# Cantonese Prosody: Sentence-final Particles and Prosodic Focus

Wing Li WU

Department of Speech, Hearing and Phonetic Sciences

Division of Psychology and Language Sciences

Submitted for the degree of Doctor of Philosophy

to

UCL

# Declaration

I hereby confirm that the work presented in this dissertation is my own work. Where information has been derived from other sources, I confirm that this has been acknowledged in the dissertation.

Wing Li WU

# Abstract

Fundamental frequency ($F_0$) is the most important feature among the components of prosody in a language, tone and intonation languages alike. In a tone language, how lexical tone and sentence intonation can both use $F_0$ as their main acoustic cues has long been an intriguing question. Will the lexical tones be so resistant to modification that no elaborate intonation is possible in the language? How much of the surface sentential $F_0$ is attributable to lexical tones and intonation? If $F_0$ modification is kept minimal, how will prosodic focus be realized? And will the lack of focus-related $F_0$ change be a disadvantage in terms of focus perception? In this dissertation, experimental studies have been made on Cantonese in some less well-understood aspects of its prosody. Firstly, the tonal characteristics of sentence-final particles (SFPs) in Cantonese as a special case of the interaction between tone and intonation are examined. Secondly, the acoustic correlates of prosodic focus in Cantonese are explored.

SFPs are a class of words known to have functions similar to intonation. It is not yet clear, however, whether the $F_0$ contours of SFPs are derived purely from lexical tones, purely intonational, or a combination of tone and intonation. As an attempt to offer a solution, a production experiment was designed in which sentences in Hong Kong Cantonese with ten different SFPs were recorded and detailed analyses of their $F_0$ contours, final $F_0$, final $F_0$ velocity and duration were performed. The results show that most of these SFPs are very similar to the lexical tones in terms of $F_0$ contours, but there are significant differences in durations in more than half the cases. In addition, the occurrence of an SFP does not give rise to differences in $F_0$ and duration in the syllables preceding the SFP in most cases. But differences can be seen in sentences with question SFPs, which indicates that the prosody of the SFPs may be partly due to intonational meanings. One of the SFPs, however, exhibits a component $F_0$ contour that seems to be sequentially attached to the end of the lexical tonal component. These findings suggest that Cantonese SFPs have underlying tonal targets just like those of lexical tones, but they also carry intonational meanings by modifying the lexical tonal contours.

Previous research has shown that Beijing Mandarin, a tone language, marks focus not only by on-focus prosodic expansion like many other languages, but also by post-focus compression of pitch range and intensity (PFC). However, recently it is found that PFC is absent in Taiwanese and Taiwan Mandarin, two languages closely related to Beijing Mandarin. This finding both highlights the non-universality of PFC and raises questions about its origin. The present study explores these issues by investigating focus production in Cantonese by native Cantonese speakers born and raised in Hong Kong, and in English and Cantonese by bilingual speakers who were born and raised in Southern England. Results from the Hong Kong speakers show that, just as in Taiwanese and Taiwan Mandarin, PFC is absent in Cantonese, and mean $F_0$, duration, intensity and excursion size were found to be higher in on-focus words. Results from the bilingual speakers show that their Cantonese also lacks PFC. More remarkably, out of the fifteen bilinguals tested, only one-third show PFC in all their English test sentences. These findings suggest that PFC is hard to transmit across languages through bilingualism. Moreover, the differential prosodic patterns among the bilingual speakers suggest that in the bilingual community, PFC may be subject to gradual loss.

Although tone-intonation relationship in SFPs and the acoustic correlates have previously been studied, most of the discussion lacked supporting evidence from phonetic experiments. The present study is distinguished in its systematic experimental design and detailed acoustic analyses, and it is hoped that the results will lay the foundations for future investigations into Cantonese phonetics.

# Acknowledgements

Writing this dissertation has been one of my most challenging experiences. I would have never been able to complete this task without the ongoing support and help of many people.

First and foremost, I would like to express my deepest gratitude to my primary supervisor Dr. Yi Xu for his excellent guidance, patience, insightful advice and continuous encouragement. With his support, I had the valuable opportunities to present papers in international conferences and teach in undergraduate lab classes at UCL, which helped prepare me for my academic career. It was my pleasure and privilege to learn and work with Dr. Xu, and I look forward to future collaborations with him.

I would like to express my heartfelt gratitude to my dissertation examiners, Prof. Anne Wichmann and Dr. Jill House, for their valuable time and insightful comments. Their questions and suggestions during the viva helped improve the dissertation. Of course, all remaining errors and omissions are my own.

Thanks also go to my secondary supervisors Drs. Jana Dankovičová and Jyrki Tuomainen, and my upgrade committee members Drs. Paul Iverson and Michael Ashby for their comments and support all along.

I will not forget the tremendous help from Mr. Steve Nevard, who guided me through the use of the Anechoic Chamber at UCL, Ms. Helen Wan, who recruited some of the informants for my research, and Ms. Lisa Chung, who shared her data files with me. To all the participants of my experiments I express my sincere gratitude.

My beloved wife Wai-Mun has been my constant inspiration and source of delight. I cannot thank her enough for standing by me during times of happiness and sorrow. Even though we could not see each other for many months during the course of my PhD studies, she was very understanding, and her phone and webcam calls were an important morale booster.

Many members of my family have been supportive in my pursuit of career. My parents and sister lent their full support, and my relatives at home and overseas also sent their blessings in various forms. I was very happy to share meals and chat with Uncles Jim and Mike, Marianne, Monica and Marielle in London.

During the period of staying in London, I received hospitality from Mr. and Mrs. Mohamad Blori and Mr. and Mrs. David Lim. Their homes provided me with a quiet environment which helped a lot with my studies. Mr. Albert Lee and Ms. Faith Chiu offered help on many occasions and I am very thankful to them. During my MA studies, I had the privilege of knowing and making friends with Prof. Yoshiki Nagase when he visited London, and I was happy to be able to meet him again during my PhD studies, and was honoured to have him and his wife, Prof. Emi Nagase, as guests at my wedding.

Once again I must thank Mrs. Rosie Chen and Prof. K. K. Luke, my MPhil supervisors at the University of Hong Kong, for making my start of linguistics and phonetics possible many years ago.

My home church, Truth Cornerstone Baptist Church, and the two churches I attended in London, London Chinese Lutheran Church and London Chinese Baptist Church, gave me support through prayers and gifts and I thank them for their kindness. I would like to thank Rev. and Mrs. Michael Leung, Rev. Ray Kwan, Pastors Flora Choi, Winnie Woo, Roger Ko and Michael Lo for their teaching and words of encouragement; members of the “Forever Young” group for their friendship and prayers; Shirley, Heather, Irene, Alan, Sandy and Jackie for their hard work in the bible study group and prayers for me; many brothers and sisters in Christ for their friendship; Pastor Bing Li in London for his preaching ministry.

I would also like to thank my students in various classes for their encouragement. It was my pleasure meeting them and I am honoured to be considered a friend by them. I wish all of them the best in their endeavours. In particular, I look forward to the graduation of HKBH PEN36 nursing students and HKPolyU BABS students. Their diligence in studies, dedication to life and steadfastness in the face of difficulty have been a source of inspiration to me.

I thank God for giving me the ability to complete this dissertation. As I begin a new chapter in my life, may God help me follow Jesus with all my heart, mind, soul and strength.

*Soli Deo Gloria*.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1 - Introduction

## 1.1 Background and aim of the dissertation

This dissertation studies two aspects related to the prosody of Hong Kong Cantonese, namely the phonetic features of sentence-final particles and prosodic focus. Cantonese, one of the most well-known Chinese dialects, has attracted a wealth of studies on its phonology. Its rich inventory of lexical tones renders it an important language for phonetic research on the interaction between tone and intonation. In a tone language, how lexical tone and sentence intonation can both use fundamental frequency ($F_0$) as their main acoustic cues has long been an intriguing question. Will the lexical tones, in order to be perceptually distinct, be resistant to pitch modification and so there cannot be an elaborate intonation system? If there is intonation besides tones, how much of the surface sentential $F_0$ is attributable to the tones and intonation? On a similar note, if $F_0$ is largely reserved for the lexical tones, can $F_0$ still be employed for the purpose of prosodic focus, as in many intonation languages? And if so, how similar and different is prosodic focus in a tone language compared to that in an intonation language? If not, will the lack of focus-related $F_0$ change be a disadvantage in terms of focus perception?

Although there have already been studies on tone-intonation relationship in Cantonese in general, one particular instance, namely the case of sentence-final particles, has yet to be explored in greater details and with systematic experimental designs and analyses. Similarly, the acoustic correlates of prosodic focus in Cantonese have previously been reported, but some important recent findings on cross-linguistic variability of post-focus compression (PFC) point to the necessity of in-depth experimental studies on Cantonese prosodic focus. Therefore, the main aim of this dissertation is two-fold. Firstly, it is hoped that the systematic experimental designs and analyses employed in the dissertation will provide valuable results which advance our knowledge of, and lay the foundations for future investigations into Cantonese phonetics. Secondly, by exploring and comparing the features of prosodic focus in Cantonese and other languages, it is hoped that important evidence can be provided which contributes to the discussion of cross-linguistic variability of PFC.

In this chapter, the Cantonese language and its tonal system will be introduced, followed by an introduction to SFPs and prosodic focus, and the less well studied areas concerning these two properties. The main research questions and hypotheses to be tested will then be proposed.

## 1.2 Cantonese and its phonology

### 1.2.1 The Cantonese language and its tonal system

Cantonese is spoken in south-east China, Hong Kong, Macau, and many overseas Chinese communities, with an estimated number of native speakers of over 54 million (Lewis, 2009). The variety of Cantonese studied in this dissertation is spoken in Hong Kong by over 6 million people. Hong Kong Cantonese, like other Chinese dialects, is a tonal language, with each syllable associated with a lexical tone. That is, differences in pitch height and contour of the syllable distinguish between different monosyllabic morphemes. The tonal system of Hong Kong Cantonese is distinguished in having six contrastive tones, which can be referred to by their tone numbers (Linguistic Society of Hong Kong, 2002) and descriptive names (Bauer & Benedict, 1997): Tone 1: High Level, Tone 2: High Rising, Tone 3: Mid Level, Tone 4: Mid-Low Falling, Tone 5: Mid-Low Rising, and Tone 6: Mid-Low Level. The lexical tones can also be alternatively referred to by their pitch contours, using the five-degree scale tone letters proposed by Chao (1930), with 5 representing the highest pitch and 1 the lowest pitch. The tone contour is represented by two numbers, the first being the starting pitch level and the second being the ending level. Table 1.1 gives the reported pitch levels of the tones in Hong Kong Cantonese with their iconic symbols as used in the IPA Handbook (Zee, 1999).

Table 1.1 The 6 contrastive lexical tones in Hong Kong Cantonese

| Tone numbers | Descriptive names | Pitch contours | IPA symbols | Example words |
|---|---|---|---|---|
| Tone 1 | High Level | 55 | ˥ | [ji˥] 衣 “clothes” |
| Tone 2 | High Rising | 25 | ˧˥ | [ji˧˥] 椅 “chair” |
| Tone 3 | Mid Level | 33 | ˧ | [ji˧] 意 “idea” |
| Tone 4 | Mid-Low Falling | 21 | ˨˩ | [ji˨˩] 疑 “doubt” |
| Tone 5 | Mid-Low Rising | 23 | ˨˧ | [ji˨˧] 耳 “ear” |
| Tone 6 | Mid-Low Level | 22 | ˨ | [ji˨] 二 “two” |

In some analyses of Cantonese tonal system, three additional traditional tone categories, called checked tones and numbered Tones 7, 8 and 9, are recognized. These are associated with closed syllables which have a plosive ([p], [t] and [k]) as the syllable coda. The three checked tones have the same pitch levels and contours as the three level tones Tone 1, Tone 3 and Tone 6 respectively, and so in this dissertation only six contrastive tones are to be distinguished.

Tone sandhi in Hong Kong Cantonese is impoverished despite its having a rich inventory of tones. Nevertheless, Hong Kong Cantonese does have other forms of tone changes, and the two notable ones are the diminutive tone change (Tan, 2011) and tone change due to contraction (Yip & Matthews, 2000). The diminutive tone change is the phenomenon that a lexical item whose last syllable has a tone which belongs to the *yang* category[1] may be changed to end in a High Rising tone (Tone 2). This occurs for words which are familiar items in daily lives or items of small sizes. For example, the word 人 "human being" has the citation form [jɐn˩], but it is pronounced with a High Rising tone in words such as "man" (男人 [nam˩ jɐn˥]) and "woman" (女人 [nɵy˨˩˧ jɐn˥]). Tone change in Cantonese also occurs in contracted forms in some grammatical constructions. For example, the word 一 [jɐt˥] "one" can be contracted in expressions having the construction "verb [jɐt˥] verb" which means doing the action for a short while; the expression 去一去 [hɵy˧ jɐt˥ hɵy˧], where 去 [hɵy˧] means "to go" and the whole expression means "to go for a while", can be pronounced as [hɵy˥ hɵy˧] where the word 一 [jɐt˥] is contracted and its meaning is conveyed by the changed tone.

[1] In traditional Chinese phonological analysis, tones are categorized into two groups, *yin* 陰 and *yang* 陽, and in each group there are four classes, in order, *ping* 平, *shang* 上, *qu* 去 and *ru* 入, which are sometimes translated as "Level", "Rising", "Leaving" and "Entering" respectively. For Cantonese, Tones 1, 2, 3 correspond to *yin ping*, *yin shang* and *yin qu*; Tones 4, 5, 6 correspond to *yang ping*, *yang shang* and *yang qu*; Tone 7 corresponds to upper *yin ru*; Tone 8 corresponds to lower *yin ru*; Tone 9 corresponds to *yang ru*.

### 1.2.2 Cantonese tone and intonation

Since the main acoustic correlate of both tone and intonation is fundamental frequency ($F_0$), how tone and intonation interact in a tone language has interested many scholars. Questions regarding this interaction include, if the $F_0$ of the tones is changed as a result of intonation, how may the lexical meanings be maintained? If the lexical tonal system has to be rigid, maintaining the overall patterns of $F_0$, what features of $F_0$ are at the disposal of the speaker for signalling intonation functions? Pike (1945, 1948) thinks that tone languages can only have limited intonation and their intonation patterns do not form a highly organized contrastive system. Furthermore, since Cantonese is rich in sentence-final particles (more discussion in the next section) which can serve the functions of intonation, a logical question to ask is whether intonation still plays a significant role in Cantonese. Fry (1968) thinks that as Cantonese uses a lot of particles for "syntactic purposes, the phonological tones are less modified by sentence intonation" (p. 368) than are in the case of intonation languages such as English, pitch differences in tone languages would be restricted in use for higher linguistic levels (that is, grammatical and affective levels). However, Chao (1933) points out that in Mandarin Chinese, the intonation and tone systems are distinct in both form and meaning, and that for every Chinese utterance, the $F_0$ changes on the syllables are the result of the interplay of three factors, namely the lexical tones, the influence of neighbouring syllables, and the expressive intonation. He suggests that the pitches on the syllables in an utterance are the algebraic sums of lexical tones and sentence intonation.

Is there an organized contrastive intonation system in Cantonese? Researches on

Cantonese intonation have not been many, but a number of studies have tried to define the forms of Cantonese intonation. The studies to date mainly focus on Cantonese intonation at the sentence level. The scholars who describe the forms of Cantonese sentential intonation usually divide the sentence types into four, namely declarative, interrogative, imperative and exclamatory, e.g., Mai (1998), Lam (2002). The general intonation pattern is that a declarative is produced with a falling intonation, and an interrogative a rising intonation. Mai (1998) divides Cantonese intonation into two components: sentence-segment intonation and sentence-final intonation. He uses the terminology "simultaneous addition" and "successive addition", proposed by Chao (1933), to describe the Cantonese intonation patterns: sentence-segment intonation is a kind of simultaneous addition, and sentence-final intonation is a type of successive intonation. For simultaneous addition, the $F_0$ range of the whole segment may be raised or lowered but the relative pitch level and pitch contour of each syllable is basically unchanged. For successive intonation, the tone of each syllable is followed immediately by the intonation. He identifies five patterns of sentence-segment intonation patterns (rising, falling, low, high, normal); for sentence-final intonation patterns, three patterns for sentences without final particles (rising, falling, unchanged), and five patterns for sentences with final particles (rising, falling, high, low, unchanged), are distinguished. His description is very elaborate but is not accompanied by experimental data.

S. Chow (2001) analyses experimentally the intonation of Cantonese sentences with and without SFPs. She shows that the lexical tones in the sentence are the basic building elements of the intonation; downdrift is a characteristic of the intonation of declarative sentences; and for other sentence types, the $F_0$ contours of the sentence-final syllables are the tokens which carry the intonation. Fox, Luke, & Nancarrow (2008), based on their acoustic studies, observe that intonation divides a Cantonese utterance into phrases, and in "neutral" Cantonese utterances (as opposed to "expressive" ones), the pitch of the first phrase starts at the middle pitch range, and declines along the phrase, and is raised in the next phrase to a level lower that the starting pitch of the previous phrase, and such pattern repeats.

As pointed out by Fry (1968), lexical tones "are likely to be modified by sentence intonation, but not totally subordinated to the latter, and the degree of subordination varies from language to language" (p. 368). What kind of modifications can intonation impose on tones in Cantonese? Kwok (1984) believes that modifications of tones may occur as a result of the superimposition of a rising or falling intonation, and such modifications are more prominent in the sentence-final position. In other words, for non-rising lexical tones, if the sentence intonation is rising, these tones will be made to rise, and for rising lexical tones, the rise in pitch will be much higher. She also points out that those tones in the middle pitch range would sound alike after modifications. As for the modification by a falling intonation, a rising tone will be modified into having a rise-fall contour.

S. Chow (2001), Fok-Chan (1974) and Vance (1976) have found that Cantonese, similar to other languages, shows a falling $F_0$ contour for statements, a rising $F_0$ contour for questions, and $F_0$ declination in statements. Vance (1976) reports that the relative pitch relationships among different lexical tones are preserved in different phonetic contexts. However, Lee (2004) and Ma, Ciocca & Whitehill (2006) have found that in questions all the Cantonese tonal contours are changed to rising contours. This is different from other languages where it is mainly the tonal height that is affected.

In essence, past studies on Cantonese intonation have shown that intonation in Cantonese affects the pitch range of the lexical tones, more noticeably in the final syllables, by raising or lowering it, and the overall pattern of the contrastive levels is well preserved. On the other hand, experimental studies on Cantonese intonation are limited, and much more has to be done in clarifying the phonetic details of intonation patterns and as well as their functions in communication.

## 1.3 Sentence-final particles (SFPs)

### 1.3.1 SFPs in general

SFPs are also referred to as "utterance-final particles", "utterance particles", "modal particles" and "final particles" in the literature, and are a group of morphemes which are believed to convey "emotive and/or epistemic nuances within a particular discourse context" (Li, 2006, p. 1). They are attached to the end of sentences, clauses, phrases and words and do not have any semantic content (Luke, 1990), and are described to have similar functions as intonation in intonation languages in conveying the speaker's attitudes and emotions, and serving grammatical functions such as indicating the sentence type. It is worth noting that similar sentence particles in some languages, e.g., some yes/no question particles in Yoruba (Bamgbose, 1966), occur in the sentence-initial position.

### 1.3.2 Cantonese SFPs

Although SFPs exist in many languages, e.g., Mandarin, Japanese, Thai, and Vietnamese, the SFPs in Cantonese are known for their huge number, diverse discourse functions and high frequency of occurrence. Owing to the prominence of Cantonese SFPs, a complete description of Cantonese grammar (e.g., Cheung, 2007; Matthews & Yip, 2011) will discuss the SFPs in detail. In fact, a substantial amount of research has been dedicated to analysing selected Cantonese SFPs in various dimensions, including their historical features and development (e.g., Cheung, 2009; W. M. Leung, 2010); syntactic and phonological properties (e.g., Law, 1990; Li, 2006); semantic contents and pragmatic functions (C. S. Leung, 1992; Fang, 2003; Gibbons, 1980; Kwok, 1984; Luke, 1990).

Cantonese SFPs can be divided into monosyllables (e.g., [a˧], [wɔ˧]), particle-clusters consisting of two syllables (e.g., [a˥ ma˧]) and those consisting of three syllables (e.g., [ka˧ tsa˧ pɔ˧]). There are more than thirty monosyllabic SFPs (Kwok, 1984; C. S. Leung, 1992) and as many as forty-five particle-clusters have been identified (C. S. Leung, 1992), but the most frequently occurring ones are monosyllabic. The SFPs may appear in different sentence types, like statements, questions and requests. Some of them have $F_0$ contours similar to what is expected of their intonational functions, for example, a question SFP may have rising $F_0$, while an assertion SFP may have level or falling $F_0$ (Kwok, 1984; Luke, 1990; Matthews & Yip, 2011). The following sentences illustrate briefly the kind of meanings SFPs in Cantonese can convey:

| | | | | |
|---|---|---|---|---|
| Sentence 1: | [ji˩ ka˥ | sam˥ | dim˧˥] | |
| | now | three | o'clock | |
| | "It's three o'clock now." | | | |
| Sentence 2: | [ji˩ ka˥ | sam˥ | dim˧˥ | **a**˩] |
| | now | three | o'clock | SFP |
| | "It's three o'clock now?" | | | |
| Sentence 3: | [ji˩ ka˥ | sam˥ | dim˧˥ | **wɔ**˨˧] |
| | now | three | o'clock | SFP |
| | "It's said that it's three o'clock now." | | | |
| Sentence 4: | [ji˩ ka˥ | sam˥ | dim˧˥ | **kʷa**˧] |
| | now | three | o'clock | SFP |
| | "I guess it's three o'clock now." | | | |

The sentences above, which differ only in the presence or absence of a different SFP, show that an SFP can change a statement into a question, or turn a statement into reported speech, or tell the listener that the speaker is unsure of the truth of the statement. If the SFP [a˩] is used with the statement [ji˩ ka˥ sam˥ dim˧˥] (Sentence 2), the sentence becomes a question showing that the speaker thinks it is a surprise that it is three o'clock now, or that the speaker is seeking confirmation that the time is correct. But if the SFP [wɔ˨˧ ] is used (Sentence 3), then the statement adds the overtones that the information about the current time is according to someone other than the speaker. And if [kʷa˧] is used (Sentence 4), the speaker explicitly tells the listener that he or she is not certain about the truth of the information (Kwok, 1984; C. S. Leung, 1992; Matthews & Yip, 2011).

### 1.3.3 Tonal characteristics of SFPs

It has long been noticed that the sentence-final location exhibits intonation patterns that carry important communicative information (Palmer, 1922). And the importance of sentence-final intonation has been reported for various languages (Beckman & Pierrehumbert, 1986; Bolinger, 1978; Eady & Cooper, 1986; House, 2005; Kohler, 2005; Lindsay & Ainsworth, 1985; Niebuhr, 2007; Prieto, Shih, & Nibert, 1996; Sadock & Zwicky, 1985). It has also been known that sentence-final intonation exists in tone languages as well, even though pitch is already used for carrying lexical tonal contrasts. In fact, much of the intonation research on tone languages has been focused on how tones are affected by sentence-final intonation (e.g., Connell, Hogan, & Rozsypal, 1983; Ho, 1976, 1977; Ma, Ciocca, & Whitehill, 2006; Shen, 1990; Vance, 1976). For tone languages, however, there is an intriguing issue that has yet to be fully addressed. That is, it is theoretically possible to achieve sentence-final intonation directly through tone, e.g., to use a lexically rising tone to indicate questions and a level or falling tone to indicate statements. This possibility is especially plausible when a morpheme seems to serve no other purpose but to carry sentential information, and this is the case with SFPs in Cantonese.

Tone-intonation interaction in the case of SFPs has been studied in some languages, and many of the studies have treated the SFPs as having inherent lexical tones. In Thai, a generalization can be made that the same SFP has a surface $F_0$ similar to a "falling tone" in statements and a "high tone" in questions, but exceptions do occur. For example, the SFP [kháp] in statements is always perceived to have a "high tone", whereas the SFP [lâ] has a “falling tone” in wh-questions. In view of this, Pittayaporn (2007) suggests that some SFPs in Thai are tonal while the others are toneless, and that for the tonal SFPs in Thai, lexical tones always override boundary tones. In contrast, in Northern Vietnamese, intonation seems to override lexical tones. Hạ (2010) reports that a falling or low-level pitch contour is used on affirmative or acknowledgement SFPs [ờ] and [ừ], consistent with their falling lexical tone contours, whereas the SFP [vâng], having a high-level tone contour, assumes a falling or low-level pitch when it appears in affirmative sentences. In addition, when these three SFPs appear in requests, a final rising/mid-level pitch contour is used. For Singaporean English, D. Chow (2010) proposes that since the pitch contours on SFPs are limited, intonational melody rather than lexical tone better accounts for the surface $F_0$ of the SFPs. He further suggests that the intonational effect on tones of the SFPs in Singaporean English gives rise to a form of tonogenesis where the resultant pitch contour is reinterpreted as a lexical tone associated with a semantic content.

Conventionally, SFPs in Cantonese are referred to as having lexical tones like other ordinary syllables in the language, some being identical in phonetic segments but different in tones, for example, [a˥], [a˧], and [a˩], and [wɔ˧], [wɔ˩], and [wɔ˨˧], and this kind of phonetic notation is used in the examples in this dissertation. One important observation is that among the six contrastive tones in Cantonese (Table 1.1), no SFPs are associated with the Mid-Low Level Tone (Fang, 2003). Most studies on Cantonese SFPs treat them as having lexical tones without reference to what exactly the surface $F_0$ may be like. In their analysis of Cantonese utterance-final intonation, Fox et al. (2008) treat the SFPs like other lexical syllables as having tones and observe the various effects of intonation on the SFPs, such as the preservation of the original tones of SFPs in “neutral” utterances, and the change of the tones into high-wide rising contours in “question” utterances. Li (2006) analyzes the Cantonese SFPs into smaller meaningful phonetic units, and ascribes different discourse functions to the initials, rhymes and tones of the SFPs. Law (1990) holds a different view towards the claim that all Cantonese SFPs have inherent tones. She analyzes the phonology of Cantonese SFPs using the autosegmental approach and proposes that several particles ([la], [ɛ], [a] and [lɔ]) do not have inherent tones but can combine with tonal particles to derive the surface forms which differ in degrees of strength, i.e., how strongly the speaker believes in the truth of the assertion of the statement that he or she is making. It has never been investigated systematically, however, whether the $F_0$ contours of SFPs are derived directly from lexical tones, purely intonational, or due to a combination of tone and intonation. Thus there is no consensus as to the tonal nature of Cantonese SFPs and how they are linked to intonational meanings.

## 1.4 <u>Prosodic focus</u>

### 1.4.1 Focus

In verbal communication there is often a need to emphasize part of an utterance so as to introduce changes in the common ground shared by speak and listener (Krifka, 2008; Lambrecht, 1994). Such an emphasis can be marked by morpho-syntactic means, such as lexical choice, word order, sentence structure, etc. (Lambrecht, 1994; Zerbian, Downing & Kügler, 2009), but it can also be prosodically marked, in which case it is variably referred to as contrastive stress, nuclear tone, nuclear accent or sentence stress (Chen, Robb, Gilbert, & Lerman, 2001; Cooper, Eady & Mueller, 1985; Crystal, 1969). In recent years there has emerged a common practice of referring to prosodic emphasis simply as focus (Cooper, et al., 1985; Fernald & Mazzie, 1991; Ishihara, 2001; Pell, 2001; Rump

& Collier, 1996), which will also be followed in the present dissertation. However, it is important to point out that the term ‘prosodic emphasis’ is not synonymous with ‘focus’, but is one of the latter’s possible manifestations.

Much of the phonetic research on focus has mainly concentrated on examining the focused component itself, and many acoustic correlates of focus have been reported (de Jong & Zawaydeh, 2002; O'Shaughnessy, 1979; Sluijter & van Heuven, 1996; Turk & Sawusch, 1997). Meanwhile, some studies have reported phonetic variations beyond the focused components. The traditional nuclear tone analysis has described relatively low pitch tails after the nuclear tone in British English (Crystal, 1969; O’Connor & Arnold, 1961; Palmer, 1922). Bruce (1982) observed greater declination after focus than before focus in Swedish. Cooper et al. (1985) reported that in American English post-focus $F_0$ peaks are dramatically lowered while pre-focus peaks are comparable to those of the neutral-focus sentences. They have also noted that sentence-final focus is not marked as clearly as initial and medial focus. This asymmetry is confirmed by later studies (Pell, 2001; Xu & Xu, 2005). Also consistent with this asymmetry is the finding that perceptual identification of final focus is not nearly as good as non-final focus in American English, Greek, Swedish, Finnish and Dutch (Botinis, Fourakis, & Gawronska, 1999; Mixdorff, 2004; Rump & Collier, 1996), because no post-focus $F_0$ is available at that location. Such prosodic asymmetry involving a lowering of $F_0$ in the post-focus components, which can be referred to as post-focus compression (PFC), has also been shown directly or indirectly in many other languages, including Indo-European, e.g., French (Dohen & Lœvenbruck, 2004), German (Féry & Kügler, 2008), Greek (Botinis et al., 1999), Altaic, e.g., Turkish

(Ipek, 2011), Japanese (Ishihara, 2002), and Korean (Lee & Xu, 2010), and Uralic, e.g., Finnish (Mixdorff, 2004).

