participants came from different provinces of Northern China and may speak other Northern Chinese dialects, they reported that their dominant language is Mandarin only but not other Northern dialects. The participants all arrived in Hong Kong after puberty (average age: 22.73 ± 4.21) and the average length of residence was 7.41 ± 3.11 years. To assess their language profile of Cantonese and Mandarin, the immigrants completed a language background questionnaire prior to the recording session. The questionnaire was an adapted version of Bilingual Language Profile (Birdsong, Gertken, & Amengual, 2012), which collected information on language history, language use, language proficiency and language attitudes, and converted the results to scores for each subsection. The scores suggest that, the participants are fluent speakers of Cantonese, although they were more dominant in Mandarin at the time of attending the experiment. No participants had any history of speaking, hearing or language difficulties.

3.2.2 Experiment design

The same stimuli for focus production were prepared for the immigrants. The immigrants attended both the Mandarin and Cantonese experiments, and they all attended the Cantonese recording prior to the Mandarin recording. For the immigrants, each sentence was repeated only once. In total, 1,540 target sentences (3 Mandarin sentences * 5 focus types * 2 repetitions * 22 immigrants + 4 Cantonese sentences * 5 focus types * 2 repetitions * 22 immigrants) were collected.

3.2.3 Data collection and analysis

As described in the previous chapter on focus production of native speakers, the same procedures for data collection and analysis were performed for the data of immigrants' focus production.

3.3 Results

3.3.1 Mandarin focus marking by immigrants

3.3.1.1 Broad focus vs subject focus

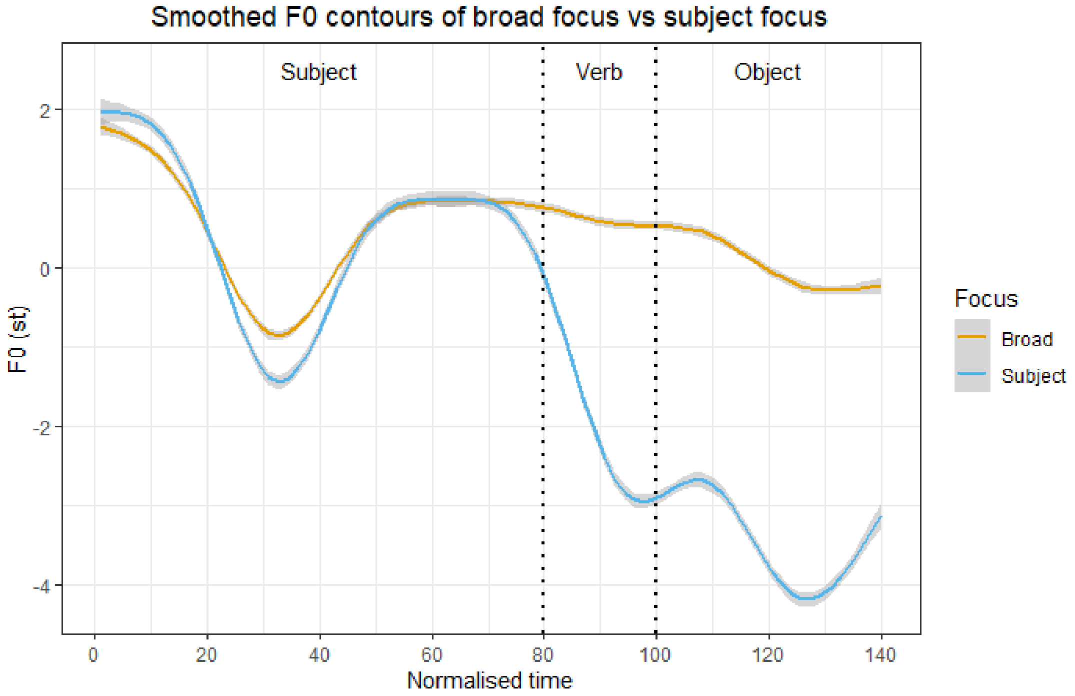

Figure 3-1: F0 contours of Mandarin broad focus vs subject focus by immigrants.

F0 contour: Figure 3-1 shows the smoothed F0 contours of Mandarin SVO sentences under broad focus and subject focus produced by immigrants. The F0 contours generally overlapped in the subject position, suggesting no on-focus increase of F0. For the subject focus condition, the F0 contour started to drop sharply in the verb and object positions.

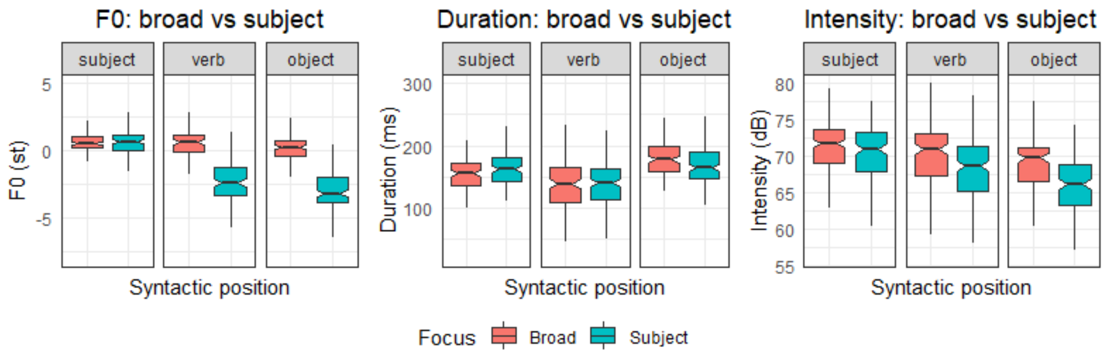

Figure 3-2: Mean values of Mandarin broad focus vs subject focus by immigrants.

Mean F0: There was no main effect of focus on mean F0 in the subject position ($\chi^2(1) = .019$, $p = .890$), as has been observed in the F0 contours. The focus effect was significant in the verb ($\chi^2(1) = 237.480$, $p < .001$) and object ($\chi^2(1) = 209.070$, $p < .001$) positions, lowering the mean F0 by 2.934 ± 0.146 st and 3.617 ± 0.199 st, respectively.

Duration: There was a main effect of focus on duration in the subject position ($\chi^2(1) = 33.945$, $p < .001$), lengthening one syllable by 9.360 ± 1.553 ms. No effect of focus was found on the duration in the verb position ($\chi^2(1) = .092$, $p = .762$). In the object position, there was a post-focus shortening of 12.264 ± 2.184 ms for each syllable ($\chi^2(1) = 29.740$, $p < .001$).

Intensity: There were main effects of focus on intensity in the subject ($\chi^2(1) = 239.000$, $p = .001$), verb ($\chi^2(1) = 67.853$, $p < .001$) and object ($\chi^2(1) = 119.870$, $p < .001$) positions, lowering the intensity by 0.515 ± 0.160 dB, 2.081 ± 0.236 dB and 2.919 ± 0.234 dB, respectively.

3.3.1.2 Broad focus vs verb focus

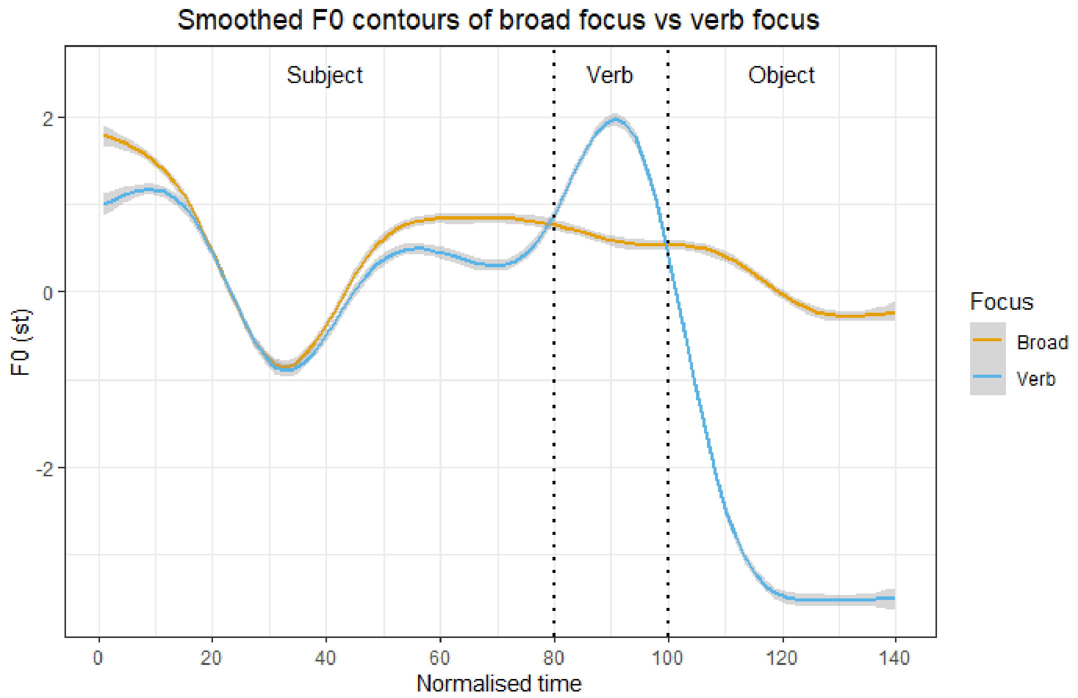

Figure 3-3: F0 contours of Mandarin broad focus vs verb focus by immigrants.

F0 contour: Figure 3-3 presents the smoothed F0 contours of Mandarin SVO sentences under broad focus and verb focus produced by immigrants. In the subject position, the contour for verb focus was different from the contour for broad focus only in the initial Points 1 to 15 and in the final Points 45 to 80. The verb focus raised the F0 contour in the verb position and lowered the F0 contours in the object position.

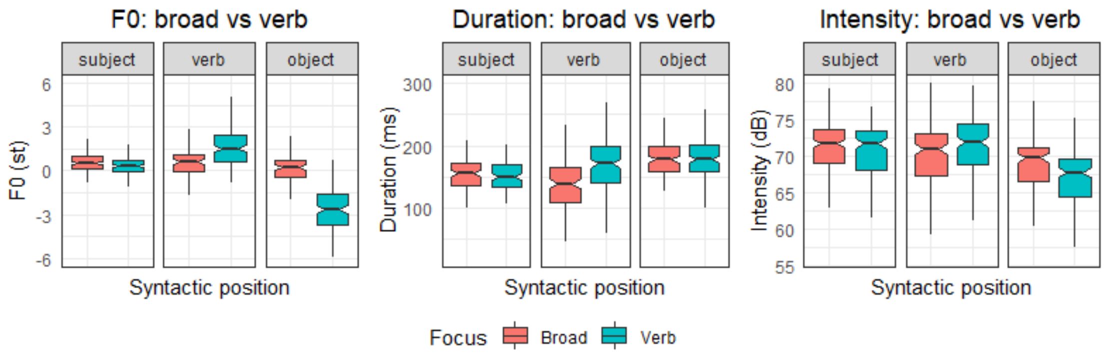

Figure 3-4: Mean values of Mandarin broad focus vs verb focus by immigrants.

Mean F0: There was a pre-focus lowering of mean F0 in the subject position (0.231 ± 0.075 st; $\chi^2(1) = 9.367$, $p = .002$), followed by an on-focus raising of mean F0 in the verb (1.213 ± 0.107 st; $\chi^2(1) = 102.590$, $p < .001$) position. There was also a main effect of focus on mean F0 in the object position ($\chi^2(1) = 211.79$, $p < .001$), lowering it by 3.006 ± 0.164 st.

Duration: The main effect of focus on duration was non-significant in the subject ($\chi^2(1) = 3.806$, $p = .051$) and object ($\chi^2(1) = .400$, $p = .527$) positions. There was a main effect of focus on duration in the verb position ($\chi^2(1) = 125.19$, $p < .001$), lengthening one syllable by 31.518 ± 2.462 ms.

Intensity: No effect of focus was found on intensity in the subject position ($\chi^2(1) = 2.281$, $p = .131$). There was a main effect of focus on intensity in the verb position ($\chi^2(1) = 27.216$, $p < .001$), which increased the intensity by 1.233 ± 0.230 dB. The main effect of focus on intensity also reached significance in the object position ($\chi^2(1) = 62.617$, $p < .001$), which lowered the intensity by 1.835 ± 0.217 dB.

3.3.1.3 Broad focus vs object focus

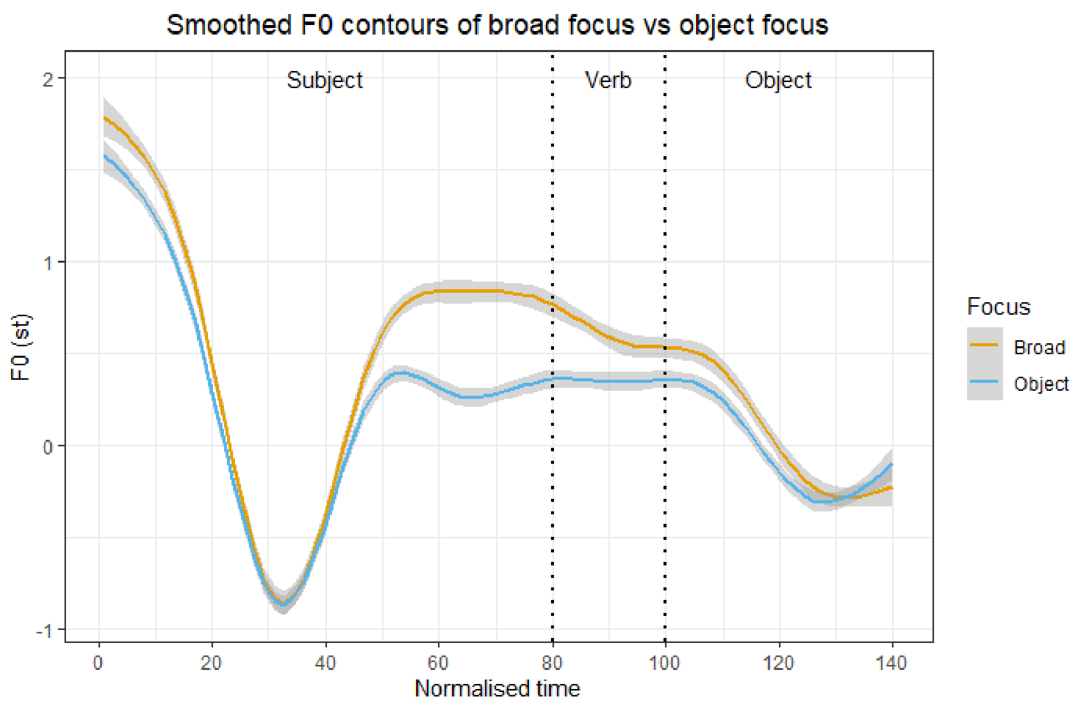

Figure 3-5: F0 contours of Mandarin broad focus vs object focus by immigrants.

F0 contour: The contours of broad focus and object focus did not differ from each other in the object position, but there was pre-focus lowering for object focus in the subject and verb positions, particularly in the head noun position (Points 40 to 80).

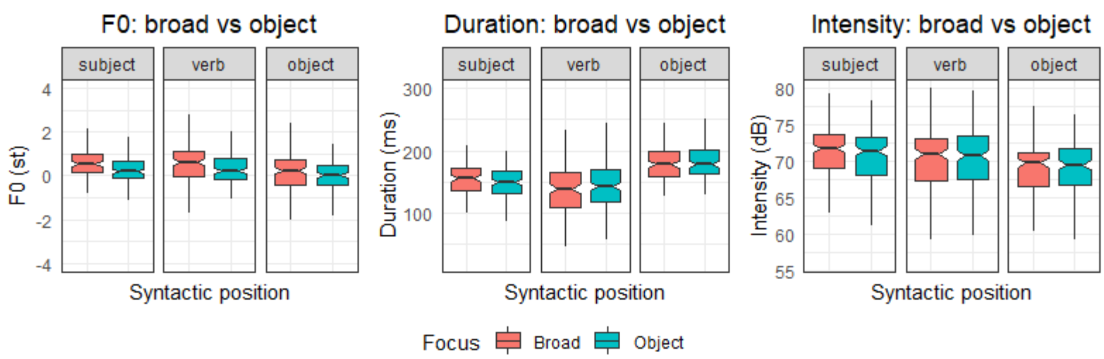

Figure 3-6: Mean values of Mandarin broad focus vs object focus by immigrants.

Mean F0: There was a main effect of focus on mean F0 in the subject ($\chi^2(1) = 15.239$,

$p < .001$) and verb ($\chi^2(1) = 6.064$, $p = .014$) positions, lowering the mean F0 by 0.291 ± 0.073 st and 0.227 ± 0.092 st, respectively. No effect of focus was found in the object ($\chi^2(1) =.852$, $p = .356$).

Duration: In the subject position, the main effect of focus on duration was significant ($\chi^2(1) = 11.638$, $p < .001$), which shortened one syllable by 5.020 ± 1.457 ms. The effect of focus did not reach significance in the verb ($\chi^2(1) = 3.447$, $p = .063$) or object ($\chi^2(1) = 1.928$, $p = .165$) positions.

Intensity: There was a main effect of focus on intensity in the subject position ($\chi^2(1) = 4.199$, $p = .040$), decreasing the intensity by 0.310 ± 0.151 dB. There was no main effect of focus on intensity in the verb ($\chi^2(1) = .021$, $p = .884$) or object ($\chi^2(1) = .441$, $p = .507$) positions.

3.3.1.4 Broad focus vs VP focus

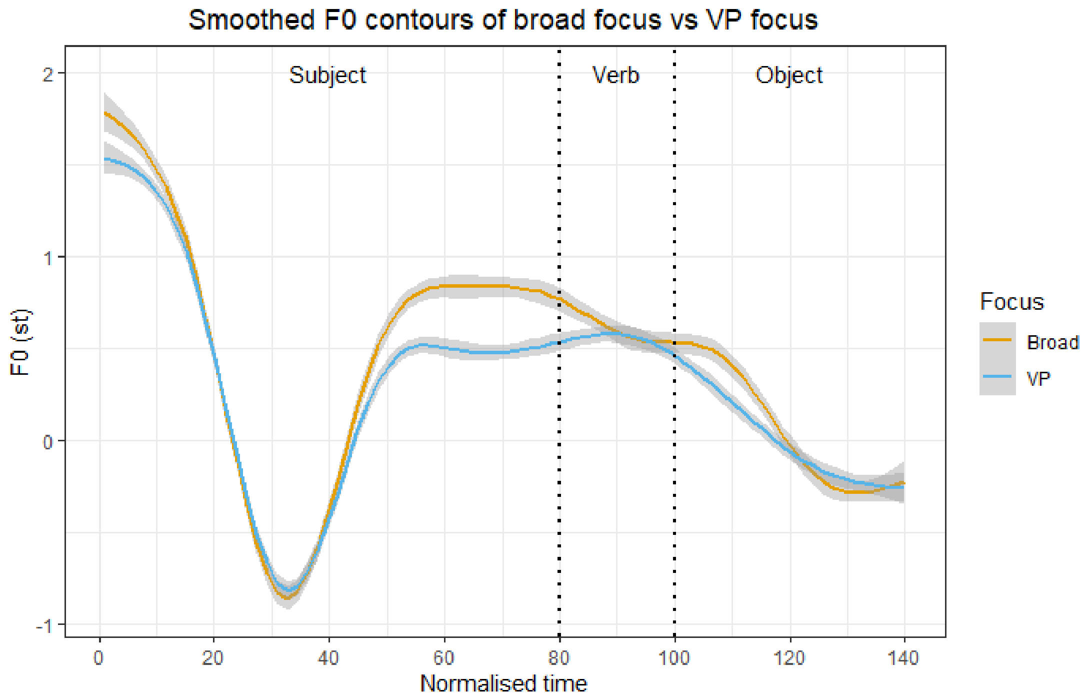

Figure 3-7: F0 contours of Mandarin broad focus vs VP focus by immigrants.

F0 contour: The contours of broad focus and VP focus were very similar, except for that there was pre-focus lowering of F0 in the final Points 50 to 80 of the subject position, which also corresponds to the head nouns.

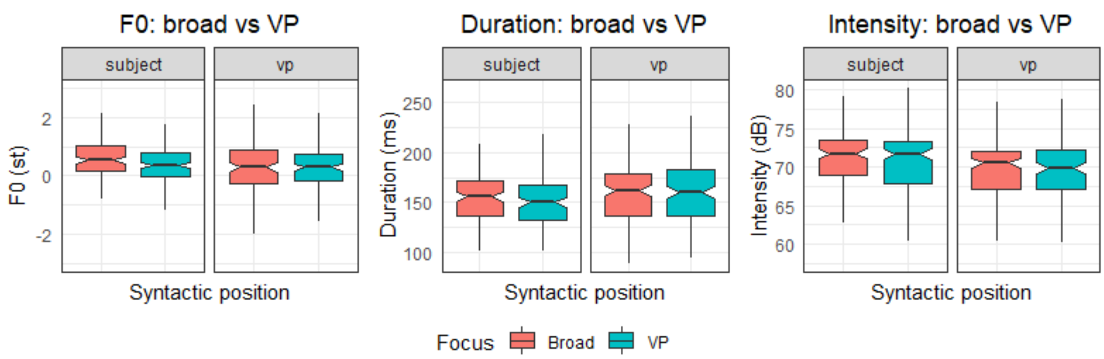

Figure 3-8: Mean values of Mandarin broad focus vs VP focus by immigrants.

Mean F0: In the subject position, there was a main effect of focus on mean F0 in the

VP ($\chi^2(1) = 5.629$, $p = .018$), which lowered the mean F0 by 0.177 ± 0.074 st. No effect of focus was found in the VP position ($\chi^2(1) = .191$, $p = .662$).

Duration: There was a main effect of focus on duration in the subject position ($\chi^2(1) = 7.460$, $p = .006$), shortening each syllable by 4.247 ± 1.546 ms. No effect of focus was found in the VP position ($\chi^2(1) = 1.859$, $p = .173$).

Intensity: The main effect of focus on intensity was found in the subject position ($\chi^2(1) = 4.620$, $p = .032$), which decreased the intensity by 0.362 ± 0.168 dB. No effect of focus was found in the VP position ($\chi^2(1) = .684$, $p = .408$).

3.3.1.5 Comparing Mandarin focus marking by native speakers and immigrants

In this section, we compared the production of Mandarin prosodic focus by native speakers and immigrants, with a summary of the acoustic cues used by native speakers and immigrants presented in Table 3-1.

Focus pair	Group	Subject position	Verb position	Object position
Broad vs subject	Natives	F0 ↑, duration ↑	F0 ↓, intensity ↓	F0 ↓, duration ↓, intensity ↓
	Immigrants	Duration ↑, intensity ↓	F0 ↓, intensity ↓	F0 ↓, duration ↓, intensity ↓
Broad vs verb	Natives	F0 ↓, intensity ↓	F0 ↑, duration ↑, intensity ↑	F0 ↓, intensity ↓

	Immigrants	F0 ↓	F0 ↑, duration ↑, intensity ↑	F0 ↓, intensity ↓
Broad vs object	Natives	F0 ↓, duration ↓, intensity ↓	F0 ↓, duration ↑	N/A
	Immigrants	F0 ↓, duration ↓, intensity ↓	F0 ↓	N/A
Broad vs VP	Natives	F0 ↓, duration ↓, intensity ↓	N/A	
	Immigrants	F0 ↓, duration ↓, intensity ↓	N/A	

Table 3-1: Acoustic cues that marked prosodic focus in Mandarin.

There are several differences between native speakers and immigrants in marking Mandarin focus. First, the immigrants did not use F0 to mark focus as much as the native speakers. For example, when marking subject focus, the immigrants showed no change in the subject position. Second, for subject focus and verb focus, which belong to non-sentence-final focus types, the native speakers showed a higher degree of PFC than the immigrants. Third, for verb focus, object focus and VP focus, the native speakers consistently had pre-focus lowering of F0, but the immigrants only lowered the F0 in some portions of the contour. These differences suggest that, although the immigrants also mark prosodic focus in Mandarin, the distinction between broad focus and other focus types is not as clear as the native speakers.

Next, we presented the results of the differences of the acoustic cues between broad focus and each of the other focus types (subject, verb and object). The comparisons were conducted on the verb position only.

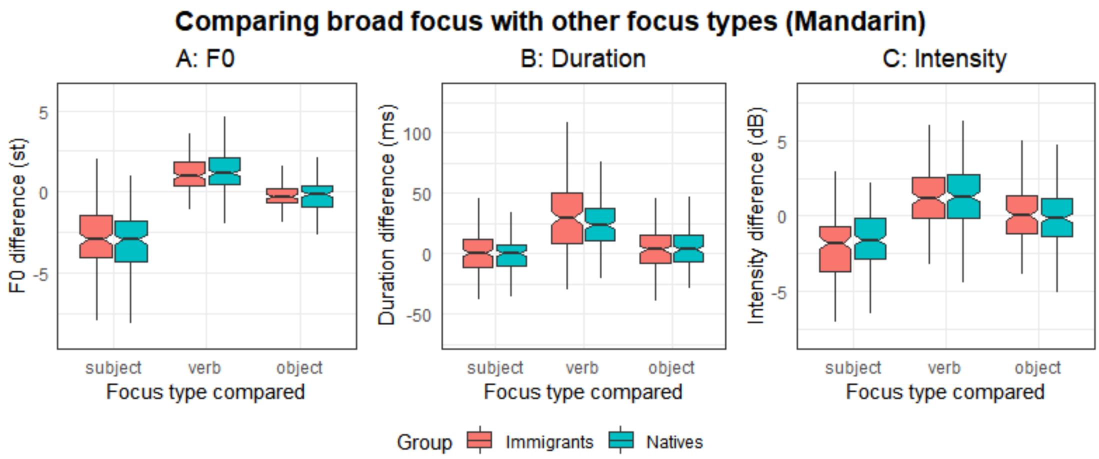

Figure 3-9: Differences between broad and other focus types in Mandarin.

Mean F0 difference: Figure 3-9-a shows the differences in mean F0 between broad and other focus types by native speakers and immigrants. Generally, the native speakers and immigrants had similar patterns, where there was obvious on-focus increase and post-focus decrease of mean F0 and there was no much pre-focus change of mean F0. The observation was also confirmed by statistics. There was an effect of focus on mean F0 difference ($\chi^2(1) = 814.800$, $p < .001$) but no effect of group on mean F0 difference ($\chi2(1) = .069$, $p = .793$). Post-hoc tests suggest the following: 1) native speakers showed 0.346 ± 0.188 st larger mean F0 differences between broad focus and verb focus than immigrants ($\chi2(1) = 3.085$, $p = .079$); 2) the differences between each pair of the calculated differences all reached significance (ps < .001).

Duration difference: Figure 3-9-b represents the differences in duration between broad and other focus types by native speakers and immigrants. According to the linear mixed-effects models, there was a main effect of focus on duration difference ($\chi^2(1) = 311.370$, $p < .001$) but no effect of speaker group ($\chi2(1) = .406$, $p = .524$). Post-hoc tests suggest the following: 1) there was almost no difference between broad focus and subject focus in both groups of speakers for the post-focus condition, because the difference values were very close to zero; 2) there was apparent on-focus lengthening in both groups, and the immigrants had 3.303 ± 2.565 ms longer duration than the native speakers, although the difference did not reach significance ($\chi2(1) = 1.669$, $p = .199$); 3) there was slight pre-focus lengthening of duration in both groups; 4) the differences between each pair reached significance (ps < .013).

Intensity difference: Figure 3-9-c shows the differences in intensity between broad and other focus types by native speakers and immigrants. There were main effects of focus ($\chi^2(1) = 6.055$, $p < .001$) and speaker group ($\chi^2(1) = 261.695$, $p < .001$) on intensity difference. The interaction of these two variables also reached significance ($\chi^2(4) = 272.940$, $p < .001$). Specifically, there was on-focus increase of intensity and post-focus decrease of intensity in both groups, but the natives showed a higher degree of increase and the immigrants showed a higher degree of decrease. No pre-focus decrease of intensity was found in either group.

3.3.1.6 Linear discriminant analysis

Also, we conducted linear discriminant analyses to each focus pair to test the prediction accuracy of the acoustic cues in the immigrants' Mandarin data. The results are presented in Table 3-2.

Focus pair	Accuracy of prediction		
	Mean F0	Duration	Intensity
Broad vs subject	90.1%	62.2%	73.3%
Broad vs verb	87.5%	69.9%	75.9%
Broad vs object	58.0%	51.1%	56.0%
Broad vs VP	54.8%	53.1%	56.5%

Table 3-2: Predicted accuracy from the LDA (immigrants' Mandarin).

In the immigrants' Mandarin data, the LDA results generally echoed the LDA results of the Mandarin native speakers. In other words, mean F0 was very useful in distinguishing subject focus and verb focus from broad focus, and the accuracy for teasing apart the object focus and VP focus from broad focus was just slightly above chance level. However, for the immigrants, the prediction accuracies of mean F0 were lower and those of duration were higher than the native speakers, which should be due to the influence from the immigrants' L2 Cantonese.

3.3.1.7 An interim summary

This section presented data from Mandarin focus production by immigrants and conducted pairwise comparisons of broad focus with other focus types. After that, we also compared the production data by native speakers and immigrants. We have shown that, the immigrants did not use F0 and intensity in focus marking as much as the native speakers of Mandarin did, but they made use of duration to a greater extent compared with the native speakers.

3.3.2 Cantonese focus marking by immigrants

3.3.2.1 Broad focus vs subject focus

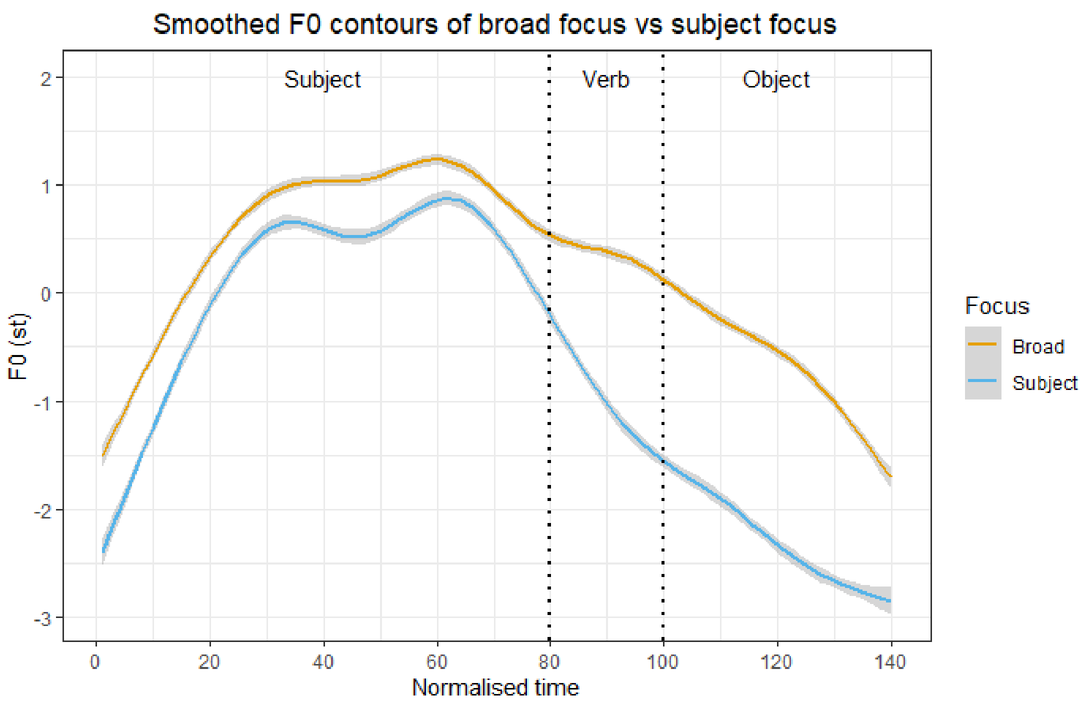

Figure 3-10: F0 contours of Cantonese broad focus vs subject focus by immigrants.

F0 contour: Figure 3-10 shows the F0 contours of broad focus and subject focus in immigrants' Cantonese. The F0 of subject focus was always lower than that of broad focus, but there was a decrease of F0 in the post-focus verb and object positions, which suggested post-focus compression of F0.

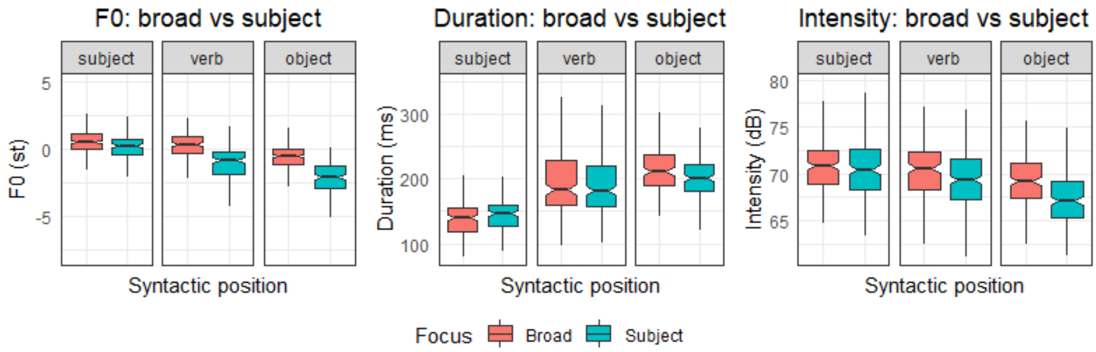

Figure 3-11: Mean values of Cantonese broad focus vs subject focus by immigrants.

Mean F0: There was a main effect of focus on the mean F0 in the subject ($\chi^2(1)$ = 22.784, $p < .001$), verb ($\chi^2(1) = 91.193$, $p < .001$) and object ($\chi^2(1) = 144.800$, $p < .001$) positions, and the mean F0 was lowered by 0.426 ± 0.088 st, 1.524 ± 0.149 st and 1.638 ± 0.122 st in each position, respectively.

Duration: The main effect of focus on duration was significant in the subject position ($\chi^2(1) = 13.990$, $p < .001$), lengthening each syllable by 6.931 ± 1.836 ms. There was also an effect of focus on duration in the object position ($\chi^2(1) = 24.903$, $p < .001$), which shortened each syllable by 12.235 ± 2.409 ms. No effect of focus was found in the verb position ($\chi^2(1) = .195$, $p = .659$).

Intensity: There was a main effect of focus on the intensity in the subject ($\chi^2(1) = 5.352$,

$p = .021$), verb ($\chi^2(1) = 32.305$, $p < .001$) and object ($\chi^2(1) = 93.362$, $p < .001$) positions, which decreased the intensity by 0.356 ± 0.154 dB, 1.231 ± 0.212 dB and 1.923 ± 0.185 dB, respectively.

3.3.2.2 Broad focus vs verb focus

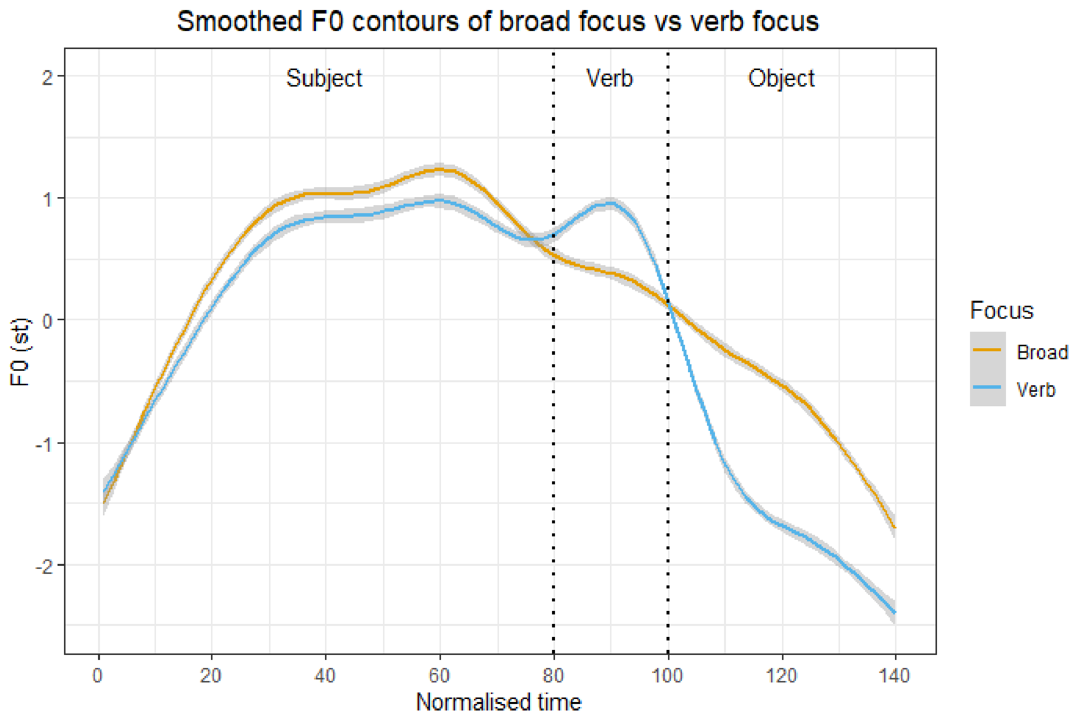

Figure 3-12: F0 contours of Cantonese broad focus vs verb focus by immigrants.

F0 contour: Figure 3-12 plots the F0 contours of broad focus and verb focus. There was evidence of pre-focus lowering, on-focus raising and post-focus lowering of F0 under verb focus.

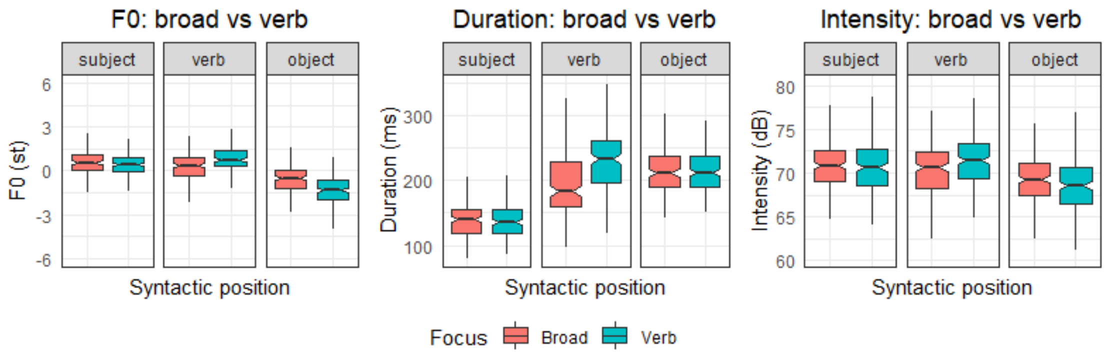

Figure 3-13: Mean values of Cantonese broad focus vs verb focus by immigrants.

Mean F0: There was a pre-focus lowering of mean F0 in the subject position (0.190 ± 0.083 st; $\chi^2(1) = 5.272$, $p = .022$), followed by an on-focus raising of mean F0 in the verb position (0.531 ± 0.087 st; $\chi^2(1) = 35.748$, $p < .001$) and a post-focus decreasing of mean F0 in the object position (0.927 ± 0.115 st; $\chi^2(1) = 59.186$, $p < .001$).

Duration: There was a main effect of focus on duration in the verb ($\chi^2(1) = 97.974$, $p < .001$), and the duration of one syllable was lengthened by 43.942 ± 4.116 ms. No focus effect was found in the subject ($\chi^2(1) = .595$, $p = .441$) or verb ($\chi^2(1) = .034$, $p = .854$) positions.

Intensity: The main effect of focus on intensity was significant in the subject ($\chi^2(1) = 5.802$, $p = .016$), verb ($\chi^2(1) = 19.346$, $p < .001$) and object ($\chi^2(1) = 21.986$, $p < .001$) positions. Specifically, focus decreased the intensity by 0.369 ± 0.153 dB and 0.896 ± 0.188 dB in the subject and object positions, respectively, and increased the intensity by 0.938 ± 0.210 dB in the verb position.

3.3.2.3 Broad focus vs object focus

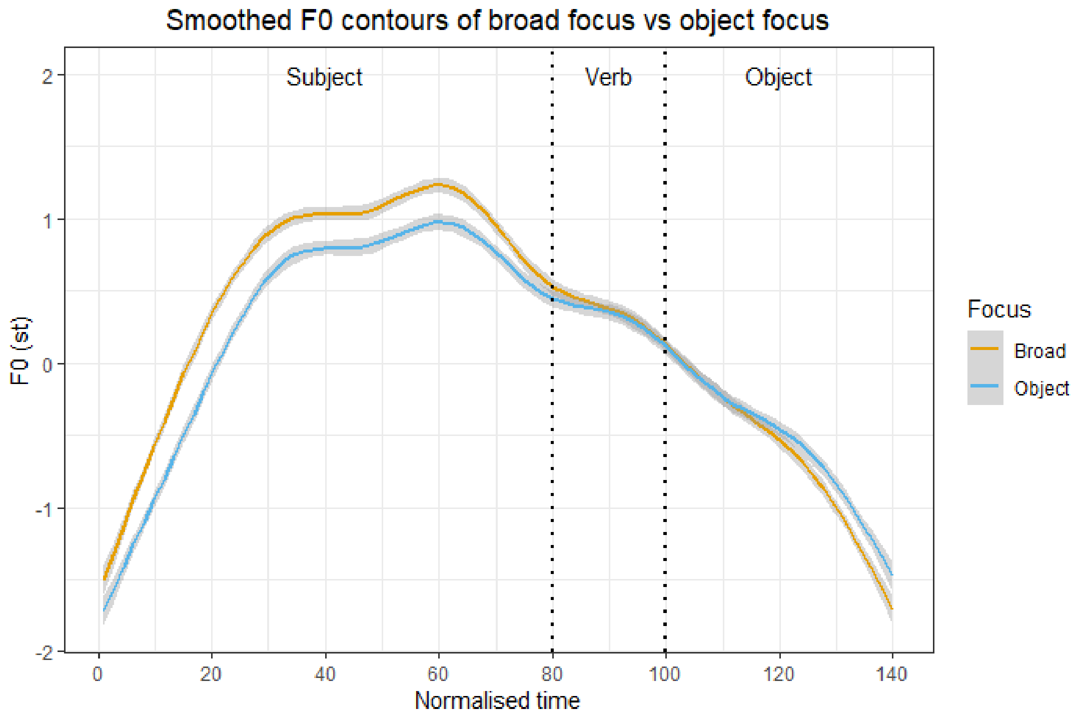

Figure 3-14: F0 contours of Cantonese broad focus vs object focus by immigrants.

F0 contour: Figure 3-14 shows the F0 contours of broad focus and object focus. Although the two contours generally overlapped, there was pre-focus lowering in the subject position and on-focus raising in the object position.

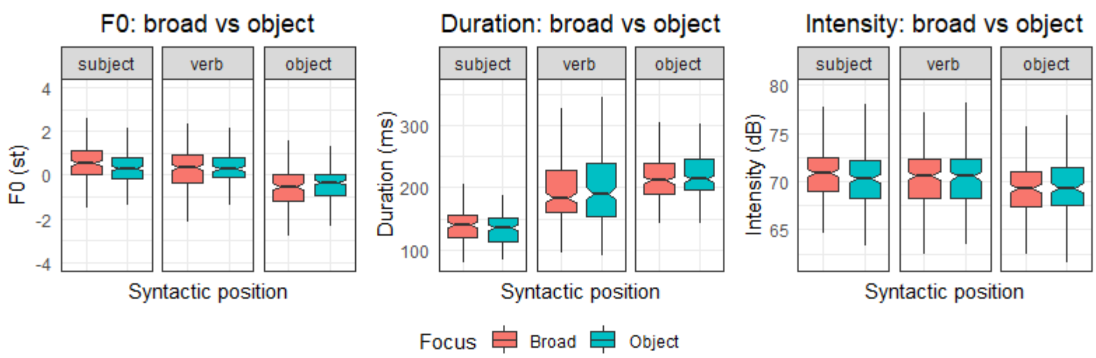

Figure 3-15: Mean values of Mandarin broad focus vs object focus by immigrants.