### 1.4.2 Prosodic focus in tone languages and post-focus compression (PFC)

In tone languages, since $F_0$ variations are used for lexical distinction, there is the likelihood that they are not used to encode focus. However, experimental studies have demonstrated that in some tone languages in China, e.g., Beijing Mandarin (Jin, 1996; Liu & Xu, 2005; Xu, 1999; Xu, Chen, & Wang, 2012) and Nanchang (Wang, Wang, & Qadir, 2011), $F_0$ variations are an important means of realizing focus, and both $F_0$ and intensity show PFC as observed in the non-tone languages mentioned above.

However, there is also evidence that PFC is absent in other tone languages, some of which are closely related to Mandarin. Xu et al. (2011) examined focus realization in Beijing Mandarin, Taiwanese (also known as Southern Min) and Taiwan Mandarin (Mandarin spoken in Taiwan as one of the official languages). They found that PFC is present in Beijing Mandarin, as found previously (Jin, 1996; Xu, 1999), but absent in Taiwanese spoken by both Taiwanese monolinguals and Taiwanese-Mandarin bilinguals. More surprisingly, they found that PFC is absent also in Taiwan Mandarin, spoken by both Taiwan Mandarin monolinguals and Taiwanese-Mandarin bilinguals. PFC therefore seems to vary across even closely related tonal languages.

A number of other recent studies have found that PFC is also absent in many other languages that are either tonal or non-tonal, including Deang, Wa and Yi (Wang, et al., 2011), Hausa, Chichewa, Wolof, Northern Sotho and Buli (Zerbian, Genzel,

& Kügler, 2010). Thus lack of PFC does not seem to be a rare feature across languages.

### 1.4.3 Cross-linguistic variability of PFC

The finding of cross-linguistic variability of PFC, especially between closely related tone languages, raises a critical question. That is, given that PFC is absent in some Chinese languages/dialects and many other languages, how did it get into Beijing Mandarin and other PFC languages in the first place? Xu et al. (2012) proposed three hypotheses: (1) *local emergence* — PFC emerged locally in each language, possibly in connection with certain other properties of the language, such as lexical stress, lexical tone or non-prosodic means of marking focus; (2) *spreading* — PFC was spread into a language through contact with other PFC languages; and (3) *inheritance* — PFC was inherited from a protolanguage where PFC originated. Among these the *local emergence* hypothesis may be the most easily conceivable at the first glance. However, known cross-linguistic distribution patterns already show that PFC is present or absent in both tonal and non-tonal languages, and its existence also goes across languages with or without lexical stress (as reviewed in Xu et al., 2012). So, while local emergence can never be fully ruled out, the other two possibilities are also worth considering.

The spreading hypothesis seems quite plausible at least in the case of Beijing Mandarin, as it has historically been in close contact with Manchurian and Mongolian, both being Altaic languages. Although neither Manchurian nor Mongolian has yet been systematically examined for PFC, all the other Altaic languages examined so far show clear evidence of PFC (Wang et al., 2011; Ipek, 2011; Lee & Xu, 2010). Even given a plausible source, another crucial issue for the spreading hypothesis, however, is whether PFC is actually *spreadable* across languages via contact, for example, through bilingualism. This may at first seem to be a nonissue, as it is well established that the speech of a second language learner typically exhibits various properties of their first language, including syntactic (Kormos, 2006), segmental (Flege, 1995; Piske, MacKay, & Flege, 2001) and suprasegmental (Guion, Flege, Liu, & Yeni-Komshian, 2000; Huang & Jun, 2011; Lee, Guion, & Harada, 2006) features. Hence the properties of the first language are indeed often transferred into the second language. The case of Taiwan Mandarin reported by Xu et al. (2011) also shows that, if the lack of PFC is a property of Taiwanese, it could be considered as having been spread into Mandarin spoken as a second language, or, for many of the speakers in that study, a second native language. What has never been shown, however, is that the *presence* of PFC as a prosodic feature is also spreadable from one language to another through bilingualism.

The inheritance hypothesis, according to which all PFC languages are descendants of a common protolanguage, is perhaps the least plausible. For one thing it would imply that Mandarin is descended from a different protolanguage than Taiwanese, which would deviate from the widely held consensus that all Chinese languages/dialects are descendants of a common language (Bodman, 1980). For another, it would place Mandarin genetically closer to other PFC languages, such as English, Japanese, Persian, than to some other Chinese languages. Such an extreme hypothesis therefore should be considered only if there is clear evidence against both *local emergence* and *spreading*.

## 1.5 Main research questions and method

### 1.5.1 Tone-intonation interaction in SFPs

What is not yet clear is whether tone or intonation is the underlying source of these $F_0$ contours. It could be the case that, because lexical tone is obligatory in the language, even intonational meanings are conveyed directly through tones, as long as suitable contours are available. It could also be the case that because these SFPs are bearers of intonational meaning, their $F_0$ contours are purely intonational, bearing little resemblance to any of the lexical tones. A third possibility is that $F_0$ contours of the SFPs are a combination of tone and intonation, because tonal contours, despite their partial resemblance, cannot fully account for the detailed $F_0$ contours.

Production experiments have therefore been designed to specifically test which of the three possibilities is the most likely in Cantonese. Short sentences in Hong Kong Cantonese with ten different SFPs were recorded and detailed analyses of their $F_0$ contours, final $F_0$, final $F_0$ velocity and duration were performed. Details about the ten SFPs and the important findings will be presented in Chapter 2.

### 1.5.2 Post-focus compression and its transferability

As reviewed in Section 1.4, recent cross-linguistic comparison of the presence and absence of PFC has led to the search for the explanation as to why languages believed to be in the same language family have different behaviours in terms of post-focus $F_0$ behavour. Chapters 3 and 4 will be respectively devoted to the investigation into whether Cantonese has PFC or not and whether PFC is a transferable feature across languages in the case of English-Cantonese bilinguals, with the aim to address both the local emergence hypothesis and the spreading hypothesis (Section 1.4.3).

Regarding the local emergence hypothesis, our findings will help to establish whether Hong Kong Cantonese is another language like Taiwanese, i.e., being tonal and closely related to Mandarin but lacking PFC. Previous studies of Hong Kong Cantonese have suggested that duration is the most significant acoustic correlate of prosodic focus (Bauer, Cheung, Cheung, & Ng, 2004), and that $F_0$ tracings of sentences with focus do not show clear evidence of PFC (Gu & Lee, 2007). To further confirm the absence of PFC with certainty, we will systematically analyze prosodic focus realization in all the lexical tones in Hong Kong Cantonese using methods similar to those of Xu et al. (2012).

With regard to the spreading hypothesis, this study will investigate how prosodic focus is realized in English and Cantonese spoken by native English-Cantonese bilinguals living in England. The aim is to find out, assuming absence of PFC in Hong Kong Cantonese, whether PFC is transferred from English to Cantonese as spoken by English-Cantonese simultaneous bilinguals (as opposed to sequential bilinguals; Bhatia & Ritchie, 2004). This may shed light on the transferability of PFC between languages. In addition, we will also examine whether PFC is affected in any way in the English sentences produced by the English-Cantonese bilinguals.

In general, therefore, three hypotheses will be tested: (1) Prosodic focus is realized in Hong Kong Cantonese without PFC, (2) PFC can be transferred from English to Cantonese through simultaneous bilingualism, and (3) PFC is unaffected in the English of bilingual speakers.

# Chapter 2 - Cantonese SFPs

## 2.1 Rationale for the present study

A common problem with previous accounts of SFP in Cantonese and other languages is that they are not based on data obtained with systematic experimental control and detailed acoustic analysis. Much of the existing disagreement may therefore be due to the impossibility of comparing non-experimental data across the studies. This state of the art also makes it difficult for anyone to see a clear picture of how intonation and SFPs interact, or even to judge the relative importance of SFPs in these languages. To overcome this difficulty, the present study adopts an approach that combines systematic experimental control, direct $F_0$ contour comparison facilitated by time-normalization, and statistical analysis of acoustic measurements that best reflect underlying tonal targets based on previous research (Chen & Xu, 2006; Liu & Xu, 2005; Xu, 1997, 1999; Xu & Wang, 2009; Wang & Xu, 2011). Although the systematic experimental control may carry the risk of not guaranteeing full naturalness of the materials used (which is not unusual in phonetic studies in general), the approach has been shown to be effective in generating findings that are replicable and therefore generalizable (Xu, 2010, 2011).

## 2.2 The SFPs being tested

The overall strategy of the present study is to explore the nature of SFPs by phonetically comparing a limited set of monosyllabic SFPs in Hong Kong

Cantonese, [a˥, a˧, a˩, kɛ˧˥, kɛ˧, la˥, la˧, wɔ˧, wɔ˩, wɔ˩˧], to ordinary tonal syllables in the language with similar segmental compositions as controls. These ten monosyllabic SFPs contain all the five possible "lexical tones" out of the six HK Cantonese tones that SFPs are associated with (the Mid-Low Level Tone is not associated with any SFP, Fang, 2003). They are also among the SFPs which can form pairs or triplets that share the same segmental phonemes but differ in pitch contours (as can be seen in the phonetic transcriptions). Finally they are among the most frequently used SFPs in Hong Kong Cantonese[2], which not only guarantees their representativeness, but also ensures that native subjects would have common knowledge about their meanings as intended in the test sentences.

Among these SFPs, some can be used in both statements and questions, while others serve as statement SFPs or question SFPs only. For [a˧] and [kɛ˧˥], which may be used in statements or questions, only one of their functions was examined in this study (statement for the former, question for the latter) out of consideration for experimental control, as we would like to use the same carrier sentence for all the ten SFPs: the carrier cannot be combined with [a˧] to form a question without the inclusion of additional words, and the use of [kɛ˧˥] with the carrier is less natural semantically. The linguistic functions of the ten SFPs are summarized in Table 2.1 (based on Kwok, 1984; C. S. Leung, 1992; Luke, 1990; Matthews & Yip, 2011).

---

[2] The frequency of use of these SFPs as studied by the Hong Kong University Cantonese Corpus [University of Hong Kong, n.d], a 180,000-word corpus consisting of 52 spontaneous conversations and 42 radio programmes, is as follows (numbers represent the number of times the word appears in the corpus): a˥ 617, a˧ 3674, a˩ 384, kɛ˧˥ 352, kɛ˧ 183, la˥ 1578, la˧ 664, wɔ˧ 928, wɔ˩ no data, wɔ˩˧ no data. The most frequently occurring word is the verb hɐi˨, 5222 times, and the SFP a˧, 3674, comes in second.

Table 2.1 The 10 SFPs studied and their main linguistic functions.

| SFPs | Main functions and characteristics |
|---|---|
| [a˥] | statement - polite request/suggestion, lively/animated |
| [a˧] | statement - softening |
| [a˩] | question - surprised, suspicious, seeking confirmation |
| [kɛ˧˥] | question - surprised |
| [kɛ˧] | statement - affirmative, state of affairs |
| [la˥] | statement - request |
| [la˧] | statement - “current relevance”, action completed |
| [wɔ˧] | statement - reminder, noteworthiness |
| [wɔ˩] | statement - discovery |
| [wɔ˨˧] | statement - hearsay, reported speech |

Because of the phonetic nature of the study, it is not our aim, nor is it possible, to comprehensively examine the meaning-form relation in the SFPs. Nevertheless, because the discourse functions of the SFPs need be considered when we discuss the findings of the acoustic data, we shall briefly discuss how the ten chosen SFPs differ in functions. As aforementioned, for each SFP we will only consider the one function which is intended in the test sentences. Of the ten SFPs, two are question SFPs and the rest are statement SFPs. The question SFP [a˩] is used when the speaker wants to seek confirmation or clarification of information, and can suggest a sense of surprise, scepticism or disapproval (Matthews & Yip, 2011). The question SFP [kɛ˦˥] is often used in why-questions, and can be used with other wh-questions and A-not-A questions. Similar to [a˨˩], it conveys a sense of surprise or unexpectedness (Li, 2006). The statement SFPs [a˥] and [la˥] are similar in that they are used in imperative sentences expressing a suggestion. [a˥] is usually described as making the utterances sound more lively (Kwok, 1984; Law, 1990), and [la˥] "elicits the recipient's agreement on some proposed course of action" (Luke, 1990, p. 95). The statement SFP [a˧] is notable for its function as a softener, making the statement sound less abrupt (Matthews & Yip, 2011). The statement SFP [kɛ˧] is used for assertion of facts (Matthews & Yip, 2011), whereas the statement SFP [la˧] conveys "realization of a state" (Fung, 2000, p. 93), that is, there is a change of state of affairs that the listener may not be aware of. The statement SFP [wɔ˧] serves as a reminder, highlighting some noteworthy information that the speaker wants the listener to realize (Matthews & Yip, 2011). The statement SFP [wɔ˨˩] signals a sense of surprise due to sudden awareness of previously unknown or unexpected information (Li, 2006). Lastly, the statement SFP [wɔ˩˧], a hearsay SFP, is used when the speaker explicitly indicates that he or she is quoting someone else (Kwok, 1984; Luke, 1990; Matthews & Yip, 2011). It is clear from the above descriptions that the SFPs serve diverse discourse functions similar to intonation in conveying the speaker's attitudes and emotions, and some SFPs serve grammatical functions in changing a statement to a question with attitudinal overtones.

### 2.3 Main hypothesis

In Cantonese, besides the use of SFPs to indicate questions, a rising intonation superimposed on a statement is also used to form echo questions without SFPs (Matthews & Yip, 2011). Since some SFPs are question SFPs, this makes it possible for us to compare the $F_0$ contours of the questions with and without question SFPs to see if the SFP $F_0$ contours are purely tonal, purely intonational, or a combination of the two, as mentioned earlier. There is some evidence from previous research to suggest that the third possibility, namely, tone-intonation combination, is highly likely. It has been shown that in Thai (Abramson & Svastikula, 1983), Chengdu Chinese (Chang, 1958), Mandarin (Connell et al., 1983; Ho, 1976; Liu & Xu, 2005) and Hausa (Lindau, 1986) sentence final tone is affected by question intonation, but tonal distinctions remain largely intact. In the case of Mandarin, the neutral tone has often been regarded as toneless, and its surface $F_0$ is thought to be derived from the preceding tone. But Chen & Xu (2006) showed with detailed acoustic analysis that the neutral tone actually has a static mid target. Liu & Xu (2007b) further argued, based on comparisons of the High tone and the neutral tone in question intonation, that the manifestation of the intonation of sentence type in Mandarin "has to be achieved through the articulatory implementation of local tonal pitch target specified by the lexical tones" (p. 633), because the surface $F_0$ contours reflect the characteristics of both the intonational component and the tone, including the neutral tone. Liu & Xu (2005) also found that the prosodic contrast between question and statement was not restricted only to the sentence-final location, but rather distributed, somewhat unevenly, throughout the sentence, with a sharp increase in the differences toward the end of the sentence. Previous evidence therefore suggests that the possibility of SFPs in Cantonese being both tonal and intonational is worth exploring. Thus the main hypothesis to be tested in the present study is that the surface $F_0$ of each SFP is a combination of both the lexical tonal target and intonational variation. Meanwhile, the detailed acoustic analyses necessitated by the testing of this hypothesis will also provide precise characterizations of SFPs in Cantonese, which will be valuable in their own right. To test the main hypothesis, the study will try to answer three specific questions: (1) Is the $F_0$ of SFPs the same as that of the lexical tones? (2) If not, how do they differ? (3) Does the presence of an SFP have any effect on the sentence prosody? To answer these questions, we will examine the $F_0$ and duration of both SFPs and the disyllabic units preceding the SFPs throughout a sentence.

## 2.4 <u>Method</u>

### 2.4.1 Stimuli

The experiment was designed to compare the $F_0$ contours of the ten monosyllabic SFPs, introduced in Section 2.2, with those of their (near) homophonic monosyllables in both statement and question intonations (Table 2.2). The target sentences consist of seven syllables. The carrier part in each target sentence is [ kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ ] (嗰個地方叫三) meaning "That place is called [sam˥]". There are ten sets of target sentences (Sets A to J), corresponding to the ten SFPs under study, and each set contains three targets (Table 2.3): Target 1 is the carrier plus a final ordinary lexical syllable, which, when combined with the last syllable of the carrier, will form a word which is to be interpreted as a made-up place name, and the target sentence is a statement indicated by the use of a full stop at the end. Target 2 is the intonation question formed by Target 1, indicated by a question mark. Target 3 is the carrier plus the SFP. In each set, the ordinary lexical syllable was chosen such that it is either a homophone having identical segmental phonemes and lexical tone, or a near homophone having the same rhyme and lexical tone, as the SFP. Since not all SFPs have homophones which are meaningful or which can be written down, near homophones are needed in some cases. There was a precursor sentence to every target sentence, which helped to provide the subject with the context of the intended SFPs. Since some SFPs do not have standardized written forms and some SFPs may be used as either statement or question SFPs, the context helped to signal which SFP was to be used.

The total number of stimuli is:

$$10 \text{ (SFPs)} \times 3 \text{ (final syllables)} \times 5 \text{ (repetitions)} = 150.$$

Table 2.2 The (near) homophones of the 10 SFPs

| SFPs | Homophones or near homophones |
|---|---|
| [a˥] | [a˥] 丫 “bifurcation” |
| [a˧] | [a˧] 亞 “second” |
| [a˩] | [ŋa˩] 牙 “tooth” |
| [kɛ˧˥] | [tsɛ˧˥] 者 “person” |
| [kɛ˧] | [tsɛ˧] 借 “to borrow” |
| [la˥] | [la˥] 喇 “Lama” |

| | |
|---|---|
| [la˧] | [la˧] 罅 "gap" |
| [wɔ˧] | [kɔ˧] 個 *(a measure word)* |
| [wɔ˩] | [wɔ˩] 禾 "cereal plant" |
| [wɔ˨] | [ŋɔ˨] 我 *(first person singular pronoun)* |

Table 2.3 The precursor and target sentences for the 10 sets of final syllables.

| Set A | |
|---|---|
| (1) precursor<br>嗰 個 地方 叫 咩 名 啊？<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ a˧<br>that place call what name SFP<br>What's the name of that place? | (1) target (statement)<br>嗰個 地方 叫 三 丫。<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ sam˥ a˥<br>that place call three bifurcation<br>That place is called "Three-Bifurcation". |
| (2) precursor<br>話 你 知， 嗰個 地方 叫 三 丫。<br>wa˧ nei˨ tsi˥ kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ a˥<br>tell you know that place call three bifurcation<br>Let me tell you, that place is called "Three-Bifurcation". | (2) target (intonation question)<br>嗰個 地方 叫 三 丫？<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ sam˥ a˥<br>that place call three bifurcation<br>That place is called "Three-Bifurcation"? |
| (3) precursor<br>嗰個 地方 叫 咩 名 好 啊？<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ hou˧˥ a˧<br>that place call what name good SFP<br>What shall we call that place? | (3) target (with the target SFP)<br>嗰個 地方 叫 三 吖。<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ sam˥ a˥<br>that place call three SFP<br>Let's call that place "Three". |
| Set B | |
| (1) precursor<br>嗰 個 地方 叫 咩 名 啊？<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ a˧ | (1) target (statement)<br>嗰個 地方 叫 三 亞。<br>kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ a˧ |

| | |
|---|---|
| that place call what name SFP<br>What’s the name of that place? | that place call three second<br>That place is called “Three-Second”. |
| (2) precursor<br>話 你 知， 嗰個 地方 叫 三 亞。<br>wa˧ nei˩˧ tsi˥ kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ a˧<br>tell you know that place call three second<br>Let me tell you, that place is called “Three-Second”. | (2) target (intonation question)<br>嗰個 地方 叫 三 亞？<br>kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ a˧<br>that place call three second<br>That place is called “Three-Second”? |
| (3) precursor<br>嗰個 地方 叫 咩 名 話？<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ wa˧˥<br>that place call what name SFP<br>What did you say that place is called? | (3) target (with the target SFP)<br>嗰個 地方 叫 三 啊。<br>kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ a˧<br>that place call three SFP<br>That place is called “Three”. |
| Set C | |
| (1) precursor<br>嗰 個 地方 叫 咩 名 啊？<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ a˧<br>that place call what name SFP<br>What’s the name of that place? | (1) target (statement)<br>嗰個 地方 叫 三 牙。<br>kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ ŋa˩<br>that place call three tooth<br>That place is called “Three-Tooth”. |
| (2) precursor<br>話 你 知， 嗰個 地方 叫 三 牙。<br>wa˩ nei˩˧ tsi˥ kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu sam˥ ŋa˩<br>tell you know that place call three tooth<br>Let me tell you, that place is called “Three-Tooth”. | (2) target (intonation question)<br>嗰個 地方 叫 三 牙？<br>kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ ŋa˩<br>that place call three tooth<br>That place is called “Three-Tooth”? |
| (3) precursor<br>嗰個 地方 叫 三 啊。<br>kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ a˧<br>that place call three SFP<br>That place is called “Three”. | (3) target (with the target SFP)<br>嗰個 地方 叫 三 呀？<br>kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ a˩<br>that place call three SFP<br>That place is called “Three”? |
| Set D | |
| (1) precursor<br>嗰 個 地方 叫 咩 名 啊？<br>kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ a˧ | (1) target (statement)<br>嗰個 地方 叫 三 者。<br>kɔ˧˥ kɔ˧ tei˧ fɔŋ˥ kiu˧ sam˥ tsɛ˧˥ |

| | |
|---|---|
| that place call what name SFP<br>What's the name of that place? | that place call three person<br>That place is called "Three-Person". |
| (2) precursor<br>話 你 知，　嗰個　地方　叫　三　者。<br>wa˨ nei˩˧ tsi˥　kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥<br>tsɛ˧˥<br>tell you know　that　place　call three<br>person<br>Let me tell you, that place is called<br>"Three-Person". | (2) target (intonation question)<br>嗰個　地方　叫　三　者？<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥　kiu˧　sam˥　tsɛ˧˥<br>that　place　call　three　person<br>That place is called "Three-Person"? |
| (3) precursor<br>嗰個　地方　叫　三。<br>kɔ˧˥ kɔ˧　tei˨fɔŋ˥　kiu˧　sam˥<br>that　place　call　three<br>That place is called "Three". | (3) target (with the target SFP)<br>嗰個　地方　叫　三　嘅？<br>kɔ˧˥ kɔ˧　tei˨fɔŋ˥　kiu˧　sam˥　kɛ˧˥<br>that　place　call　three　SFP<br>Why is that place called "Three"? |
| Set E | |
| (1) precursor<br>嗰 個　地方　叫　咩　名　啊？<br>kɔ˧˥ kɔ˧　tei˨fɔŋ˥　kiu˧　mɛ˥ mɛŋ˧˥　a˧<br>that　place　call　what name　SFP<br>What's the name of that place? | (1) target (statement)<br>嗰個　地方　叫　三　借。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥　kiu˧　sam˥　tsɛ˧<br>that　place　call　three　borrow<br>That place is called "Three-Borrow". |
| (2) precursor<br>話 你 知，　嗰個　地方　叫　三　借。<br>wa˨ nei˩˧ tsi˥　kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥<br>tsɛ˧<br>tell you know　that　place　call three<br>borrow<br>Let me tell you, that place is called<br>"Three-Borrow". | (2) target (intonation question)<br>嗰個　地方　叫　三　借？<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥　kiu˧　sam˥　tsɛ˧<br>that　place　call　three　borrow<br>That place is called "Three-Borrow"? |
| (3) precursor<br>你　知道　嗰個　地方　□咩　啊？<br>nei˩˧　tsi˥ dou˧　kɔ˧˥ kɔ˧　tei˨fɔŋ˥ di˥ mɛ˥<br>a˧<br>you　know　that　place　what　SFP<br>What do you know about that place? | (3) target (with the target SFP)<br>嗰個　地方　叫　三　□。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥　kiu˧　sam˥　kɛ˧<br>that　place　call　three　SFP<br>That place is called "Three". |

| Set F | |
|---|---|
| (1) precursor<br>嗰 個　地方　叫　咩 名　啊？<br>kɔ˧˥ kɔ˧　tei˧fɔŋ˥　kiu˧　mɛ˥ mɛŋ˧˥　a˧<br>that　place　call　what name　SFP<br>What’s the name of that place? | (1) target (statement)<br>嗰個　地方　叫　三　喇。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥　kiu˧　sam˥　la˥<br>that　place　call　three　Lama<br>That place is called “Three-Lama”. |
| (2) precursor<br>話 你 知，　嗰個　地方　叫　三　喇。<br>wa˧ nei˩˧ tsi˥　kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ sam˥　la˥<br>tell you know　that　place　call three<br>Lama<br>Let me tell you, that place is called<br>“Three-Lama”. | (2) target (intonation question)<br>嗰個　地方　叫　三　喇？<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥　kiu˧　sam˥　la˥<br>that　place　call　three　Lama<br><br>That place is called “Three-Lama”? |
| (3) precursor<br>嗰個　地方　叫　咩　名　好<br>啊？<br>kɔ˧˥ kɔ˧　tei˧fɔŋ˥　kiu˧　mɛ˥　mɛŋ˧˥　hou˧˥<br>a˧<br>that　place　call　what　name　good<br>SFP<br>What shall we call that place? | (3) target (with the target SFP)<br>嗰個　地方　叫　三　啦<br><br>kɔ˧˥ kɔ˧　tei˧fɔŋ˥ kiu˧　sam˥　la˥<br><br>that　place　call　three　SFP<br><br>Let’s call that place “Three”. |
| Set G | |
| (1) precursor<br>嗰 個　地方　叫　咩 名　啊？<br>kɔ˧˥ kɔ˧　tei˧fɔŋ˥　kiu˧　mɛ˥ mɛŋ˧˥　a˧<br>that　place　call　what name　SFP<br>What’s the name of that place? | (1) target (statement)<br>嗰個　地方　叫　三　罅。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥　kiu˧　sam˥　la˧<br>that　place　call　three　gap<br>That place is called “Three-Gap”. |
| (2) precursor<br>話 你 知，　嗰個　地方　叫　三　罅。<br>wa˨ nei˩˧ tsi˥　kɔ˧˥ kɔ˧ tei˧fɔŋ˥ kiu˧ sam˥　la˧<br>tell you know　that　place　call three　gap<br>Let me tell you, that place is called<br>“Three-Gap”. | (2) target (intonation question)<br>嗰個　地方　叫　三　罅？<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥　kiu˧　sam˥　la˧<br>that　place　call　three　gap<br>That place is called “Three-Gap”? |
| (3) precursor<br>改　佢個　名　做　三　啦。 | (3) target (with the target SFP)<br>嗰個　地方　叫　三　o 罅。 |

| | |
|---|---|
| kɔi˧˥ kʰøy˩˧ kɔ˧ mɛŋ˧˥ tsou˨ sam˥ la˥<br>change its name be three SFP<br>Let’s call it “Three”. | kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ la˧<br>that place call three SFP<br>That place is already called “Three”. |
| Set H | |
| (1) precursor<br>嗰 個 地方 叫 咩 名 啊？<br>kɔ˧˥ kɔ˧ tei˨fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ a˧<br>that place call what name SFP<br>What’s the name of that place? | (1) target (statement)<br>嗰個 地方 叫 三 個。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ kɔ˧<br>that place call three piece<br>That place is called “Three-Piece”. |
| (2) precursor<br>話 你 知， 嗰個 地方 叫 三 個。<br>wa˨ nei˩˧ tsi˥ kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ kɔ˧<br>tell you know that place call three piece<br>Let me tell you, that place is called “Three-Piece”. | (2) target (intonation question)<br>嗰個 地方 叫 三 個？<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ kɔ˧<br>that place call three piece<br>That place is called “Three-Piece”? |
| (3) precursor<br>嗰個 地方 叫 四。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sei˧<br>that place call four<br>That place is called “Four”. | (3) target (with the target SFP)<br>嗰個 地方 叫 三 喎。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ wɔ˧<br>that place call three piece<br>That place is actually called “Three”. |
| Set I | |
| (1) precursor<br>嗰 個 地方 叫 咩 名 啊？<br>kɔ˧˥ kɔ˧ tei˨fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ a˧<br>that place call what name SFP<br>What’s the name of that place? | (1) target (statement)<br>嗰個 地方 叫 三 禾。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ wɔ˩<br>that place call three cereal<br>That place is called “Three-Cereal”. |
| (2) precursor<br>話 你 知， 嗰個 地方 叫 三 禾。<br>wa˨ nei˩˧ tsi˥ kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ wɔ˩<br>tell you know that place call three cereal<br>Let me tell you, that place is called “Three-Cereal”. | (2) target (intonation question)<br>嗰個 地方 叫 三 禾？<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ wɔ˩<br>that place call three cereal<br>That place is called “Three-Piece”? |

| | |
|---|---|
| (3) precursor<br>有 咩 咁 特別 啊？<br>jɐu˩˧ mɛ˥ kɐm˧ tɐk˨ pit˨ a˧<br>what is there so special SFP<br>What’s so special? | (3) target (with the target SFP)<br>嗰個 地方 叫 三 o和。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ wɔ˨˩<br>that place call three SFP<br>That place is called “Three”. |
| Set J | |
| (1) precursor<br>嗰 個 地方 叫 咩 名 啊？<br>kɔ˧˥ kɔ˧ tei˨fɔŋ˥ kiu˧ mɛ˥ mɛŋ˧˥ a˧<br>that place call what name SFP<br>What’s the name of that place? | (1) target (statement)<br>嗰個 地方 叫 三 我。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ ŋɔ˩˧<br>that place call three me<br>That place is called “Three-Me”. |
| (2) precursor<br>話 你 知， 嗰個 地方 叫 三 我。<br>wa˨ nei˩˧ tsi˥ kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥<br>ŋɔ˩˧<br>tell you know that place call three me<br>Let me tell you, that place is called<br>“Three-Me”. | (2) target (intonation question)<br>嗰個 地方 叫 三 我？<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ ŋɔ˩˧<br>that place call three me<br>That place is called “Three-Me”? |
| (3) precursor<br>佢 話 咩 話？<br>kʰøy˩˧ wa˨ mɛ˥ wa˧˥<br>he say what SFP<br>What did he say? | (3) target (with the target SFP)<br>嗰個 地方 叫 三 o禍。<br>kɔ˧˥ kɔ˧ tei˨ fɔŋ˥ kiu˧ sam˥ wɔ˩˧<br>that place call three SFP<br>He said that place is called “Three”. |

### 2.4.2 Subjects

Native Hong Kong Cantonese speakers born and raised in Hong Kong participated as subjects. They were two females and five males, aged 19-33. They were recruited from University College London, UK and were paid for their participation. They received primary and secondary education in Hong Kong and used Cantonese as their main spoken language in daily life. All of them were beginning the first semester of undergraduate or postgraduate studies in the UK, and had been away from Hong Kong for only a few months. Their fields of study are varied, including linguistics, psychology, arts, and law. None of them reported having any speech disorders.

### 2.4.3 Recording procedure

The recording was conducted in an anechoic chamber at University College London, UK. Before the start of the recording, the experimenter went through the list of precursors and test sentences, and information about the larger contexts in which the hypothetical dialogues happened, with each subject. For target sentences with SFPs, the subjects were briefed on which exact nuance of the SFP was to be used; this is necessary since there are no standard written forms for all the SFPs, and specifying which particular meaning to use is to minimize confusion and to get all the subjects to use the same SFPs. After the briefing, the subjects understood which SFPs were intended and did not find it difficult or unnatural to produce the target sentences. During the recording, the target sentences and their precursor sentences, presented in Chinese characters and punctuations, were shown on a computer screen one pair at a time in random order, and a different list of precursors and target sentences in different order was used for each subject. The subjects read aloud the sentences at a normal conversational rate, and were instructed not to pause in the middle of a sentence. Each subject recorded both the precursor and target sentences. If the subject made a mistake in producing the target sentence, he or she was asked by the experimenter to repeat the sentence. A sound level meter (Bruel & Kjaer 2231) was used as the microphone and the speech was recorded directly onto a computer hard disk using the software Adobe Audition with a sampling rate of 44.1 kHz. The recording session for each subject lasted for about 20 minutes.

### 2.4.4 Data processing

The acoustic analysis was done using a custom-written Praat (Boersma, 2001) script, which was an early version of ProsodyPro (Xu, 2005-2012). The script allowed manual rectification of the vocal pulse markings (e.g., missing cycles) generated by the autocorrelation algorithm in Praat, and displayed a TextGrid which allowed labelling of intervals of the voiced segments to be analyzed. The labelling in the TextGrid was done for the voiced part of each syllable. For syllables with initial voiceless consonants, the labelling onset was the first full vocal cycle of the vowel. For syllables with initial nasal or lateral consonants, the labelled onset was the first vocal cycle; in cases where the target word following the carrier word [sam˥] begins with a velar nasal [ŋ] (in test sentence sets 3 and 10), although the two consecutive nasals become partially geminated, there was an obvious spectral shift which helped delineate the boundary between the two nasals; for syllables with initial glides, the labelling onset was the point along the continuous voicing where amplitude is the lowest. The labelled offset for open syllables was the last full vocal cycle of the vowel, and that for closed syllables was the final nasal cycle of the nasal consonant or the last full vocal cycle of the vowel in cases where the final syllables are voiceless consonants.

The Praat script generated two types of output. The first is time-normalized continuous $F_0$ contours, which are used *only for graphical analysis*. For the present study, the time-normalized $F_0$ contours consisted of 10 $F_0$ points, equally spaced in the labelled interval for each of the syllables to be analysed. Time-normalization is an established method that makes it possible to average $F_0$

contours across repetitions as well as speakers (Chen & Xu, 2006; Chen & Gussenhoven, 2008; Gu & Lee, 2007; Liu & Xu, 2005; Ma et al., 2006; Potisuk, Gandour, & Harper, 1997; Wang & Xu, 2011; Xu, 1997, 1999; Xu & Wang, 2009)[3]. The second type of output is various measurements taken from raw $F_0$ contours (thus without loss of any timing information). These measurements were divided into those for the final syllable and those for the pre-final syllables.

For the final syllables, the following measurements were taken:

Final $F_0$ (Hz and semitones): the final fundamental frequency value of the last syllable of each test sentence, taken at 30 ms (from the raw $F_0$ contour) before the offset of the marked voice interval. The values in Hz were used for plotting $F_0$ contours for graphical analyses, and the values in semitones were used for statistical analyses.

Final $F_0$ velocity (semitone/s): the instantaneous rate of $F_0$ change obtained at 30 ms before the end of the final syllable, which is an indicator of the steepness of the $F_0$ contour. It is based on raw rather than time-normalized $F_0$ contours. The instantaneous velocity profile was computed with a two-point central differentiation algorithm ($F_0' = (F_{0i+1} - F_{0i-1}) / (t_{i+1} - t_{i-1})$) (Bahill, Kallman, & Lieberman, 1982), implemented in the Praat script, with a sampling rate of 100 Hz. Final $F_0$ velocity is taken as a sample point along the continuous velocity

[3] Although time-normalization may sound like some kind of destructive data reduction, it is in fact a sampling method no different from taking one, two or three data points from each tone or syllable, except usually at a much higher sampling rate (e.g. 10 instead of 3 points per tone). Also it is mainly used for graphical display, while statistical analyses can be always based on raw data. See description of ProsodyPro for more detailed explanations (Xu, 2005-2012).

profile.

Duration (milliseconds): time interval between the onset and offset labels, corresponding to the voiced section of each syllable. For Sets C, H and J, the different initial consonants between the SFPs and their near homophones (Table 2.2) will affect the genuineness of duration comparison, and so for these three sets of final syllables only the vowel portions were measured rather than the whole voiced section of the syllable. Duration measurements were needed to check for potential differences reflecting the functional meanings of specific SFPs, and for possible sequentially attached utterance-final edge tone for intonational purposes, as reported for Mandarin (Chao, 1968).

For the pre-final syllables, the following measurements were taken from each disyllabic unit:

Mean $F_0$ (Hz and semitones): the average of all the fundamental frequency values of the disyllabic unit. The values in Hz and semitones were used for graphical and statistical analyses respectively.

Mean duration (milliseconds): the average duration of the disyllabic unit

Since we would like to see whether there is any effect on the sentence prosody with the presence of an SFP, we compared the mean $F_0$ and duration of the part of the sentence preceding the final syllable across the three conditions (sentence without SFP, intonation question and sentence with SFP). Disyllabic units rather than the sentence as a whole or single syllables were chosen for analysis because we would like to be able to see generally if there is any difference in different parts of the sentence (initial, medial, final) under the three conditions. On the other hand, detailed syllable-by-syllable $F_0$ contour comparison was done by examining the time-normalized $F_0$ plots.

The measurements final $F_0$ and final $F_0$ velocity are motivated by previous research on Mandarin tones (Chen & Xu, 2006; Xu, 1997, 1999). It is found that tone production is synchronized with the syllable (Gandour, Siripong, & Sumalee, 1994 for Thai; Xu, 1998, 1999, 2001 for Mandarin; Gu & Lee, 2007, Wong, 2006a and Wong & Xu, 2007 for Cantonese), and there is evidence that there are two kinds of underlying pitch targets, static and dynamic, the former being specified in terms of $F_0$ height and the latter in terms of both height and slope (Xu & Wang, 2001 for Mandarin, and Wong, 2006b for Cantonese). As a result, an underlying tonal target is best approached near the end of the syllable in both Cantonese and Mandarin (Gu & Lee, 2007; Wong, 2006a; Xu, 1997, 1999), and therefore final $F_0$ and final $F_0$ velocity are the closest indicators of the underlying tonal target. The 30 ms adjustment is to avoid irregularities that often occur at the end of an utterance due to various strategies speakers use to terminate an utterance. Both of these measurements have been used in previous studies (Chen & Xu, 2006; Liu & Xu, 2007a; Xu & Xu, 2005).

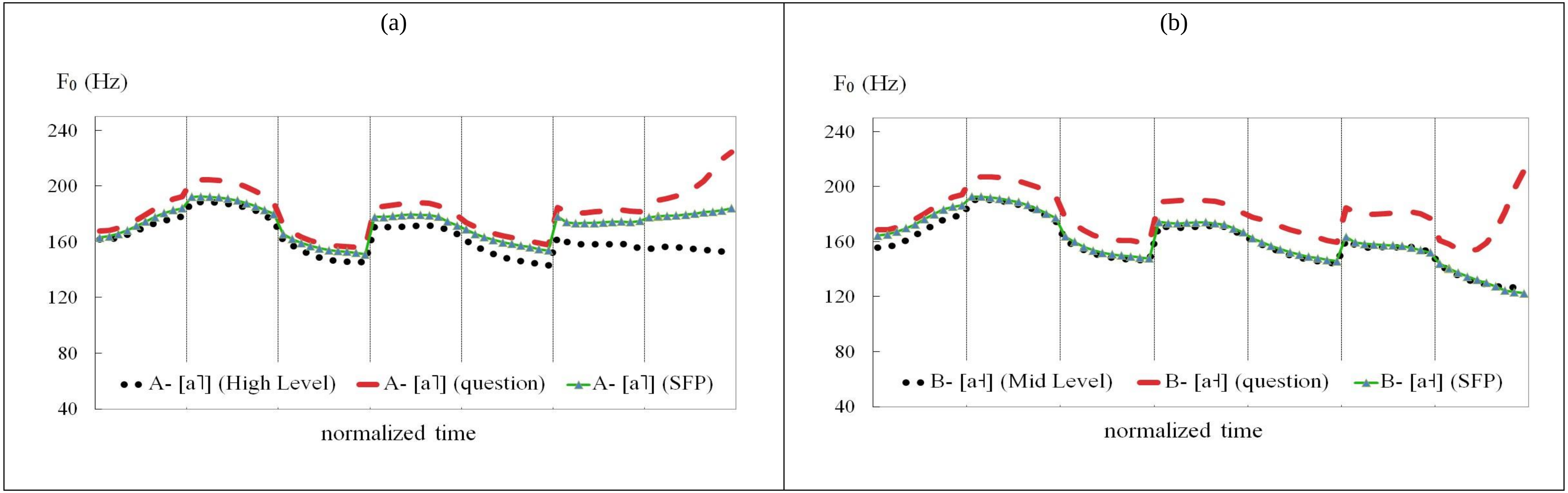
## 2.5 Analyses and Results

### 2.5.1 Final syllables

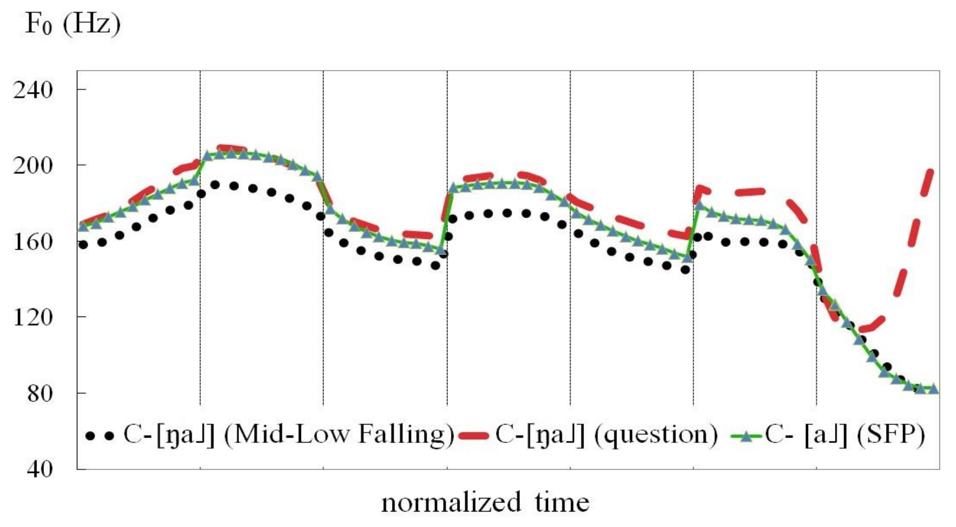
#### *2.5.1.1 Graphical analyses*

Figure 2.1 displays time-normalized $F_0$ contours of the ten sets of test sentences, with mean $F_0$ averaged across all 7 speakers plotted against normalized time. These plots allow quick inspection of the salient features of the $F_0$ contours. The averaging across speakers was done on a logarithmic scale, but $F_0$ was then converted back to Hz so as to make the plots readily comparable to most $F_0$ displays in publication which are typically in Hz.

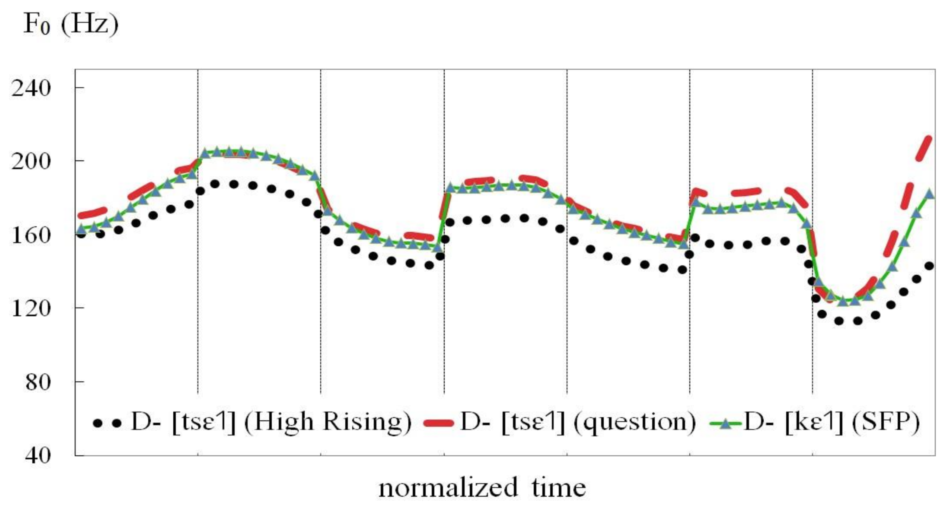
Figure 2.1 Time-normalized $F_0$ graphs averaged across all 7 speakers of the 10 sets of test sentences. Each data point of each $F_0$ curve was obtained by taking the antilogarithm of the average logarithmic $F_0$ values of the 7 speakers. The vertical lines indicate syllable boundaries. The time gaps corresponding to the voiceless consonants are not shown in the plots.



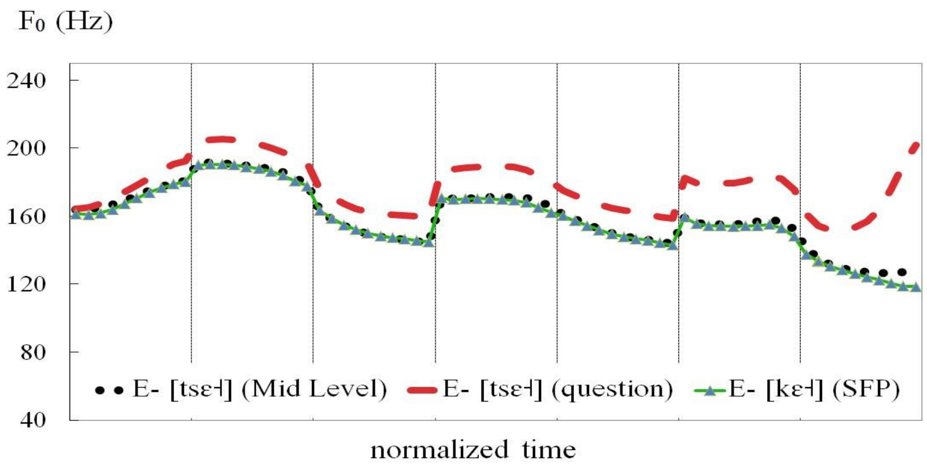
(c)


F0 (Hz)
240
200
160
120
80
40
C-[ŋa˩] (Mid-Low Falling)
C-[ŋa˩] (question)
C- [a˩] (SFP)
normalized time

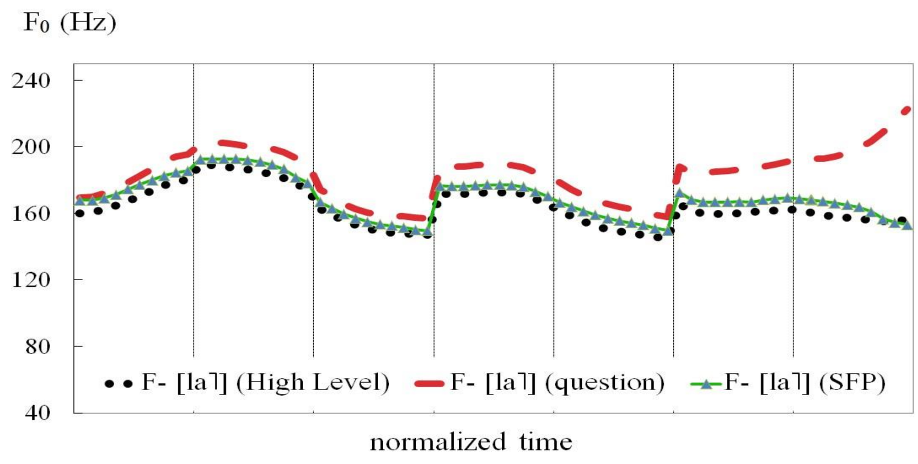
(d)


F0 (Hz)
240
200
160
120
80
40
D- [tsɛ˥] (High Rising)
D- [tsɛ˥] (question)
D- [kɛ˥] (SFP)
normalized time

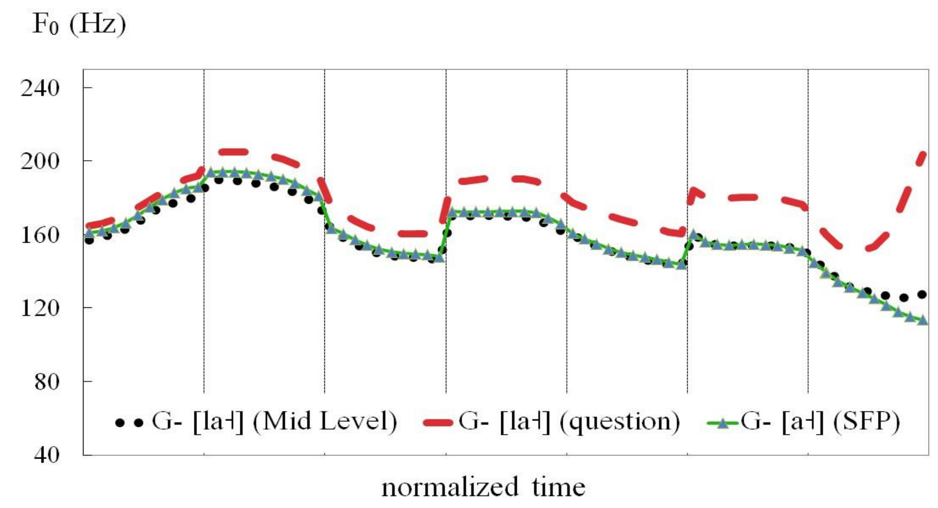
(e)


F0 (Hz)
240
200
160
120
80
40
E- [tsɛ˧] (Mid Level)
E- [tsɛ˧] (question)
E- [kɛ˧] (SFP)
normalized time

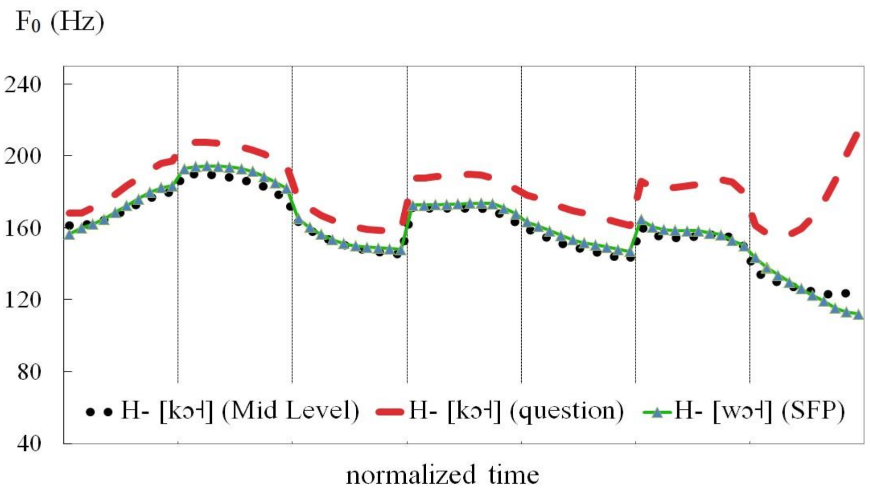
(f)


F0 (Hz)
240
200
160
120
80
40
F- [la˥] (High Level)
F- [la˥] (question)
F- [la˥] (SFP)
normalized time

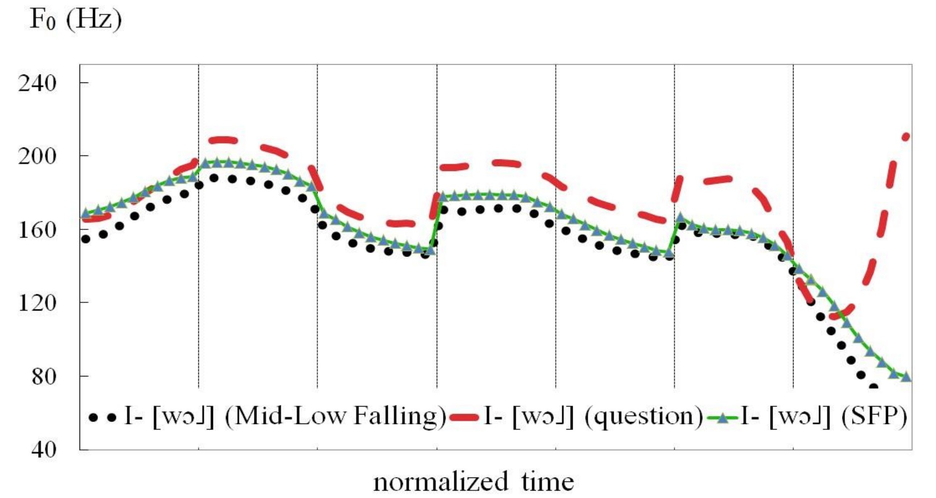
(g)


F0 (Hz)
240
200
160
120
80
40
G- [la˧] (Mid Level)
G- [la˧] (question)
G- [a˧] (SFP)
normalized time

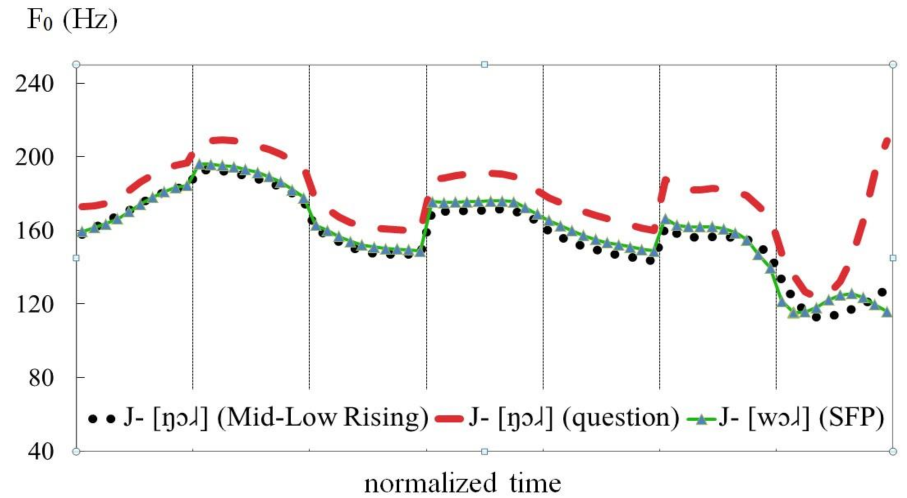
(h)


F0 (Hz)
240
200
160
120
80
40
H- [kɔ˧] (Mid Level)
H- [kɔ˧] (question)
H- [wɔ˧] (SFP)
normalized time

(i)


F0 (Hz)
240
200
160
120
80
40
I- [wɔ˩] (Mid-Low Falling)
I- [wɔ˩] (question)
I- [wɔ˩] (SFP)
normalized time

(j)


F0 (Hz)
240
200
160
120
80
40
J- [ŋɔ˩˧] (Mid-Low Rising)
J- [ŋɔ˩˧] (question)
J- [wɔ˩] (SFP)
normalized time

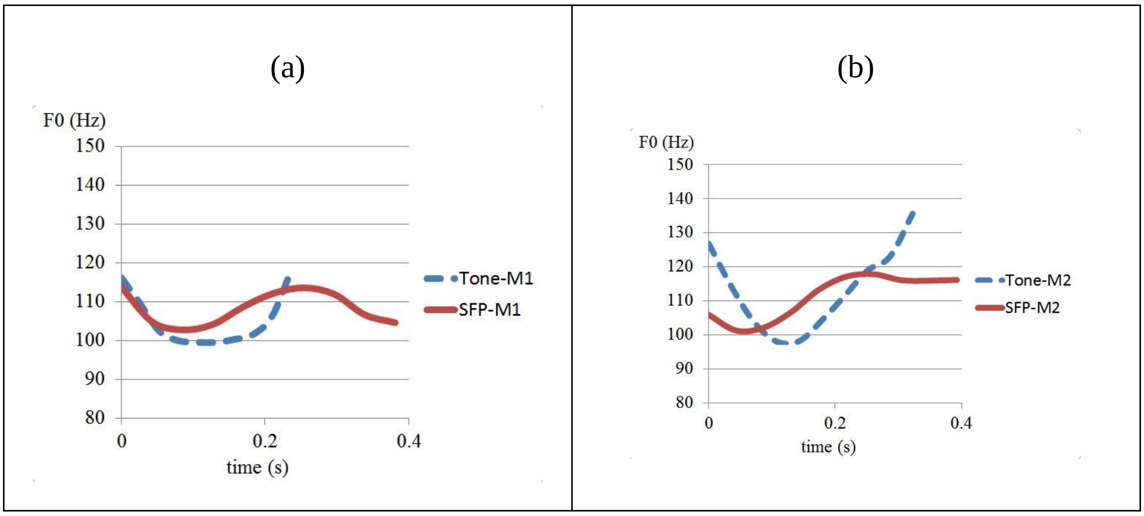
In the following discussion, when the final syllables are discussed, the sentence-final ordinary lexical syllables in statements are referred to as Tones-in-statement, and the sentence-final ordinary lexical syllables in intonation questions are referred to as Tones-in-question. The statistical analyses were performed using the measurements taken from raw $F_0$ contours without time normalization, as explained earlier.

From Figure 2.1, it is obvious that the $F_0$ contours of Tones-in-question (dashed lines) are very different from those of the SFPs (solid lines with triangles) and Tones-in-statement (dotted lines): they show steep rises in $F_0$, which agree well with previously reported auditory impressions (Matthews & Yip, 2011) and acoustic data (S. Chow, 2001; Ma et al., 2006). In each test sentence, the penultimate syllable was a High Level tone. The High Level tone-in-question shows an immediate rise at the beginning of the syllable (Figures 2.1a and 2.1f), whereas Tones-in-question with the other tones show a dip followed by a sharp rise. The only case where the three final syllables have very similar $F_0$ tracings is Set D sentences, containing the High Rising tone and the SFP [kɛ˧˥] (Figure 2.1d). The question SFP [kɛ˧˥] and the High Rising Tone-in-statement have similar rising $F_0$ contours, but the SFP has higher $F_0$ values over the whole course of the contour, although not as high as those of Tone-in-question in the same test sentence set.

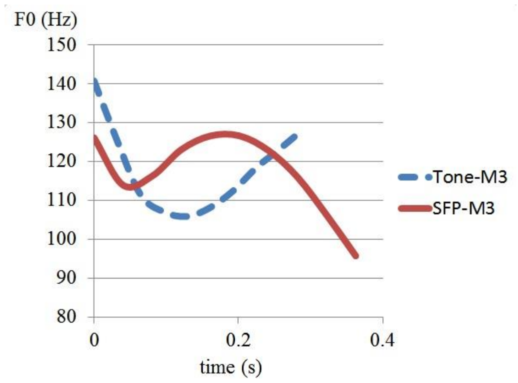
Comparing the $F_0$ contours of the SFPs and Tones-in-statement, which is the primary concern of the present study, the general observation is that in most cases the SFPs are similar to Tones-in-statement. But most importantly there are noticeable differences in the cases of the SFPs [a˥] (Figure 2.1a) and [wɔ˨˩]

(Figure 2.1j). The statement SFP [a˥] has a slightly rising contour and is much higher in $F_0$ values than the High Level Tone-in-statement, which is level in contour (Figure 2.1a). The statement SFP [wɔʎ] has a convex contour (Figure 2.1j), which is very different from the Mid-Low Rising Tone-in-statement. To supplement our analysis, the average $F_0$ contours of the SFP [wɔʎ] and the Mid-Low Rising Tone-in-statement of the 7 speakers using *actual time* data were plotted and are shown in Figure 2.2. We can see clearly that for all the speakers, the $F_0$ of the SFP consists of an initial low-rise, which resembles that of the Mid-Low Rising Tone-in-statement except for the higher $F_0$ values (with the exception of speaker M5, Figure 2.2e), and a final fall. Also, the duration of the initial low-rise of [wɔʎ] is comparable to that of the Mid-Low Rising Tone-in-statement.