Mean F0: There was a main effect of focus on mean F0 in the subject position ($\chi^2(1) = 8.995$, $p = .003$), and the mean F0 was lowered by 0.275 ± 0.091 st. No effect of focus was found in the verb ($\chi^2(1) = .013$, $p = .910$) or object ($\chi^2(1) = .718$, $p = .367$) positions.

Duration: In the subject position, the main effect of focus on duration was significant ($\chi^2(1) = 11.912$, $p < .001$), which shortened each syllable by 6.253 ± 1.798 ms. The effect of focus on duration was also found in the object position ($\chi^2(1) = 6.919$, $p = .009$), lengthening one syllable by 15.171 ± 1.147 ms. No effect of focus was found in the verb position ($\chi^2(1) = 1.774$, $p = .183$).

Intensity: There was a main effect of focus on intensity in the subject position ($\chi^2(1) = 13.398$, $p < .001$), decreasing the intensity by 0.553 ± 0.150 dB. There was no main effect of focus on intensity in the verb ($\chi^2(1) = .257$, $p = .612$) or object ($\chi^2(1) = .093$, $p = .760$) positions.

3.3.2.4 Broad focus vs VP focus

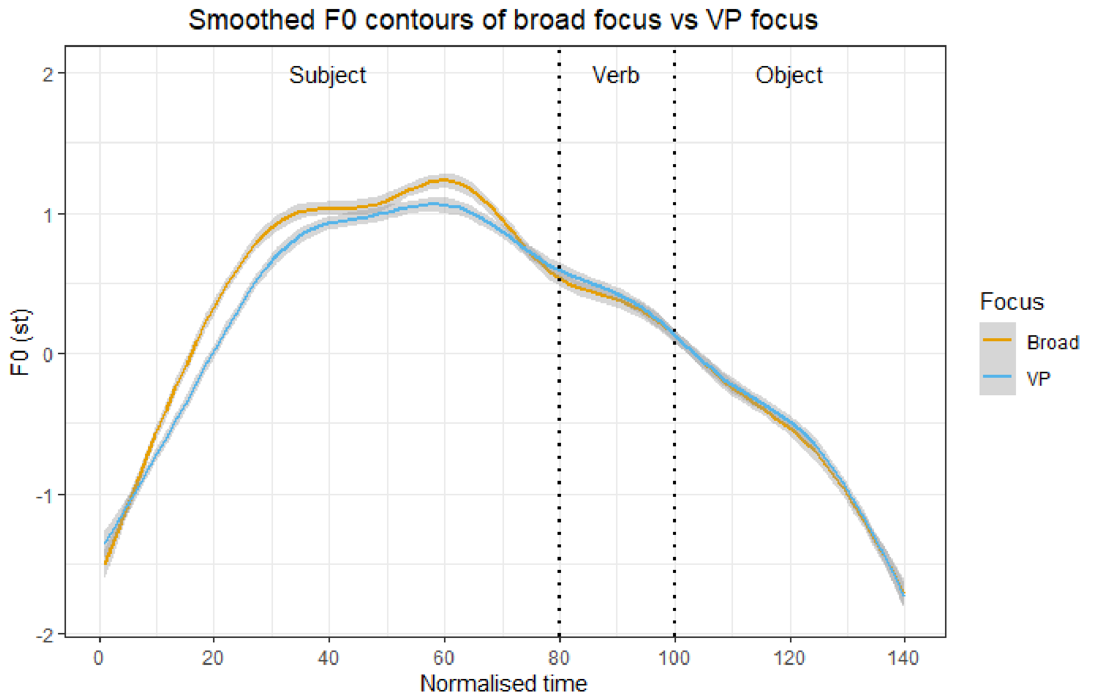

Figure 3-16: F0 contours of Cantonese broad focus vs VP focus by immigrants.

F0 contour: In the subject position, the contour of VP focus was slightly lower than that of broad focus, suggesting a sign of pre-focus F0 lowering. The two contours overlapped in the verb and object positions.

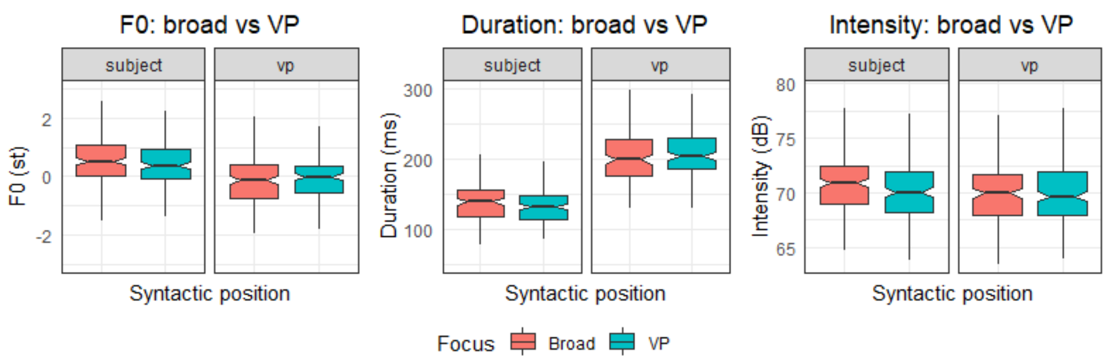

Figure 3-17: Mean values of Cantonese broad focus vs VP focus by immigrants.

Mean F0: No effect of focus was found in the subject ($\chi^2(1) = 3.428$, $p = .064$) or VP

($\chi^2(1) = .139$, $p = .710$) positions.

Duration: The main effect of focus on duration was significant in the subject ($\chi^2(1) = 19.276$, $p < .001$) and VP ($\chi^2(1) = 6.627$, $p = .010$) positions. Focus shortened each syllable by 7.770 ± 1.746 ms in the subject position and lengthened each syllable 6.168 ± 2.394 ms in the VP position.

Intensity: There was a main effect of focus on the intensity in the subject position ($\chi^2(1) = 20.542$, $p < .001$), which decreased the intensity by 0.685 ± 0.149 dB. No effect of focus was found in the VP position ($\chi^2(1) = .133$, $p = .716$).

3.3.2.5 Comparing Cantonese focus marking by native speakers and immigrants

In this section, we compared the production of Cantonese prosodic focus by native speakers and immigrants. A summary of the acoustic cues used by native speakers and immigrants has been presented in Table 3-3.

Focus pair	**Group**	**Subject position**	**Verb position**	**Object position**
Broad vs subject	Natives	F0 ↓, duration ↑	F0 ↓, duration ↓, intensity ↓	F0 ↓, duration ↓, intensity ↓
	Immigrants	F0 ↓, duration ↑, intensity ↓	F0 ↓, intensity ↓	F0 ↓, duration ↓, intensity ↓
Broad vs verb	Natives	N/A	duration ↑, intensity ↑	N/A

	Immigrants	F0 ↓, intensity ↓	F0 ↑, duration ↑, intensity ↑	F0 ↓, intensity ↓
Broad vs object	Natives	Duration ↓	N/A	Duration ↑
	Immigrants	F0 ↓, duration ↓, intensity ↓	N/A	Duration ↑
Broad vs VP	Natives	Duration ↓	N/A	
	Immigrants	Duration ↓, intensity ↓	Duration ↑	

Table 3-3: Acoustic cues that marked prosodic focus in Cantonese.

From the above table, it is clear that the immigrants employed more acoustic cues in focus marking, especially F0 and intensity, which the native speakers did not use often, at least in our data. Apart from that, the immigrants were able to use duration to mark focus most of the time, except for marking VP focus, suggesting that the immigrants' Cantonese focus marking was native-like to some extent.

Next, we presented the results of the differences between broad focus with each of the other focus types (subject, verb and object).

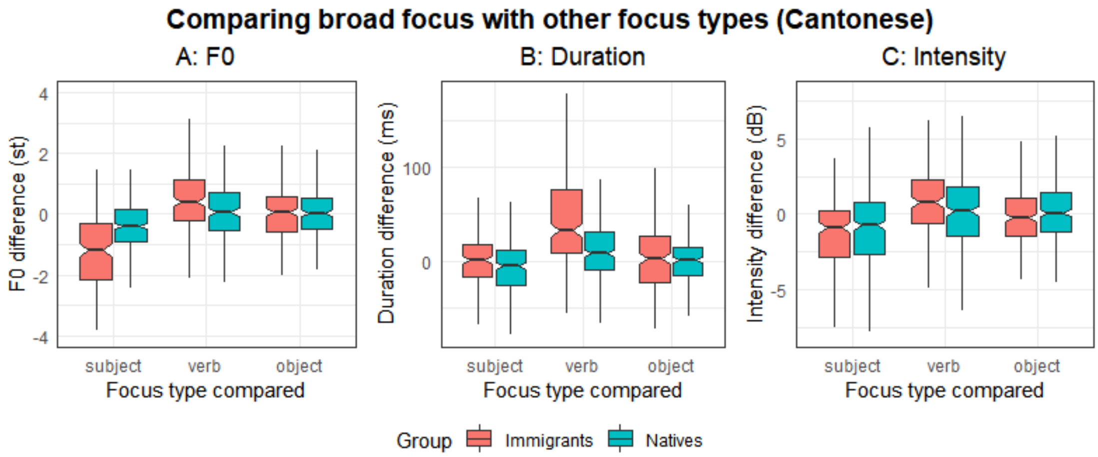

Figure 3-18: Differences between broad and other focus types in Cantonese.

Mean F0 difference: Figure 3-18-a shows the differences in mean F0 between broad and other focus types by native speakers and immigrants. There were main effects of focus ($\chi^2(1) = 184.009$, $p < .001$) and group ($\chi^2(1) = 7.765$, $p = .005$) on mean F0 difference. The interaction of these two variables also reached significance ($\chi^2(3) = 99.819$, $p < .001$). Post-hoc analyses suggest the following: 1) native speakers of Cantonese did not make use of mean F0 to mark on-focus and pre-focus regions, as the difference values were close to zero; 2) there seemed to be post-focus decrease of F0 in Cantonese speakers, but it was not the case if we check the contours in Figure 2-9, which showed parallel contours of broad focus and subject focus; 3) the immigrants showed clear on-focus and post-focus change of mean F0, although they did not have pre-focus change of mean F0.

Duration difference: Figure 3-18-b presents the differences in duration between broad and other focus types by native speakers and immigrants. There were main effects of focus ($\chi^2(1) = 120.062$, $p < .001$) and group ($\chi^2(1) = 19.781$, $p < .001$) on duration difference. There was also an interaction of focus and group on duration difference ($\chi^2(3) = 35.318$, $p < .001$). Specifically, the native speakers marked both the on-focus and post-focus regions with duration difference, while the immigrants marked the on-focus region only. Both groups of speakers did not mark the pre-focus region. Also, although native speakers lengthened the duration under focus, the extent of the lengthening was not as large as that of the immigrants.

Intensity difference: Figure 3-18-c plots the differences in intensity between broad and other focus types by native speakers and immigrants. There was a main effect of focus on intensity difference ($\chi^2(1) = 77.108$, $p < .001$). Although no effect of speaker group was found ($\chi^2(1) = .074$, $p = .785$), there was an interaction of focus and group on intensity difference ($\chi^2(3) = 12.207$, $p = .007$). Both speaker groups marked the on-focus and post-focus regions but did not mark the pre-focus region with intensity. The effect of focus was more robust for the immigrants than the native speakers.

3.3.2.6 Linear discriminant analysis

We also conducted a linear discriminant analysis to each focus pair to test the prediction accuracy of the acoustic cues in the immigrants' Cantonese data. The results are presented in Table 3-4.

Focus pair	Accuracy of prediction

	Mean F0	Duration	Intensity
Broad vs subject	75.0%	61.9%	67.9%
Broad vs verb	75.9%	68.5%	69.6%
Broad vs object	59.3%	57.1%	58.0%
Broad vs VP	56.3%	57.7%	58.5%

Table 3-4: Predicted accuracy from the LDA (immigrants' Cantonese).

For the immigrants, the LDA results showed higher prediction accuracies than native speakers of Cantonese, especially the contribution from mean F0 for distinguishing subject focus and verb focus from broad focus. However, the accuracy rates for distinguishing object focus and VP focus from broad focus were slightly above chance level only.

3.3.2.7 An interim summary

This section presented data from Cantonese focus production by immigrants and conducted pairwise comparisons of broad focus with other focus types. After that, we also compared the production data by native speakers and immigrants. We have shown that, the immigrants made more use of F0 and intensity in focus marking than the native speakers did, and they also used duration to mark focus in a near-native way.

3.3.3 Comparing Mandarin and Cantonese focus marking by immigrants

3.3.3.1 F0 contours

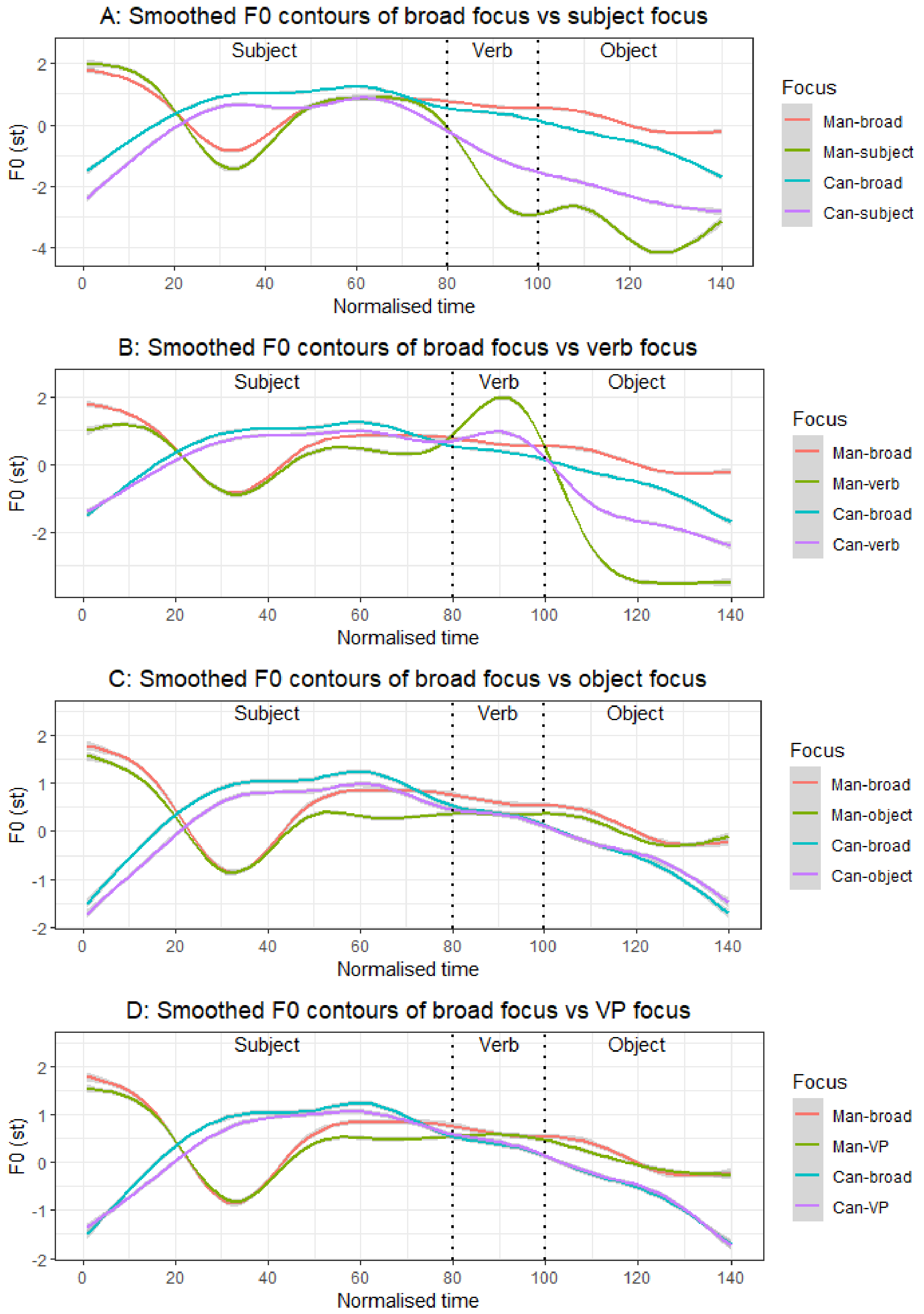

Figure 3-19: F0 contours of focus marking by immigrants.

Man= Mandarin speakers; Can = Cantonese speakers.

Figure 3-19 shows the smoothed F0 contours of Mandarin and Cantonese focus production by the immigrants. The red and blue lines are the contours for broad focus for immigrants' Mandarin and Cantonese, which have very different patterns. The Mandarin contour is generally a flat one while the Cantonese contour is declining gradually, the pattern of the latter being similar to what has been observed for native Cantonese speakers' production.

The green and purple lines from Figure 3-19 represent F0 contours of different focus conditions in the Mandarin and Cantonese of the immigrants. Although produced by the same group of speakers, the Mandarin sentences showed more focus-induced F0 changes than the Cantonese sentences.

3.3.3.2 Differences in the acoustic cues

In this subsection, we presented the results of the differences between broad focus with each of the other focus types (subject, verb and object) for each language of the immigrants.

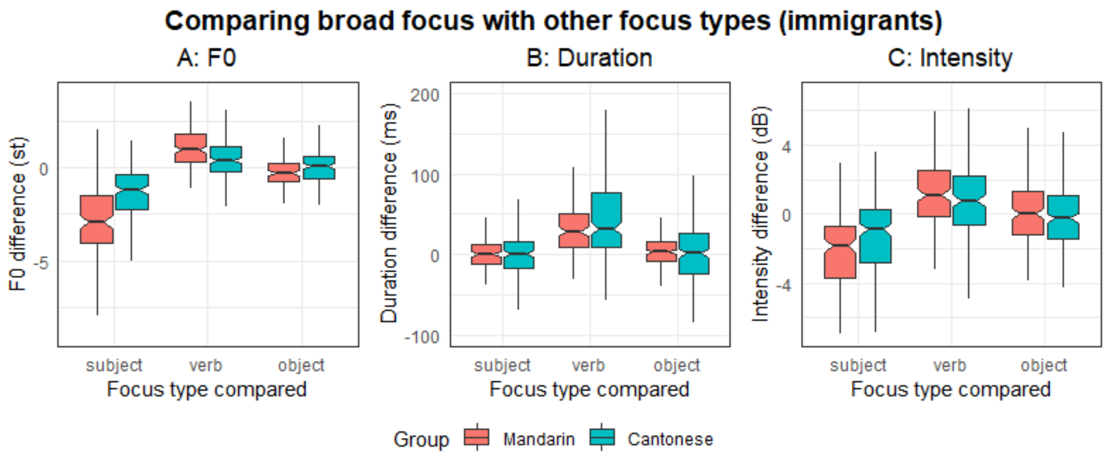

Figure 3-20: Differences between broad and other focus types (immigrants).

Mean F0 difference: Figure 3-20-a presents the differences in mean F0 between broad and other focus types in the Mandarin and Cantonese of the immigrants. There were main effects of focus ($\chi^2(1) = 464.455$, $p < .001$) and language ($\chi^2(1) = 5.922$, $p = .015$) on mean F0 difference. The interaction of focus and language also reached significance ($\chi^2(3) = 84.761$, $p < .001$). Post-hoc tests showed that the immigrants marked focus with F0 in both languages, but the extent to which they used F0 was larger in their L1 Mandarin than their L2 Cantonese.

Duration difference: Figure 3-20-b shows the differences in duration between broad and other focus types in the Mandarin and Cantonese of the immigrants. According to the linear mixed-effects models, there was an effect of focus ($\chi^2(1) = 180.433$, $p < .001$) and an interaction of focus and language ($\chi^2(3) = 9.668$, $p = .022$) on duration difference. Post-hoc tests suggested the following: 1) the effect of on-focus lengthening was found in both languages but was more robust in Cantonese; 2) the immigrants did not mark pre- or post-focus duration.

Intensity difference: Figure 3-20-c shows the differences in intensity between broad and other focus types in the Mandarin and Cantonese of the immigrants. There was a main effect of focus ($\chi^2(1) = 200.090$, $p < .001$) on intensity difference. The interaction of focus and language was also found ($\chi^2(3) = 13.069$, $p = .004$). Specifically, there was on-focus increase of intensity and post-focus decrease of intensity in both languages, but the immigrants' Mandarin showed a higher degree of increase and decrease. No pre-focus decrease of intensity was found for the immigrants' Mandarin or Cantonese.

3.4 Discussion

3.4.1 Summary of findings

This chapter examined how immigrants mark prosodic focus in Mandarin and Cantonese. The Mandarin and Cantonese data produced by the immigrants were presented separately and then compared with the data produced by native speakers, which showed some divergence from native speakers in both their L1 Mandarin and L2 Cantonese. The immigrants' use of acoustic cues in Mandarin and Cantonese was also compared, and a summary of results is presented in Table 3-5.

Focus pair	Language	Subject position	Verb position	Object position
Broad vs subject	Mandarin	Duration ↑, intensity ↓	F0 ↓, intensity ↓	F0 ↓, duration ↓, intensity ↓

	Cantonese	F0 ↓, duration ↑, intensity ↓	F0 ↓, intensity ↓	F0 ↓, duration ↓, intensity ↓
Broad vs verb	Mandarin	F0 ↓	F0 ↑, duration ↑, intensity ↑	F0 ↓, intensity ↓
	Cantonese	F0 ↓, intensity ↓	F0 ↑, duration ↑, intensity ↑	F0 ↓, intensity ↓
Broad vs object	Mandarin	F0 ↓, duration ↓, intensity ↓	F0 ↓	N/A
	Cantonese	F0 ↓, duration ↓, intensity ↓	N/A	Duration ↑
Broad vs VP	Mandarin	F0 ↓, duration ↓, intensity ↓	N/A	
	Cantonese	Duration ↓, intensity ↓	Duration ↑	

Table 3-5: Acoustic cues for focus marking in Mandarin and Cantonese (immigrants).

The table reveals some differences in the immigrants' marking of focus between Mandarin and Cantonese. First, the immigrants used F0 more in Mandarin than in Cantonese (e.g., in the pre-focus marking of subject under VP focus). On the other hand, the immigrants used duration more often in Cantonese than in Mandarin, as is shown in the on-focus lengthening of object focus and VP focus. We also compared the immigrants' focus production with the native speakers of each language. In Mandarin, the immigrants used less F0 and intensity but more duration in focus marking compared to native speakers. In Cantonese, the immigrants made more use of F0 and intensity than the native speakers, and they also used duration to mark focus in a near-native way.