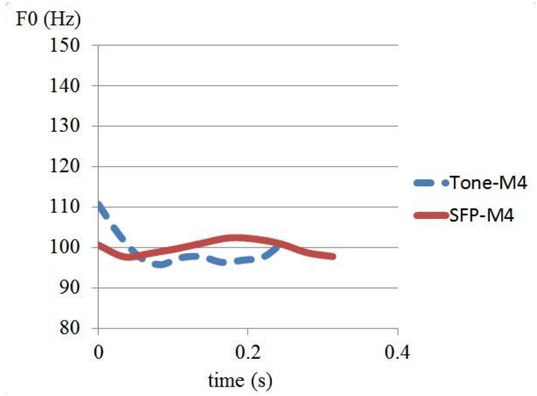
Figure 2.2 $F_0$ plots in real time of the SFP [wɔʎ] and low-rising (Tone 5) tone-in-statement [ŋɔʎ] for all 7 speakers. Each plot is an average over the 5 repetitions by each speaker. Figures (a) to (e) belong to the male speakers; (f) and (g) belong to the female speakers. Different y-axes are used for the two groups to show the $F_0$ contours more clearly.



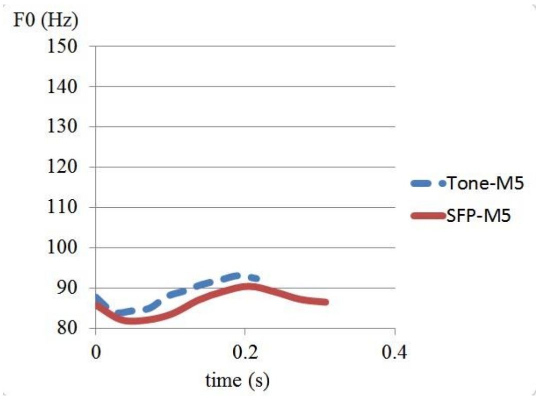
(c)


F0 (Hz)
150
140
130
120
110
100
90
80
0
0.2
0.4
time (s)
Tone-M3
SFP-M3

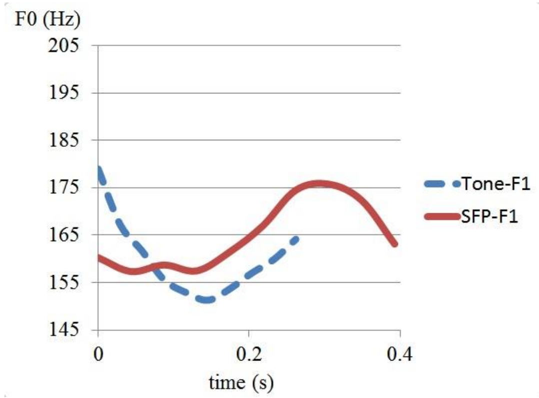
(d)


F0 (Hz)
150
140
130
120
110
100
90
80
0
0.2
0.4
time (s)
Tone-M4
SFP-M4

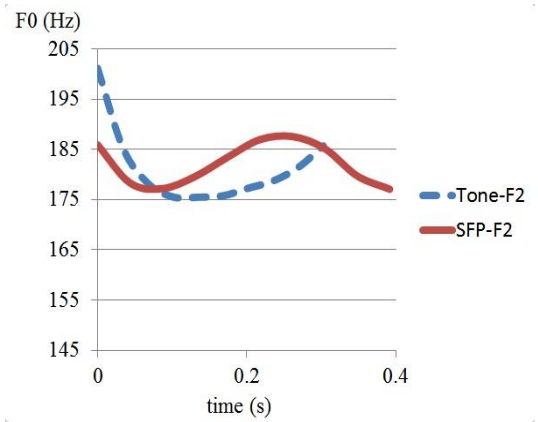
(e)


F0 (Hz)
150
140
130
120
110
100
90
80
0
0.2
0.4
time (s)
Tone-M5
SFP-M5

(f)


F0 (Hz)
205
195
185
175
165
155
145
0
0.2
0.4
time (s)
Tone-F1
SFP-F1

(g)

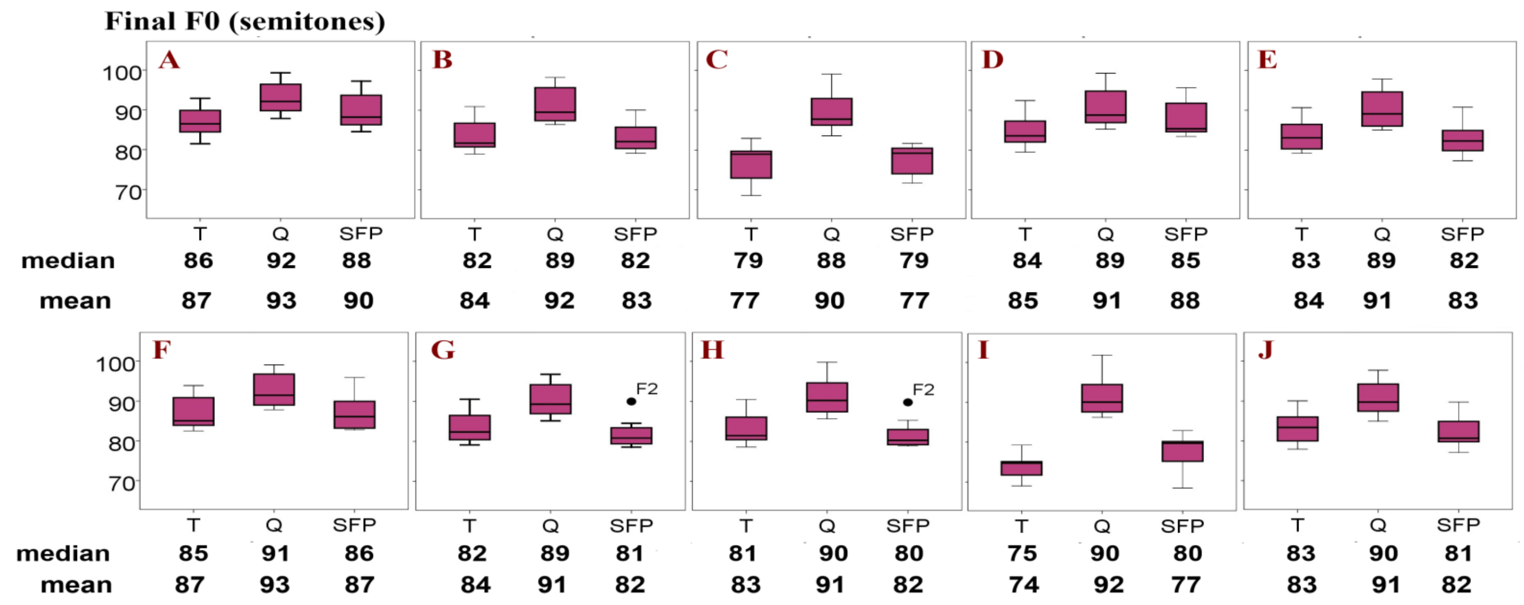

F0 (Hz)
205
195
185
175
165
155
145
0
0.2
0.4
time (s)
Tone-F2
SFP-F2

### *2.5.1.2 Statistical analyses*

To verify the differences observed in the mean $F_0$ tracings, one-way repeated measures ANOVAs were performed on final $F_0$ and final $F_0$ velocity of the three groups of final syllables (Tone-in-statement, Tone-in-question, SFP). The boxplots showing the $F_0$ measurements are displayed in Figure 2.3, with the median and mean values shown under each boxplot, and the F and *p* values of the ANOVAs are shown in Table 2.4.

Figure 2.3 Boxplots of final $F_0$ and final $F_0$ velocity of the final syllables for all 7 speakers. The ten sets of final syllables are labelled A to J, and in each set S refers to Tone-in-statement, and Q refers to Tone-in-question. For final $F_0$, the same Y-axis is used for all boxplots; for final $F_0$ velocity, different scales of Y-axis are used so that the plots are shown more clearly. The median and mean values are given under each boxplot.

(a)



(b)

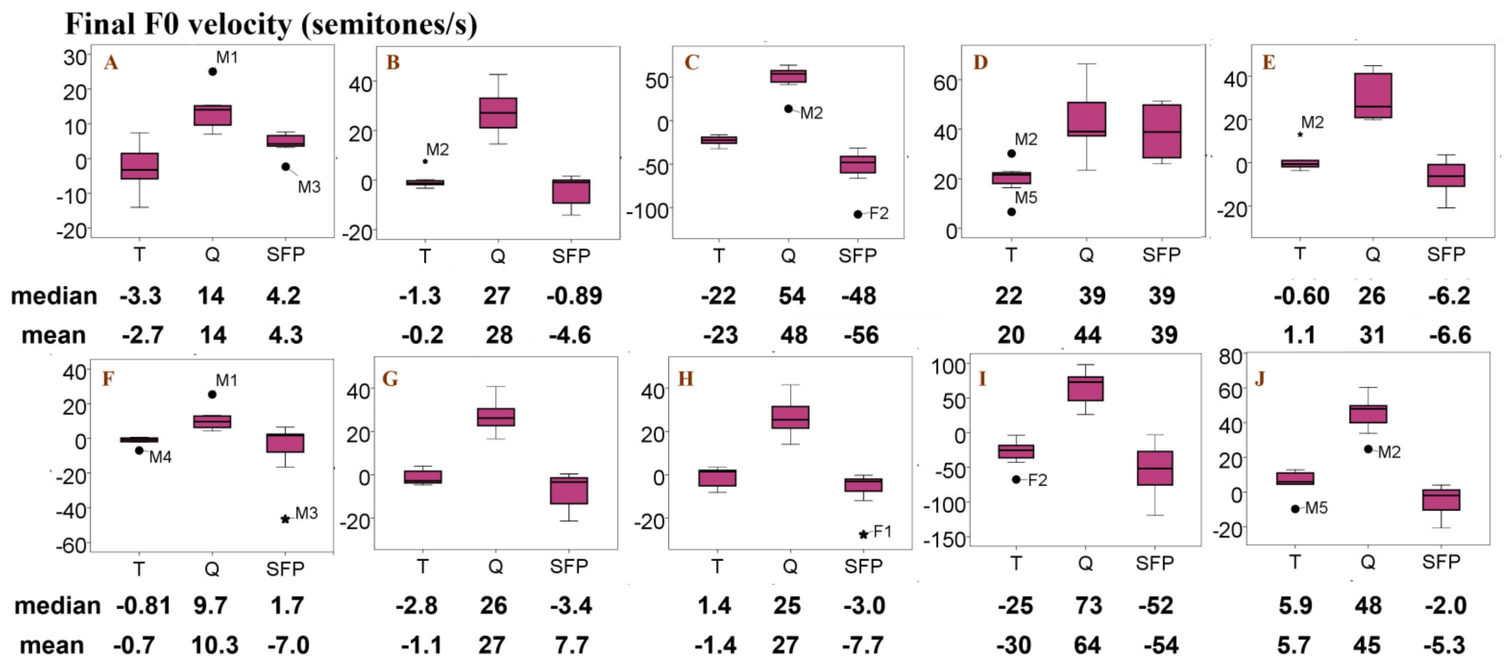

Final F0 velocity (semitones/s)
A
B
C
D
E
T
Q
SFP
M1
M2
M3
M5
F2
median -3.3 14 4.2 -1.3 27 -0.89 -22 54 -48 22 39 39 -0.60 26 -6.2
mean -2.7 14 4.3 -0.2 28 -4.6 -23 48 -56 20 44 39 1.1 31 -6.6
F
G
H
I
J
M4
F1
median -0.81 9.7 1.7 -2.8 26 -3.4 1.4 25 -3.0 -25 73 -52 5.9 48 -2.0
mean -0.7 10.3 -7.0 -1.1 27 7.7 -1.4 27 -7.7 -30 64 -54 5.7 45 -5.3

Table 2.4 One-way repeated measures ANOVAs on final $F_0$ and final $F_0$ velocity. Significant *p* values are presented in boldface.

| Sentence set | SFP | Tone-in-statement / Tone-in-question | Final $F_0$ | | Final $F_0$ velocity | |
|---|---|---|---|---|---|---|
| | | | F(2,12) | *p* | F(2,12) | *p* |
| A | [a˥] | [a˥] | 57.2 | ***<0.001*** | 15.9 | ***<0.001*** |
| B | [a˧] | [a˧] | 70.5 | ***<0.001*** | 44.8 | ***<0.001*** |
| C | [a˩] | [ŋ˩] | 37.7 | ***<0.001*** | 54.0 | ***<0.001*** |
| D | [kɛ˧˥] | [tsɛ˧˥] | 43.5 | ***<0.001*** | 8.1 | ***<0.001*** |
| E | [kɛ˧] | [tsɛ˧] | 86.8 | ***<0.001*** | 33.7 | ***<0.001*** |
| F | [la˥] | [la˥] | 26.3 | ***<0.001*** | 5.0 | ***<0.001*** |
| G | [la˧] | [la˧] | 58.9 | ***<0.001*** | 51.1 | ***<0.001*** |
| H | [wɔ˧] | [kɔ˧] | 62.4 | ***<0.001*** | 27.1 | ***<0.001*** |
| I | [wɔ˩] | [wɔ˩] | 31.0 | ***<0.001*** | 38.4 | ***<0.001*** |
| J | [wɔ˩˧] | [wɔ˩˧] | 92.0 | ***<0.001*** | 40.2 | ***<0.001*** |

From Figure 2.3, it can be seen that for final $F_0$, Tones-in-question have higher values than both Tones-in-statement and SFPs; Tones-in-statement and SFPs seem to differ less in most sentence sets. For final $F_0$, Tones-in-statement show very little individual variation, but SFPs show more individual variation. As seen in Table 2.4, in all ten cases the three types of final syllables are statistically different from each other with respect to final $F_0$ and final $F_0$ velocity. This is consistent with what we see in Figure 2.1 that Tones-in-question have the distinctive rise in $F_0$ contour, which is different from both Tones-in-statement and

SFPs except the SFP [kɛ˧˥] (Figure 2.1d). Since the one-way repeated ANOVAs only tell us whether in each case there is a difference among the three final syllables (Tone-in-statement, Tone-in-question, and SFP), post hoc tests (Bonferroni) were done to check which pairs (SFP vs. Tone-in-statement, Tone-in-statement vs. Tone-in-question, SFP vs. Tone-in-question) were statistically different (Table 2.5; *p* values smaller than $<0.05$ are taken as significant, and this same threshold is used for all the statistical tests in this dissertation). We will begin with the comparisons involving Tones-in-question. For Tone-in-question vs. Tone-in-statement and Tone-in-question vs. SFP comparisons, the post hoc test results show that final $F_0$ of Tones-in-question is significantly higher than that of Tone-in-statement and SFP in all cases (Table 2.5a), and final $F_0$ velocity of Tone-in-question is significantly greater (that is, having a more positive slope) than that of Tone-in-statement and SFP in all but three cases (Table 2.5b, both Tone-in-statement [tsɛ˧˥] and SFP [kɛ˧˥] of Set D, Figure 2.1d; and SFP [la˥] of Set F, Figure 2.1f).

Table 2.5 Post hoc test (Bonferroni) results on the final syllables comparing the final $F_0$ and final $F_0$ velocity of Tone-in-question vs. Tone-in-statement, and Tone-in-question vs. SFP. Significant *p* values ($<0.05$) are presented in boldface.

(a) Post hoc tests on final $F_0$ (in semitones).

| Sentence set | Mean difference of tone-in-question - tone-in-statement | *p* | Mean difference of tone-in-question - SFP | *p* |
|---|---|---|---|---|
| A | 6.020 | ***<0.001*** | 3.125 | ***0.001*** |
| B | 7.630 | ***<0.001*** | 8.072 | ***0.001*** |
| C | 13.276 | ***<0.001*** | 12.440 | ***0.004*** |

| | | | | |
|---|---|---|---|---|
| D | 6.088 | ***0.001*** | 2.790 | ***0.014*** |
| E | 6.665 | ***<0.001*** | 7.596 | ***<0.001*** |
| F | 5.550 | ***<0.001*** | 5.545 | ***0.009*** |
| G | 6.799 | ***<0.001*** | 8.339 | ***0.001*** |
| H | 8.025 | ***<0.001*** | 9.497 | ***0.001*** |
| I | 17.919 | ***0.002*** | 14.336 | ***0.008*** |
| J | 7.509 | ***<0.001*** | 8.422 | ***<0.001*** |

(b) Post hoc tests on final $F_0$ velocity (in semitones/s).

| Sentence set | Mean difference of Question - Statement | *p* | Mean difference of Question - SFP | *p* |
|---|---|---|---|---|
| A | 16.372 | ***0.006*** | 9.413 | ***0.027*** |
| B | 27.691 | ***0.002*** | 32.075 | ***0.001*** |
| C | 71.038 | ***<0.001*** | 103.764 | ***<0.001*** |
| D | 23.664 | 0.065 | 4.596 | 1.000 |
| E | 29.590 | ***0.004*** | 37.286 | ***0.002*** |
| F | 12.597 | ***0.011*** | 18.146 | 0.106 |
| G | 28.172 | ***<0.001*** | 34.766 | ***0.001*** |
| H | 28.043 | ***0.003*** | 33.825 | ***0.006*** |
| I | 93.698 | ***<0.001*** | 118.440 | ***0.002*** |
| J | 38.949 | ***0.001*** | 49.947 | ***0.001*** |

The post hoc results of $F_0$ measurements show that Tones-in-question have different $F_0$ contours than Tones-in-statement and SFPs. As far as the non-significant cases of final $F_0$ velocity are concerned, since the High Rising tone [tsɛ˧˥] and SFP [kɛ˧˥] have similar rising $F_0$ contours in statement as their counterparts in question (Figure 2.1d), it is not surprising that their final $F_0$ velocity values are similar; for the SFP [la˥], the reason why the final $F_0$ velocity does not show statistical difference from Tone-in-question despite the obvious difference observed in graphical analysis (Figure 2.1f) became evident when we examined the individual data: there is individual variation for this SFP, with some speakers producing a falling $F_0$ contour while others producing a level contour.

The use of falling intonation by some speakers for the SFP [la˥] (Set F) was noticed during the recording, and it was felt by author (who is a native speaker of Hong Kong Cantonese) as a natural and valid way of pronouncing the sentence, which suggests that both contours are free variants of this SFP.

For SFP vs. Tone-in-statement comparisons, which are the main concern in the present study, the post hoc results for the $F_0$ measurements (final $F_0$ and final $F_0$ velocity) are presented in Table 2.6.

Table 2.6 Post hoc test (Bonferroni) results on final $F_0$ and final $F_0$ velocity of the final syllables comparing the SFPs and Tones-in-statement. Only the statistically significant cases are presented.

(a) Post hoc tests on final $F_0$ (in semitones).

| Sentence set | SFP | Tone-in-statement | Mean difference of SFP - Tone-in-statement | *p* |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| A | [a˥] | [a˥] | 2.895 | ***0.016*** |
| D | [kɛ˧˥] | [tsɛ˧˥] | 3.298 | ***0.001*** |
| I | [wɔ˨˩] | [wɔ˨˩] | 3.583 | ***0.041*** |

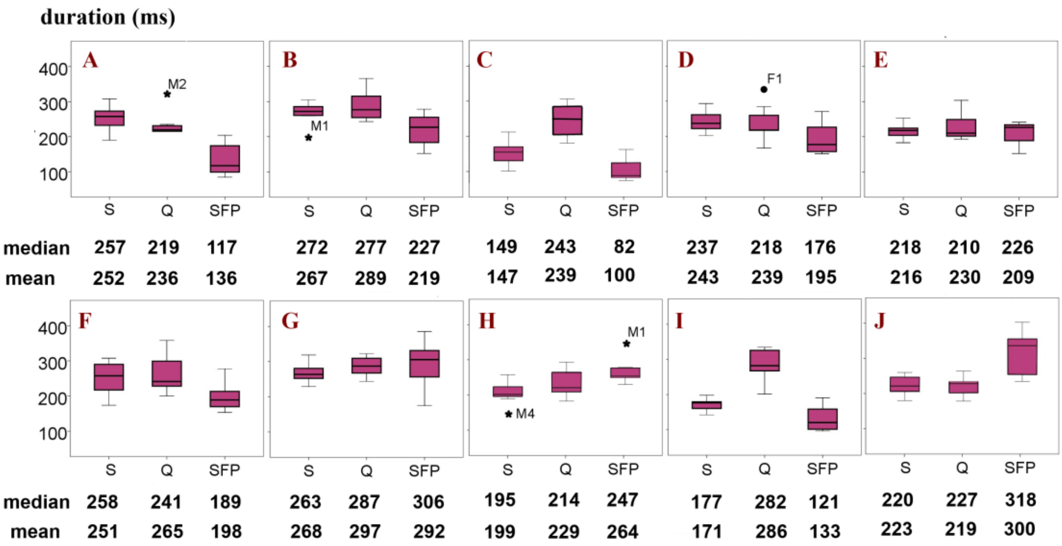
(b) Post hoc tests on final $F_0$ velocity (in semitones/second).

| Sentence set | SFP | Tone-in-statement | Mean difference of SFP - Tone-in-statement | *p* |
|---|---|---|---|---|
| D | [kɛ˧˥] | [tsɛ˧˥] | 19.068 | ***0.044*** |

With respect to final $F_0$, only three SFPs, [a˥], [kɛ˧˥] and [wɔ˨˩], exhibit statistical differences from their tone-in-statement counterparts (Table 2.6a), and in fact all these SFPs have higher final $F_0$ than Tones-in-statement. This is consistent with the patterns seen in the graphical analysis. Regarding final $F_0$ velocity, only one SFP shows statistical difference from its Tone-in-statement counterpart (Table 2.6b): the SFP, [kɛ˧˥] (Figure 2.1d) has a greater final $F_0$ velocity, that is, a steeper rise in final $F_0$, than the High Rising Tone-in-statement.

The descriptive statistics of the duration data are presented as boxplots in Figure 2.4. From the boxplots, we see duration differences between Tones-in-statement and SFPs in some cases, between Tones-in-question and Tones-in-statement in two cases, and between Tones-in-question and SFPs in many cases. Firstly, comparing Tones-in-statement and SFPs, SFPs in sets A and C ([a˥] and [a˧] respectively) show shorter duration, whereas SFPs in sets H and J ([wɔ˧] and [wɔ˨˧] respectively) show longer duration than their tone-in-statement counterparts. Secondly, when comparing Tones-in-statement and Tones-in-question (that is, the same final syllable in a statement versus when it is in an intonation question), we see the latter is longer than the former in Sets C and I, and in both cases the lexical tone is Mid-Low Falling. Finally, comparing Tones-in-question and SFPs, there seem to be differences in Sets A, C, F, H, I and J, and among these sets the SFPs in Sets H and J ([wɔ˧] and [wɔ˩˧] respectively) are longer whereas the rest are shorter than their Tone-in-question counterparts.

Figure 2.4 Boxplots of duration of the final syllables for all 7 speakers. The 10 sets of final syllables are labelled A to J, and in each set S refers to Tone-in-statement, and Q refers to Tone-in-question. The same Y-axis is used for all boxplots. The median and mean values are given under each boxplot.




One-way repeated measures ANOVAs were performed on the duration data, and the post hoc test results comparing Tones-in-statement and SFPs, Tones-in-statement and Tones-in-question, and Tones-in-question and SFPs are shown in Table 2.7. Six SFPs are statistically different in duration from Tones-in-statement (Table 2.7a), four being shorter (Set A [a˥]; Set C [a˩], Set D [kɛ˧˥] and Set I [wɔ˩]), and the other two longer (Set H [wɔ˧] and Set J [wɔ˩˧]).

Only two Tones-in-statement, Sets C and I, show differences from Tones-in-question. In both cases the lexical tones are Mid-Low Falling, and Tones-in-question are longer than Tones-in-statement (Table 2.7b). Six Tone-in-question-SFP pairs show differences in duration: Sets A, C, D, F, I and J (Table 2.7c). Therefore, from Table 2.7a and c, the SFP in Set J ( [wɔ˩˧]) is longer than both its statement and question counterparts, and the SFPs in Sets A, C and D ([a˥], [a˩] and [kɛ˧˥]) are shorter than both their respective statement and question counterparts. A summary of the post hoc test results is given in Table 2.7d.

Table 2.7 Post hoc test (Bonferroni) results on duration of the final syllables (in milliseconds) comparing the three types of final syllables (Tones-in-statement, SFPs, Tones-in-question). Only the significant (*p* values < 0.05) and marginally significant (*p* values <0.06) results are presented (*p* values < 0.05 are shown in boldface).

(a) SFP vs. Tone-in-statement

| Sentence set | SFP | Tone-in-statement | Mean difference of SFP - Tone-in-statement | *p* |
|---|---|---|---|---|
| A | [a˥] | [a˥] | -116.496 | ***0.006*** |
| C | [a˩] | [ŋa˩] | -46.926 | ***0.004*** |
| D | [kɛ˧˥] | [tsɛ˧˥] | -48.000 | ***0.036*** |
| H | [wɔ˧] | [kɔ˧] | 64.620 | 0.058 |
| I | [wɔ˩] | [wɔ˩] | -38.339 | 0.055 |
| J | [wɔ˩˧] | [ŋɔ˩˧] | 89.868 | ***0.031*** |

(b) Tone-in-statement vs. Tone-in-question

| Sentence set | Tone-in-statement | Tone-in-question | Mean difference of tone-in-statement - tone-in-question | *p* |
|---|---|---|---|---|
| C | [a˩] | [a˩] | -92.480 | ***0.005*** |
| I | [wɔ˩] | [wɔ˩] | -115.383 | ***0.001*** |

(c) Tone-in-question vs. SFP

| Sentence set | Tone-in-question | SFP | Mean difference of tone-in-question - SFP | *p* |
|---|---|---|---|---|
| A | [a˥] | [a˥] | 99.767 | ***0.012*** |
| C | [a˩] | [ŋa˩] | 139.406 | **<*0.001*** |
| D | [kɛ˦] | [tsɛ˦] | 44.080 | ***0.026*** |
| F | [la˥] | [la˥] | 67.067 | ***0.040*** |
| I | [wɔ˩] | [wɔ˩] | 153.722 | ***0.001*** |
| J | [wɔ˨˩˦] | [ŋɔ˨˩˦] | -93.561 | **<*0.017*** |

(d) Summary of the significant ($p < 0.05$) and marginally significant ($p < 0.06$) differences in duration, showing the direction of difference in each pair of final syllables under comparison in sentence sets which show significant results (– means the former is shorter than the latter; + means the former is longer than the latter; n.s. means non-significant difference).

| Sentence set | SFP - tone-in-statement | Tone-in-statement - tone-in-question | SFP - tone-in-question |
|---|---|---|---|
| A | – | n.s. | – |
| B | n.s. | n.s. | n.s. |
| C | – | – | – |
| D | – | n.s. | – |
| E | n.s. | n.s. | n.s. |
| F | n.s. | n.s. | – |
| G | n.s. | n.s. | n.s. |
| H | marginal + | n.s. | n.s. |
| I | marginal – | – | – |
| J | + | n.s. | + |

### 2.5.2 Pre-final syllables

#### *2.5.2.1 Graphical analyses*

Comparing the sentences with SFPs and statements without SFPs, the $F_0$ contours of the part of the sentence preceding the final syllables are largely the same in many cases. For example, in Figures 2.1b, 1e, 1g, 1h and 1j, we see that in nearly all of the six pre-SFP syllables, the $F_0$ contours of the sentences with SFPs are identical to the statements without SFPs. Minor differences in $F_0$ contours are observed in Figures 2.1f and 2.1i. In the case of the statement SFP [la˥] (Figure 2.1f), the $F_0$ values are slightly higher in all the six pre-SFP syllables, whereas in the case of the statement SFP [wɔ˩] (Figure 2.1i), higher $F_0$ values are found in the first five pre-SFP syllables than in the statements without SFPs. Noticeable differences are found in the sentences with the statement SFP [a˥] and the question SFPs [a˩] and [kɛ˦] (Figures 2.1a, 2.1c and 2.1d respectively). For the sentence with the statement SFP [a˥] (Figure 2.1a), we see a raised $F_0$ contour towards the second half of the sentence, whereas for the sentences containing the question SFPs [a˩] and [kɛ˦] (Figure 2.1c and Figure 2.1d), the $F_0$ contour of the whole sentence is raised.

*2.5.2.2 Statistical analyses*

One-way repeated measures ANOVAs were performed on the disyllabic units preceding the final syllables, with mean $F_0$ and duration in all ten sets of test sentences as dependent variables. The summary of the statistically significant results are presented in Table 2.8. For mean $F_0$, all the differences in each of the disyllabic units across the three types of final syllables are significant, meaning that in each set of test sentences, each disyllabic unit preceding the final syllables has different mean $F_0$ values in the sentence with and without SFP and intonation question; for duration, significant differences are observed in five sets of test sentences (Sets A, B, G, H and J), and the differences are in the third disyllabic unit in all the five sets, and there is also a difference in the second disyllabic unit in Set H (Table 2.8a). In order to identify which pairs (sentence with SFP vs. sentence without SFP; sentence with SFP vs. intonation question; sentence without SFP vs. intonation question) are significantly different, post hoc tests (Bonferroni) were performed, and the significant results are summarized in Table 2.8b.

Table 2.8 Summary of the ANOVA and post hoc test (Bonferroni) results of the disyllabic units preceding the final syllables.