3.4.2 Bidirectional influences of L1 and L2 in focus marking

Our data provide evidence of bidirectional cross-linguistic influences of the immigrants' L1 Mandarin and L2 Cantonese. The L2 influence on L1 can be obviously seen in the marking of subject focus. For subject focus, the native speakers of Mandarin marked the subject position with higher F0 and longer duration, and the native speakers of Cantonese marked the subject position with lower F0 and longer duration. While it requires further examination as to why Cantonese speakers used lower F0 for subject focus, it is clear that Mandarin and Cantonese are very different in marking subject focus, namely, Mandarin shows an increase of F0 in the subject position and Cantonese exhibits a lowering of F0 in the subject position. When the immigrants were required to mark subject focus in Mandarin, they chose neither way; rather, they did not manipulate F0 and leave it unchanged, resulting in similar F0 values for subject focus and broad focus in the subject position, which shows an intermediate stage in the use of F0 (Ying Chen et al., 2014). This is an example of category assimilation in L1

Mandarin, as the immigrants' Mandarin has been moving towards Cantonese in the use of F0, although it did not resemble either Mandarin or Cantonese produced by the native speakers.

The influence of L1 Mandarin on L2 Cantonese is most evident in the use of F0, which is rarely used in Cantonese but is frequently used in Mandarin. For example, the immigrants marked Cantonese verb focus with pre-, on- and post-focus manipulations of F0, which is similar to Mandarin speakers but does not like Cantonese speakers, who do not mark verb focus with F0. This is another case of category assimilation because the immigrants' L2 Cantonese had undergone assimilation to their L1 Mandarin in the use of F0. Moreover, for VP focus, while both Mandarin and Cantonese native speakers did not show any on-focus changes, the immigrants marked the Cantonese VP focus with lengthened duration. In this case, the participants had shifted the duration values of their L2 towards neither the L1 or L2, namely, the duration values were different from both Mandarin and Cantonese, which provides an example of category dissimilation in L2 Cantonese.

In this section, we have shown evidence of interactions in the immigrants' L1 and L2 in manipulating various acoustic cues to mark focus. The immigrants failed to mark prosodic focus in the same way as native speakers in either their L1 Mandarin and L2 Cantonese. The ways the immigrants marked prosodic focus in their L1 and L2 became more similar to each other, although there were also rare cases of dissimilation.

3.4.3 Transferability of PFC

Unlike the native speakers of Cantonese, who only had some PFC of duration and intensity, the immigrants showed consistent PFC of both F0 and intensity for subject focus and verb focus in their Cantonese. These results indicate that, the feature of PFC can be transferred to an L2, even if this L2 does not have clear PFC, which is in line with studies that support the learnability of PFC between the bilinguals' two languages. Ying Chen (2015) showed that PFC in English was acquirable by Mandarin-speaking learners, although the use of Mandarin learners' PFC in English was not identical to that of the native English speakers'. Moreover, Ying Chen et al. (2014) investigated Southern Min-Mandarin speakers' focus production and found that, although the bilingual speakers did not show PFC in their L1 Southern Min, the younger group, who had more exposure to Mandarin, did have PFC in their L2 Mandarin and showed nativelike production. These two studies suggest that, as long as there is enough exposure to the target language, the participants are able to acquire PFC in L2 to some extent, regardless of whether the L1 has PFC or not.

There is also evidence against the transferability account of PFC, mainly from English-Cantonese bilinguals' focus production (Wu, 2013; Wu & Chung, 2011). In these studies, the bilingual speakers' production varied in English, where the participants either showed clearly PFC of F0 in all sentences, in some sentences, or showed no PFC of F0. In the participants' production of Cantonese, however, there was no sign of PFC at all. The authors then proposed that PFC may not be easily transferred.

Note that the participants in the latter studies are all simultaneous bilingual speakers of English and Cantonese, namely, they speak English and Cantonese from birth and can thus be regarded as native speakers of both languages. It is reasonable that they are able use Cantonese in a nativelike manner. The participants from the former two studies all learned their L1 first and are late learners of an L2. The fact that they could acquire PFC in the target language (at least to some extent) indicates the learnability of PFC for late learners. Another study on Korean-speaking learners of English compared the production of English focus by learners with various proficiency levels (J. Liu, Xu, & Lee, 2019). This cross-sectional investigation showed that the learners at lower levels were more deviant from native English speakers and the advanced learners were more native-like in both on-focus and post-focus manipulation of acoustic cues, indicating that PFC may not be transferrable from the native language. In our study, where the target language Cantonese does not use much PFC to mark focus, the immigrants extensively showed PFC of F0 and intensity, suggesting the transferability of PFC to an L2 and also the incompleteness in the immigrants' acquisition of Cantonese prosody. Because we lack data from immigrants with lower

Cantonese proficiency, it is impossible to further compare our results with those of J. Liu et al. (2019). A preliminary explanation to the observed divergence is that PFC may be a property that will be transferred to the L2 for advanced learners only, as PFC is at a higher level than segments in learners' mental representation.

To sum up, this chapter compared the immigrants' focus marking in Mandarin and Cantonese, and also compared the immigrants' marking of focus with native speakers of each language. It is shown that the immigrants used F0 more frequently in Mandarin and used duration more often in Cantonese. In Mandarin, the immigrants used less F0 and intensity but more duration in focus marking compared to native speakers. In Cantonese, the immigrants made more use of F0 and intensity than the native speakers, and they also used duration to mark focus in a near-native way. Taken together, these results indicate evidence of cross-linguistic influences from the immigrants' L1 Mandarin to L2 Cantonese as well as from their L2 Cantonese to L1 Mandarin in their marking of prosodic focus. We also discussed the issue of transferability of PFC in focus marking.

Chapter 4: Perception of Mandarin and Cantonese Focus by Native Speakers

4.1 Introduction

As has been shown in the previous chapters, prosodic focus is realised acoustically with various cues on the focused components and possibly also on the pre- and post-focus components, but the mapping between prosody and information structure may not be straightforward. In Mandarin, while verb focus is marked by on-focus increase of F0, duration and intensity, object focus is marked by pre-focus decrease of F0 only, without any on-focus change. In Cantonese, however, object focus is realised with marginal on-focus lengthening of duration. These indicate both within-language and cross-language variations of focus marking. Additionally, native speakers of the same language may differ in how they map prosody and information structure. For example, cross-speaker variations are found in how native speakers of British English use prosody to express different linguistic functions, although there are clear patterns when all the speakers are included as one group (Peppé, Maxim, & Wells, 2000). Research has also identified other factors that bring more variations to the actual realisation of speech, such as context (Wouters & Macon, 2002) and speaking rate (Allen, Miller, & DeSteno, 2003).

A question then arises as to whether and how native speakers of a language can correctly map acoustic cues and intended meaning in their perception when there exist variations in speech production. It is proposed that the variations of a given category are generally in a normal distribution, which allows listeners to make use of the probabilistic acoustic information to infer speakers' intention (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008). A recent study reveals the difficulty in mapping focus and meaning for English native speakers (Roettger, Mahrt, & Cole, 2019). Also, native speakers of English accept a narrow-focused sentence as an appropriate response to either a question of narrow focus (such as 'Where are you from?') or a question of broad focus (such as 'What did you say?') (Welby, 2003). More research is needed to understand the mapping of prosody and meaning, particularly within the scope of this study, the mapping of acoustic cues and focus.

Besides, the dichotomy of bottom-up and top-down processing mechanisms has been well studied in speech comprehension in general, but there still lacks research on this issue when acoustic cues and information structure are involved. Bottom-up processing is a data-driven process, which retrieves the sensory information (for speech processing, acoustic cues) from the external environment (Gibson, 1966), while top-down processing makes predictions from the available information (usually not complete) based on prior knowledge and experiences (Gregory, 1970). As will be detailed in the methodology section, two types of mapping (from acoustic cues to focus and from focus to acoustic cues) are designed, each corresponding to the bottom-up and top-down processing mechanisms.

In this chapter, we examine how native speakers of Mandarin and Cantonese map focus and meaning in their native language. Unlike Mandarin, which does not mark on-focus components for object focus and VP focus, Cantonese marks all the on-focus components by lengthening the duration. It is thus predicted that the differences in the accuracy may not be as large as those in Mandarin.

Below are the research questions this chapter aims to address:

1) Can native speakers of Mandarin and Cantonese correctly map prosodic form and information structure in question-answer dialogues?

2) Will the participants' performance differ when the direction of the mapping is reversed?

4.2 Methodology

4.2.1 Participants

Two groups of participants were invited to attend the experiment at the Speech and Language Sciences Laboratory of the Hong Kong Polytechnic University. The Mandarin group consisted of 30 native speakers of Mandarin (18 females, 12 males; aged: 27.33 ± 2.17), who were born and raised up in northern China and do not speak any Cantonese. The Cantonese group consisted of 31 native speakers of Cantonese (17 females, 14 males; aged: 21.29 ± 2.41), all born and brought up in Hong Kong, where Cantonese is the native and dominant language. No participants had any history of speaking, hearing or language difficulties.

4.2.2 Materials

Two native speakers of Mandarin were invited to record the stimuli in Mandarin. The male speaker always asked the questions and the female speaker always produced the answers. Consistent with the production experiment, T1 syllables were used as much as possible in this experiment, namely, all the content words are in Tone 1 and only the determiners and classifiers may be in other tones. There were six target sentences, all prepared with a specific picture to provide the necessary information to the speakers. The same focus conditions were manipulated (broad focus, subject focus, verb focus, object focus and VP focus) and were elicited with a precursor question and a picture on the computer screen. The male speaker asked the questions as shown on the screen, and the female speaker was instructed to answer each question with the information provided in the figures. The trials appeared randomly so that the speakers did not know which sentence or focus condition the next page would be. This method was adopted to make the female's utterances as natural as possible. Each trial occurred three times in the recording and only one clear token was selected as the stimuli for the perception experiment. Figure 4-1 shows the plotted F0 contours of the Mandarin female speaker, which generally resemble the results from our production experiment (with pre-focus, on-focus and post-focus manipulation of F0 in most of the cases, although not very strong for object focus and VP focus). Before the recording, the participants were first presented with the sentence list for them to get familiar with the sentences.

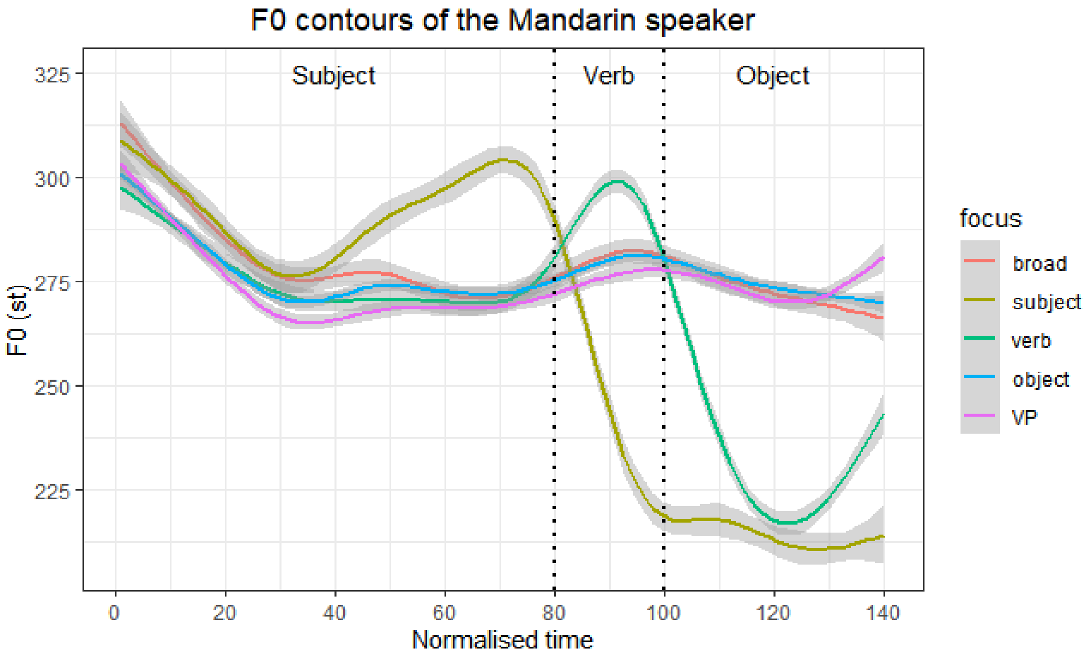

Figure 4-1: F0 contours of the Mandarin speaker.

Two native speakers of Cantonese were invited to record the stimuli in Cantonese. The male speaker always asked the questions and the female speaker always produced the answers. The same stimulus design and focus condition were manipulated. Again, each trial was recorded three times by the native speakers and only one clear token was selected as the stimuli for the perception experiment. Figure 4-2 shows the averaged syllable duration of the Cantonese female speaker, which generally resemble the results from our production experiment.

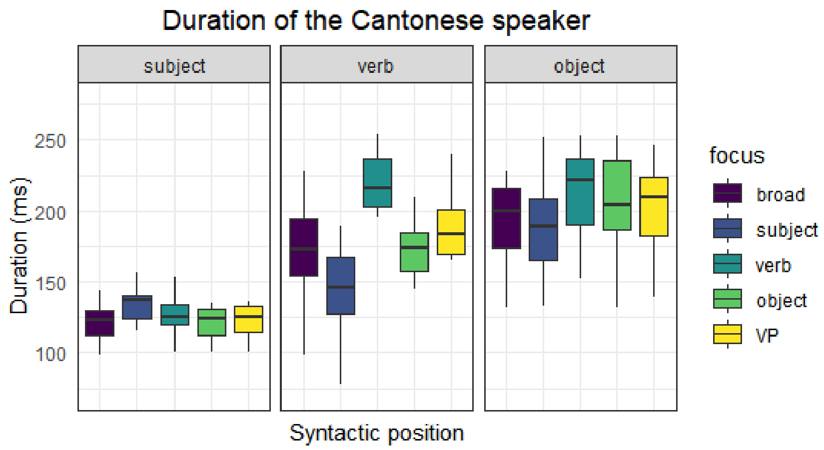

Figure 4-2: Duration of the Cantonese speaker.

The selected question and answer pairs were first extracted from the recording, and the intensity of the extracted questions and answers was normalised to 75 dB separately. The interval between the question and answer within each pair was fixed to 500 ms. No further adjustment was made to the recordings.

Two tasks were designed in this experiment. In each task, there were two dialogues played to the participants in one trial, with an interval of 1,000 ms between the two dialogues. Task 1 requires the mapping from context to prosodic form, in which the participants heard two dialogues with the same question and different focus conditions. One dialogue matches the focus context (e.g., (9a)) while the other does not (e.g., (9b)). Task 2 requires the mapping from prosodic form to context. The two dialogues in one trial share the same answer but have different questions, such as in (9b) and (9c). In each trial, broad focus appeared as either the target focus condition or the competitor focus condition (to the remaining four conditions). Exactly the same conditions were prepared for the tasks in Cantonese. If the participants rely more on the acoustic details in the processing (bottom-up), they will show better performance in Task 1. If they are more dependent on the higher level linguistic knowledge (top-down), they will have better performance in Task 2.

(9) a. Congruous context

Q: *ni3 shuo1 shen2me0?* (Broad focus)

What did you say?

A: *na4 wei4 yi1sheng1 he1 ka1fei1.* (Broad focus)

The doctor drank coffee.

b. Incongruous context

Q: *ni3 shuo1 shen2me0?* (Broad focus)

What did you say?

A: ***na4 wei4 yi1sheng1*** *he1 ka1fei1.* (Subject focus)

The doctor drank coffee.

c. Congruous context

Q: *shei2 he1 ka1fei1?* (Subject focus)

Who drinks coffee?

A: ***na4 wei4 yi1sheng1*** *he1 ka1fei1.* (Subject focus)

The doctor drank coffee.

In total, there were 192 target trials for each language (6 sentences*4 focus conditions*2 competitor conditions*2 tasks*2 orders). To eliminate the order effect, the congruent dialogue appears once in the first dialogue and once in the second dialogue (hence two orders).

4.2.3 Procedures

Before the experiment, the participants were first briefed with the procedures and then provided with practice trials to get themselves familiar with the tasks. The sentences in the practice trials were different from those in the actual experiments. Each task was divided into two blocks and the four blocks were randomly assigned to the participants. The trials within each block were also randomly presented, and the participants were required to make their choice (which dialogue sounds more natural) within five seconds after the second dialogue ended. They were allowed to take a break between the blocks. During the experiment, the participants heard two dialogues within each trial, a congruous one as in (9a) or (9c) and an incongruous one as in (9b). Either the question or the answer was exactly the same for the two dialogues. The participants were then required to decide which dialogue sounded more natural to them by pressing '1' or '2', each representing the former dialogue or the latter one.

4.2.4 Data analysis

The accuracy of each response was labelled as either correct or incorrect and then submitted to mixed-effects logistic regression models (Agresti, 2013). We chose logistic modelling because our outcome variable was dichotomous, and the results might be misleading if we treated the response accuracy as a continuous variable. In the regression models, response accuracy was the dependent variable, focus condition and task were the predictors, and subject was included as the random effect. Models were fitted with the 'lme4' package in R. The plots were generated based on the R code from Roettger et al. (2019).

4.3 Results

4.3.1 Perception of Mandarin focus by native speakers

We first fitted logistic regression models with all the response data from native speakers of Mandarin. There was a main effect of focus condition ($\chi^2(7) = 975.470$, $p < .001$), task ($\chi^2(1) = 5.600$, $p = .018$) and an interaction of focus condition and task ($\chi^2(7) = 43.263$, $p < .001$). Post-hoc analysis showed higher predicted accuracy in Task 1 than in Task 2. Next, we fitted logistic regression models for each task for a closer examination of the data.

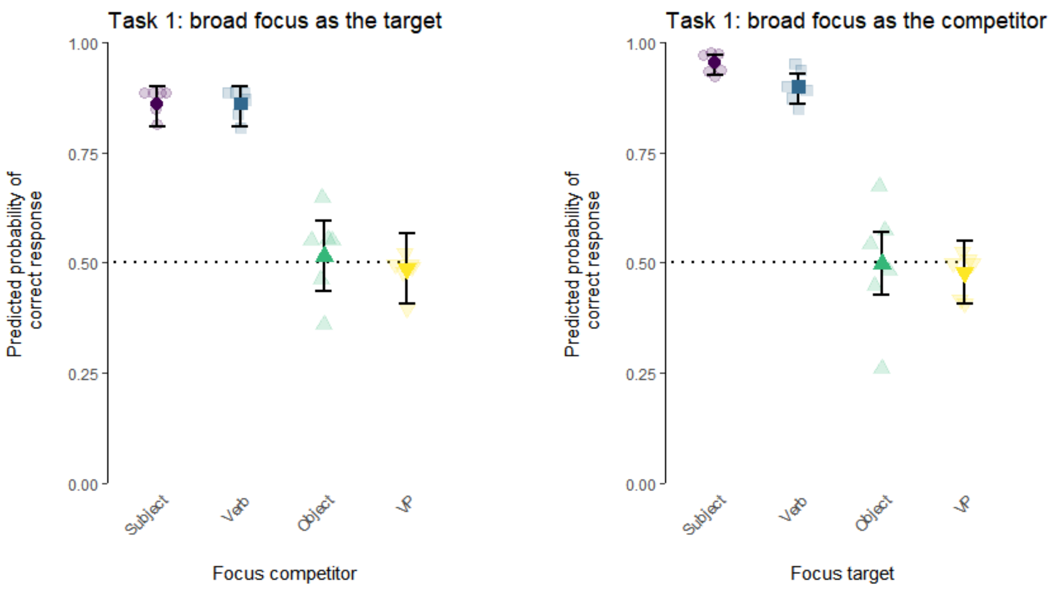

Figure 4-3: Results for Task 1 from native speakers of Mandarin.

Notes: The bar stands for means and 95% confidence intervals. The transparent points represent the mean values of each sentence.

Target	Competitor	Predicted accuracy	SE	95% CI
Broad	Subject	85.99%	19.45%	(80.74%, 89.99%)
Broad	Verb	85.99%	19.45%	(80.74%, 89.99%)
Broad	Object	51.51%	16.39%	(43.51%, 59.42%)
Broad	VP	48.78%	16.38%	(40.86%, 56.76%)
Subject	Broad	95.31%	25.70%	(92.48%, 97.12%)
Verb	Broad	89.91%	19.63%	(85.84%, 92.90%)
Object	Broad	49.94%	14.54%	(42.87%, 57.02%)
VP	Broad	47.87%	14.54%	(40.84%, 54.97%)

Table 4-1: Predicted accuracy for Task 1 (native speakers of Mandarin).

For Task 1, a main effect of focus condition was significant ($\chi^2(7) = 559.040$, $p < .001$). The predicted probability of correct response for different focus conditions was obtained from the final model and has been plotted in Figure 4-3 and Table 4-1. The figure on the left shows the cases where broad focus is the target and other focus types are the competitors. The participants' predicted accuracy is very high when subject focus and verb focus co-occur with broad focus (85.99% for both focus types), suggesting that they could clearly distinguish broad focus from these two focus types. When object focus and VP focus are the competitors, the participants' predicted performance is at chance level (51.51% and 48.78%), which indicates that they fail to notice the differences between broad focus and object focus and between broad focus and VP focus. For the figure on the right side, broad focus is always the competitor and the other focus types are the targets. When subject focus is the target, the participants are predicted to be accurate at 95.31% of the time. The predicted accuracy is also high when verb focus is the target (89.91%). Again, the participants have difficulty when object focus or VP focus are the targets, with predicted accuracy of 49.94% and 47.87%.

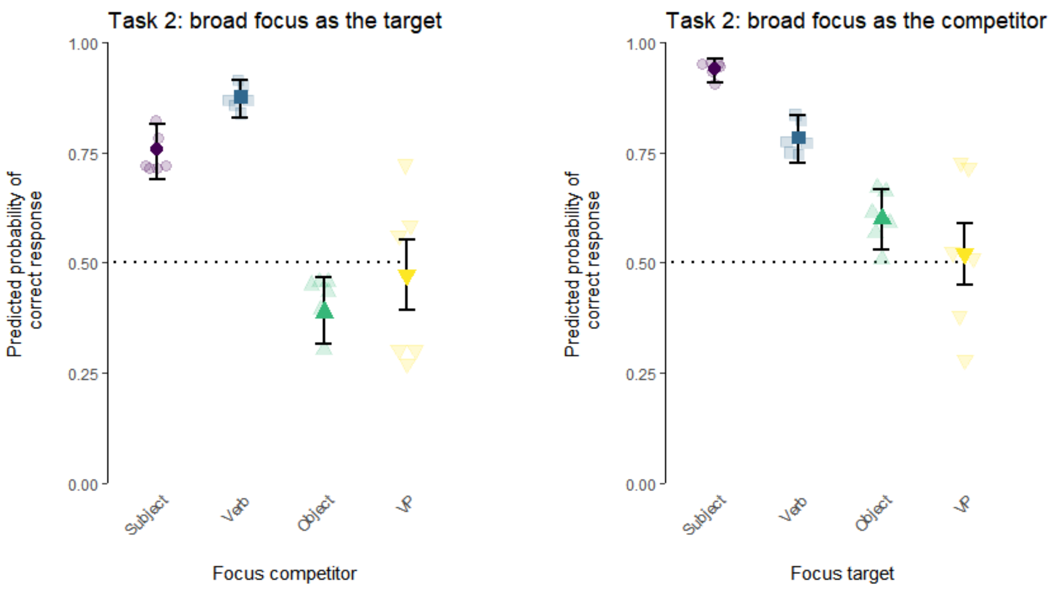

Figure 4-4: Results for Task 2 from native speakers of Mandarin.

Target	Competitor	Predicted accuracy	SE	95% CI
Broad	Subject	75.69%	17.54%	(68.82%, 81.45%)
Broad	Verb	87.63%	20.01%	(82.72%, 91.29%)
Broad	Object	38.90%	16.50%	(31.55%, 46.80%)
Broad	VP	47.27%	16.38%	(39.41%, 55.27%)
Subject	Broad	94.05%	23.48%	(90.89%, 96.16%)
Verb	Broad	78.49%	16.13%	(72.67%, 83.34%)
Object	Broad	60.01%	14.69%	(52.94%, 66.67%)
VP	Broad	52.02%	14.54%	(44.91%, 59.04%)

Table 4-2: Predicted accuracy for Task 2 (native speakers of Mandarin).

For Task 2, there was also a main effect of focus condition ($\chi^2(7) = 433.330$, $p < .001$). The predicted probability of correct response for different focus conditions was obtained from the final model and has been plotted in Figure 4-4 and Table 4-2. In general, the predicted accuracy of the participants is lower in Task 2 compared with Task 1. When target is broad focus, the predicted accuracy is the highest for verb focus as the competitor (87.63%), followed by subject focus (75.69%). The predicted accuracy for object focus and VP focus is even below chance level (38.90% and 47.27%), suggesting that participants tend not to choose broad focus in the experiment. When broad focus is the competitor, the performance is estimated to be better. Subject focus (94.05) has higher accuracy than verb focus (78.49%). The predicted accuracy for object focus is 60.01% and that for VP focus is 52.02%.

The data from native speakers of Mandarin reveal that, the predicted accuracy of response varies according to focus types and tasks. The predicted accuracy is higher when: 1) subject focus and verb focus are paired with broad focus than object focus and VP focus are paired with broad focus; 2) the accuracy is obtained from Task 1 than from Task 2; and 3) broad focus is the competitor rather than the target.

4.3.2 Perception of Cantonese focus by native speakers

We fitted logistic regression models with all the response data from native speakers of Cantonese. There was a main effect of focus condition ($\chi^2(7) = 528.910$, $p < .001$), task ($\chi^2(1) = 5.179$, $p = .023$) and an interaction of focus condition and task ($\chi^2(7) = 47.552$, $p < .001$). Post-hoc analysis showed higher predicted accuracy in Task 1 than in Task

2. Next, we fitted logistic regression models for each task for a closer examination of the data.

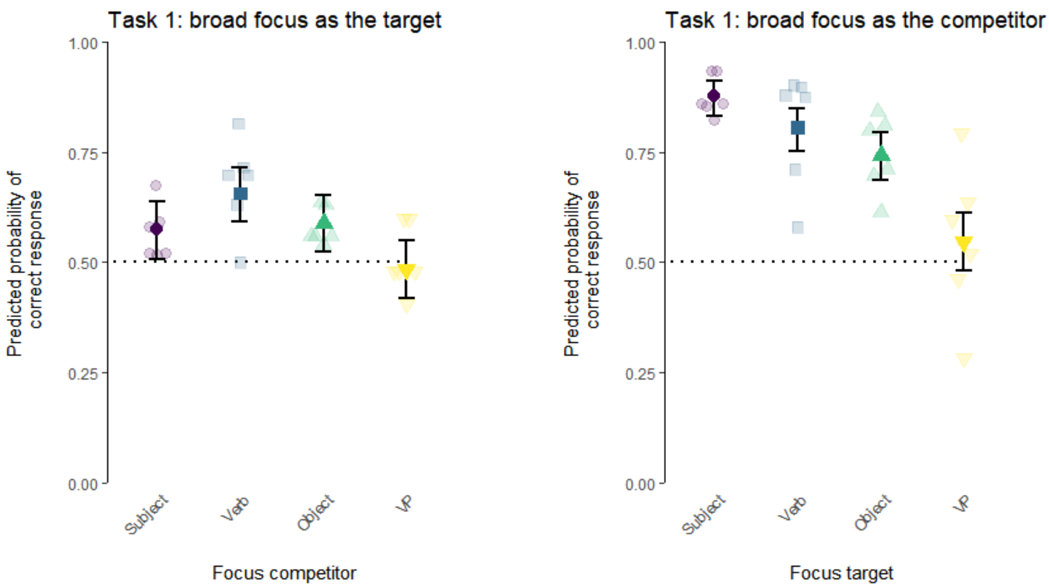

Figure 4-5: Results for Task 1 from native speakers of Cantonese.

Target	Competitor	Predicted accuracy	SE	95% CI
Broad	Subject	57.38%	13.73%	(50.70%, 63.79%)
Broad	Verb	65.65%	14.09%	(59.19%, 71.58%)
Broad	Object	58.97%	13.77%	(52.32%, 65.31%)
Broad	VP	48.41%	13.64%	(41.80%, 55.08%)
Subject	Broad	87.59%	18.11%	(83.19%, 90.96%)
Verb	Broad	80.56%	15.70%	(75.28%, 84.93%)
Object	Broad	74.35%	14.60%	(68.53%, 79.42%)
VP	Broad	54.74%	13.33%	(48.22%, 61.09%)

Table 4-3: Predicted accuracy for Task 1 (native speakers of Cantonese).

For Task 1, a main effect of focus condition was found ($\chi^2(7) = 230.560$, $p < .001$). The predicted probability of correct response for different focus conditions was obtained from the final model and has been plotted in Figure 4-5 and Table 4-3. When broad focus is the target, the predicted accuracy of response ranges from 48.41% (VP focus as the competitor) to 65.65% (verb focus as the competitor), and the accuracy is very similar for subject focus (57.38%) and object focus (58.97%) as the competitors. When broad focus is the competitor, the participants are able to clearly distinguish subject focus (87.59%), verb focus (80.56%) and object focus (74.35%) from broad focus, but they cannot notice the differences between broad focus and VP focus (54.74%).

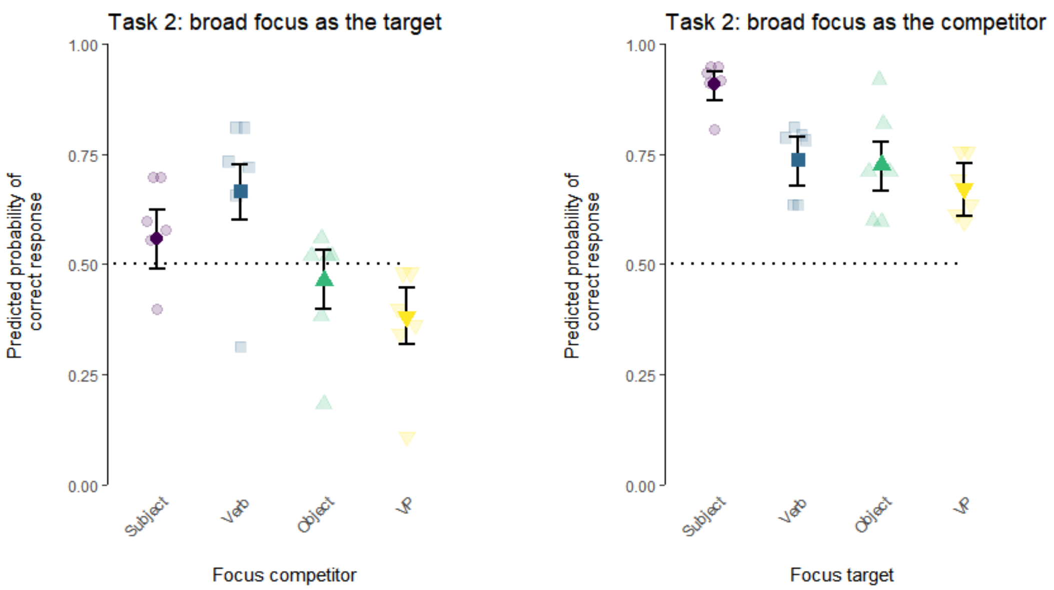

Figure 4-6: Results for Task 2 from native speakers of Cantonese.

Target	Competitor	Predicted accuracy	SE	95% CI
Broad	Subject	55.78%	13.69%	(49.10%, 62.26%)
Broad	Verb	66.60%	14.15%	(60.18%, 72.46%)
Broad	Object	46.49%	13.65%	(39.93%, 53.17%)
Broad	VP	38.18%	13.89%	(31.99%, 44.78%)
Subject	Broad	90.90%	20.28%	(87.03%, 93.69%)
Verb	Broad	73.72%	14.52%	(67.85%, 78.85%)
Object	Broad	72.47%	14.37%	(66.52%, 77.72%)
VP	Broad	67.12%	13.87%	(60.87%, 72.82%)

Table 4-4: Predicted accuracy for Task 2 (native speakers of Cantonese).

For Task 2, a main effect of focus condition also reached significance ($\chi^2(7) = 356.450$, $p < .001$). The predicted probability of correct response for different focus conditions was obtained from the final model and has been plotted in Figure 4-6 and Table 4-4. When broad focus is the target, the participants could only distinguish it from verb focus, but the accuracy of response is not very high (66.60%). When other focus types are the competitors, the predicted accuracy of response is very low (55.78% for subject focus, 46.49% for object focus, and 38.18% for VP focus), revealing the incapability of the participants in telling the difference between these pairs. When broad focus is the competitor, the participants' accuracy is predicted to be higher. More specifically, subject focus has the highest predicted response accuracy of 90.90%, and the remaining three focus types are also distinguishable from broad focus, with similar predicted response accuracy rates (73.72% for verb focus, 72.47% for object focus, 67.12% for VP focus).

The data from native speakers of Cantonese reveal that, the predicted accuracy of response varies according to focus types and tasks. The predicted accuracy is higher when: 1) subject focus and verb focus are paired with broad focus than object focus and VP focus are paired with broad focus; 2) the accuracy is obtained from Task 1 than from Task 2; and 3) broad focus is the competitor rather than the target.

4.4 Discussion

4.4.1 Summary of findings

The results above suggest that Mandarin-speaking listeners could correctly map prosody and information structure when subject focus or verb focus is paired with broad focus, wherein noticeable differences in the acoustic cues (particularly F0) are available. When object focus or VP focus is paired with broad focus, wherein such noticeable differences are absent and the focus types do not acoustically differentiate from each other (apart from slight differences in the pre-focus regions), the predicted accuracy is at the chance level. The native speakers of Cantonese could clearly map prosody and information structure for subject focus and verb focus, and they are able to distinguish object focus from broad focus for most of the time. The predicted accuracy for Cantonese speakers is not as high as that for Mandarin speakers, which is probably due to the fact that Cantonese does not make much use of prosody in focus marking as Mandarin does.

Also, the native speakers of Mandarin and Cantonese showed better performance in Task 1 (form to meaning) than in Task 2 (meaning to form). Recall that our Task 1 requires the mapping from form to meaning and is more data-driven, and Task 2 requires the mapping from meaning to form and relies more on prior knowledge. The results thus indicate that, although the participants employed both bottom-up and top-down processing mechanisms, they should have paid more attention to the acoustic cues than top-down expectations in this perception experiment. Our findings are generally in line with Roettger et al. (2019), which also shows that the mapping between acoustic form and focus condition is not one-to-one but varies across focus pairs and that there is also a bias against broad focus.

4.4.2 The relationship between speech production and perception

There have been debates over the relationship between speech production and perception. Regarded as different processes in traditional approaches, speech production and speech perception are divided into separate lines of inquiry: speech production concerns direct observations of the acoustic or articulatory aspects of speech, and speech perception examines how speech is processed and interpreted by human listeners (Casserly & Pisoni, 2010; Schmitz, Díaz, Fernández Rubio, & Sebastian-

Galles, 2018). More research, on the other hand, points to the idea that speech production and perception are somehow linked to each other, although the exact relationship between them is unclear (Lotto, Hickok, & Holt, 2009; Pulvermüller & Fadiga, 2010). The influential motor theory of speech perception, for instance, highlights the role of speech production on speech perception (Liberman & Mattingly, 1985). According to motor theory, speech perception is based on the listener's knowledge of the articulatory gestures, namely, a listener can produce the speech and is aware of the articulator movements and positions. In addition, studies on speech accommodation, or phonetic convergence, further support a link between speech production and perception (Babel, 2009; Pardo, 2006; Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013).

Our data on focus production and perception in Mandarin and Cantonese also offer some insights to the discussion of this issue. Our production results showed on-focus changes for subject focus and verb focus conditions but not (obviously) for object focus and VP focus in both languages, and the perception data mirrored the production data in that native speakers of Mandarin and Cantonese had higher accuracy rates for distinguishing subject focus and verb focus from broad focus than the other focus pairs. Thus, we propose that, there is some link between the production and perception of focus in Mandarin and Cantonese, but we do not yet know whether perception or production is more essential for prosodic focus, which should be tested in the future. Besides, there are differences in the use of the acoustic cues in focus marking in each language. Although we have employed LDA to show the weights of each acoustic cue, further studies are required to confirm the LDA results and to check whether the listeners are equally sensitive to the acoustic cues in speech processing. The role of F0 in focus perception will be discussed later in this section. Given the different weights of the acoustic cues in focus production, another issue worth examination is whether there would be any compensation mechanism involved in focus perception.

4.4.3 Evidence for focus projection theories

In the production experiments, the native participants did not clearly distinguish object focus or VP focus from broad focus in both Mandarin and Cantonese. In the perception experiments, the native listeners also showed poor performance when object focus and VP focus were paired with broad focus. The results are in line with the claim from theories of focus projection that object focus, VP focus and broad focus are ambiguous in both production and perception (Arregi, 2016; Bishop, 2017). It is proposed that if a language marks focus with prosody, then a sentence with a narrowly focused component can have more than one reading, namely, it can be interpreted as either a narrowly focused sentence or a broadly focused sentence. A dominant approach within focus projection theories is the Default Prosody approach, suggesting that prosodic prominence is determined by default principles of prosody within a focused constituent (Arregi, 2016). By default, the most prominent element of a constituent in English is the rightmost one (Chomsky & Halle, 1968), and consequently, when focus is on the object, on the VP, or on the full sentence, the prominence is always on the object position. According to this claim, English does not acoustically differentiate object focus, VP focus and broad focus, which resemble our production and perception results in Cantonese and Mandarin. Together with similar findings for Mandarin and Cantonese in previous studies (Wu, 2013; Xu et al., 2012), our data provide further evidence for focus projection theories. Yet it still remains to be explored as to why this is case.

4.4.4 The role of F0 in focus perception

Chapter 2 suggests the role of F0 in focus marking in Mandarin, and the results from the perception tasks further reveal the role of F0 in focus perception in Mandarin and in general. First, the Mandarin modal speaker, just like the 21 Mandarin speakers in our production experiment, marks subject focus and verb focus mostly with F0 (and also duration and intensity, but to a lesser extent) but does not clearly mark object focus and VP focus. The Mandarin listeners' high accuracy rates in distinguishing subject focus and verb focus from broad focus and low accuracy rates in distinguishing object focus and VP focus from broad focus indicate that they also largely rely on F0 to make judgements about which element is more prominent within an utterance.

Besides, the Mandarin listeners generally outperform the Cantonese listeners in the perception tasks, which is probably due to the fact that Cantonese speakers tend not to employ F0 as much as Mandarin speakers, but they tend to use duration to mark focus. A number of studies have identified F0 as the primary acoustic cue (higher than duration in the hierarchy) to mark prominence (e.g., German: Andreeva, Barry, & Steiner, 2007; Mandarin: Yang & Chen, 2020a), and it is recently shown that F0 plays a more important role than duration in German prominence perception (Niebuhr & Winkler, 2017). Specifically, Niebuhr and Winkler (2017) manipulated the values of F0 and duration and found out that an increase of less than 1 st in F0 could outweigh a lengthening of 30% in duration. Although the structure Niebuhr and Winkler worked on is syllables, their results can be interpreted together with our findings, postulating the argument that F0 is the most noticeable acoustic cue for prominence perception cross-linguistically.

To conclude, this chapter investigated the perception of focus in Mandarin and Cantonese and showed that the mapping between acoustic form and focus condition varied across focus pairs. The Mandarin and Cantonese listeners showed high accuracy rate when subject focus and verb focus were paired with broad focus, but the accuracy rate was very low when object focus and VP focus were paired with broad focus. Given the observed acoustic overlaps between object focus and broad focus and between VP focus and broad focus, we proposed that speech production and perception are related. We also provided evidence for theories of focus projection from our perception data and identified the role of F0 in focus perception.

Chapter 5: Perception of Mandarin and Cantonese Focus by Immigrants

5.1 Introduction

Although infants are born sensitive to all categories in speech sounds (being able to discriminate non-native speech contrasts), such ability declines gradually within the first year of life (Werker & Tees, 1984); consequently, knowledge of the native languages is usually reported to have hindered the perception of L2 speech sounds (e.g., Guion, Flege, Akahane-Yamada, & Pruitt (2000)).

Investigations on perception of L2 intonation are very limited. Baker (2010) tests the extent to which L1 and L2 (Chinese and Korean) speakers of English are able to determine whether an English sentence has context-appropriate prosody and results suggest the role of L1 on L2 perception. More specifically, the Korean speakers performed better than the Mandarin speakers in different perception tasks, which was attributed to L1 transfer by the author.

This chapter examines whether the immigrants are able to map acoustic cues and focus in their L1 Mandarin and L2 Cantonese. Also, we will compare the native data with the immigrants' data to test whether such ability can be maintained in the immigrants' L1 Mandarin and be acquired in their L2 Cantonese.

Below are the research questions this chapter aims to address:

1) Can the immigrants correctly map prosody and focus in their L1 Mandarin and L2 Cantonese?

2) Are there differences in the mapping between the immigrants and native speakers in each language?

5.2 Methodology

5.2.1 Participants

We invited 33 immigrants (27 females, 6 males; aged: 30.67 ± 4.95) who speak Mandarin as the L1 and have been exposed to Cantonese since their arrival in Hong Kong. Again, although the participants originated from different provinces of Northern China and may speak other Northern Chinese dialects, they reported that, in terms of the frequency of language use and the proficiency level, their dominant language is Mandarin only but not other Northern dialects. The participants all arrived in Hong Kong after puberty (average age: 23.48 ± 4.95) and the average length of residence was 7.18 ± 3.24 years. The immigrants completed an adapted version of Bilingual Language Profile prior to the experiment, the scores of which showed that, the participants were fluent speakers of Cantonese, although they were more dominant in Mandarin at the time of attending the experiment. No participants had any history of speaking, hearing or language difficulties.

5.2.2 Experiment design and data analysis

As described in the previous chapter, the same stimuli were prepared for the immigrants. The immigrants completed the perception tasks in Cantonese before the tasks in Mandarin. The same procedures for data collection and analysis were performed for the data of immigrants' focus perception.

5.3 Results

5.3.1 Perception of Mandarin focus by immigrants

Logistic regression models with all the response data from immigrants showed a main effect of focus condition ($\chi^2(7) = 1224.9$, $p < .001$) and an interaction of focus condition and task ($\chi^2(8) = 94.097$, $p < .001$), but the effect of task did not reach significance ($\chi^2(1) = 1.756$, $p = .185$). Next, we fitted logistic regression models for each task for a closer examination of the data.

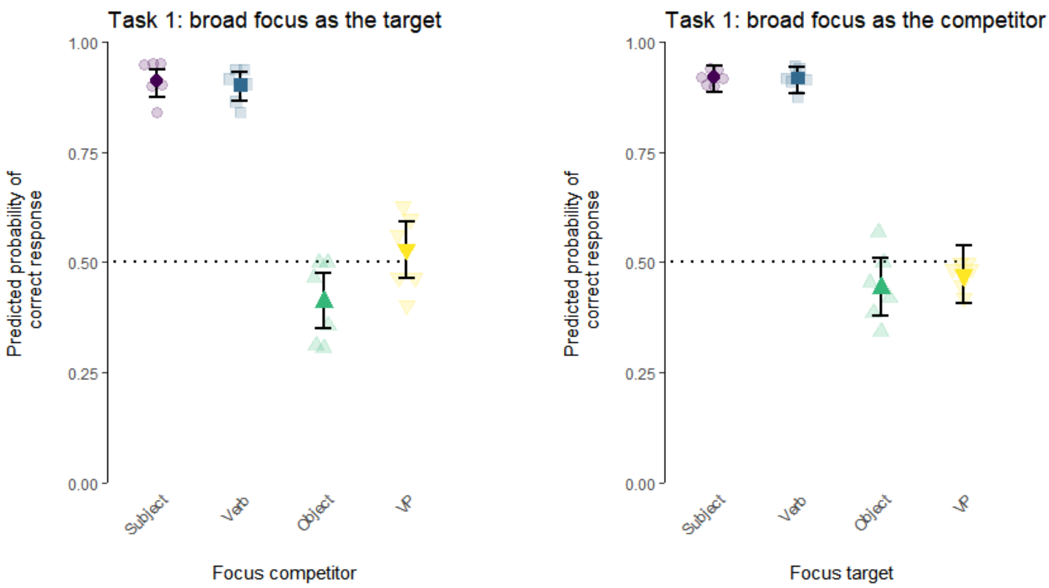

Figure 5-1: Results for Task 1 from immigrants (Mandarin).