(a) ANOVA results:

(i) mean $F_0$ (in semitones): significantly different in all cases of the three disyllabic units

(ii) duration (in ms): n.s. refers to non-significant results; sentence sets with no significant results in any of the disyllabic units are not shown:

| Sentence set | 1st disyll. unit | | 2nd disyll. unit | | 3rd disyll. unit | |
|---|---|---|---|---|---|---|
| | F(2,12) | *p* | F(2,12) | *p* | F(2,12) | *p* |
| A (Fig. 1a) | n.s. | | n.s. | | 4.304 | ***0.039*** |
| B (Fig. 1b) | n.s. | | n.s. | | 6.263 | ***0.014*** |
| G (Fig. 1g) | n.s. | | n.s. | | 5.232 | ***0.023*** |
| H (Fig. 1h) | n.s. | | 6.132 | ***0.015*** | 8.233 | ***0.006*** |
| J (Fig. 1j) | n.s. | | n.s. | | 12.650 | ***0.001*** |

(b) post hoc test results

(i) mean $F_0$ (in semitones)

(1) statement without SFP vs. intonation question: mostly statistically significant

(exceptions include the 1st disyllabic unit in sentence sets E (Fig. 1e), G (Fig. 1g) and H (Fig. 1h))

(2) intonation question vs. sentence with SFP: the mean difference and *p* value of each comparison is shown for the significant cases; n.s. refers to non-significant results:

| Sentence set | SFP | 1st disyll. unit | | 2nd disyll. unit | | 3rd disyll. unit | |
|---|---|---|---|---|---|---|---|
| | | diff. | *p* | diff. | *p* | diff. | *p* |
| A (Fig. 1a) | [a˥] | n.s. | | n.s. | | n.s. | |
| B (Fig. 1b) | [a˧] | n.s. | | 1.375 | ***0.014*** | 2.062 | ***0.002*** |
| C (Fig. 1c) | [a˩] | n.s. | | n.s. | | 1.120 | ***0.011*** |
| D (Fig. 1d) | [kɛ˧˥] | n.s. | | n.s. | | n.s. | |
| E (Fig. 1e) | [kɛ˧] | n.s. | | 1.626 | ***0.004*** | 2.214 | ***<0.001*** |
| F (Fig. 1f) | [la˥] | n.s. | | n.s. | | 1.418 | ***0.006*** |
| G (Fig. 1g) | [la˧] | n.s. | | 1.431 | ***0.010*** | 2.251 | ***0.001*** |
| H (Fig. 1h) | [wɔ˧] | n.s. | | n.s. | | 2.177 | ***0.007*** |
| I (Fig. 1i) | [wɔ˩] | n.s. | | n.s. | | 1.955 | ***0.017*** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| J (Fig. 1j) | [wɔ˩˧] | 1.212 | ***0.021*** | 1.254 | ***0.002*** | 1.826 | ***0.001*** |

(3) statement without SFP vs. sentence with SFP: the mean difference and p value of each comparison is shown for the significant cases; n.s. refers to non-significant results; sentence sets with no significant results in any of the disyllabic units are not shown:

| Sentence set | SFP | 1st disyll. unit | | 2nd disyll. unit | | 3rd disyll. unit | |
|---|---|---|---|---|---|---|---|
| | | diff. | *p* | diff. | *p* | diff. | *p* |
| A (Fig. 1a) | [a˥] | n.s. | | -0.714 | ***0.023*** | -1.402 | ***0.017*** |
| C (Fig. 1c) | [a˩] | -1.389 | ***0.003*** | -1.311 | ***<0.001*** | -1.048 | ***0.001*** |
| D (Fig. 1d) | [kɛ˧˥] | -1.305 | ***0.011*** | -1.505 | ***0.001*** | -1.958 | ***<0.001*** |

(ii) duration (in ms)

(1) statement without SFP vs. intonation question: significant differences are found in sentence set H (Fig. 1h) only in the 2nd (diff. and *p*: 25.565, ***0.003***) and 3rd (diff. and *p*: 21.446, ***0.004***) disyllabic units, and in both cases the duration in the statement without SFP is longer than that in the intonation question.

(2) intonation question vs. sentence with SFP: significant differences are found in four cases, namely the 3rd disyllabic units of sentence sets B (Fig. 1b; diff. and *p*: -51.504, ***0.037***), H (Fig. 1h; diff. and *p*: -52.519, ***0.0038***) and J (Fig. 1j; diff. and *p*: -59.946, ***0.009***), and the 1st disyllabic unit of sentence set J (Fig. 1j; diff. and *p*: -22.120, ***0.022***); in all these four significant cases,

the disyllabic units in sentences with SFPs are longer than those in the intonation questions.

(3) statement without SFP vs. sentence with SFP: no cases of statistical significant difference are found.

The results show that, compared to statements without SFPs, the intonation questions have higher mean $F_0$ in all disyllabic units except for three cases (Table 2.8bi1; sentence sets E, G and H; Figures 2.1e, 2.1g and 2.1h) where only the second and third disyllabic units show differences. Together with the $F_0$ tracings in Figure 2.1, this result shows that intonation questions in Hong Kong Cantonese are best described as having both raised and expanded $F_0$ from the very beginning of the question, rather than just raised $F_0$ on the final syllable.

Comparing mean $F_0$ of the intonation questions with sentences with SFPs in the ten sets of sentences (Table 2.8bi2), we can identify the following types of sentences with SFPs, (1) where there is no difference in mean $F_0$ of all penultimate disyllabic units from the intonation question: the statement SFP [a˥] (Set A, Figure 2.1a) and question SFP [kɛ˧˥] (Set D, Figure 2.1d); (2) where there are differences in only the third disyllabic unit: the question SFP [a˩] (Set C, Figure 2.1c), the statement SFP [la˥] (Set F, Figure 2.1f), the statement SFP [wɔ˧] (Set H, Figure 2.1h), and the statement SFP [wɔ˩] (Set I, Figure 2.1i); (3) where differences are found in the second and third disyllabic units: the statement SFP [a˧] (Set B, Figure 2.1b), the statement SFP [kɛ˧] (Set E, Figure 2.1e), and the statement SFP [la˧] (Set G, Figure 2.1g); and (4) where all three disyllabic units are different: the statement SFP [wɔ˩˧] (Set J, Figure 2.1j).

The post hoc results comparing mean $F_0$ of sentences with and without SFPs (Table 2.8bi3) show that only three sets have significant differences, involving the statement SFP [a˥] (Set A, Figure 2.1a), question SFP [a˨] (Set C, Figure 2.1c), and question SFP [kɛ˧˥] (Set D, Figure 2.1d). In all three sets, where in each at least the last two disyllabic units show significant differences, the mean $F_0$ in the disyllabic units of sentences with SFPs are higher than that of the sentences without SFPs.

With respect to duration, the post hoc results show that it is not much different in the three types of sentences (Table 2.8bii). Comparing intonation questions and their counterparts without SFPs, we see significant differences in only the last two disyllabic units in one set (Table 2.8bii1), where the intonation question has shorter durations. Comparing intonation questions and their counterpart sentences with SFPs, in the ten sets of sentences, only the third disyllabic units in three sets and the first disyllabic unit in one set show significant differences. And in all the significant cases the intonation questions have shorter durations (Table 2.8bii2). Finally, no difference in duration in disyllabic units is found when sentences with and without SFPs are compared (Table 2.8bii3).

## 2.6 <u>Discussion</u>

### 2.6.1 $F_0$ and duration of sentence-final particles

The general questions raised at the beginning of the present study were: (1) Is the $F_0$ of SFPs the same as that of the lexical tones? (2) If not, how do they differ? (3) Does the presence of an SFP have any effect on the sentence prosody? The graphical analyses (Figure 2.1; Section 2.5.1.1) indicate that many of the ten SFPs are very similar to the ordinary lexical tones in Cantonese. The statistical analyses (post hoc test comparing SFPs and Tones-in-statement in Table 2.6) further indicate that the $F_0$ contours of SFPs are similar to Tones-in-statement. We saw that only three SFPs ([a˥], [kɛ˦˥], [wɔ˩]) differ from Tones-in-statement in terms of final $F_0$ (all three SFPs having higher values), and only one SFP ([kɛ˦˥]) differs from Tone-in-statement with respect to final $F_0$ velocity (the SFP rising more steeply than Tone-in-statement). Since final $F_0$ and final $F_0$ velocity indicate the underlying pitch target of the syllable (Chen & Xu, 2006; Liu & Xu, 2007a; Xu & Wang, 2001), together with the $F_0$ tracings, they provide evidence that the pitch targets of most of the SFPs in the present study are similar to those of Tones-in-statement in the language. This provides a positive answer to the first question.

Regarding the second question, our data do show various discrepancies between SFPs and Tones-in-statement. Interestingly, the patterns of the differences seem interpretable given the known meanings of specific SFPs. Three of the SFPs, [a˥, kɛ˦˥, wɔ˩], have significantly higher final $F_0$ than the corresponding Tones-in-statement. This seems to be consistent with the idea of the frequency code (Ohala, 1984) or size code (Gussenhoven, 2004; Chuenwattanapranithi, Thipakorn, & Maneewongvatana, 2008). According to this idea, there is a correlation between $F_0$ and interactional intentions (based on a body size projection principle): other things being equal, a lower fundamental frequency signals aggression, dominance or confidence, while a higher $F_0$ signals friendliness, subordination or uncertainty. Firstly, SFP [a˥], with higher $F_0$ than a corresponding High Level Tone-in-statement, is used for showing suggestion or request. It has also been suggested that compared to another SFP [la˥] which also expresses suggestion or request, [a˥] is more “consultative” (Kwok, 1984), meaning that the speaker has a more inviting attitude. Secondly, [kɛ˦˥] is the only SFP in the present study that differs in final $F_0$, final $F_0$ velocity and duration from Tone-in-statement. These differences may be attributed to the fact that it is used as a question SFP showing surprise or seeking confirmation. Lastly, [wɔ˨˩] indicates that the speaker is having sudden awareness of some unexpected information, and this surprise component may be the cause for the overall upward shift in the $F_0$ in the low-falling contour of the SFP (Figure 2.1i).

Apart from $F_0$ features, the duration data also help us understand the nature of the SFPs. The durations of six of the SFPs ([a˥], [a˨˩], [kɛ˦˥], [wɔ˧], [wɔ˨˩] and [wɔ˨˧], in sentence sets A, C, D, H, I and J respectively) under study are different from Tones-in-statement. Of the four SFPs that are shorter in duration than the corresponding Tones-in-statement (Table 2.7a), [a˨˩] and [kɛ˦˥] are used as question SFPs and [a˥] and [wɔ˨˩] as statement SFPs. The two question SFPs and the statement SFP [wɔ˨˩] share the commonality of showing surprise, and so the shorter durations may be related to this meaning. The SFP [a˥] is heard as lively or animated (Kwok, 1984; Law, 1990), and its shorter duration than the corresponding High Level tone is likely a contributing factor. It is notable that these six SFPs which show durational differences from Tones-in-statement are also significantly different from Tones-in-question, except Set H (involving SFP [wɔ˧]). Importantly, as seen in Table 2.7d where the directions of differences are summarized, these SFPs which differ from Tones-in-statement have the same direction of difference from Tones-in-question, and such relationship does not change when the direction of difference between Tones-in-statement and Tones-in-questions in the respective cases is considered. On the whole, the significantly different durations between SFPs and Tones-in-statement are consistent with their previously described discourse functions.

The SFPs [wɔ˧] and [wɔ˨˧] are longer than Tones-in-statement (Table 2.7a), and these two SFPs have the common function of reminding: the former reminds the hearer of the fact or opinion, and the latter, the so-called 'hearsay SFP', draws the hearer's attention to the fact that the utterance just produced is not based on the speaker's own knowledge or information but is according to someone else (Kwok, 1984; Luke, 1990; Matthews & Yip, 2011). The lengthening of the final syllable could thus be a manifestation of the reminding function. As we have seen in Figure 2.2, [wɔ˨˧] has a falling $F_0$ component attached to the low-rising contour, and this may explain the increased durations of [wɔ˨˧] as compared to the Mid-Low Rising Tone-in-statement. The additional falling $F_0$ contour would be its distinctive reminding function as a hearsay SFP. The sequential movements in [wɔ˨˧] are reminiscent of the proposal by Chao (1968) that Mandarin sometimes uses an extra tonal component attached to the end of a sentence-final tone for intonation purposes, for which there has been recent experimental evidence (Li, Fang, & Dang, 2011; Mueller-Liu, 2006). Such sequential extra tone is realized on the final portion of a lengthened sentence final syllable, i.e., without inserting an additional syllable. The $F_0$ and duration pattern of [wɔ˩˥˩] therefore seem similar to the Mandarin sentence final extra tone. However, because this is a surprise finding as no one has reported before, further studies are needed to look into it in a systematic manner.

One additional point of interest as far as duration is concerned is related to those final syllables which show a double movement in $F_0$ contour, namely the Tones-in-question of all but Sets A and F (Figure 2.1) and the SFP [wɔ˩˥˩] (Set J, Figure 2.2). In these Tones-in-question, there is an initial fall followed by a rise in $F_0$ contour, whereas in the SFP [wɔ˩˥˩] (Set J), the double movement consists of an initial rise followed by a fall. We have seen that the SFP [wɔ˩˥˩] has a longer duration than the Tone-in-statement. What is interesting is that only two Tones-in-question with double $F_0$ movement, Sets C and I, show significantly longer duration than the Tones-in-statement, and in fact, only these two show significant differences (Table 2.7b). The Tones-in-question in Sets C and I are both Mid-Low Falling, but those in the other Tones-in-question involve High Rising (Set D), Mid Level (Sets B, E G, and H) and Mid-Low Rising (Set J) tones. Based on this observation, we speculate that the Mid-Low Falling tone in an intonation question, in order to preserve its identity by having a recognizable falling $F_0$ contour towards the lowest level of the speaker's voice range, is more resistant to integration with intonation than other lexical tones, giving rise to an overall longer duration. Such a difference may explain why Tones-in-question in Sets C and I are longer than their respective Tones-in-statement but in the other Sets show no differences. In view of the above duration comparisons, the longer duration of SFP [wɔ˩] than both Tone-in-statement and Tone-in-question (Table 2.7) thus appears even more likely to be related to its inherent discourse function.

Finally, it is worth mentioning that the $F_0$ contours shown in Figure 2.1 corroborate the finding of Ma et al. (2006) that all Cantonese tones at the final position of questions have rising contour, regardless of their canonical form. Although not the main focus of the present study, this phenomenon seems to be indication of an extra weight given to the sentence-final intonation in Cantonese, which is consistent with the existence of an elaborate set of SPFs in the language. It also sets Cantonese apart from Mandarin where sentence-final tones are modified by question intonation mainly in terms of $F_0$ height but little in shape (Ho, 1976; Liu & Xu, 2005). This language difference also suggests that functional meanings are only partial predictors of the phonetic patterns of Cantonese SFPs or their equivalents in other languages, because other factors, particularly diachronic changes, may also play an important role in shaping the phonetic details of various prosodic patterns (Xu et al., 2012).

To summarize, the $F_0$ and duration patterns of the ten SFPs proper have shed some light on the nature of the prosody of the SFPs in Cantonese: Firstly, the acoustic properties are consistent with previously reported auditory impressions (e.g., Fang, 2003; Kwok, 1984; Law, 1990; Luke, 1990; Matthews & Yip, 2011). Those lively and animated SFPs are produced with shorter duration, higher $F_0$ and greater $F_0$ velocity; those SFPs which function as reminder are produced with longer duration. Secondly, the $F_0$ contours of many of the SFPs under study are similar to Tones-in-statement, suggesting that SFPs have underlying tonal targets.

Yet, the fact that some SFPs ([a˥], [kɛ˦˥], [wɔ˩]) differ from Tones-in-statement in having higher final $F_0$ values suggests that the surface $F_0$ of SFPs may be a combination of lexical tonal targets and intonation. Furthermore, the convex $F_0$ contour of the SFP [wɔ˨˧˩], where the falling part of the contour seems to be an additional intonation component attached after the Mid-Low Rising tonal target (Figure 2.2), suggests this may indeed be the case. Furthermore, the acoustic data of the pre-final syllables, as discussed next, give further support to this interpretation.

### 2.6.2 Sentence-body $F_0$

As seen in Section 2.5.2.2 and Table 2.8bi2, the two sentences with the SFPs [a˥] (sentence set A, Figure 2.1a) and [kɛ˦˥] (sentence set D, Figure 2.1d) do not show difference in mean $F_0$ from intonation questions. That is, the pre-final syllables have the same $F_0$ height and contour values as an intonation question. This is of importance to our discussion of the possibility that the surface $F_0$ of the SFPs is due to both lexical tonal targets and intonation. As mentioned earlier, the high $F_0$ and short duration of [a˥] are likely linked to its function of showing polite suggestion or request. It is thus possible that such function is encoded by the entire sentence intonation. [kɛ˦˥] is a question SFP and its high-rising contour may then be part of the question intonation. Also, the pre-final syllables of the sentence with the question SFP [kɛ˦˥] are the same in mean $F_0$ as the question intonation, and this gives support to our suspicion. Further perception tests in future research involving the pre-final syllables as stimuli should give better confirmation about the importance of the raised $F_0$ in identifying the sentence as a question.

Besides the SFPs [a˥] and [kɛ˦], from Table 2.8bi2 we also see that the sentences with the question SFP [a˩] (Set C, Figure 2.1c) and statement SFP [wɔ˩] (Set I, Figure 2.1i), both being SFPs that show surprise, also exhibit a high degree of similarity in terms of $F_0$ contours to intonation questions: the mean $F_0$ of the pre-final syllables of these sentences differ from the intonation questions in the last disyllabic units only. Such similarity gives another hint that the discourse functions traditionally attributed to the SFPs may in fact be also partly carried out by the sentence-body intonation. Again, further perception tests involving the pre-final syllables as stimuli should help determine the contribution of sentence-body $F_0$ in the overall discourse function.

With the above information, it is reasonable to postulate that the system of intonation is working simultaneously with the SFPs in determining the discourse function of the sentence, and that each SFP has its tonal target with its surface $F_0$ being a result of realizing both the tonal target and the sentence intonation, rather than being purely a lexical tone or purely a localized intonational event in the final syllabic position.

# Chapter 3 - Prosodic focus in Hong Kong Cantonese

In this chapter, we will investigate the acoustic correlates of prosodic focus in Hong Kong Cantonese. Our focus will be to experimentally test whether post-focus compression is present in the language, and to show whether the presence of absence of PFC in the language affects the perception of focus.

## 3.1 Production of focus in Cantonese by Hong Kong speakers

A production experiment was designed to examine the acoustic features of prosodic focus in short sentences spoken by Hong Kong Cantonese speakers.

### 3.1.1 Stimuli

Six declarative sentences (Sets C1a, C2 to C6), each consisting of syllables of the same lexical tone, were composed (Table 3.1). For the five-syllable Tone 1 sentence (C1a), the first two syllables form a word, the third syllable forms a word on its own, and the last two syllables form a word, and this is the Cantonese equivalent of the Mandarin test sentence used in Chen et al. (2009). In the other test sentences, each disyllable forms a word. In the experiment, all these three-word sentences were assigned four different focus locations: no focus, initial (the first word), medial (the second word), and final (the last word), which were elicited by precursor questions (Table 3.2) each asking for a specific piece of information related to the target sentence. For each sentence, which was produced five times by each speaker, there are four versions that differ only in focus: neutral focus, initial focus, medial focus and final focus. The total number of stimuli is: 6 (lexical tones) × 4 (focus conditions) × 5 (repetitions) = 120.

The test materials were designed as being composed of only one tone in each sentence so that any focus-related behaviour can be seen clearly in the $F_0$ plots. Since Cantonese contains both static and dynamic tones, there may be differences in focus-related $F_0$ behaviours depending on whether the lexical tone is static or dynamic, and such differences, if any, will be shown more clearly if all the syllables in the sentence have the same tone. The test sentence set C1a, which is the Cantonese equivalent of that used in Chen et al. (2009), was used because comparisons can be made with reference to the Mandarin one with respect to both production and perception.

Table 3.1 Cantonese target sentences.

| Set | Test sentences | Type of focus | | |
|---|---|---|---|---|
| | | Initial | Medial | Final |
| C1a | [ma˥ ma˥ mɔ˥ mau˥ mi˥]<br>Mother touches the cat. | [ma˥ ma˥] | [mɔ˥] | [mau˥ mi˥] |
| C2 | [siu˧˥ tsɛ˧˥ hou˧˥ tsou˧˥ tsɐu˧˥ tsɔ˧˥]<br>The young lady has left very early. | [siu˧˥ tsɛ˧˥] | [hou˧˥ tsou˧˥] | [tsɐu˧˥ tsɔ˧˥] |
| C3 | [a˧ tʰai˧ tsɔi˧ tsʰi˧ him˧ tsai˧]<br>Tai is in debt again. | [a˧ tʰai˧] | [tsɔi˧ tsʰi˧] | [him˧ tsai˧] |

| | | | | |
|---|---|---|---|---|
| C4 | [jɐn˩ kʷʰɐn˩ wɔ˩ pʰɪŋ˩ jɐu˩ hɐŋ˩]<br><br>The crowd marched peacefully. | [jɐn˩ kʷʰɐn˩] | [wɔ˩ pʰɪŋ˩] | [jɐu˩ hɐŋ˩] |
| C5 | [lou˨˧ lei˨˧ man˨˧ man˨˧ mai˨˧ hai˨˧]<br><br>Mr. Li buys crabs every night. | [lou˨˧ lei˨˧] | [man˨˧ man˨˧] | [mai˨˧ hai˨˧] |
| C6 | [hɔk˨ hau˨ tsʊŋ˨ si˨ wɐn˨ tʊŋ˨]<br><br>The school values sports. | [hɔk˨ hau˨] | [tsʊŋ˨ si˨] | [wɐn˨ tʊŋ˨] |

Table 3.2 The precursor questions for the Cantonese target sentences.

| Set | | Precursor questions | Focus to be elicited |
|---|---|---|---|
| C1a | (i) | [nei˨˧ hɔn˧ tou˧ sɐm˨ mɔ˥]<br><br>What do you see? | none |
| | (ii) | [sɵy˩ tsɔi˨ mɔ˥ mau˥ mi˥]<br><br>Who touches the cat? | initial |
| | (iii) | [ma˥ ma˥ tɵy˧ mau˥ mi˥ tsou˨ sɐm˨ mɔ˥]<br><br>What does Mother do to the cat? | medial |
| | (iv) | [ma˥ ma˥ tsɔi˨ mɔ˥ sɐm˨ mɔ˥]<br><br>What is Mother touching? | final |
| C2 | (i) | [jɐu˨˧ mɛ˥ siu˥ sɪk˥ a˧]<br><br>What's the news? | none |
| | (ii) | [pin˥ kɔ˧ hou˧˥ tsou˧˥ tsɐu˧˥ tsɔ˧˥]<br><br>Who has left very early? | initial |
| | (iii) | [siu˧˥ tsɛ˧˥ kei˧˥ si˧ tsɐu˧˥ tsɔ˧˥]<br><br>When did the young lady left? | medial |

| | | | |
|---|---|---|---|
| | (iv) | [siu˦ tsɛ˦ hou˦ tsou˦ tsou˩ mɛ˥]<br>What did the young lady do very early? | final |
| C3 | (i) | [jɐu˩˧ mɛ˥ siu˥ sɪk˥ a˧]<br>What’s the news? | none |
| | (ii) | [pin˥ kɔ˧ tsɔi˧ tsʰi˧ him˧ tsai˧]<br>Who is in debt again? | initial |
| | (iii) | [a˧ tʰai˧ mɛ˥ him˧ tsai˧]<br>What’s that about Tai and debt? | medial |
| | (iv) | [a˧ tʰai˧ tsɔi˧ tsʰi˧ tsou˩ mɛ˥]<br>What’s the situation of Tai again? | final |
| C4 | (i) | [jɐu˩˧ mɛ˥ siu˥ sɪk˥ a˧]<br>What’s the news? | none |
| | (ii) | [pin˥ kɔ˧ wɔ˩ pʰɪŋ˩ jɐu˩ hɐŋ˩]<br>Who marched peacefully? | initial |
| | (iii) | [jɐn˩ kʷʰɐn˩ tim˦ jœŋ˦ jɐu˩ hɐŋ˩]<br>How did the crowd march? | medial |
| | (iv) | [jɐn˩ kʷʰɐn˩ wɔ˩ pʰɪŋ˩ tsou˩ mɛ˥]<br>What did the crowd do peacefully? | final |
| C5 | (i) | [jɐu˩˧ mɛ˥ siu˥ sɪk˥ a˧]<br>What’s the news? | none |
| | (ii) | [pin˥ kɔ˧ man˩˧ man˩˧ mai˩˧ hai˩˧]<br>Who buys crabs every day? | initial |
| | (iii) | [lou˩˧ lei˩˧ kei˦ si˩ mai˩˧ hai˩˧]<br>When did Mr. Li buy crabs? | medial |
| | (iv) | [lou˩˧ lei˩˧ man˩˧ man˩˧ tsou˩ mɛ˥]<br>What did Mr. Li do every night? | final |
| C6 | (i) | [jɐu˩˧ mɛ˥ siu˥ sɪk˥ a˧]<br>What’s the news? | none |

| | | | |
|---|---|---|---|
| | (ii) | [pin˥ kɔ˧ tsʊŋ˨ si˨ wɐn˨ tʊŋ˨]<br>Who values sports? | initial |
| | (iii) | [hɔk˨ hau˨ mɛ˥ wɐn˨ tʊŋ˨]<br>What's about the school and sports? | medial |
| | (iv) | [hɔk˨ hau˨ tsʊŋ˨ si˨ mɛ˥]<br>What does the school value? | final |

### 3.1.2 Subjects

Eight Cantonese speakers (four males and four females) from Hong Kong, aged between 19 and 33, took part in the production tests. They were born and raised in Hong Kong, and received their primary and secondary education in Hong Kong. They were university undergraduate and postgraduate students studying at University College London, UK at the time of the experiment, but they used mainly Cantonese in their daily life outside school. These subjects had been in the UK for five to eight months prior to the recording. None of them reported having any speech disorders.

### 3.1.3 Recording procedure

The recording was conducted in an anechoic room at University College London, UK. Before the start of the recording, the experimenter went through the list of precursor and test sentences with the subjects, who were instructed to read aloud both the precursor and target sentences at a normal conversational rate without any pause in the middle of a sentence. During the recording, both the target sentences and their precursors were presented in Chinese characters on a computer screen one pair at a time in random order, and a different list of precursor and target sentences in different order was used for each subject, who recorded both the precursor and target sentences. Five repetitions of each precursor-target sentence pair were recorded by each speaker, in randomized blocks. If a mistake was made, the subject was asked to repeat the sentence. A sound level meter (Bruel & Kjaer 2231) was used as the microphone and the speech was recorded directly onto a computer hard disk using the software Adobe Audition with a sampling rate of 44.1 kHz. Each recording session lasted for about 20 minutes.

### 3.1.4 Data extraction and measurements

The acoustic analysis was done using ProsodyPro, a Praat (Boersma, 2001) script (Xu, 2005-2012). The script allows users to manually rectify the vocal cycle markings generated by Praat based on autocorrelation. Continuous $F_0$ tracks were then computed from the rectified vocal cycles. Also facilitated by the script, the voiced intervals of each syllable were labelled in Praat's TextGrid for analysis. For syllables with initial voiceless consonants, the onset of a voiced interval was the first full vocal cycle of the vowel; for syllables with initial nasal or lateral consonants, the onset was the first nasal vocal cycle; for syllables with initial glides, the onset was the point along the continuous voicing where amplitude is the lowest; for open syllables, the offset was the last full vocal cycle of the vowel; for closed syllables with final nasal consonants, the offset was the final nasal cycle of the nasal consonant; for closed syllables with final voiceless consonants, the offset was the last full vocal cycle of the vowel.

ProsodyPro generated two types of output, time-normalized continuous $F_0$ contours and measurements from raw $F_0$ data. The time-normalized $F_0$ contours in each voiced interval consisted of 10 equally spaced $F_0$ points, which were used for graphical analysis only. The measurements taken from raw $F_0$ contours of each voiced interval are as follows:

Mean $F_0$ (Hz and semitones) — average of all the fundamental frequency values of each word. The values in Hz and semitones were used for graphic analysis and statistical analysis respectively.

Mean duration (milliseconds) — average duration of each word; the measurement is restricted to the voiced portions of each word only.

Mean intensity (dB) — average intensity of each word.

Excursion size — average difference between maximum and minimum $F_0$ within each word

As Cantonese contains both static and dynamic tones, mean $F_0$ and $F_0$ excursion size can reflect the focus-related behaviours in $F_0$ and so are important in identifying the presence of PFC. Since it was reported that duration is the most significant correlate of focus in Cantonese (Bauer, et al., 2004), duration was also measured in this study so as to provide a complete description of the acoustic correlates of focus. Mean intensity was also measured following Chen et al. (2009).

### 3.1.5 Results and analysis

Figure 3.1 displays time-normalized mean $F_0$ contours of all the target sentences produced by all eight speakers. Each of the $F_0$ contours was obtained in two steps. In step 1, mean $F_0$ contours of each speaker were obtained by averaging over the

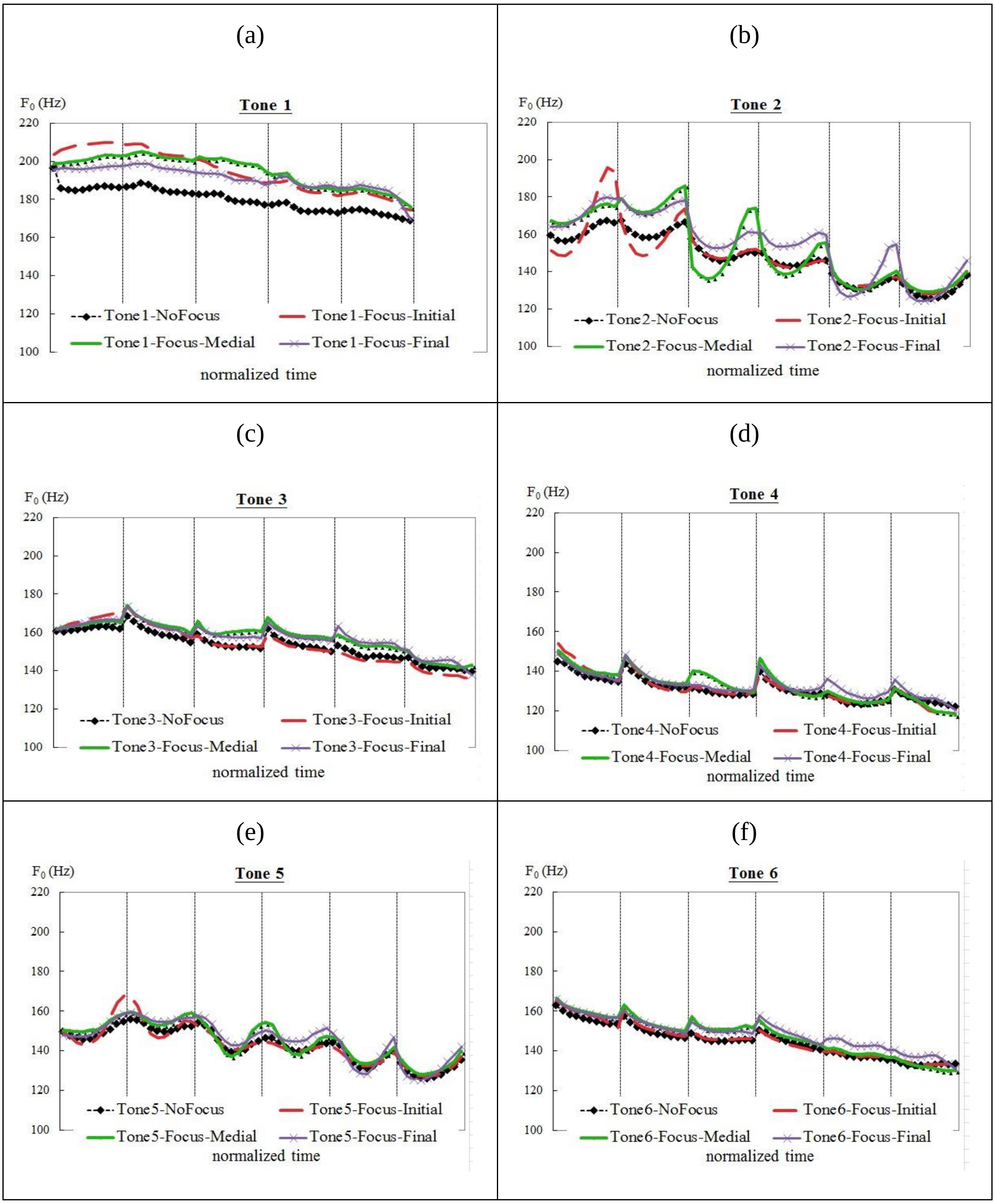
$F_0$ values of the five repetitions at each of the 10 normalized time points. In step 2, mean $F_0$ contours of all speakers were obtained by taking the log-average of the mean $F_0$ contours by individual speakers (cf. Xu, 2011 for the need to use logarithmic scale) using the following formula:

$$meanf_0 = \exp(\frac{1}{n}\sum_{i=1}^{n} \ln f_{0i}) \qquad (1)$$

where $meanf_0$ is the log-average $F_0$, $f_{0i}$ is the speaker mean $F_0$, and $n$ is the number of speakers.

The vertical lines in Figure 3.1 indicate syllable boundaries. The time gaps corresponding to the voiceless consonants are not shown in the plots. As can be seen, except for Tone 1 (Figure 3.1a), the static tones (Tones 3, 4 and 6, Figures 3.1c, 3.1d and 3.1f) show little difference in pitch contours between on-focus words and neutral-focus words in all three word positions. The pitch contours of Tone 1 (Figure 3.1a) show an increase in the pitch level of the focused words in all three word positions, and the increase is more pronounced when the first word was in focus, and the pitch level of the whole sentence is shifted upwards in all the three sentences with focus. For the dynamic tones (Tones 2 and 5, Figures 3.1b and 3.1e), the pitch contours of the on-focus words for Tone 2 appear different from those of the neutral-focus sentences, with larger excursion size and higher pitch level in on-focus words, but Tone 5 does not show obvious differences. In general the graphs show that the on-focus words exhibit some $F_0$ variations, but there is little focus-related $F_0$ change in the words before or after focus. Importantly, in none of the six tones can we see post-focus $F_0$ contours going below the neutral focus contours.

Figure 3.1 Time-normalized $F_0$ graphs of the Cantonese test sentences by the Hong Kong speakers. Each graph shows the averaged $F_0$ contours across all 8 speakers.



To statistically corroborate the observations from the mean $F_0$ contours in Figure 3.1, one-way repeated measures ANOVAs were performed on mean $F_0$, mean intensity, mean duration and excursion size of each word. For each target sentence, the values of on-focus and post-focus disyllabic units were compared with the corresponding unit in the same sentence in neutral focus. The results are shown in Table 3.3.

Table 3.3 Results of one-way repeated measures ANOVAs on mean $F_0$, mean intensity and duration in the focus production experiment. ↑ and ↓ signify that the value is significantly greater or smaller than in the neutral-focus counterpart.

(a) on-focus vs. neutral-focus

| | mean $F_0$ | | mean intensity | | duration | | excursion size | |
|---|---|---|---|---|---|---|---|---|
| Tone | F | *p* | F | *p* | F | *p* | F | *p* |
| 1 | 6.665 | ***0.036***↑ | 157.909 | ***<0.001***↑ | 20.686 | ***0.003***↑ | 7.030 | ***0.033***↑ |
| 2 | 1.116 | 0.326 | 31.419 | ***0.001***↑ | 31.386 | ***0.001***↑ | 7.000 | ***<0.001***↑ |
| 3 | 7.494 | ***0.029***↑ | 26.461 | ***0.001***↑ | 30.079 | ***0.001***↑ | 8.746 | ***0.021***↑ |
| 4 | 9.554 | ***0.018***↑ | 37.099 | ***<0.001***↑ | 29.926 | ***0.001***↑ | 15.228 | ***0.006***↑ |
| 5 | 3.287 | 0.113 | 33.459 | ***0.001***↑ | 31.055 | ***0.001***↑ | 89.505 | ***<0.001***↑ |
| 6 | 9.480 | ***0.018***↑ | 35.625 | ***0.001***↑ | 49.622 | ***<0.001***↑ | 69.325 | ***<0.001***↑ |

(b) post-focus vs. neutral-focus

| Tone | mean $F_0$ | | mean intensity | | duration | | excursion size | |
|---|---|---|---|---|---|---|---|---|
| | F | *p* | F | *p* | F | *p* | F | *p* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.662 | 0.443 | 3.383 | 0.108 | 0.479 | 0.511 | 9.750 | ***0.017***↑ |
| 2 | 0.229 | 0.647 | 2.192 | 0.182 | 0.511 | 0.498 | 0.278 | 0.614 |
| 3 | 0.001 | 0.979 | 3.447 | 0.106 | 5.646 | ***0.049***↓ | 0.029 | 0.869 |
| 4 | 0.044 | 0.840 | 7.933 | ***0.026***↓ | 10.129 | ***0.015***↓ | 0.316 | 0.592 |
| 5 | 0.118 | 0.741 | 1.027 | 0.345 | 7.873 | ***0.026***↓ | 0.015 | 0.906 |
| 6 | 0.044 | 0.840 | 2.816 | 0.137 | 1.385 | 0.278 | 0.420 | 0.538 |

In terms of mean $F_0$, the on-focus words show increased value compared to neutral-focus words in Tones 1, 3, 4 and 6. There are also significant differences in duration, intensity and excursion size, which are greater in magnitude for the on-focus words of all six tones.

## 3.2 Perception of prosodic focus in Hong Kong Cantonese

### 3.2.1 Stimuli

The stimuli used in the perception experiment were sentences taken from the production experiment (Table 3.1) following a method used in Xu et al. (2012). The recordings of the three speakers who showed maximum, median and minimum standard deviations of all $F_0$ points across all the focus conditions were selected. Consequently, there were a total of 360 tokens in the listening test (3 speakers x 5 repetitions x 4 foci x 6 sentence sets).

### 3.2.2 Subjects

16 native speakers of Hong Kong Cantonese, recruited at the Hong Kong Institute of Education, participated in the experiment. The experiment was carried out in a quiet room using the MFC function of Praat. Each subject wore a pair of

headphones and listened to the randomized test sentences and decided on which one or none of the three words in each test sentence was emphasized. Note that here the identification of emphasis was regarded as equivalent to the identification of focus.

### 3.2.3 Results and analysis

The confusion matrices are shown in Table 3.4, which shows that the overall focus identification rate is fairly high, except for Tone 1, for which the average rate is less than 60%. In addition, the identification rate of initial focus was generally poorer than that of medial or final focus, which is different from the pattern seen in Beijing Mandarin (Xu et al., 2012).

When focus was heard wrongly, in most cases the listeners heard no focus in the sentence. In cases when there was no focus in the sentence but the listeners thought that there was one, any of the three locations (initial, medial, final) could be perceived as having the focus, but in five out of the six tones the perceived focus was in the initial position.

Table 3.4 Confusion matrices of focus perception (%). The rates of correct identifications are indicated by boldface.

| Sentence | heard as / original | None | Initial | Medial | Final |
|---|---|---|---|---|---|
| Tone 1 (high level) | None | **77.92** | 12.08 | 6.25 | 3.75 |
| | Initial | 34.17 | **57.92** | 7.08 | 0.83 |
| | Medial | 35.00 | 5.42 | **58.33** | 1.25 |
| | Final | 27.50 | 2.08 | 18.33 | **52.08** |
| Tone 2 | None | **84.17** | 8.33 | 3.75 | 3.75 |
| | Initial | 20.42 | **73.33** | 3.75 | 2.50 |

| | | | | | |
|---|---|---|---|---|---|
| (high rising) | Medial | 17.92 | 2.08 | **78.75** | 1.25 |
| | Final | 9.58 | 0.83 | 2.08 | **87.50** |
| Tone 3 (mid-level) | None | **81.67** | 7.50 | 7.50 | 3.33 |
| | Initial | 20.83 | **69.58** | 8.33 | 1.25 |
| | Medial | 7.08 | 2.92 | **89.58** | 0.42 |
| | Final | 12.92 | 0.42 | 2.08 | **84.58** |
| Tone 4 (low falling) | None | **86.25** | 7.50 | 5.00 | 1.25 |
| | Initial | 28.33 | **69.58** | 2.08 | 0.83 |
| | Medial | 13.33 | 1.67 | **84.58** | 0.42 |
| | Final | 13.33 | 0.83 | 6.67 | **79.17** |
| Tone 5 (low rising) | None | **82.50** | 8.33 | 5.00 | 4.17 |
| | Initial | 10.83 | **87.08** | 2.08 | 0.00 |
| | Medial | 11.25 | 1.67 | **86.25** | 0.83 |
| | Final | 18.33 | 0.83 | 3.75 | **77.08** |
| Tone 6 (low-mid level) | None | **81.25** | 5.00 | 9.17 | 4.58 |
| | Initial | 33.33 | **57.50** | 6.67 | 2.50 |
| | Medial | 16.67 | 1.67 | **80.83** | 0.83 |
| | Final | 9.58 | 0.83 | 2.50 | **87.08** |

## 3.3 Discussion

The high recognition rate of focus for most of the tone indicates that, despite the lack of PFC, there are other acoustic cues to mark focus in Cantonese.

Xu et al. (2012) showed that with the compression in both pitch range and intensity of post-focus words in Beijing Mandarin, higher focus recognition rates were achieved by Beijing Mandarin speakers than speakers of Taiwanese and Taiwan Mandarin (recognition rate of >90% versus <75%). Their test sentences consist of five syllables all with the high level tone and essentially the same words as the Tone 1 test sentence (C1a) used in the production test by the Hong Kong Cantonese speakers. The present study on Cantonese shows that focus identification was poor for the high level tone, with a recognition rate of less than 60% in all three word locations. However, the high focus recognition rates of approaching 70-80% in most other tones suggest that there may be other properties besides PFC as effective cues for focus perception. Nevertheless, the direct comparison of the Cantonese data with the data of Beijing Mandarin, Taiwan Mandarin and Taiwanese suggest the importance of PFC, at least for words with the high level tone.

There are several possibilities for the poor performance of focus identification in Tone 1 sentences. Firstly, the acoustic analysis reveals that $F_0$ excursion of on-focus words is significantly different from that of words in neutral focus in all tones except Tone 1, and so this distinction between Tone 1 and other tones can explain the poorer focus recognition rate. Secondly, increases in both duration and intensity are the most prominent feature of focus in Cantonese, and if intensity differences are indeed important for focus recognition, then Tone 1 syllables would be the least advantageous, because a higher $F_0$ is associated with a higher level of intensity, and the already high $F_0$ in Tone 1 syllables would exhibit relatively smaller percentage increase in intensity when they are on-focus than syllables of other lexical tones. Thirdly, the Tone 1 syllables all begin with the sonorant consonant [m], whereas in the other five test sentences, there is a mixture of sonorant and obstruent consonants in syllable initial positions. This difference may explain why Tone 1 sentences had the poorest focus recognition rate, i.e., it could be due to a lack of consonantal cue present in other test sentences. Since acoustic properties such as voice onset time, stop burst energy and consonantal length may act as cues for focus perception, they might have contributed to the higher recognition rates of focus in sentences of Tones 2 to 5 in the experiment, and this will require a further experiment to confirm. Regardless of whether consonantal effects are significant, the direct comparison of the Cantonese data with the data of Beijing Mandarin, Taiwan Mandarin and Taiwanese suggest the importance of PFC, at least for words with the high level tone.

# Chapter 4 - Production of focus in Cantonese and English by bilingual speakers

The production experiment detailed in this chapter examines the prosodic focus in short sentences in Cantonese and English spoken by English-Cantonese bilinguals living in England, with the aim to test whether PFC, which is present in English, appears in the Cantonese of the bilinguals, and whether PFC is affected in any way in the English of the bilinguals. In this chapter, the Cantonese and English speech data come from Chung (2010), who performed a preliminary investigation into the PFC of the recruited 15 bilingual speakers and adopted the Cantonese target and test sentences in Chapter 3 (which were used in a conference paper) in her experiments. All the data extraction and measurements presented in this chapter were done by the author of this dissertation.

## 4.1 Stimuli

The Cantonese sentences were the same as those in the production experiment by the Hong Kong speakers (Section 3.1.1), except the Tone 1 sentence (Sets C1b, C2 to C6, Table 4.1), which consists of six syllables instead of five. All these six Cantonese sentences consist of three disyllabic units in which each unit is a word, and each sentence can have neutral, initial, medial or final focus. The total number of stimuli is: 6 (lexical tones) × 4 (focus conditions) × 5 (repetitions) = 120.

Table 4.1 Cantonese target sentences for the bilingual speakers

| Set | Test sentences | Focus | | |
|---|---|---|---|---|
| | | Initial | Medial | Final |
| C1b | [tsɪŋ˥ tsɪŋ˥ fan˥ fan˥ tʊŋ˥ kɪŋ]<br>Jing-jing has returned to Tokyo. | [tsɪŋ˥ tsɪŋ˥] | [fan˥ fan˥] | [tʊŋ˥ kɪŋ] |
| C2 | [siu˧˥ tsɛ˧˥ hou˧˥ tsou˧˥ tsɐu˧˥ tsɔ˧˥]<br>The young lady has left very early. | [siu˧˥ tsɛ˧˥] | [hou˧˥ tsou˧˥] | [tsɐu˧˥ tsɔ˧˥] |
| C3 | [a˧ tʰai˧ tsɔi˧ tsʰi˧ him˧ tsai˧]<br>Tai is in debt again. | [a˧ tʰai˧] | [tsɔi˧ tsʰi˧] | [him˧ tsai˧] |
| C4 | [jɐn˨˩ kʷʰɐn˨˩ wɔ˨˩ pʰɪŋ˨˩ jɐu˨˩ hɐŋ˨˩]<br>The crowd marched peacefully. | [jɐn˨˩ kʷʰɐn˨˩] | [wɔ˨˩ pʰɪŋ˨˩] | [jɐu˨˩ hɐŋ˨˩] |
| C5 | [lou˩˧ lei˩˧ man˩˧ man˩˧ mai˩˧ hai˩˧]<br>Mr. Li buys crabs every night. | [lou˩˧ lei˩˧] | [man˩˧ man˩˧] | [mai˩˧ hai˩˧] |
| C6 | [hɔk˨ hau˨ tsʊŋ˨ si˨ wɐn˨ tʊŋ˨]<br>The school values sports. | [hɔk˨ hau˨] | [tsʊŋ˨ si˨] | [wɐn˨ tʊŋ˨] |

For English, four target sentences (Sets E1 to E4), two longer ones and two shorter ones, were designed by Chung (2010) (Table 4.2; the precursor questions are shown in Table 4.3). For the longer sentences, focus can be medial or final, whereas in the shorter sentences, focus can be initial or final. For each set, therefore, there were three versions of the same sentence: one with neutral focus, one with initial/medial focus, and one with final focus. The total number of stimuli is: 4 (sentences) × 3 (focus conditions) × 5 (repetitions) = 60.

According to Chung (2010), the sentences were designed to have "maximum continuation of $F_0$ contours" (p. 16). The use of both short and long sentences was intended to be an investigation into whether the length of sentence will have an effect on focus; overall four focus conditions (no focus, initial, medial and final focus) were tested using the four sentences.

Table 4.2 English target sentences

| Set | Test sentences | Focus | | |
|---|---|---|---|---|
| | | Initial | Medial | Final |
| E1 | She saw Mona in the morning. | | *Mona* | *morning* |
| E2 | I'll play Naomi in a minute. | | *Naomi* | *minute* |
| E3 | Nina lost the money. | *Nina* | | *money* |
| E4 | Emma comes from Mali. | *Emma* | | *Mali* |

Table 4.3 The precursor questions for the English target sentences

| Set | Precursor questions | Target sentences (the focused word is bracketed) |
|---|---|---|
| E1 | What did you say? | She saw Mona in the morning. |
| | Who did she see in the morning? | She saw [Mona] in the morning. |
| | When did she see Mona? | She saw Mona in the [morning]. |
| E2 | What did you say? | I'll play Naomi in a minute. |
| | Who will you play in a minute? | I'll play [Naomi] in a minute. |
| | When will you play? | I'll play Naomi in a [minute] |
| E3 | What did you say? | Nina lost the money. |

| | | |
|---|---|---|
| | Who lost the money? | [Nina] lost the money. |
| | What did Nina lose? | Nina lost the [money]. |
| E4 | What did you say? | Emma comes from Mali. |
| | Who comes from Mali? | [Emma] comes from Mali. |
| | Where does Emma come from? | Emma comes from [Mali]. |

### 4.2 Subjects

Fifteen English-Cantonese bilingual speakers (seven males and eight females) from the UK, aged 20-35, were recruited by Chung (2010) as subjects. They were university undergraduates or graduates at the time of the experiments. All these speakers were born and raised in South-east England around London, and their parents were immigrants from Hong Kong to England. They speak both Southern British English (SBE) and Cantonese as native languages, and none of them reported having any speech disorders. All the bilingual subjects acquired Cantonese through conversations with their parents. Twelve of the bilingual subjects also formally learnt to read and speak Cantonese in Chinese schools when they were younger and achieved GCSE level. The duration of study in Chinese schools varied from 6 to 14 years (the numerical values indicate the approximate numbers of years): M1: 11; M3: 11; M4: 6; M6: 7; F1: 8; F2: 9; F3: 7; F4: 11; F5: 14; F6: 12; F7: 9; F8: 6. Only some of the subjects were exposed to TV programmes in Cantonese on a regular basis.

### 4.3 Recording procedure

The recording procedure was similar to the one reported in Chapter 3 for the Hong Kong speakers (Section 3.1.3), except that the recordings took place in a quiet room using a head-worn microphone (Countryman Isomax hypercardiod) that recorded the signal directly into a computer at a sampling rate of 44.1 kHz using Praat. Eight of them recorded the Cantonese sentences first and the other seven the English sentences first. Each recording session lasted for about 50 minutes.

### 4.4 Data extraction and measurements

All sound files were collected by Chung (2010), but the data extraction, measurements and analyses were carried out by the author of this dissertation. Data extraction and measurements for the Cantonese sentences were done using the same methods as those for the Hong Kong speakers (Section 3.1.4). For the English sentences, the following measurements were used for the on-focus and post-focus words and the equivalent words in the neutral focus condition:

Max $F_0$ (Hz and semitones): the maximum fundamental frequency level of the word. The values in Hz were used for plotting $F_0$ contours for graphical analysis, and the values in semitones were used for statistical analysis.

Mean intensity (dB): the average intensity of the word.

Different measurements for Cantonese and English sentences were chosen out of the consideration that the respective measurements best serve the purpose of research in the present dissertation. As presented in Chapter 3, the measurements mean $F_0$, $F_0$ excursion, duration and intensity on the one hand enable us to confirm whether PFC is present, and on the other hand provide a more complete description of the acoustic correlates of focus in Cantonese based on past research and the author’s native speaker intuition. For English, the measurements max $F_0$ and mean intensity are enough to show whether $F_0$ and intensity are lowered post-focally. Using both graphical and statistical analyses, the presence/absence of PFC in Cantonese and English can be verified and compared.

## 4.5 Graphical analysis

The graphical analysis involved two steps for each target sentence. The first step is the same as that for the Hong Kong speakers, i.e., obtaining mean $F_0$ contours of each speaker by averaging over the five repetition $F_0$ values at each of the normalized time point. In the second step, average $F_0$ contours of the five repetitions by *all speakers* who showed very similar $F_0$ patterns were plotted. Since the Cantonese target sentences produced by all the individual speakers were very similar, we will only present the group data. For the English target sentences, we observed that the speakers could be divided into three groups, (1) those with evidence of PFC in all the four sentences, (2) those with evidence of PFC in some of the sentences, and (3) those without PFC in any of the sentences. Therefore, both $F_0$ contours by individual speakers and by speaker groups will be presented.

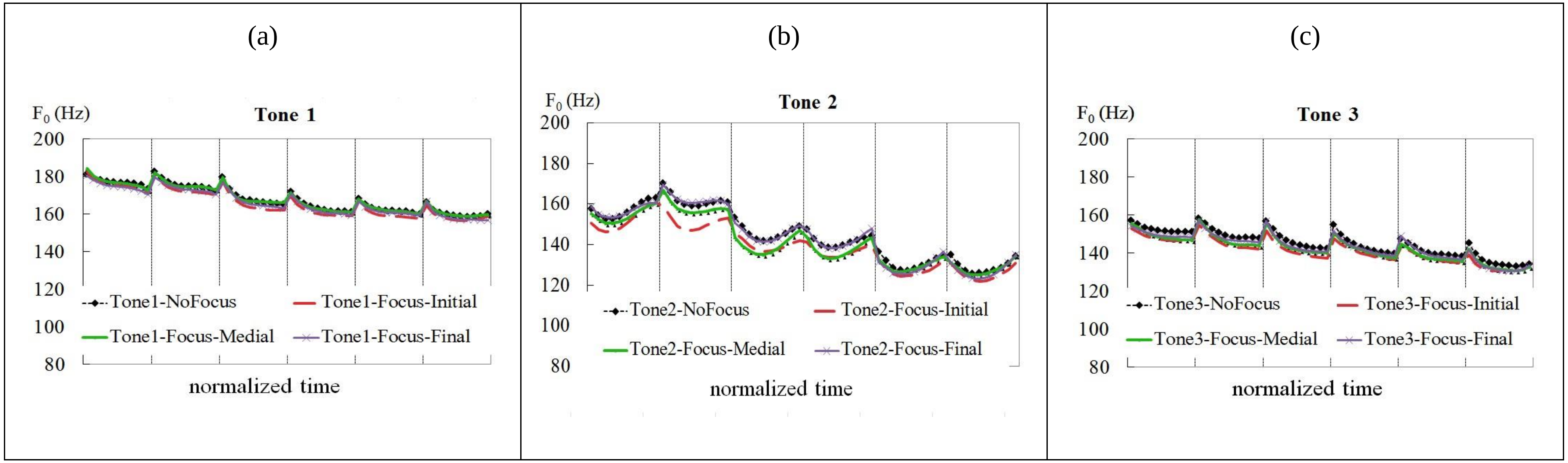
## 4.6 Cantonese

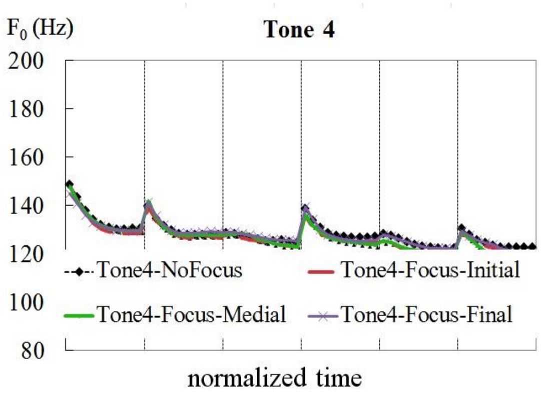
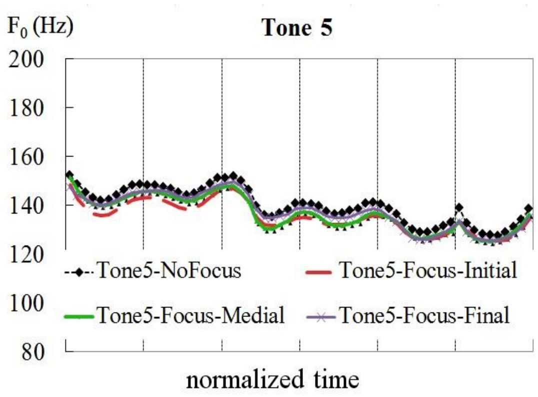
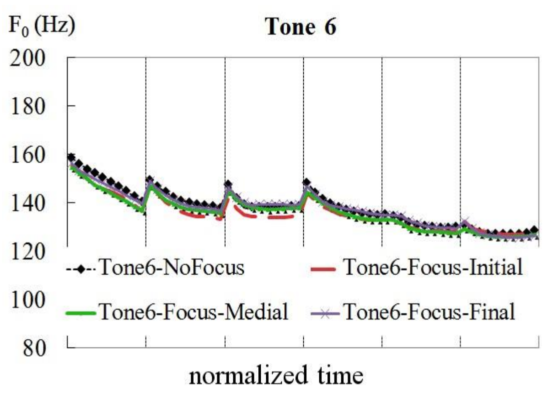
### 4.6.1 Time-normalized $F_0$ curves

The Cantonese sentences produced by the bilingual speakers, shown in Figures 4.1a to 4.1f, look similar to those by the Hong Kong speakers we saw in Figure 3.1. In all six test sentences corresponding to the six lexical tones, there are only subtle changes in on-focus $F_0$, and there is no noticeable variation in $F_0$ after the focused word. In Tones 2 and 5 sentences, the on-focus pitch range is slightly expanded, and in other tones, the $F_0$ contours do not show much change across focus conditions. One major difference from the Hong Kong Cantonese speakers in

Experiment 1 is that here Tone 2 and Tone 5 sentences show much smaller on-focus changes in $F_0$ (Figures 4.1b and 4.1e, compared to Figures 3.1b and 3.1e). Another noticeable difference is that Tone 1 sentences do not show much change in $F_0$ in initial focus (compare Figure 4.1a and Figure 3.1a).

Figure 4.1 Time-normalized $F_0$ graphs of the Cantonese test sentences by the bilingual speakers. Each graph shows the averaged $F_0$ contours across all 15 speakers.



| (d) | (e) | (f) |
| --- | --- | --- |
|  |  |  |


Tone 4
F0 (Hz)
200
180
160
140
120
100
80
Tone4-NoFocus
Tone4-Focus-Initial
Tone4-Focus-Medial
Tone4-Focus-Final
normalized time
Tone 5
Tone5-NoFocus
Tone5-Focus-Initial
Tone5-Focus-Medial
Tone5-Focus-Final
Tone 6
Tone6-NoFocus
Tone6-Focus-Initial
Tone6-Focus-Medial
Tone6-Focus-Final

### 4.6.2 Statistical analyses

The same statistical tests as those performed on the Cantonese data of the Hong Kong speakers (Section 3.1.5) were done, and the results are presented in Table 4.4. As Table 4.4a shows, there is no on-focus increase of mean $F_0$ and intensity. Rather, for both mean $F_0$ and intensity, whenever there were statistically significant differences between the neutral focus words and on-focus or post-focus words, the neutral focus words have higher values. Therefore, consistent with what we see in the graphical analysis, the Cantonese sentences show an overall downward shift in mean $F_0$ of the whole sentence when there is focus. In addition, on-focus duration is lengthened in two tones (Tones 1 and 2) only, and no difference is found for excursion size. Post-focally, there is a decrease in intensity in all six tones, but a lowering of mean $F_0$ is found in two tones only (Tones 3 and 5; Table 4.4b).

Table 4.4 One-way repeated measures ANOVA results of the Cantonese data of the bilingual speakers. ↑ and ↓ signify that the value is higher and lower than that of the neutral counterpart, respectively.