Target	Competitor	Predicted accuracy	SE	95% CI
Broad	Subject	90.95%	19.06%	(87.36%, 93.59%)
Broad	Verb	90.23%	18.58%	(86.51%,93.00%)
Broad	Object	41.17%	13.40%	(34.99%, 47.64%)
Broad	VP	52.88%	13.30%	(46.37%, 59.29%)
Subject	Broad	92.00%	20.05%	(88.59%, 94.45%)
Verb	Broad	91.75%	19.84%	(88.28%, 94.25%)
Object	Broad	44.29%	13.61%	(37.84%, 50.93%)
VP	Broad	47.15%	13.57%	(40.61%, 53.79%)

Table 5-1: Predicted accuracy for Task 1 (immigrants' Mandarin).

For Task 1, a main effect of focus condition was significant ($\chi^2(7) = 791.700$, $p < .001$), suggesting huge variations of the predicted accuracy among different focus pairs. The predicted probability of correct response for different focus conditions was obtained from the final model and has been plotted in Figure 5-1 and Table 5-1. When broad focus is the target, the predicted response accuracy is equally high for subject focus (90.95%) and verb focus (90.23%), and chance level accuracy of response is predicted for object focus (41.17%) and VP focus (52.88%). When broad focus is the competitor, similar patterns are also predicted by the model: the participants have higher accuracy rates for subject focus (92.00%) and verb focus (91.75%) than object focus (44.29%) and VP focus (47.15%).

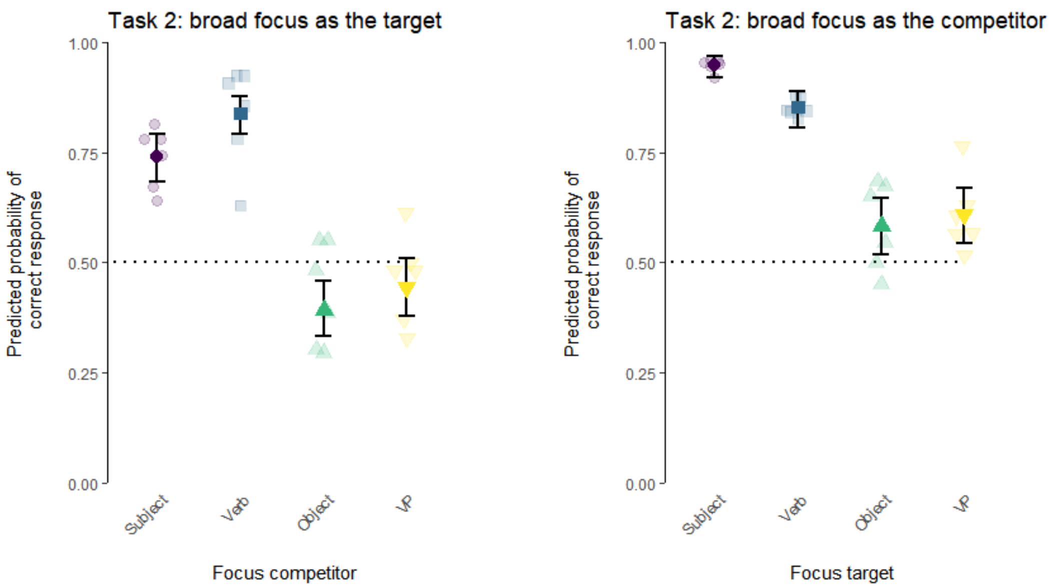

Figure 5-2: Results for Task 2 from immigrants (Mandarin).

Target	Competitor	Predicted accuracy	SE	95% CI
Broad	Subject	73.97%	14.31%	(68.22%, 79.00%)
Broad	Verb	83.86%	15.98%	(79.16%, 87.67%)

Broad	Object	39.31%	13.46%	(33.23%, 45.75%)
Broad	VP	44.35%	13.33%	(38.03%, 50.86%)
Subject	Broad	94.80%	23.45%	(92.00%, 96.65%)
Verb	Broad	85.22%	16.60%	(80.64%, 88.87%)
Object	Broad	58.40%	13.67%	(51.78%, 64.73%)
VP	Broad	60.82%	13.74%	(54.25%, 67.02%)

Table 5-2: Predicted accuracy for Task 2 (immigrants' Mandarin).

For Task 2, there was also a main effect of focus condition ($\chi^2(7) = 530.19$, $p < .001$). The predicted probability of correct response for different focus conditions was obtained from the final model and has been plotted in Figure 5-2 and Table 5-2. When broad focus is the target, the participants have higher predicted accuracy of response for verb focus (83.86%) than subject focus (73.97%) as the competitor. The predicted accuracy for object focus (39.31%) and VP focus (44.35%) is well below chance level. When broad focus is the competitor, the participants are estimated to perform much better in the task. The best performance is predicted for subject focus as the target (94.80%), followed by verb focus (85.22%). The accuracy for object focus (58.40%) and VP focus (60.82%) as targets is higher than they are competitors, but is still relatively low.

Compared with data in Chapter 4, the native speakers and immigrants are predicted to perform similarly in the Mandarin mapping tasks. Further statistical evidence with speaker group as a fixed effect also confirmed this observation: there is no main effect of group for all the data ($\chi^2(1) = .049$, $p = .824$), Task 1 ($\chi^2(1) = .248$, $p = .619$), or Task 2 ($\chi^2(1) = .730$, $p = .383$).

5.3.2 Perception of Cantonese focus by immigrants

Logistic regression models with all the responses from immigrants revealed a main effect of focus condition ($\chi^2(7) = 418.450$, $p < .001$), a main effect of task ($\chi^2(1) = 15.897$, $p < .001$) and an interaction of focus condition and task ($\chi^2(7) = 112.820$, $p < .001$). Post-hoc analysis showed higher predicted accuracy in Task 1 than in Task 2. Next, we fitted logistic regression models for each task for a closer examination of the data.

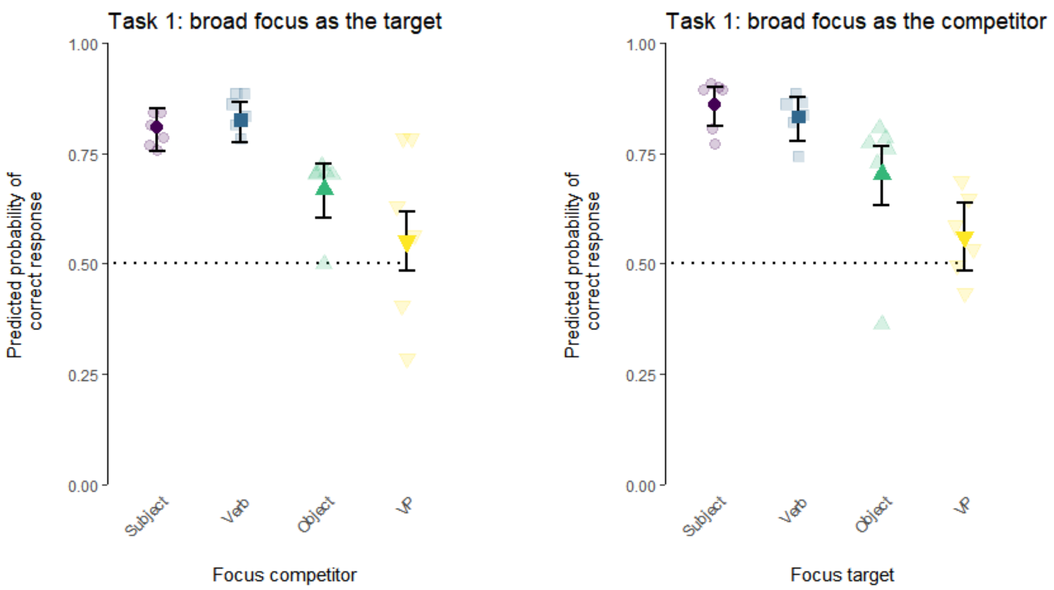

Figure 5-3: Results for Task 1 from immigrants (Cantonese).

Target	Competitor	Predicted accuracy	SE	95% CI
Broad	Subject	80.78%	15.66%	(75.56%, 85.10%)
Broad	Verb	82.53%	16.01%	(77.54%, 86.61%)
Broad	Object	66.82%	14.19%	(60.40%, 72.68%)
Broad	VP	55.09%	13.80%	(48.35%, 61.66%)
Subject	Broad	86.02%	18.56%	(81.05%, 89.85%)
Verb	Broad	83.31%	17.91%	(77.84%, 87.64%)
Object	Broad	70.26%	16.40%	(63.14%, 76.52%)
VP	Broad	56.21%	15.89%	(48.46%, 63.67%)

Table 5-3: Predicted accuracy for Task 1 (immigrants' Cantonese).

For Task 1, a main effect of focus condition was found ($\chi^2(7) = 204.150$, $p < .001$). The predicted probability of correct response for different focus conditions was obtained from the final model and has been plotted in Figure 5-3 and Table 5-3. When broad focus is the target, the participants are expected to have higher accuracies for subject focus (80.78%) and verb focus (82.53%) as the competitors than object focus (66.82%) and VP focus (55.09%) as the competitors. When broad focus is the competitor, the participants show high predicted accuracy for subject focus (86.02%), verb focus (83.31%) and object focus (70.26%) but could not distinguish VP focus from broad focus (56.21%).

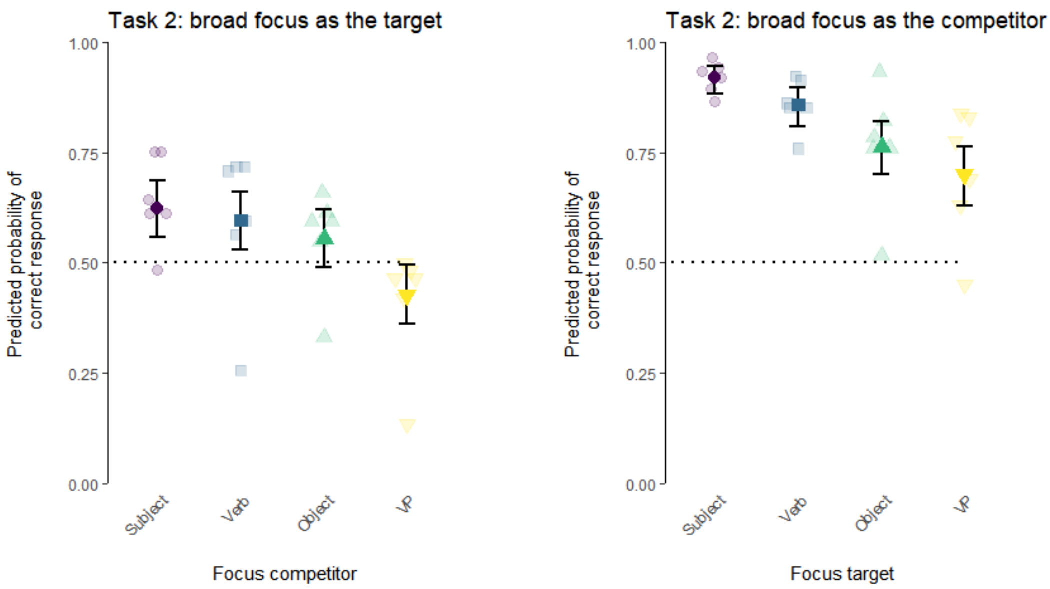

Figure 5-4: Results for Task 2 from immigrants (Cantonese).

Target	Competitor	Predicted accuracy	SE	95% CI
Broad	Subject	62.32%	13.98%	(55.70%, 68.50%)
Broad	Verb	59.65%	13.90%	(52.96%, 66.00%)
Broad	Object	55.63%	13.81%	(48.89%, 62.17%)
Broad	VP	42.74%	13.84%	(36.26%, 49.47%)
Subject	Broad	91.91%	21.19%	(88.23%, 94.51%)
Verb	Broad	85.77%	18.49%	(80.76%, 89.65%)
Object	Broad	76.39%	16.90%	(69.91%, 81.84%)
VP	Broad	69.99%	16.39%	(62.85%, 76.28%)

Table 5-4: Predicted accuracy for Task 2 (immigrants' Cantonese).

For Task 2, a main effect of focus condition also reached significance ($\chi^2(7)$ = 334.890, p < .001). The predicted probability of correct response for different focus conditions was obtained from the final model and has been plotted in Figure 5-4 and Table 5-4. When broad focus is the target, the predicted accuracy is very low for all the other focus types as competitors. When broad focus is the competitor, the predicted accuracy for each focus type as targets is well above chance level, again suggesting a bias against broad focus.

Finally, we fitted logistic regression models with speaker group (natives vs immigrants) as the fixed effect. The native speakers consistently show lower predicted accuracy than the immigrants for all the data ($\chi^2(1) = 40.372$, $p < .001$), Task 1 ($\chi^2(1) = 25.852$, $p < .001$), or Task 2 ($\chi^2(1) = 15.857$, $p < .001$).

5.4 Discussion

5.4.1 Summary of findings

For the immigrants' Mandarin, their overall predicted accuracy is similar to that of the native speakers across different focus pairs, providing no evidence of attrition with regard to mapping prosody and meaning. Unlike the native speakers of Mandarin who had better performance in Task 1 than in Task 2, the immigrants had comparable performance in Task 1 (form to meaning) and Task 2 (meaning to form), suggesting that they relied on bottom-up and top-down mechanisms equally in their processing. These findings may be accounted for by the immigrants' long-term exposure to Cantonese, leading to a compromised strategy in using acoustic cues due to the bilingual background. Recall that F0 is employed more in focus marking in Mandarin compared to Cantonese, and the use of duration is more prevalent in the focus marking of Cantonese. We have shown that, due to the interaction of Mandarin and Cantonese in the immigrants' mind, the immigrants have already started to use duration more often to mark focus in both languages, as is evidenced in our data from Chapter 3. Consequently, the immigrants become more sensitive to the acoustic cues than the native speakers, namely the immigrants are more sensitive to duration than Mandarin speakers and more sensitive to F0 than Cantonese speakers, which makes them rely on the bottom-up processing mechanism. Meanwhile, because the participants were exposed to different systems of focus marking, they should have become better at predicting the focus condition with limited available acoustic information, which in turn enhanced their performance in the meaning to form mapping task.

For the mapping in Cantonese, the immigrants have higher predicted accuracy than the native speakers of Cantonese across different focus pairs, although both speaker groups show similar patterns in the experiment (subject focus and verb focus are better differentiated than object focus and VP focus for both groups of speakers). The immigrants' better performance might be due to the fact that the immigrants have learned to use F0 from their L1 Mandarin and duration from their L2 Cantonese to mark focus and are better at distinguishing the acoustic details than the native speakers of

Cantonese. This can be supported by our results from Chapter 3, which shows clear transfer in the use of different acoustic cues: in L1 Mandarin, the immigrants used duration more often than the native speakers (transferred from L2 to L1); in L2 Cantonese, the immigrants used F0 more often than the native speakers (transferred from L1 to L2). As discussed in 4.4.2, there should be some relationship between speech production and speech perception. Here we provide further support. The immigrants are able to use more acoustic cues in focus production (although non-nativelike, especially in L2 Cantonese) as well as in focus perception. Due to the homogeneity nature of our participants (all are advanced learners of Cantonese), however, we cannot decide on whether production or perception comes first in the language development.

5.4.2 A bilingual advantage

The results in this chapter suggest a bilingual advantage in speech processing. Strictly speaking, all the participants in our study belong to bilingual speakers, because the native speakers of Mandarin and Cantonese also speak English as their second language and the Cantonese speakers learned Mandarin at school as well. The difference between the immigrants and the native speakers is that only the immigrants are expose to an L2 on a daily basis and the two languages of the immigrants are thus activated more often than the participants from the native groups.

The bilingual advantage was first proposed to account for the cases where bilinguals show better results than monolinguals in cognitive tasks (Bialystok, 2001), particularly executive functions (e.g., Bialystok, Craik, & Luk, 2008; Prior & MacWhinney, 2010). More recently, there have been some documentations of the bilingual advantage in language related tasks. For example, bilingual infants are able to discriminate a Dutch native vowel contrast (/i/ and /I/) earlier than monolingual infants (L. Liu & Kager, 2016). Moreover, bilingual speakers are found to be better at speech learning (Antoniou, Liang, Ettlinger, & Wong, 2015) and talker identification (Levi, 2018) than monolingual speakers. There is also evidence that heritage speakers show perceptual benefits in their L2 (Chang, 2016). To the best of our knowledge, no studies have provided evidence of the bilingual advantage in perception of prosodic cues. We propose that bilingualism, particularly regular contact with an L2 that differs from the L1 in prosodic cues, will make the bilinguals more sensitive to the acoustic cues and improve the bilingual speakers' performance in perception/processing of prosody.

Also, it remains to be answered why the immigrants outperformed the Cantonese speakers only but not the Mandarin speakers. One possible explanation is that the native speakers of Mandarin already showed a ceiling effect in the mapping tasks in Mandarin (Hessling, Traxel, & Schmidt, 2004), so it is unlikely that the immigrants would have better performance than the native speakers in Mandarin. On the other hand, the fact that the immigrants were better at mapping focus and prosody than the native speakers of Cantonese may be accounted for by the Cantonese speakers' insensitivity to the acoustic cues.

This chapter investigated the perception of Mandarin and Cantonese focus by the immigrants and compared the immigrants' data with the native speakers' data in Chapter 4. For Mandarin, the immigrants had comparable performance with the native speakers, which showed no clear evidence of L1 attrition in the perception of focus. An unexpected result is that the immigrants outperformed the native speakers in the Cantonese tasks, based on which we propose a bilingual advantage in speech perception of the bilingual speakers.

Chapter 6: Concluding Remarks

6.1 Summary of findings

This thesis set out to address four main research questions.

Research question 1: How do native speakers of Mandarin and Cantonese mark prosodic focus in their native languages when complex subject NPs are involved? How are Mandarin and Cantonese different in terms of focus marking?

Research question 2: How do immigrants mark prosodic focus in their L1 Mandarin and L2 Cantonese? Does their focus production resemble that of native speakers?

Research question 3: Can native speakers of Mandarin and Cantonese map prosody and focus correctly in their native languages? Do these participants show different accuracy rates when the direction of the mapping is reversed (namely, from form to meaning and from meaning to form)?

Research question 4: Can immigrants map prosody and focus correctly in their L1 Mandarin and L2 Cantonese?

To address Question 1, research was conducted to examine how native speakers of Mandarin and Cantonese marked prosodic focus in speech production (Chapter 2). The Mandarin speakers clearly distinguished broad focus from subject focus and verb focus by manipulating F0, duration and intensity. However, when comparing broad focus with object focus and VP focus, Mandarin native speakers only marked pre-focus

F0 and duration and did not show on-focus changes in the acoustic cues. The Cantonese native speakers, on the other hand, did not use F0 or intensity in focus marking most of the time, but they did manipulate duration to mark subject focus and verb focus. The Cantonese native speakers also showed pre-focus shortening of the duration for object focus and VP focus. We therefore propose the existence of a typological difference in Mandarin and Cantonese focus marking, even though the two languages are closely related.

To answer Question 2, focus production data was collected in Mandarin and Cantonese from native Mandarin-speaking immigrants in Hong Kong and compared with that of Mandarin and Cantonese native speakers (Chapter 3). In Mandarin focus marking, the immigrants did not use F0 and intensity to the same degree as the native speakers, but they made more use of duration than the native speakers (more Cantonese-like). In Cantonese focus marking, the immigrants made more use of F0 and intensity in focus marking than the native speakers (more Mandarin-like), and they also used duration to mark focus in a near-native way. Taken together, these results provide evidence of cross-linguistic influences from the immigrants' L1 Mandarin to L2 Cantonese as well as from their L2 Cantonese to L1 Mandarin in their marking of prosodic focus.

With regard to Question 3, native Mandarin-speaking listeners could correctly map prosody and information structure when subject focus or verb focus was paired with broad focus, wherein noticeable differences in the acoustic cues (particularly F0) are available (Chapter 4). When object focus or VP focus is paired with broad focus, wherein such noticeable differences are absent and the focus types are not acoustically different from one another (apart from slight differences in the pre-focus regions), the predicted accuracy is at chance level. The native speakers of Cantonese could clearly map prosody and information structure for subject focus and verb focus, and they were able to distinguish object focus from broad focus most of the time. The predicted accuracy for Cantonese speakers is not as high as that for Mandarin speakers, which is probably because Cantonese makes less use of prosody in focus marking than Mandarin.

To address Question 4, Chapter 5 investigated the mapping of prosody meaning by immigrants. For the immigrants' Mandarin, their overall predicted accuracy was similar to that of native Mandarin speakers across different focus pairs, providing no evidence of attrition with respect to mapping prosody and meaning. For mapping in Cantonese, the immigrants had higher predicted accuracy than the native speakers of Cantonese across several focus pairs, although both speaker groups showed similar patterns. We thus found that the bilingual speakers showed an advantage in speech perception.

6.2 General discussion

As discussed in Section 1.2.2, L2 Intonation Learning Theory (LILt) proposed four dimensions to account for similarities and differences between an L1 and an L2: the systemic dimension, the realisational dimension, the semantic dimension and the frequency dimension. The realisational and semantic dimensions are of particular relevance to this thesis. The realisational dimension refers to the phonetic implementation and realisation of the relevant phonological categories, while the semantic dimension captures the function of the phonological categories. In this thesis, we are interested in how acoustic cues are used (realisational dimension) in focus marking (semantic dimension).