(a) on-focus vs. neutral

| Tone | mean $F_0$ | | intensity | | duration | | excursion size | |
|---|---|---|---|---|---|---|---|---|
| | F | *p* | F | *p* | F | *p* | F | *p* |
| 1 | 2.100 | 0.169 | 0.309 | 0.587 | 39.778 | ***<0.001***↑ | 1.268 | 0.279 |
| 2 | 0.010 | 0.924 | 5.620 | ***0.033***↓ | 6.499 | ***0.023***↑ | 2.906 | 0.110 |
| 3 | 0.024 | 0.880 | 1.188 | 0.294 | 2.082 | 0.171 | 2.580 | 0.131 |
| 4 | 2.254 | 0.155 | 1.612 | 0.225 | 0.000 | 0.999 | 0.593 | 0.454 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5 | 4.579 | ***0.050***↓ | 7.528 | ***0.016***↓ | 4.053 | 0.064 | 0.001 | 0.974 |
| 6 | 0.342 | 0.568 | 1.882 | 0.192 | 1.437 | 0.251 | 0.446 | 0.515 |

(b) post-focus vs. neutral

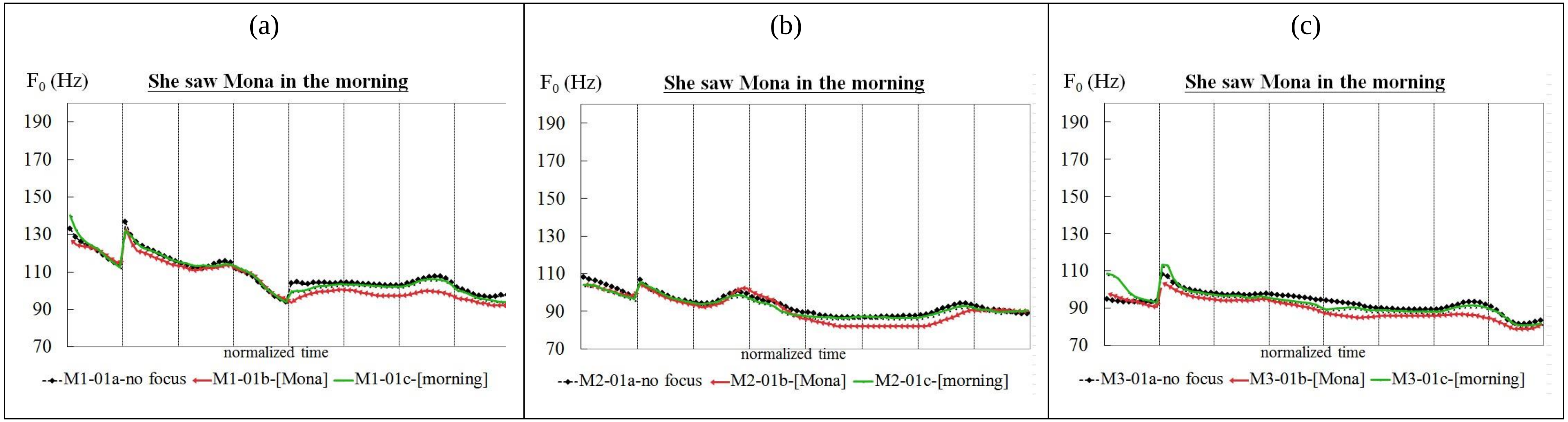
| Tone | mean $F_0$ | | intensity | | duration | | excursion size | |
|---|---|---|---|---|---|---|---|---|
| | F | *p* | F | *p* | F | *p* | F | *p* |
| 1 | 2.988 | 0.106 | 13.559 | ***0.002***↓ | 0.236 | 0.634 | 0.726 | 0.408 |
| 2 | 2.135 | 0.166 | 13.645 | ***0.002***↓ | 0.398 | 0.538 | 0.017 | 0.897 |
| 3 | 8.900 | ***0.010***↓ | 15.363 | ***0.002***↓ | 10.423 | ***0.006***↓ | 2.171 | 0.163 |
| 4 | 2.422 | 0.142 | 8.219 | ***0.012***↓ | 3.670 | 0.076 | 0.055 | 0.818 |
| 5 | 8.495 | ***0.011***↓ | 13.794 | ***0.002***↓ | 11.452 | ***0.004***↓ | 20.231 | ***0.001***↓ |
| 6 | 1.380 | 0.260 | 13.075 | ***0.003***↓ | 1.411 | 0.255 | 2.102 | 0.169 |

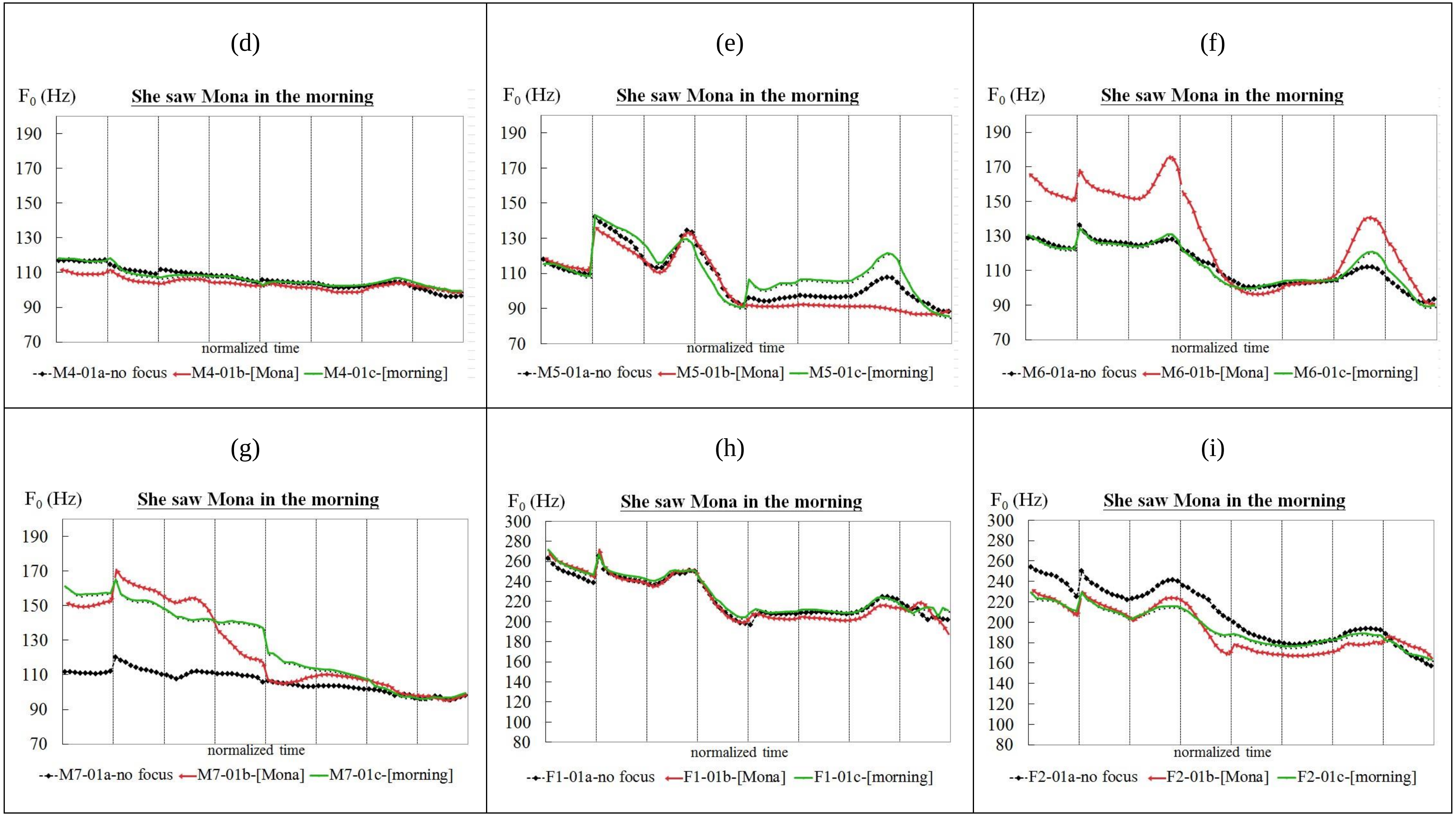
## 4.7 English

### 4.7.1 Time-normalized $F_0$ curves

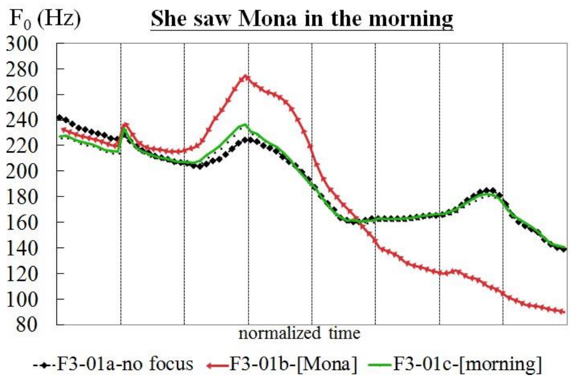
Time-normalized mean $F_0$ contours of all the four English sentences by all fifteen speakers are displayed in Figures 4.2-4.5 so as to show the individual variations. Different manifestations of focus and post-focus $F_0$ variations can be observed in these individual plots. For on-focus words, some speakers show increase in $F_0$ level and expansion of $F_0$ range, whereas others barely show any difference in $F_0$ from neutral focus. Similarly, for post-focus words, some speakers show clear lowering in $F_0$ but others show no $F_0$ lowering post-focally.

Figure 4.2 Time-normalized $F_0$ graphs of the English test sentence E1 by the 15 bilingual speakers (7 male speakers, M1 to M7, and 8 female speakers, F1 to F8).

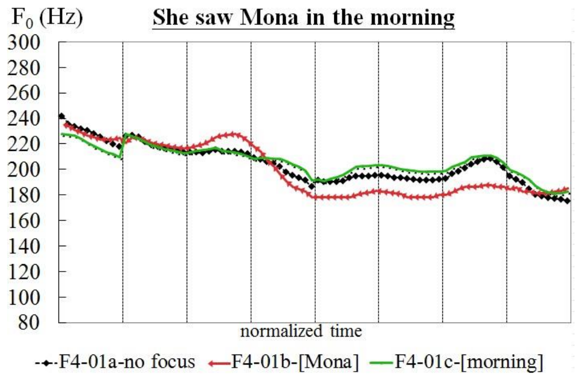



(d)
F0 (Hz)
She saw Mona in the morning
190
170
150
130
110
90
70
normalized time
M4-01a-no focus
M4-01b-[Mona]
M4-01c-[morning]
(e)
F0 (Hz)
She saw Mona in the morning
190
170
150
130
110
90
70
normalized time
M5-01a-no focus
M5-01b-[Mona]
M5-01c-[morning]
(f)
F0 (Hz)
She saw Mona in the morning
190
170
150
130
110
90
70
normalized time
M6-01a-no focus
M6-01b-[Mona]
M6-01c-[morning]
(g)
F0 (Hz)
She saw Mona in the morning
190
170
150
130
110
90
70
normalized time
M7-01a-no focus
M7-01b-[Mona]
M7-01c-[morning]
(h)
F0 (Hz)
She saw Mona in the morning
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F1-01a-no focus
F1-01b-[Mona]
F1-01c-[morning]
(i)
F0 (Hz)
She saw Mona in the morning
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F2-01a-no focus
F2-01b-[Mona]
F2-01c-[morning]

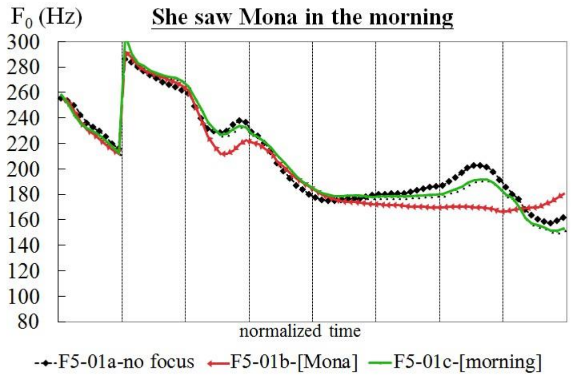
(j)


F0 (Hz)
She saw Mona in the morning
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F3-01a-no focus
F3-01b-[Mona]
F3-01c-[morning]

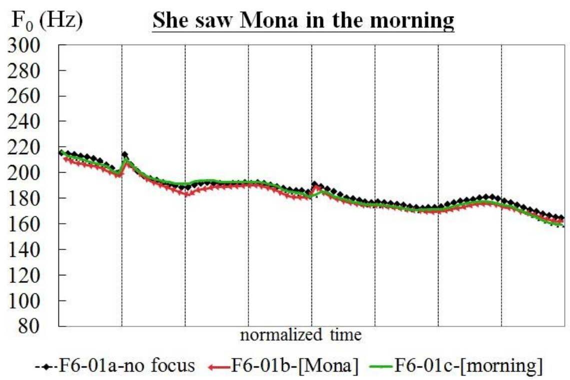
(k)


F0 (Hz)
She saw Mona in the morning
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F4-01a-no focus
F4-01b-[Mona]
F4-01c-[morning]

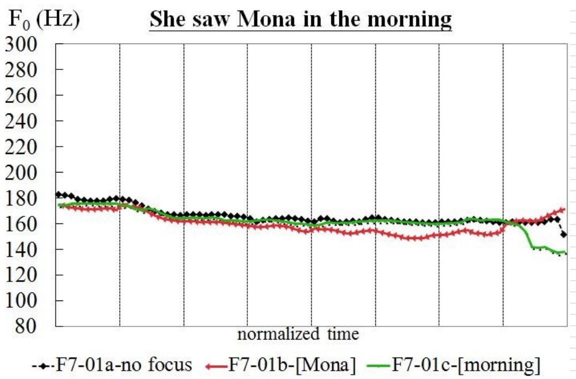
(l)


F0 (Hz)
She saw Mona in the morning
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F5-01a-no focus
F5-01b-[Mona]
F5-01c-[morning]

(m)


F0 (Hz)
She saw Mona in the morning
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F6-01a-no focus
F6-01b-[Mona]
F6-01c-[morning]

(n)


F0 (Hz)
She saw Mona in the morning
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F7-01a-no focus
F7-01b-[Mona]
F7-01c-[morning]

(o)

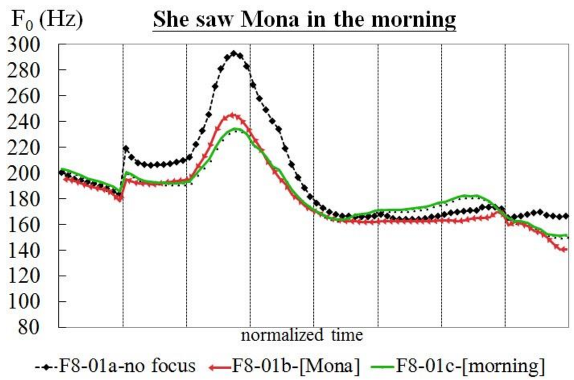
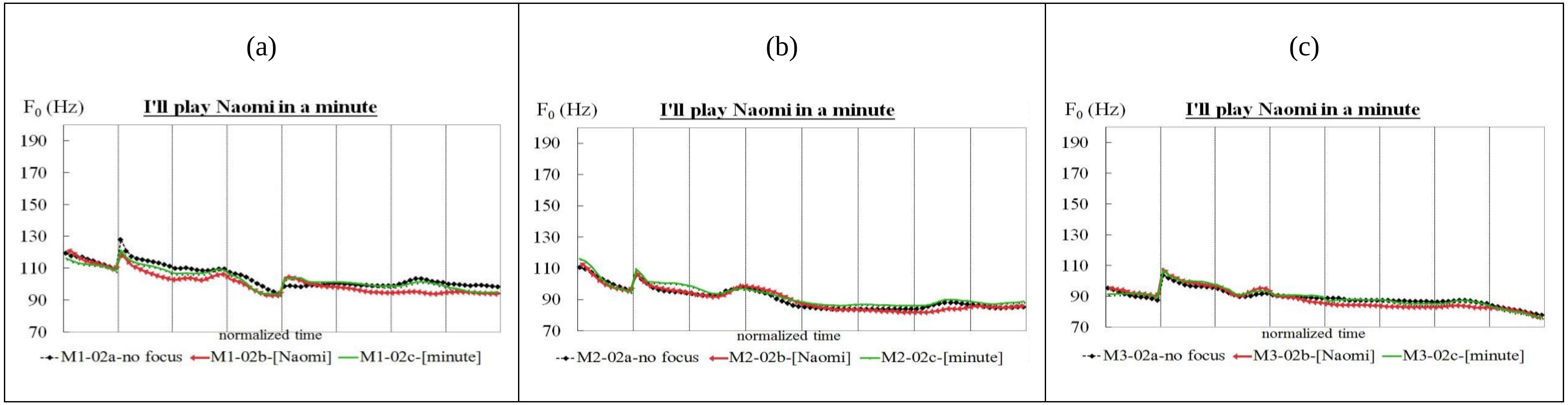
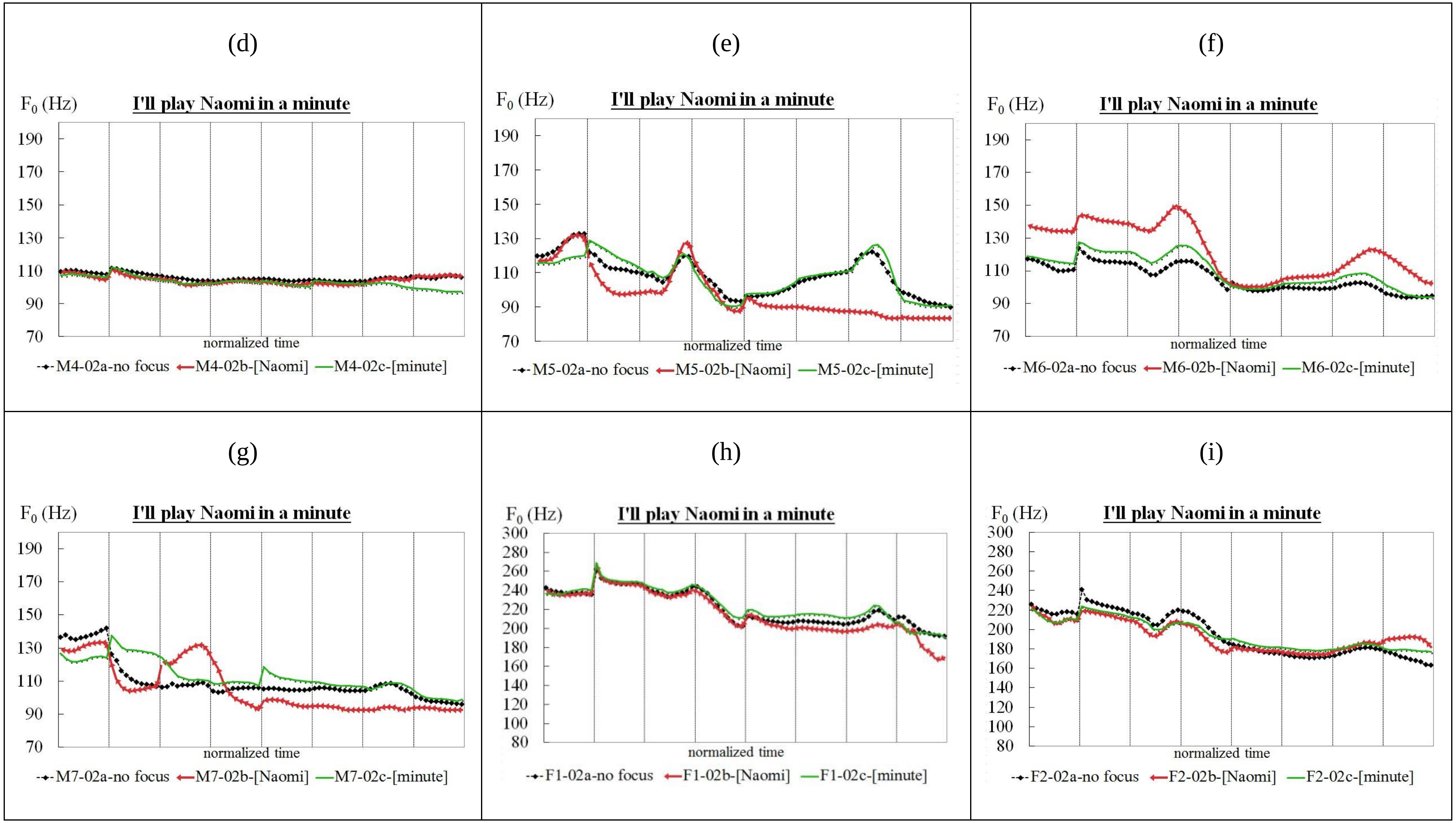
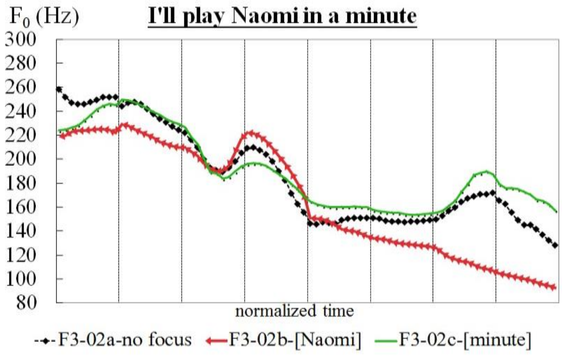

F0 (Hz)
She saw Mona in the morning
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F8-01a-no focus
F8-01b-[Mona]
F8-01c-[morning]

Figure 4.3 Time-normalized $F_0$ graphs of the English test sentence E2 by the 15 bilingual speakers (7 male speakers, M1 to M7, and 8 female speakers, F1 to F8).

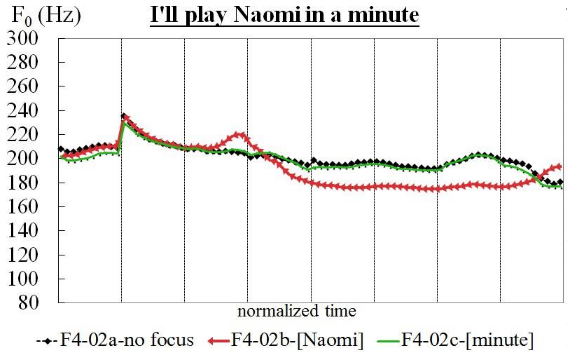



(d)
F0 (Hz)
I'll play Naomi in a minute
normalized time
M4-02a-no focus
M4-02b-[Naomi]
M4-02c-[minute]
(e)
F0 (Hz)
I'll play Naomi in a minute
normalized time
M5-02a-no focus
M5-02b-[Naomi]
M5-02c-[minute]
(f)
F0 (Hz)
I'll play Naomi in a minute
normalized time
M6-02a-no focus
M6-02b-[Naomi]
M6-02c-[minute]
(g)
F0 (Hz)
I'll play Naomi in a minute
normalized time
M7-02a-no focus
M7-02b-[Naomi]
M7-02c-[minute]
(h)
F0 (Hz)
I'll play Naomi in a minute
normalized time
F1-02a-no focus
F1-02b-[Naomi]
F1-02c-[minute]
(i)
F0 (Hz)
I'll play Naomi in a minute
normalized time
F2-02a-no focus
F2-02b-[Naomi]
F2-02c-[minute]

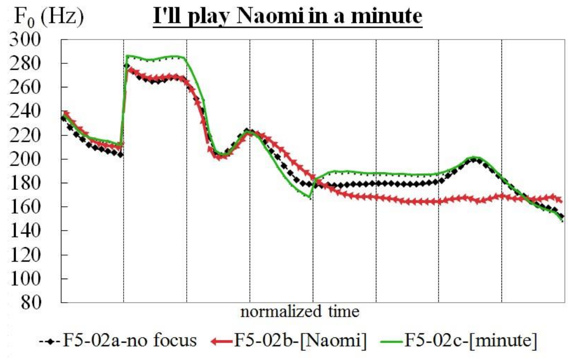
(j)


F0 (Hz)
I'll play Naomi in a minute
300 280 260 240 220 200 180 160 140 120 100 80
normalized time
F3-02a-no focus
F3-02b-[Naomi]
F3-02c-[minute]

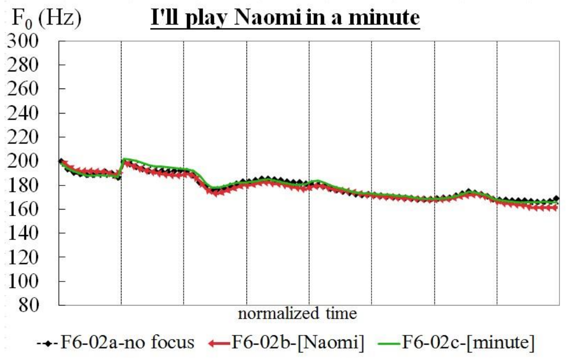
(k)


F0 (Hz)
I'll play Naomi in a minute
300 280 260 240 220 200 180 160 140 120 100 80
normalized time
F4-02a-no focus
F4-02b-[Naomi]
F4-02c-[minute]

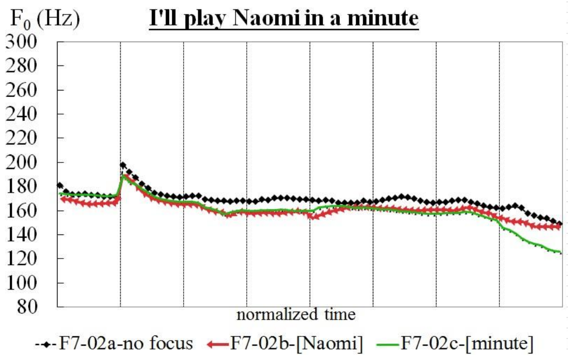
(l)


F0 (Hz)
I'll play Naomi in a minute
300 280 260 240 220 200 180 160 140 120 100 80
normalized time
F5-02a-no focus
F5-02b-[Naomi]
F5-02c-[minute]

(m)


F0 (Hz)
I'll play Naomi in a minute
300 280 260 240 220 200 180 160 140 120 100 80
normalized time
F6-02a-no focus
F6-02b-[Naomi]
F6-02c-[minute]

(n)


F0 (Hz)
I'll play Naomi in a minute
300 280 260 240 220 200 180 160 140 120 100 80
normalized time
F7-02a-no focus
F7-02b-[Naomi]
F7-02c-[minute]

(o)

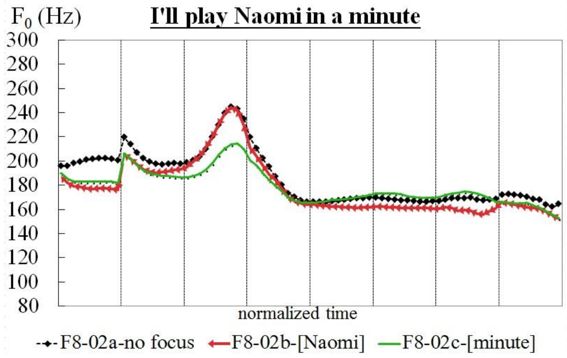
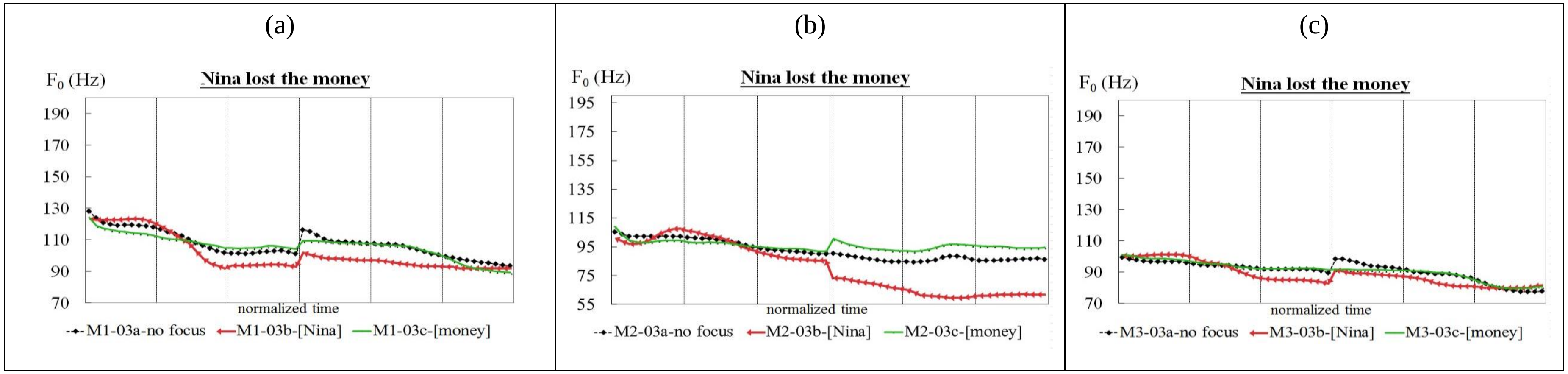
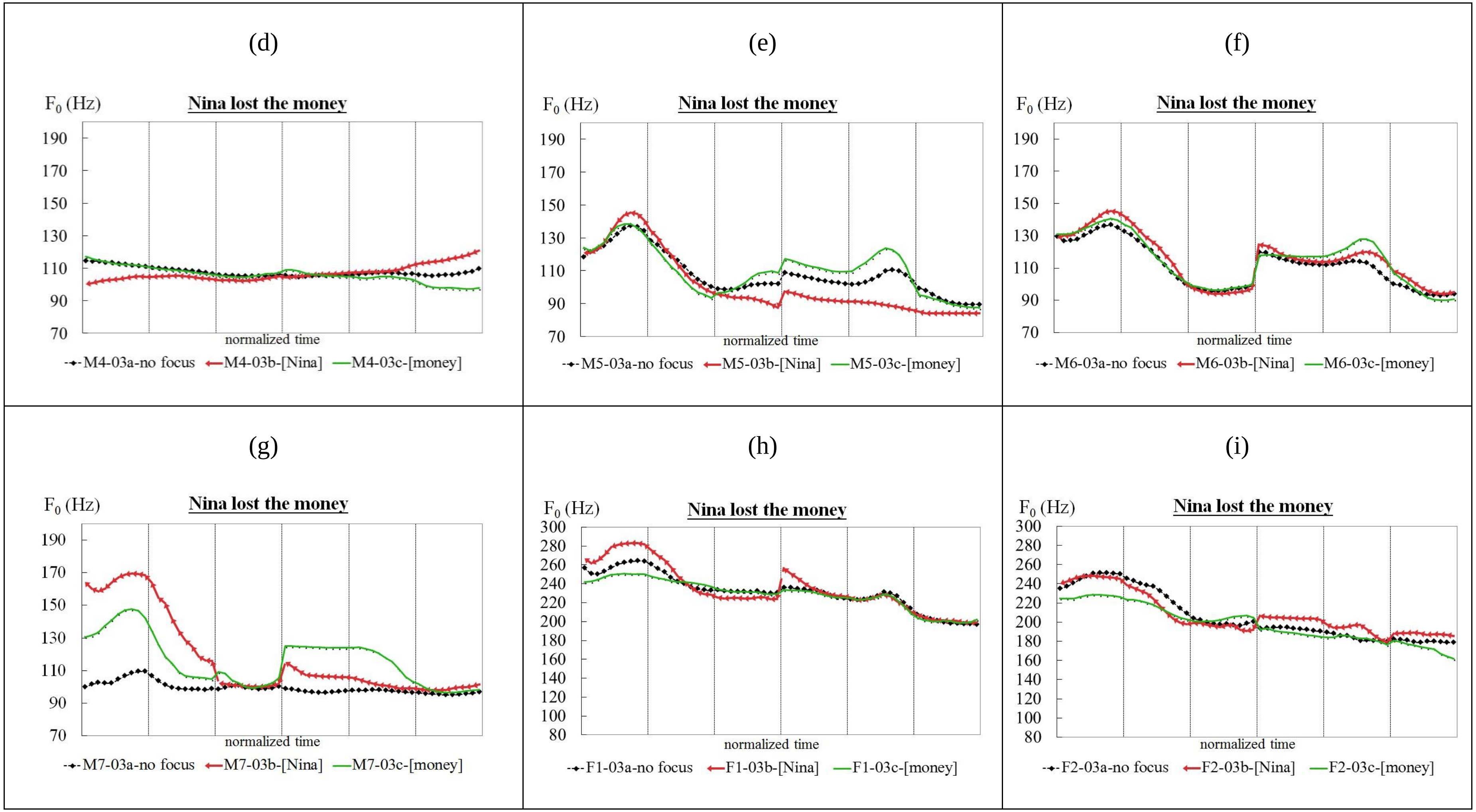

F0 (Hz)
I'll play Naomi in a minute
300 280 260 240 220 200 180 160 140 120 100 80
normalized time
F8-02a-no focus
F8-02b-[Naomi]
F8-02c-[minute]

Figure 4.4 Time-normalized $F_0$ graphs of the English test sentence E3 by the 15 bilingual speakers (7 male speakers, M1 to M7, and 8 female speakers, F1 to F8).