In the production experiments, we measured the acoustic cues (realisational dimension) of prosodic focus marking (semantic dimension) and showed the differences in prosodic focus marking in Mandarin and Cantonese. It was predicted that there would be some interactions between the immigrants' L1 and L2, given the differences in focus marking in Mandarin and Cantonese. The data from immigrants' production revealed that, compared with the native speakers, the immigrants' L1 Mandarin became more Cantonese-like and their L2 Cantonese became more Mandarin-like, suggesting cross-linguistic influences of Mandarin and Cantonese, which have led to attrition of the L1 and incomplete acquisition of the L2. In Chapter

3, we showed clear evidence of both category assimilation and category dissimilation in the immigrants' data. For example, while Mandarin typically uses F0 to mark focus and Cantonese does not, the immigrants marked focus with F0 in Cantonese. This suggests that their Cantonese focus marking had assimilated to Mandarin, at least in their use of F0. In the perception experiments, the immigrants had comparable performance to the Mandarin native speakers, providing no sign of L1 attrition in the mapping of acoustic cues and focus types. The immigrants, on the other hand, showed even higher accuracy rates than the Cantonese native speakers, which was interpreted as evidence for L1 (Mandarin) to L2 influence and the bilingual advantage in speech perception. This interpretation was also supported by our production data, wherein the immigrants used more acoustic cues (particularly F0) to mark focus in Cantonese than the native speakers.

One hypothesis of LILt is that L1 and L2 interact with each other in the common phonological space at the intonation level, just as they do at the segmental level (as proposed in SLM/SLM-r and PAM-L2). Our data provide partial support for the claim that the two languages of bilingual speakers share the same phonological space and mutually influence one another.

6.3 The Bilingual Prosody Transfer Model: A working model

As reviewed in Chapter 1, the existing models of L2 speech concern either segments (such as SLM-r and PAM-L2) or intonation (such as LILt), and therefore cannot be used to explain our data adequately. To better account for the phenomena observed, we propose a working model: the Bilingual Prosody Transfer Model (BPTM). The BPTM accounts for the results contained in this thesis and also provides a reference for future work on the prosody of bilingual speakers.

The basic claim of the BPTM is that prosodic features are transferrable between an L1 and an L2, even for late L2 learners (sequential bilinguals). Below are the initial postulates of the BPTM. Note that these postulates are drawn from the data in this thesis only, and therefore are far from conclusive or exhaustive. More investigations are required to determine whether the postulates hold for other language pairs and to update and refine the model. The current version of the BPTM applies only to the production of prosody by bilingual speakers. After more data are collected and the postulates are refined, we will extend the BPTM to the perception of prosody by late bilinguals.

Postulate 1: The L1 prosodic system is dynamic and prone to influences from L2 prosodic features.

Evidence: In Chapter 3, we showed that, unlike the native speakers of Mandarin, the immigrants did not raise on-focus F0 for subject focus. Instead, they used duration to mark subject focus in Mandarin, suggesting that their L1 prosodic system has undergone changes due to their exposure to an L2 that makes little use of F0 but considerable use of duration in focus marking. This case of L1 attrition can be well explained by Postulate 1.

Postulate 2: The L1 prosodic system is the basis upon which bilingual speakers establish their L2 prosodic system. Consequently, the L2 prosodic system is likely to exhibit prosodic features transferred from the L1.

Evidence: In our data, the native speakers of Cantonese did not mark verb focus with F0. The immigrants, however, manipulated pre-, on- and post-focus F0 to mark verb focus in Cantonese, which resembles the pattern of Mandarin verb focus marking. Also, while native speakers of Cantonese had occasional PFC of duration and intensity in our data, the immigrants showed clear PFC of F0 to mark focus in Cantonese, as did the native speakers of Mandarin. These examples demonstrate that L1 prosodic features can be transferred to an L2.

Postulate 3: The transferability of prosodic features between an L1 and an L2 may depend on acoustic or perceptual differences between the prosodic features of the L1 and the L2. The larger the acoustic or perceptual differences in a prosodic feature between the L1 and the L2, the more likely it is that this feature will be transferred between the L1 and the L2.

Evidence: For example, in the prosodic marking of focus in Mandarin and Cantonese, a clear difference is the use of F0, which is crucial in Mandarin but nearly absent in Cantonese. As a result, the immigrants showed bidirectional influences between their L1 Mandarin and L2 Cantonese in the use of F0 when marking focus in the two languages. Specifically, the immigrants employed F0 less frequently than Mandarin native speakers in Mandarin (more Cantonese-like) but much more frequently than Cantonese native speakers in Cantonese (more Mandarin-like). On the other hand, when both Mandarin and Cantonese make use of duration to mark focus and the only difference is the slightly greater extent to which Cantonese employs duration, the transfer of duration between immigrants' L1 and L2 is minimal in our data. In general, the use of duration in focus marking was similar for the immigrants and the native speakers in both languages.

6.4 Limitations and future directions

This thesis has several limitations and suggests a number of directions that could be explored in future studies. First, the studies conducted for this thesis included both production and perception tasks, but the thesis could not appropriately address the perception-production link (e.g. Mok, Fung, & Li, 2019) because of the heterogeneic nature of the participants in the immigrant group. All of the immigrants were advanced (if not near-native) speakers of Cantonese, which prevents us from carefully examining the developmental trajectories of the immigrants' L1 attrition and L2 attainment (Hansen, 2004; Kornder & Mennen, 2021). To better understand whether perception or production comes first in the acquisition of prosody, it would be necessary to include immigrants with lower Cantonese proficiency and examine the relationship between focus production and perception. Alternatively, longitudinal data could be collected from the same group of participants to track the development of L2 prosody and the changes in L1 prosody brought about by exposure to an L2.

We considered only one direction within one language pair in this thesis: Mandarin native speakers learning Cantonese as an L2. For a more complete picture of the L1 attrition and L2 attainment processes, future studies should also examine the other direction (Cantonese speakers learning Mandarin). The immigrants in our study demonstrated features of focus marking from both languages. For example, they marked Cantonese focus with PFC, which is a feature not often seen in native Cantonese. If we study Cantonese native speakers living in a Mandarin-speaking region, we can determine whether the feature of PFC is learnable in an L2 if it is originally minimal or absent in the L1. As Mandarin and Cantonese are closely related languages, data from other language families would further our understanding even more of the attrition and acquisition of speech prosody.

This study only examined prosodic features in the immigrants' L1 and L2, and this does not provide a full picture of the immigrants' phonetic development. Future studies may extend this investigation to segments, examining for example whether the L1 vowel space is affected by the L2 and whether there are cross-linguistic influences in the L1 and L2 vowel systems. A recent study has shown different effects of focus on vowels in Cantonese and Mandarin (Yang & Chen, 2022), but it is currently unknown whether these different focus effects can be found in the immigrants' L1 and L2 and which factors may shape the focus effects on L1 and L2 vowels. Mandarin and Cantonese are tone languages, so it would also be interesting to examine the interaction between lexical tones and segments. For example, one may want to explore how lexical tones and language background influence VOT in the L1 and L2. In addition to acoustic measurements, it would also be beneficial to invite native listeners to rate the accentedness of the speech samples, providing a holistic view of immigrants' speech from the perspective of native speakers.

Finally, individual differences in speech production and perception should be considered due to the diverse language backgrounds of our participants. It has been well documented that native speakers of the same language exhibit considerable variation in the acoustic properties of their speech (Allen et al., 2003; Clayards, 2018), and such variation may consequently influence the performance of listeners in speech perception

(Newman, Clouse, & Burnham, 2001; Smith & Hawkins, 2012). In our experiments, the native speakers of Cantonese are a relatively homogeneous group, as they were all born and raised in Hong Kong and speak Hong Kong Cantonese as their native language. Although we endeavoured to control the language background of the Mandarin native speaker and Mandarin-speaking immigrant groups, we only managed to restrict their place of birth to Northern China. This is an enormous geographical area containing various northern varieties of Mandarin. Following an initial attempt to explore individual production data (Yang & Chen, 2020a), we aim to have a thorough investigation of the production and perception data of both the native speakers and immigrants from the perspective of individual variation. For example, although all the immigrants in this study are native speakers of Mandarin, their local dialect may have influenced the way they speak Mandarin. If the immigrants and native speakers all have different accents in their native Mandarin, it is likely that the individual variations in the focus production data observed in Yang and Chen (2020a) are the result of local dialects. However, investigating this requires more participants to allow for further grouping analyses.

6.5 Conclusions

This thesis investigated immigrants' production and perception of prosodic focus in L1 Mandarin and L2 Cantonese. The production data suggested bidirectional influences of the immigrants' L1 and L2, which were evident when their acoustic measurements were compared with those of native speakers. The perception data, however, did not reveal any attrition of the immigrants' L1 Mandarin. The immigrants showed even higher accuracy rates than the native speakers in the Cantonese perception experiments, which we attributed to a potential bilingual advantage in speech perception, particularly for the perception of prosody. Examining the combined results, there was evidence of L1 Mandarin attrition in production but not in perception, and the immigrants were more attuned to acoustic cues than the native speakers of Cantonese. Our data thus provide partial support for the existing speech learning models in that the two languages of bilingual speakers mutually influence each other. We proposed the Bilingual Prosody Transfer Model as a working model to account for the findings (particularly for the production data), and this model requires further testing and refinement. To better understand the issues surrounding language attrition and acquisition, future research should explore the developmental sequences of immigrants' L1 and L2, include more language pairs, measure other aspects of speech and language and consider individual variations in speech production and perception.

Appendices

Appendix 1: Informants for the focus production experiment

Mandarin speakers

Code	Gender	Place of birth	Age
Sub 1	Male	Shandong	31.436
Sub 2	Female	Liaoning	23.69401
Sub 3	Male	Beijing	21.57013
Sub 4	Male	Shandong	26.14846
Sub 5	Male	Shandong	25.98966
Sub 6	Female	Liao Ning	26.46882
Sub 7	Female	Hebei	24.52908
Sub 8	Female	Tianjin	21.68298
Sub 9	Male	Shandong	24.61943
Sub 10	Male	Shandong	29.80895
Sub 11	Female	Henan	19.70222
Sub 12	Female	Shandong	25.30729
Sub 13	Female	Hebei	23.65503
Sub 14	Female	Shandong	24.11565
Sub 15	Female	Shandong	21.61668
Sub 16	Male	Shanxi	35.47981
Sub 17	Female	Heilongjiang	23.5948
Sub 18	Male	Shandong	21.27442
Sub 19	Male	Liaoning	28.78434
Sub 20	Female	Hebei	27.91579
Sub 21	Male	Beijing	22.50714

Cantonese speakers

Code	Gender	Place of birth	Age
Sub 1	Female	HK	25.29444
Sub 2	Male	HK	19.96111
Sub 3	Male	HK	23.31944
Sub 4	Male	HK	20.98889
Sub 5	Female	HK	19.53056
Sub 6	Female	HK	20.03056
Sub 7	Male	HK	20.025
Sub 8	Female	HK	21.03889
Sub 9	Male	HK	21.21111
Sub 10	Female	HK	29.06944
Sub 11	Male	HK	20.34444
Sub 12	Male	HK	18.85556
Sub 13	Male	HK	19.925
Sub 14	Male	HK	19.00833
Sub 15	Male	HK	21.79444
Sub 16	Male	HK	19.77222
Sub 17	Female	HK	18.16667
Sub 18	Female	HK	18.00556
Sub 19	Female	HK	18.29167
Sub 20	Female	HK	18.82778
Sub 21	Female	HK	22.85833

Immigrants

Code	Age	Gender	LoR	Onset	His_C	His_M	Use_C	Use_M	Pro_C	Pro_M	Att_C	Att_M	S_C	S_M	Dominance
Sub 1	33	Female	7	26	4.99	39.50	5.30	49.20	49.94	54.48	34.05	54.48	94.28	197.66	103.38
Sub 2	27	Female	9	18	9.53	38.59	13.69	40.81	52.21	54.48	47.67	54.48	123.10	188.36	65.26
Sub 3	27	Female	9	18	10.90	45.85	13.02	41.48	29.51	52.21	40.86	47.67	94.28	187.22	92.93
Sub 4	24	Female	6	18	5.90	40.41	10.54	43.96	54.48	54.48	31.78	54.48	102.70	193.33	90.63
Sub 5	35	Female	3	32	2.72	40.86	16.78	26.82	29.51	52.21	40.86	45.40	89.88	165.29	75.41
Sub 6	32	Female	8	24	6.36	44.49	13.32	41.18	34.05	52.21	40.86	49.94	94.59	187.82	93.23
Sub 7	37	Female	11	26	7.72	39.95	6.66	47.84	29.51	54.48	36.32	54.48	80.21	196.75	116.54
Sub 8	29	Female	5	24	4.54	44.04	19.18	35.32	43.13	54.48	54.48	54.48	121.33	188.31	66.98
Sub 9	35	Female	16	19	5.45	44.04	17.20	37.30	45.40	54.48	49.94	54.48	117.98	190.30	72.32
Sub 10	32	Male	6	26	3.18	42.68	11.63	42.86	38.59	54.48	49.94	54.48	103.34	194.50	91.16
Sub 11	24	Female	5	19	5.45	43.13	17.02	37.48	54.48	54.48	43.13	54.48	120.08	189.57	69.49
Sub 12	27	Female	9	18	6.36	35.41	14.17	40.33	45.40	54.48	47.67	47.67	113.60	177.89	64.30
Sub 13	23	Female	6	17	8.63	50.39	17.89	36.61	40.86	54.48	40.86	54.48	108.24	195.96	87.72
Sub 14	27	Female	5	22	4.09	42.22	18.68	35.82	38.59	54.48	43.13	49.94	104.49	182.46	77.97
Sub 15	28	Female	6	22	5.90	46.31	20.71	33.79	52.21	54.48	49.94	54.48	128.76	189.06	60.30
Sub 16	33	Female	6	27	4.99	43.13	18.65	35.85	43.13	54.48	47.67	54.48	114.44	187.94	73.49
Sub 17	36	Male	9	27	7.72	49.49	8.48	46.02	45.40	54.48	31.78	54.48	93.37	204.47	111.09
Sub 18	34	Female	12	22	9.99	44.49	17.34	37.16	52.21	54.48	40.86	49.94	120.40	186.07	65.67
Sub 19	33	Female	3	30	4.09	40.86	46.87	7.63	27.24	54.48	47.67	47.67	125.87	150.64	24.77
Sub 20	35	Female	10	25	9.08	43.13	22.30	31.11	49.94	54.48	49.94	54.48	131.26	183.20	51.94
Sub 21	25	Female	3	22	3.63	44.95	4.72	49.78	47.67	54.48	38.59	54.48	94.61	203.68	109.07
Sub 22	27	Male	9	18	6.36	36.77	6.03	48.47	29.51	49.94	27.24	52.21	69.14	187.39	118.25

Notes:

1. LoR: length of residence in Hong Kong
2. Onset: age of onset of learning Cantonese (equals to age of arrival in Hong Kong)
3. For the next eight columns, ‘C’ stands for the scores for Cantonese and ‘M’ represents scores for Mandarin from the Bilingual Language Profile. ‘His’, ‘Use’, ‘Pro’, and ‘Att’ are the scores calculated from the self-reported four blocks of language history, language use, language proficiency and language attitude, respectively. ‘S_C’ and ‘S_M’ are the total scores for Cantonese and Mandarin. Dominance score refers to language dominance, with a higher score suggesting more dominance in Mandarin.

Appendix 2: Stimuli for the focus production experiment

Mandarin target sentences:

1. na4 wei4 yi1sheng1 he1 ka1fei1
   that CL doctor drink coffee
   'The doctor drank coffee.'
2. na4 qun2 qing1wa1 ban1 xin1jia1
   that CL frog move new_house
   'The frogs moved to a new place.'
3. na4 wei4 gong1bing1 da1 fei1ji1
   that CL pioneer take airplane
   'The pioneer took the plane.'

Cantonese target sentences:

1. go2 go3 gaa1ban1 ling1 gaau1zeon1
   that CL guest carry plastic_bottle
   'The guest carried a plastic bottle.'
2. go2 go3 si1gei1 zaa1 cyun1baa1
   that CL driver drive residential_bus
   'That driver drives the residential route.'
3. go2 deoi3 fu1cai1 fei1 gwaan1sai1
   that CL couple fly Kansai
   'That couple flew to Kansai.'
4. go2 di1 can1cik1 zong1 faa1dang1
   that CL relative assemble lantern
   'The relatives assembled the lantern.'

Appendix 3: Informants for the focus perception experiment

Mandarin speakers

Code	Gender	Age	Place of birth
Sub 1	Female	28	Henan
Sub 2	Female	27	Shanxi
Sub 3	Male	31	Liaoning
Sub 4	Female	27	Henan
Sub 5	Female	25	Shanxi
Sub 6	Male	28	Inner Mongolia
Sub 7	Male	26	Heilongjiang
Sub 8	Female	27	Tianjin
Sub 9	Female	24	Shandong
Sub 10	Female	24	Shaanxi
Sub 11	Female	28	Anhui
Sub 12	Female	27	Shandong
Sub 13	Female	27	Shandong
Sub 14	Male	28	Liaoning
Sub 15	Male	25	Jilin
Sub 16	Male	30	Shandong
Sub 17	Female	30	Hebei
Sub 18	Male	26	Hebei
Sub 19	Female	30	Shanxi
Sub 20	Female	28	Heilongjiang
Sub 21	Male	22	Shandong
Sub 22	Male	25	Shandong
Sub 23	Female	27	Hebei
Sub 24	Male	30	Hebei
Sub 25	Female	30	Liaoning
Sub 26	Male	26	Jilin
Sub 27	Female	27	Gansu
Sub 28	Male	29	Henan
Sub 29	Female	31	Henan
Sub 30	Female	27	Shaanxi

Cantonese speakers

Code	Gender	Age	Place of birth
Sub 1	Male	22	Hong Kong
Sub 2	Female	22	Hong Kong
Sub 3	Male	20	Hong Kong
Sub 4	Male	20	Hong Kong
Sub 5	Male	21	Hong Kong
Sub 6	Female	20	Hong Kong
Sub 7	Male	26	Hong Kong
Sub 8	Male	20	Hong Kong
Sub 9	Male	20	Hong Kong
Sub 10	Male	20	Hong Kong
Sub 11	Female	19	Hong Kong
Sub 12	Male	23	Hong Kong
Sub 13	Female	19	Hong Kong
Sub 14	Male	20	Hong Kong
Sub 15	Female	20	Hong Kong
Sub 16	Female	20	Hong Kong
Sub 17	Female	21	Hong Kong
Sub 18	Female	23	Hong Kong
Sub 19	Female	20	Hong Kong
Sub 20	Female	30	Hong Kong
Sub 21	Male	23	Hong Kong
Sub 22	Female	19	Hong Kong
Sub 23	Female	19	Hong Kong
Sub 24	Female	23	Hong Kong
Sub 25	Female	23	Hong Kong
Sub 26	Female	19	Hong Kong
Sub 27	Female	21	Hong Kong
Sub 28	Female	25	Hong Kong
Sub 29	Male	19	Hong Kong
Sub 30	Male	23	Hong Kong
Sub 31	Male	20	Hong Kong

Immigrants

Code	Age	LoR	Gender	Place of birth	His_C	His_M	Use_C	Use_M	Pro_C	Pro_M	Att_C	Att_M	S_C	S_M	Dominance
Sub1	33	7	Female	Henan	4.994	39.498	5.29849	49.20151	49.94	54.48	34.05	54.48	94.28249	197.65951	103.377
Sub2	27	9	Female	Shannxi	9.534	38.59	13.68604	40.81396	52.21	54.48	47.67	54.48	123.1	188.36396	65.26392
Sub 3	31	8	Female	Beijing	6.356	39.498	18.8025	35.6975	47.67	54.48	40.86	54.48	113.6885	184.1555	70.467
Sub 4	24	6	Female	Shandong	5.902	40.406	10.53703	43.96297	54.48	54.48	31.78	54.48	102.699	193.32897	90.62994
Sub 5	35	3	Female	Beijing	2.724	40.86	16.78273	26.81727	29.51	52.21	40.86	45.4	89.87673	165.28727	75.41054
Sub 6	32	8	Female	Liaoning	6.356	44.492	13.32198	41.17802	34.05	52.21	40.86	49.94	94.58798	187.82002	93.23204
Sub 7	37	11	Female	Liaoning	7.718	39.952	6.66099	47.83901	29.51	54.48	36.32	54.48	80.20899	196.75101	116.542
Sub 8	29	5	Female	Shandong	4.54	44.038	19.184	35.316	43.13	54.48	54.48	54.48	121.334	188.314	66.98
Sub 9	35	16	Female	Beijing	5.448	44.038	17.19693	37.30198	45.4	54.48	49.94	54.48	117.9849	190.29998	72.31505
Sub 10	32	6	Male	Tianjin	3.178	42.676	11.63466	42.86425	38.59	54.48	49.94	54.48	103.3427	194.50025	91.15759
Sub 11	24	5	Female	Jilin	5.448	43.13	17.02035	37.47965	54.48	54.48	43.13	54.48	120.0784	189.56965	69.4913
Sub 12	27	9	Female	Jiangsu	6.356	35.412	14.17	40.33	45.4	54.48	47.67	47.67	113.596	177.892	64.296
Sub 13	23	6	Female	Jilin	8.626	50.394	17.89344	36.60547	40.86	54.48	40.86	54.48	108.2394	195.95947	87.72003
Sub 14	26	3	Female	Tianjin	2.724	36.32	16.35	38.15	31.78	54.48	29.51	29.51	80.364	158.46	78.096
Sub 15	25	3	Female	Shanxi	4.086	41.314	17.44	37.06	38.59	54.48	40.86	54.48	100.976	187.334	86.358
Sub 16	41	13	Female	Tianjin	14.074	45.4	20.24348	34.25652	47.67	54.48	36.32	54.48	118.3075	188.61652	70.30904
Sub 17	31	9	Female	Shandong	5.448	39.044	13.68495	40.81505	40.86	54.48	22.7	49.94	82.69295	184.27905	101.5861
Sub 18	30	6	Female	Heilongjiang	3.178	46.762	2.42198	52.07802	40.86	54.48	38.59	54.48	85.04998	207.80002	122.75
Sub 19	31	8	Male	Tianjin	7.264	47.216	6.72094	47.77797	38.59	54.48	22.7	54.48	75.27494	203.95397	128.679
Sub 20	27	5	Female	Henan	4.086	42.222	18.68151	35.81849	38.59	54.48	43.13	49.94	104.4875	182.46049	77.97298
Sub 21	28	6	Female	Tianjin	5.902	46.308	20.71	33.79	52.21	54.48	49.94	54.48	128.762	189.058	60.296
Sub 22	41	9	Male	Tianjin	19.068	53.572	29.06703	25.43297	47.67	54.48	54.48	54.48	150.285	187.96497	37.67994
Sub 23	27	9	Female	Henan	10.896	45.854	13.01896	41.48104	29.51	52.21	40.86	47.67	94.28496	187.21504	92.93008
Sub 24	27	9	Male	Shaanxi	6.356	36.774	6.03424	48.46576	29.51	49.94	27.24	52.21	69.14024	187.38976	118.2495
Sub 25	23	7	Male	Anhui	6.81	34.05	14.7586	39.7414	45.4	54.48	45.4	49.94	112.3686	178.2114	65.8428
Sub 26	33	6	Female	Xinjiang	4.994	43.13	18.65099	35.84901	43.13	54.48	47.67	54.48	114.445	187.93901	73.49402
Sub 27	36	11	Male	Tianjin	7.718	49.486	8.47693	46.02307	45.4	54.48	31.78	54.48	93.37493	204.46907	111.0941
Sub 28	34	12	Female	Hebei	9.988	44.492	17.34081	37.15919	52.21	54.48	40.86	49.94	120.3988	186.07119	65.67238
Sub 29	33	3	Female	Heilongjiang	4.086	40.86	46.87	7.63	27.24	54.48	47.67	47.67	125.866	150.64	24.774
Sub 30	35	10	Female	Tianjin	9.08	43.13	22.3014	31.1086	49.94	54.48	49.94	54.48	131.2614	183.1986	51.9372
Sub 31	25	3	Female	Liaoning	3.632	44.946	4.72297	49.77703	47.67	54.48	38.59	54.48	94.61497	203.68303	109.0681
Sub 32	39	1	Female	Liaoning	1.816	46.308	7.26594	47.23406	31.78	54.48	38.59	54.48	79.45194	202.50206	123.0501
Sub 33	31	5	Female	Shanxi	6.356	53.118	12.06848	42.43152	13.62	27.24	24.97	27.24	57.01448	150.02952	93.01504

Appendix 4: Stimuli for the focus perception experiment

Mandarin target sentences:

1. na4 wei4 yi1sheng1 he1 ka1fei1
   that CL doctor drink coffee
   'The doctor drank coffee.'
2. zhe4 xie1 zhuan1jia1 tie1 biaoqian1
   that CL expert stick label
   'The experts sticked some labels.'
3. na4 wei4 gong1bing1 da1 fei1ji1
   that CL pioneer take airplane
   'The pioneer took the plane.'
4. zhe4 ge4 si1ji1 kai1 ban1che1
   this CL driver drive shuttle_bus
   'This driver drives the shuttle bus.'
5. na4 dui4 fu1qi1 kai1 gong1si1
   that CL couple open company
   'That couple started their business.'
6. na4 xie1 jia1bin1 chuan1 xi1zhuang1
   that CL guest wear suit
   'The guests were in suits.'

Cantonese target sentences:

1. go2 go3 gaa1ban1 ling1 gaau1zeon1
   that CL guest carry plastic_bottle
   'The guest carried a plastic bottle.'
2. go2 go3 si1gei1 zaa1 cyun1baa1
   that CL driver drive residential_bus
   'That driver drives the residential route.'
3. go2 go3 go1sing1 ceot1 zyun1cap1
   that CL singer publish album
   'That singer published an album.'
4. go2 di1 zik1gung1 bun1 gaa1si1
   that CL staff move furniture
   'The staff moved the furniture.'
5. nei1 di1 hat1ji4 sau1 bo1saam1
   this CL beggar collect shirt
   'The beggar collected some shirts.'
6. go2 go3 si1naai5 bou1 gai1tong1
   that CL housewife cook chicken_soup
   'The housewife cooked some chicken soup.'

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Agresti, A. (2013). *Categorical Data Analysis*. Hoboken, New Jersey: John Wiley & Sons.

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *113*(1), 544–552. https://doi.org/10.1121/1.1528172

Andreeva, B., Barry, W., & Steiner, I. (2007). Producing phrasal prominence in German. In *Proceedings of ICPhS 2007* (pp. 1209–1212).

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition*, *18*(4), 683–695. https://doi.org/10.1017/S1366728914000777

Arregi, K. (2016). Focus Projection Theories. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 185–202). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642670.013.005

Audacity Team. (2019). Audacity(R): Free Audio Editor and Recorder. Retrieved from https://audacityteam.org/

Babel, M. E. (2009). *Phonetic and social selectivity in speech accommodation*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract%5Cnhttp://www.linguistics.berkeley.edu/phonlab/annual_report/documents/2009/Babel_dissertation_2009.pdf

Baker, R. E. (2010). *The Acquisition of English Focus Marking by Non-Native Speakers*. Northwestern University.

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child-Adult Differences in Second-Language Phonological Learning: The Role of Cross-Language Similarity. *Language and Speech*, *51*(4), 317–342. https://doi.org/10.1177/0023830908099068

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica*, *142*(3), 402–409.

https://doi.org/10.1016/j.actpsy.2013.01.010
Best, C. T. (1991). *The emergence of native-language phonological influences in infants: A perceptual assimilation model*. *Haskins Laboratories Status Reporl on Speech Research*.
Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.
Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511605963
Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 859–873. https://doi.org/10.1037/0278-7393.34.4.859
Birdsong, D. (2004). Second Language and Ultimate Attainment. In A. Davies & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 82–105). Oxford: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757000.ch3
Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. Retrieved from https://sites.la.utexas.edu/bilingual/
Bishop, J. (2017). Focus projection and prenuclear accents: evidence from lexical processing. *Language, Cognition and Neuroscience*, *32*(2), 236–253. https://doi.org/10.1080/23273798.2016.1246745
Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/
Bolinger, D. (1972). Accent Is Predictable (If You’re a Mind-Reader). *Language*, *48*(3), 633–644.
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
Büring, D. (2012). Focus and intonation. In G. Russell & D. G. Fara (Eds.), *The Routledge Companion to the Philosophy of Language* (pp. 103–115). London: Routledge.
Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 629–647. https://doi.org/10.1002/wcs.63
Chamorro, G., Sturt, P., & Sorace, A. (2016). Selectivity in L1 attrition: Differential

object marking in Spanish near-native speakers of English. *Journal of Psycholinguistic Research*, *45*(3), 697–715. https://doi.org/10.1007/s10936-015-9372-4

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*(2), 249–268. https://doi.org/10.1016/j.wocn.2011.10.007

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, *41*(6), 520–533. https://doi.org/10.1016/j.wocn.2013.09.006

Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, *19*(4), 791–809. https://doi.org/10.1017/S1366728914000261

Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, Ying. (2014). *Prosodic Realization of Focus in Second Language Speech: Effects of Language Experience*. University of Oregon.

Chen, Ying. (2015). Post-focus compression in English by Mandarin learners. In *Proceedings of ICPhS 2015* (pp. 1–5).

Chen, Ying, Xu, Y., & Guion-Anderson, S. (2014). Prosodic Realization of Focus in Bilingual Production of Southern Min and Mandarin. *Phonetica*, *71*(4), 249–270. https://doi.org/10.1159/000371891

Chen, Yiya. (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics*, *34*(2), 176–201. https://doi.org/10.1016/j.wocn.2005.05.002

Chen, Yiya. (2010). Post-focus F0 compression-Now you see it, now you don’t. *Journal of Phonetics*, *38*(4), 517–525. https://doi.org/10.1016/j.wocn.2010.06.004

Chen, Yiya, & 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, Yiya, Po-lun Lee, P., & Pan, H. (2016). Topic and Focus Marking in Chinese. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 733–752). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642670.013.34

Cherciov, M. (2013). Investigating the impact of attitude on first language attrition and second language acquisition from a Dynamic Systems Theory perspective. *International Journal of Bilingualism*, *17*(6), 716–733. https://doi.org/10.1177/1367006912454622

Cheung, L. Y. L. (2009). Dislocation focus construction in Chinese. *Journal of East Asian Linguistics*, *18*(3), 197–232. https://doi.org/10.1007/s10831-009-9046-z

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Clayards, M. (2018). Individual Talker and Token Covariation in the Production of Multiple Cues to Stop Voicing. *Phonetica*, *75*(1), 1–23. https://doi.org/10.1159/000448809

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

Cook, V. J. (1991). The poverty-of-the-stimulus argument and multicompetence. *Second Language Research*, *7*(2), 103–117. https://doi.org/10.1177/026765839100700203

Cook, V. J. (2016). Premises of multi-competence. In V. Cook & W. Li (Eds.), *The Cambridge Handbook of Linguistic Multi-Competence* (pp. 1–25). Cambridge University Press.

Cooper, W., Eady, S., & Mueller, P. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, *77*(6), 2142–2155.

de Leeuw, E., Mennen, I., & Scobbie, J. M. (2012). Singing a different tune in your native language: first language attrition of prosody. *International Journal of Bilingualism*, *16*(1), 101–116. https://doi.org/10.1177/1367006911405576

de Leeuw, E., Opitz, C., & Lubiska, D. (2013). Dynamics of first language attrition across the lifespan. *International Journal of Bilingualism*, *17*(6), 667–674. https://doi.org/10.1177/1367006912454618

Dik, S. C. (1980). On the typology of Focus Phenomena. In T. Hoekstra, H. van der Hulst, & M. Moortgat (Eds.), *Perspectives on Functional Grammar* (pp. 41–74). De Gruyter. https://doi.org/10.1515/9783112329603-005

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (2002). Interactions between the native and second-Ianguage phonetic systems. In T. Piske, A. Rohde, & P. Burmeister (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217–244). Trier: Wissenschaftlicher Verlag.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, *34*(2), 153–175.

https://doi.org/10.1016/j.wocn.2005.05.001
Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108886901.002
Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*, 67–83.
Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language*, *41*, 78–104. https://doi.org/10.1006/jmla.1999.2638
Ge, H., & Yuen, H. K. (2020). Syntax or prosody? Comprehension of focus by Cantonese-speaking children and adults. In *The 24th International Conference on Yue Dialects*.
Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Houghton Mifflin.
Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9–31). Tuebingen: Max Niemeyer Verlag.
Gregory, R. (1970). *The Intelligent Eye*. Weidenfeld & Nicolson.
Gu, W., & Lee, T. (2007). Effects of tonal context and focus on Cantonese F0. In *Proceedings of 16th International Congress of Phonetic Sciences* (pp. 1033–1036). Retrieved from http://icphs2007.de/conference/Papers/1689/1689.pdf
Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107*(5), 2711–2724. https://doi.org/10.1121/1.428657
Gürel, A. (2004). Selectivity in L2-induced L1 attrition: A psycholinguistic account. *Journal of Neurolinguistics*, *17*(1), 53–78. https://doi.org/10.1016/S0911-6044(03)00054-X
Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, *19*(2), 377–417. https://doi.org/10.1017/S0022226700007799
Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511616983
Gussenhoven, C. (2007). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 83–100). Dordrecht: Springer.
Gut, U., & Pillai, S. (2014). Prosodic marking of information structure by Malaysian

speakers of English. *Studies in Second Language Acquisition*, *36*(2), 283–302. https://doi.org/10.1017/S0272263113000739

Hansen, J. G. (2004). Developmental sequences in the acquisition of English L2 syllable codas: A preliminary study. *Studies in Second Language Acquisition*, *26*(1), 85–124. https://doi.org/10.1017/S0272263104026142

Harada, T. (2003). L2 Influence on L1 Speech in the Production of VOT. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1085–1088). Barcelona, Spain.

Hessling, R. M., Traxel, N. M., & Schmidt, T. J. (2004). Ceiling Effect. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *The SAGE Encyclopedia of Social Science Research Methods* (p. 106). Thousand Oaks: SAGE Publications, Inc.

Hopp, H., & Schmid, M. S. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, *34*(2), 361–394. https://doi.org/10.1017/S0142716411000737

Hsu, Y.-Y., & Xu, A. (2017). Focus Acoustics in Mandarin Nominals. In *Proceedings of Interspeech 2017* (pp. 3231–3235). ISCA: ISCA. https://doi.org/10.21437/Interspeech.2017-1167

Hsu, Y.-Y., Xu, A., & Ngai, H. (2018). Focus Prosody in Cantonese and Teochew Noun Phrases. In *Proceedings of Speech Prosody 2018* (pp. 961–965). ISCA: ISCA. https://doi.org/10.21437/SpeechProsody.2018-194

Huang, T., & Hsieh, F. (2016). Post-focus compression: All or nothing? *The Journal of the Acoustical Society of America*, *140*(4), 3224–3224. https://doi.org/10.1121/1.4970180

Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese*. The Ohio State University.

Jun, S.-A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 430–458). Oxford: Oxford University Press.

Keijzer, M. (2010). The regression hypothesis as a framework for first language attrition. *Bilingualism: Language and Cognition*, *13*(131), 9–18. https://doi.org/10.1017/S1366728909990356

Köpke, B., & Schmid, M. S. (2004). Language attrition: The next phase. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 1–43). Amsterdam: John Benjamins.

Kornder, L., & Mennen, I. (2021). Longitudinal Developments in Bilingual Second Language Acquisition and First Language Attrition of Speech: The Case of

Arnold Schwarzenegger. *Languages*, *6*(2), 61. https://doi.org/10.3390/languages6020061

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

Kügler, F., & Skopeteas, S. (2007). On the universality of prosodic reflexes of contrast: The case of Yucatec Maya. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1025–1028).

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D. R., Schepman, A., White, L., Quarmby, L. M., & Stackhouse, R. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics*, *37*(2), 145–161. https://doi.org/10.1016/j.wocn.2008.11.001

Lee, T. M. (2017). Defocalization in Cantonese Right Dislocation. *Gengo Kenkyu*, *152*, 59–87. https://doi.org/10.11435/gengo.152.0

Levi, S. (2018). Another bilingual advantage? Perception of talker-voice information. *Bilingualism: Language and Cognition*, *21*(3), 523–536. https://doi.org/10.1017/S1366728917000153

Li, J. J., & Grigos, M. (2018). The Effect of L1 Prosody in the Perception and Production of Non-native Lexical Stress. In *Proceedings of Speech Prosody 2018* (pp. 898–902). ISCA. https://doi.org/10.21437/SpeechProsody.2018-181

Li, J. J., & Grigos, M. I. (2021). Production of English lexical stress by Mandarin speakers: Acoustics and kinematics. *Second Language Research*, 1–29. https://doi.org/10.1177/02676583211066290

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Liu, J., Xu, Y., & Lee, Y. (2019). Post-focus compression is not automatically transferred from Korean to L2 English. *Phonetics and Speech Sciences*, *11*(2), 15–21. https://doi.org/10.13064/ksss.2019.11.2.015

Liu, L., & Kager, R. (2016). Perception of a native vowel contrast by Dutch monolingual and bilingual infants: A bilingual perceptual lead. *International Journal of Bilingualism*, *20*(3), 335–345. https://doi.org/10.1177/1367006914566082

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*(3), 110–114. https://doi.org/10.1016/j.tics.2008.11.008

Major, R. C. (1992). Losing English as a First Language. *The Modern Language Journal*, *76*(2), 190–208. https://doi.org/10.2307/329772

Matsuoka, K., Miyoshi, N., Hoshi, K., Ueda, M., Yabu, I., & Hirata, M. (2006). The Acquisition of Japanese Focus Particles: dake (only) and mo (also). In *Online supplement to the proceedings of the 30th Boston University conference on language development* (pp. 1–12).

McGory, J. T. (1997). *Acquisition of intonational prominence in English by Seoul Korean and Mandarin Chinese speakers*. Ohio State University.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*(4), 543–563. https://doi.org/10.1016/j.wocn.2004.02.002

Mennen, I. (2015). Beyond segments: towards a L2 intonation learning theory (LILt). In E. Delais-Roussarie (Ed.), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations* (pp. 171–188). Berlin: Springer-Verlag. https://doi.org/10.1007/978-3-662-45168-7

Mennen, I., Chen, A., & Karlsson, F. (2010). Characterising the internal structure of learner intonation and its development over time. In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech* (pp. 319–324).

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, *131*(3), 2249–2260. https://doi.org/10.1121/1.3681950

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. https://doi.org/10.1044/2018_JSLHR-S-17-0430

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, *109*(3), 1181–1196. https://doi.org/10.1121/1.1348009

Niebuhr, O., & Winkler, J. (2017). The Relative Cueing Power of F0 and Duration in German Prominence Perception. In *Proceedings of Interspeech 2017* (pp. 611–615). ISCA: ISCA. https://doi.org/10.21437/Interspeech.2017-375

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In *Proc. ICPhS 2003* (pp. 771–774). Retrieved from http://www.let.uu.nl/LOT/GraduateProgram/LotSchools/Summerschool2005/course descriptions/nolan_icphs2003.pdf

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting.

Journal of Phonetics, *39*(2), 156–167. https://doi.org/10.1016/j.wocn.2011.01.002
Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720
Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, *69*(3), 183–195. https://doi.org/10.1016/j.jml.2013.06.002
Peppé, S., Maxim, J., & Wells, B. (2000). Prosodic Variation in Southern British English. *Language and Speech*, *43*(3), 309–334. https://doi.org/10.1177/00238309000430030501
Post, B. (2000). *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.
Post, B., D'Imperio, M., & Gussenhoven, C. (2007). Fine Phonetic Detail and Intonational Meaning. In *International Congress of Phonetic Sciences (ICPhS XVI)* (pp. 191–196).
Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(2), 253–262. https://doi.org/10.1017/S1366728909990526
Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811
R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org
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
RStudio Team. (2016). RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/
Saito, K. (2015). The Role of Age of Acquisition in Late Second Language Oral Proficiency Attainment. *Studies in Second Language Acquisition*, *37*(4), 713–743. https://doi.org/10.1017/S0272263115000248
Sancier, M. L., & Fowler, C. a. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*(4), 421–436. https://doi.org/10.1006/jpho.1997.0051
Schmid, M. S. (2002). *First language attrition, use and maintenance: The case of*

German Jews in anglophone countries. Amsterdam/Philadelphia: John Benjamins Publishing. https://doi.org/10.1075/sibil.24

Schmid, M. S., & Yilmaz, G. (2018). Predictors of language dominance: An integrated analysis of first language attrition and second language acquisition in late bilinguals. *Frontiers in Psychology*, *9*, 1–19. https://doi.org/10.3389/fpsyg.2018.01306

Schmitz, J., Díaz, B., Fernández Rubio, K., & Sebastian-Galles, N. (2018). Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities. *Language, Cognition and Neuroscience*, *33*(5), 527–546. https://doi.org/10.1080/23273798.2017.1390142

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime User's Guide*. Pittsburgh: Psychological Software Tools Inc.

Selkirk, E. (2008). Contrastive focus, givenness and the unmarked status of "discourse-new." *Acta Linguistica Hungarica*, *55*(3–4), 331–346. https://doi.org/10.1556/ALing.55.2008.3-4.8

Simonet, M. (2011). Production of a catalan-specific vowel contrast by early Spanish-Catalan bilinguals. *Phonetica*, *68*(1–2), 88–110. https://doi.org/10.1159/000328847

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours. *IRAL - International Review of Applied Linguistics in Language Teaching*, *43*(4), 269–285. https://doi.org/10.1515/iral.2005.43.4.269

Smith, R., & Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics*, *40*(2), 213–233. https://doi.org/10.1016/j.wocn.2011.11.003

Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*(3), 339–368. https://doi.org/10.1191/0267658306sr271oa

Ulbrich, C., & Ordin, M. (2014). Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics*, *45*(1), 26–42. https://doi.org/10.1016/j.wocn.2014.02.008

Weinreich, U. (1957). On the Description of Phonic Interference. *Word*, *13*(1), 1–11. https://doi.org/10.1080/00437956.1957.11659624

Welby, P. (2003). Effects of Pitch Accent Position, Type, and Status on Focus Projection. *Language and Speech*, *46*(1), 53–81. https://doi.org/10.1177/00238309030460010401

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and*

Development, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer.

Wouters, J., & Macon, M. W. (2002). Effects of prosodic factors on spectral dynamics. I. Analysis. *The Journal of the Acoustical Society of America*, *111*(1), 417–427. https://doi.org/10.1121/1.1428262

Wu, W. L. (2013). *Cantonese prosody: Sentence-final particles and prosodic focus*. University College London.

Wu, W. L., & Chung, L. (2011). Post-focus compression in English-Cantonese bilingual speakers. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)* (pp. 148–151).

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

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*(1), 55–105.

Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 152–155).

Xu, Y. (2013). ProsodyPro - A tool for large-scale systematic prosody analysis. In *Proc. TRASP'2013* (pp. 7–10). Aix-en-Provence, France.

Xu, Y., Chen, S. W., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review*, *29*(1), 131–147. https://doi.org/10.1515/tlr-2012-0006

Yan, M., & Calhoun, S. (2020). Rejecting false alternatives in Chinese and English: The interaction of prosody, clefting, and default focus position. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1), 17. https://doi.org/10.5334/labphon.255

Yang, Y. (in preparation). Informational focus and contrastive focus are not distinguished in Mandarin: Evidence from production and perception.

Yang, Y., & Chen, S. (2020a). Individual differences in Mandarin focus production. In *Proceedings of 11th International Conference of Experimental Linguistics* (pp. 229–232). ExLing Society. https://doi.org/10.36505/ExLing-2020/11/0057/000472

Yang, Y., & Chen, S. (2020b). Revisiting focus production in Mandarin Chinese: Some preliminary findings. In *Proceedings of Speech Prosody 2020* (pp. 260–264). ISCA. https://doi.org/10.21437/SpeechProsody.2020-53

Yang, Y., & Chen, S. (2022). Does prosody influence segments differently in Cantonese and Mandarin? A case study of the open vowel /a/. In *Proceedings of*

Speech Prosody 2022. ISCA.
Yang, Y., Chen, S., & Li, K. (2018). Pitch realization of post-focus components in Chongming Chinese. *The Journal of the Acoustical Society of America*, *144*(3), 1938–1938. https://doi.org/10.1121/1.5068477
Yang, Y., Chen, S., & Li, K. (2019). Effects of focus on duration and intensity in Chongming Chinese. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of ICPhS 2019* (pp. 3578–3582). Canberra, Australia: Australasian Speech Science and Technology Association Inc.
Zee, E. (1978). Duration and intensity as correlates of F0. *Journal of Phonetics*, *6*(3), 213–220. https://doi.org/10.1016/S0095-4470(19)31153-2
Zhang, X. (1998). Dialect MT: A case study between Cantonese and Mandarin. In *Proceedings of COLING 1998* (Vol. 2, pp. 1460–1464). Association for Computational Linguistics.