(d)
F0 (Hz)
Nina lost the money
190
170
150
130
110
90
70
normalized time
M4-03a-no focus
M4-03b-[Nina]
M4-03c-[money]
(e)
F0 (Hz)
Nina lost the money
190
170
150
130
110
90
70
normalized time
M5-03a-no focus
M5-03b-[Nina]
M5-03c-[money]
(f)
F0 (Hz)
Nina lost the money
190
170
150
130
110
90
70
normalized time
M6-03a-no focus
M6-03b-[Nina]
M6-03c-[money]
(g)
F0 (Hz)
Nina lost the money
190
170
150
130
110
90
70
normalized time
M7-03a-no focus
M7-03b-[Nina]
M7-03c-[money]
(h)
F0 (Hz)
Nina lost the money
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F1-03a-no focus
F1-03b-[Nina]
F1-03c-[money]
(i)
F0 (Hz)
Nina lost the money
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F2-03a-no focus
F2-03b-[Nina]
F2-03c-[money]

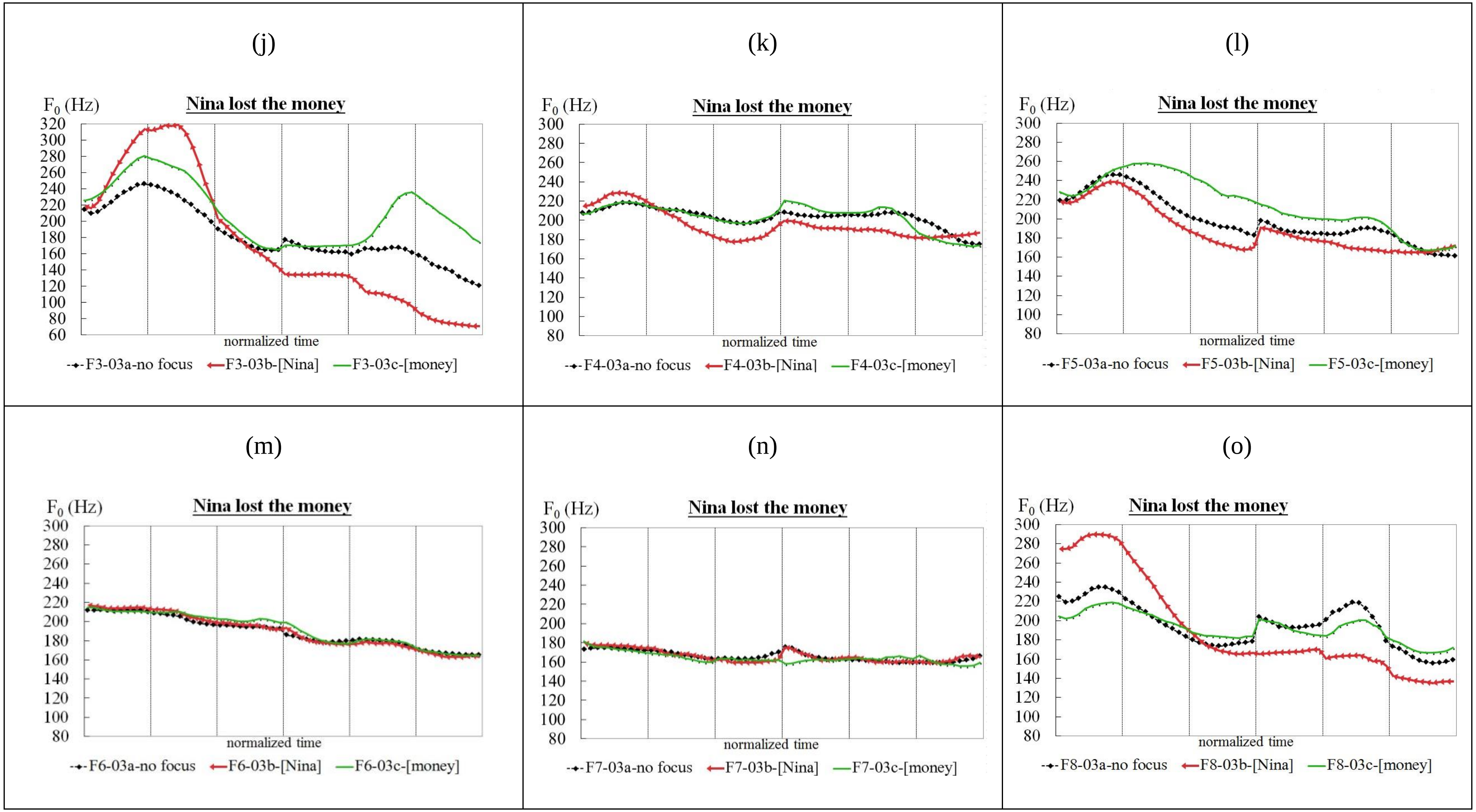
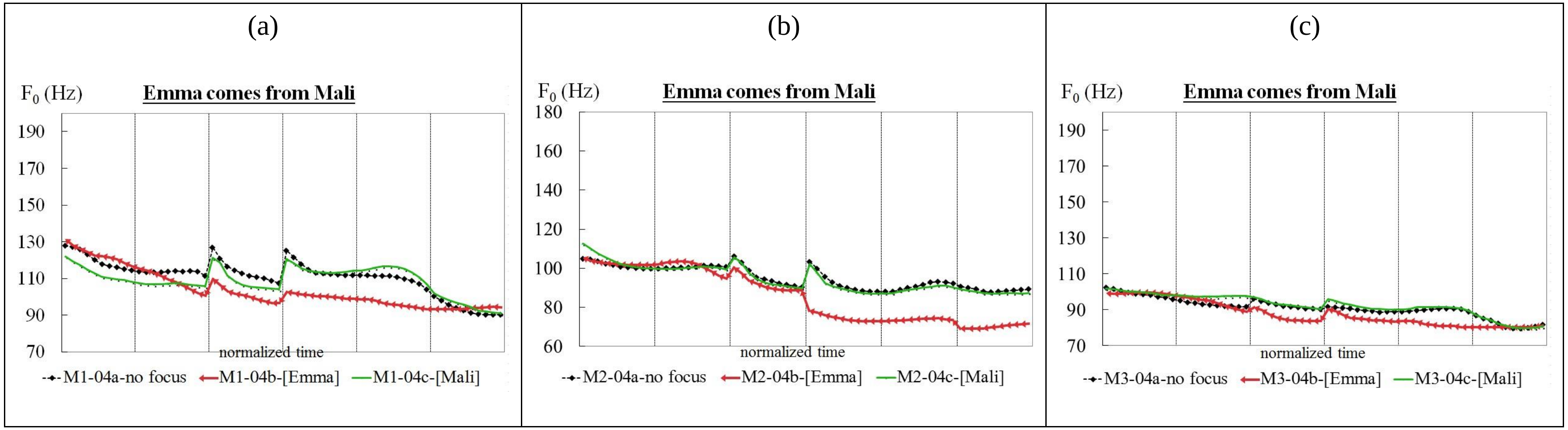
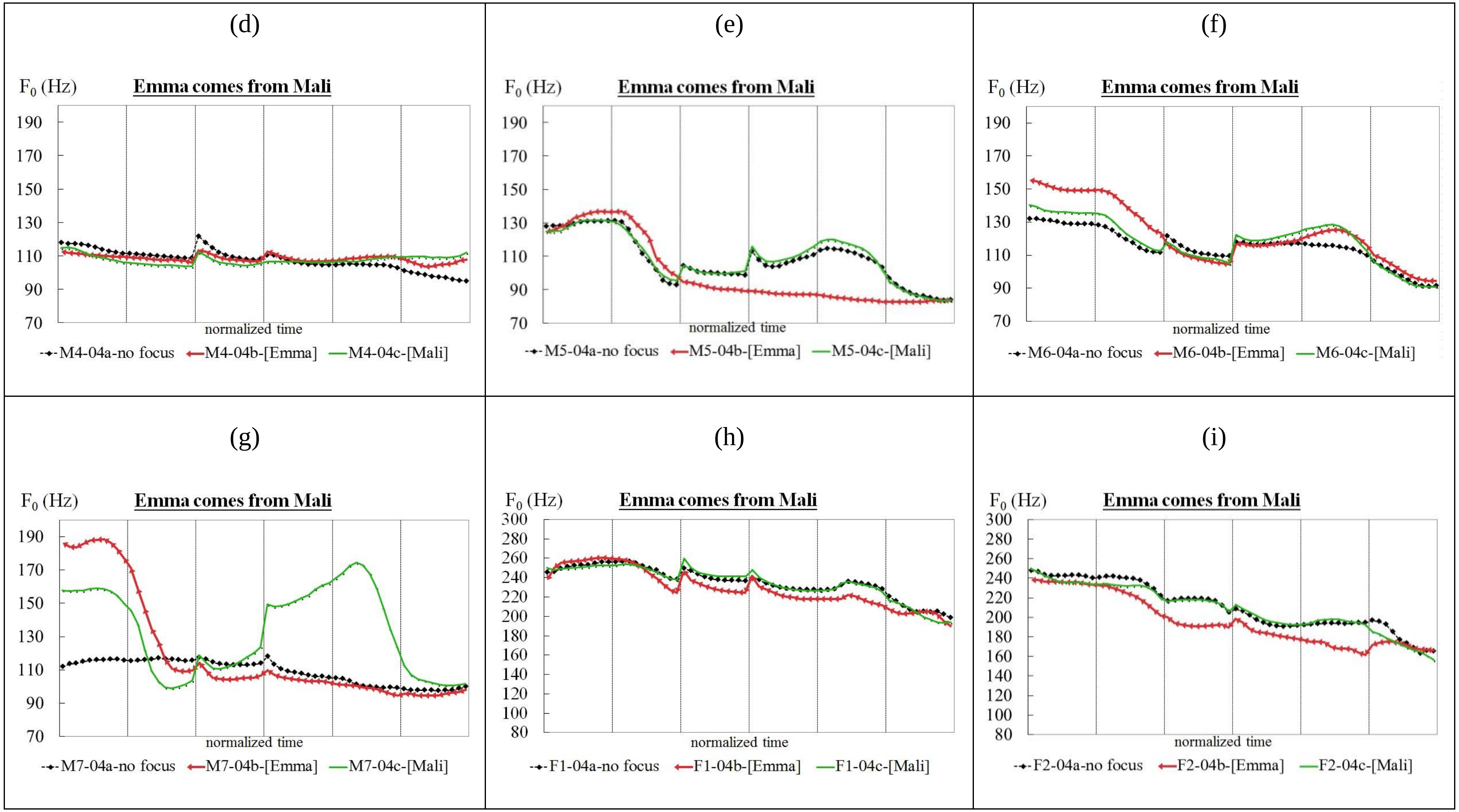
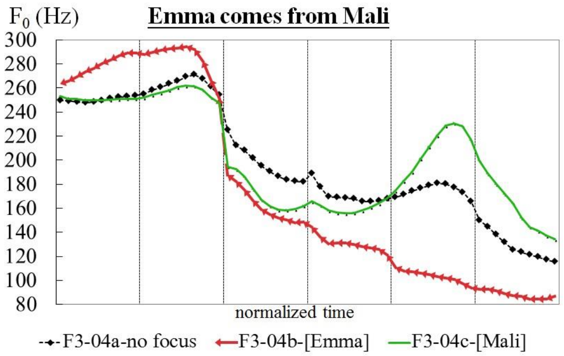

(j)
Nina lost the money
F0 (Hz)
normalized time
F3-03a-no focus
F3-03b-[Nina]
F3-03c-[money]
(k)
Nina lost the money
F0 (Hz)
normalized time
F4-03a-no focus
F4-03b-[Nina]
F4-03c-[money]
(l)
Nina lost the money
F0 (Hz)
normalized time
F5-03a-no focus
F5-03b-[Nina]
F5-03c-[money]
(m)
Nina lost the money
F0 (Hz)
normalized time
F6-03a-no focus
F6-03b-[Nina]
F6-03c-[money]
(n)
Nina lost the money
F0 (Hz)
normalized time
F7-03a-no focus
F7-03b-[Nina]
F7-03c-[money]
(o)
Nina lost the money
F0 (Hz)
normalized time
F8-03a-no focus
F8-03b-[Nina]
F8-03c-[money]

Figure 4.5 Time-normalized $F_0$ graphs of the English test sentence E4 by the 15 bilingual speakers (7 male speakers, M1 to M7, and 8 female speakers, F1 to F8).

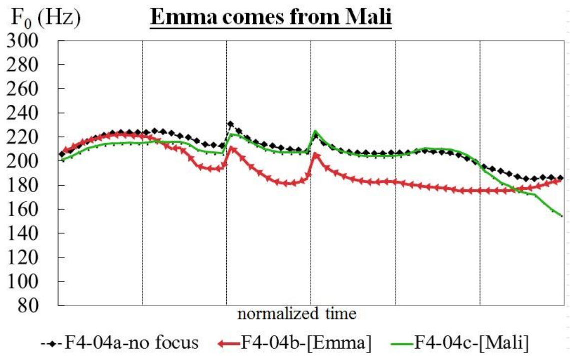



(d)
F0 (Hz)
Emma comes from Mali
190
170
150
130
110
90
70
normalized time
M4-04a-no focus
M4-04b-[Emma]
M4-04c-[Mali]
(e)
F0 (Hz)
Emma comes from Mali
190
170
150
130
110
90
70
normalized time
M5-04a-no focus
M5-04b-[Emma]
M5-04c-[Mali]
(f)
F0 (Hz)
Emma comes from Mali
190
170
150
130
110
90
70
normalized time
M6-04a-no focus
M6-04b-[Emma]
M6-04c-[Mali]
(g)
F0 (Hz)
Emma comes from Mali
190
170
150
130
110
90
70
normalized time
M7-04a-no focus
M7-04b-[Emma]
M7-04c-[Mali]
(h)
F0 (Hz)
Emma comes from Mali
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F1-04a-no focus
F1-04b-[Emma]
F1-04c-[Mali]
(i)
F0 (Hz)
Emma comes from Mali
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
F2-04a-no focus
F2-04b-[Emma]
F2-04c-[Mali]

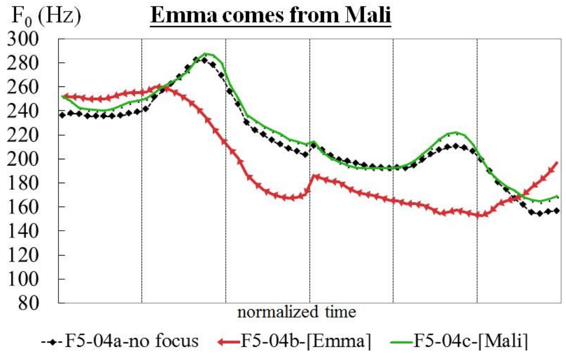
(j)


F0 (Hz)
Emma comes from Mali
normalized time
F3-04a-no focus
F3-04b-[Emma]
F3-04c-[Mali]

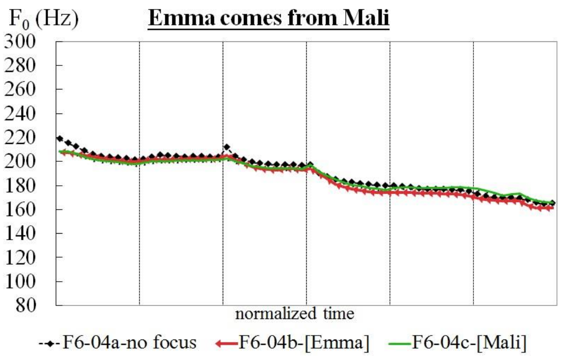
(k)


F0 (Hz)
Emma comes from Mali
normalized time
F4-04a-no focus
F4-04b-[Emma]
F4-04c-[Mali]

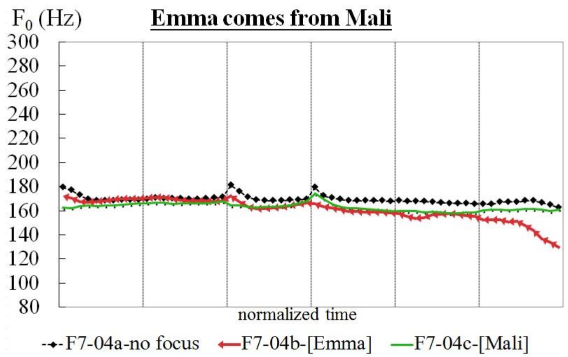
(l)


F0 (Hz)
Emma comes from Mali
normalized time
F5-04a-no focus
F5-04b-[Emma]
F5-04c-[Mali]

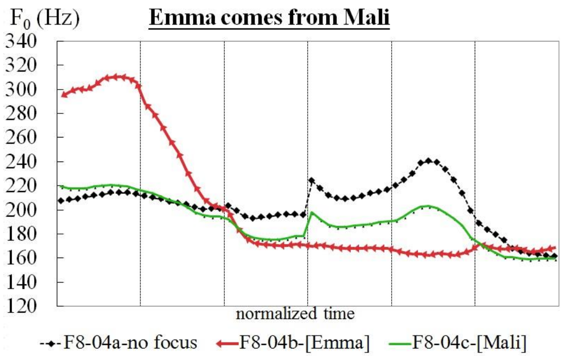
(m)


F0 (Hz)
Emma comes from Mali
normalized time
F6-04a-no focus
F6-04b-[Emma]
F6-04c-[Mali]

(n)


F0 (Hz)
Emma comes from Mali
normalized time
F7-04a-no focus
F7-04b-[Emma]
F7-04c-[Mali]

(o)


F0 (Hz)
Emma comes from Mali
normalized time
F8-04a-no focus
F8-04b-[Emma]
F8-04c-[Mali]

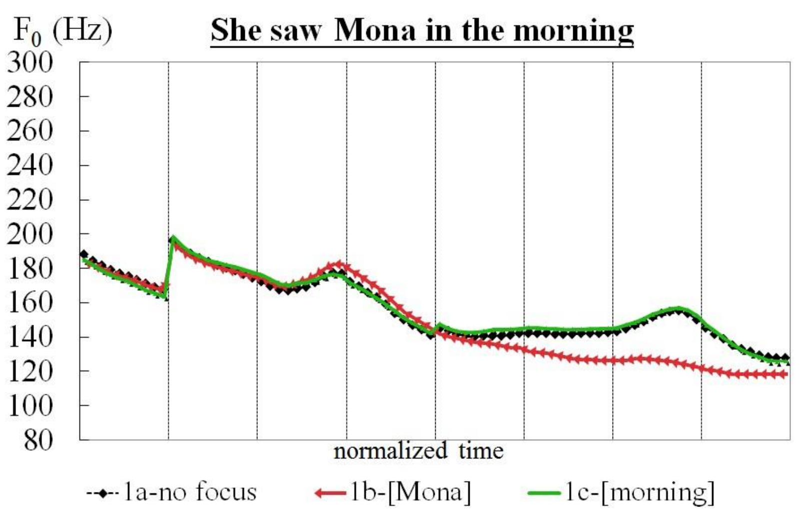
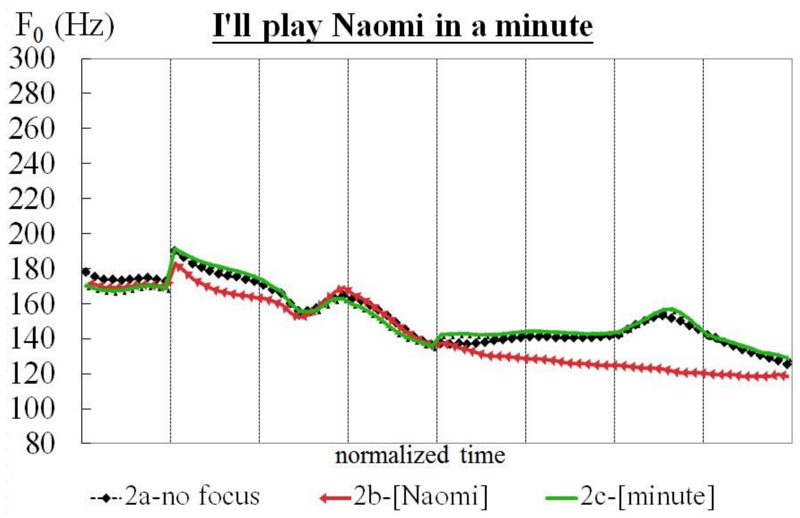
Based on the $F_0$ contours of individual speakers in Figures 4.2 to 4.5, we identified three groups of bilingual speakers according to the presence/absence of PFC, as summarized in Table 4.5. Group A (M1, 5, F3, 4, 5) shows evidence of PFC in all the English sentences; Group B (M2, 3, 7, F2, 8) shows PFC in some sentences; and Group C (M4, 6, F1, 6, 7) shows no sign of PFC in any of the sentences.

Table 4.5 Summary of the presence and absence of PFC in the English test sentences based on graphical analysis. M1 to M7 are male speakers; F1 to F8 are female speakers; + refers to the presence of PFC; – refers to the absence of PFC.

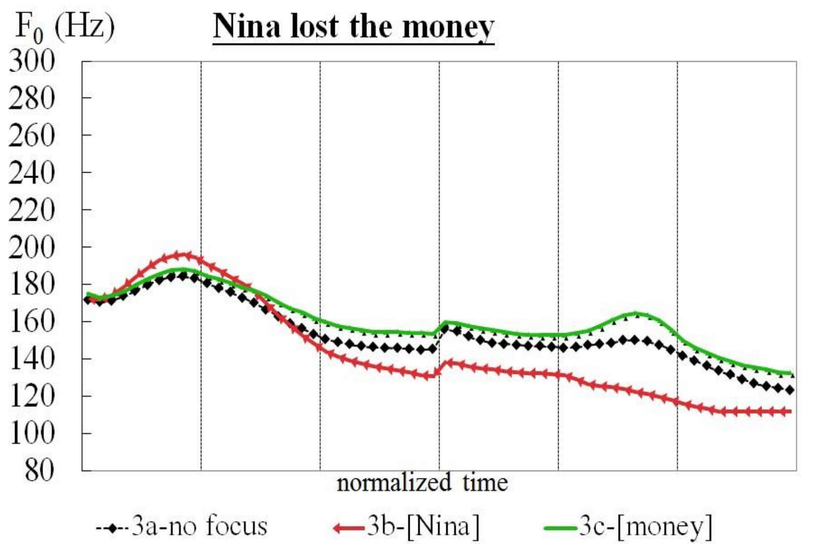
| Speaker/Sentence set | E1 | E2 | E3 | E4 |
|---|---|---|---|---|
| M1 | + | + | + | + |
| M2 | – | – | + | + |
| M3 | – | – | + | + |
| M4 | – | – | – | – |
| M5 | + | + | + | + |
| M6 | – | – | – | – |
| M7 | – | + | – | – |
| F1 | – | – | – | – |
| F2 | – | – | – | + |
| F3 | + | + | + | + |
| F4 | + | + | + | + |
| F5 | + | + | + | + |
| F6 | – | – | – | – |
| F7 | – | – | – | – |
| F8 | – | – | + | + |

We then averaged the mean time-normalized $F_0$ contours in Figures 4.2-4.5 according to the aforementioned three groups. These average contours are presented in Figures 4.6-4.8.

Figure 4.6 Time-normalized $F_0$ graphs of the 4 English test sentences (a) to (d), corresponding to Sentence Sets E1 to E4, averaged across the 5 bilingual speakers who show PFC in all the 4 test sentences in the graphical analyses (Table 4.5).

| (a) | (b) |
| --- | --- |
|  |  |

(c)


F0 (Hz)
Nina lost the money
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
3a-no focus
3b-[Nina]
3c-[money]

(d)

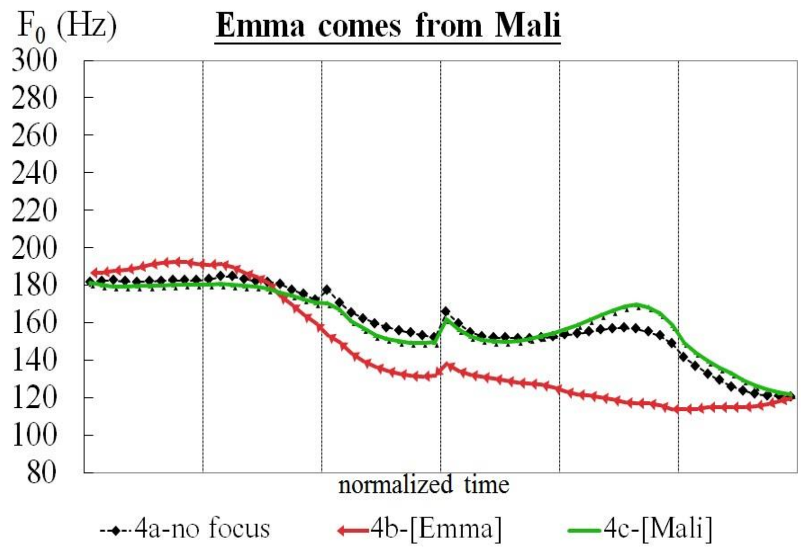
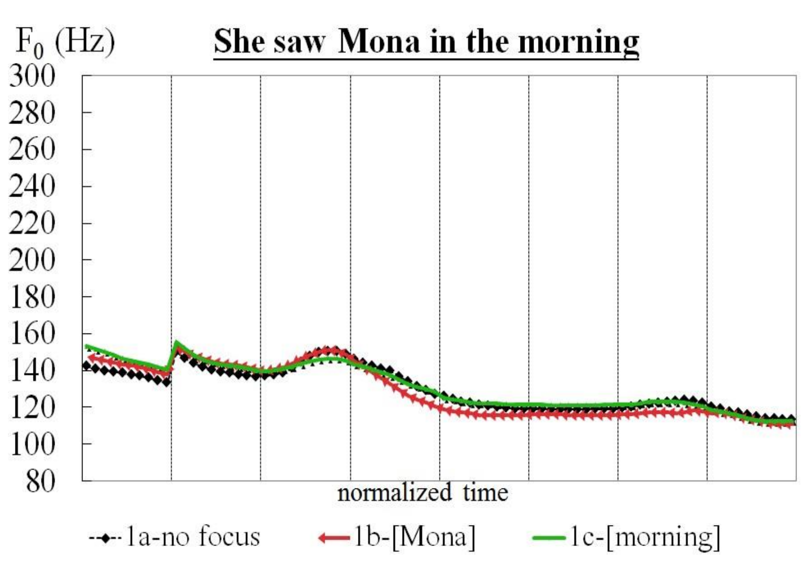
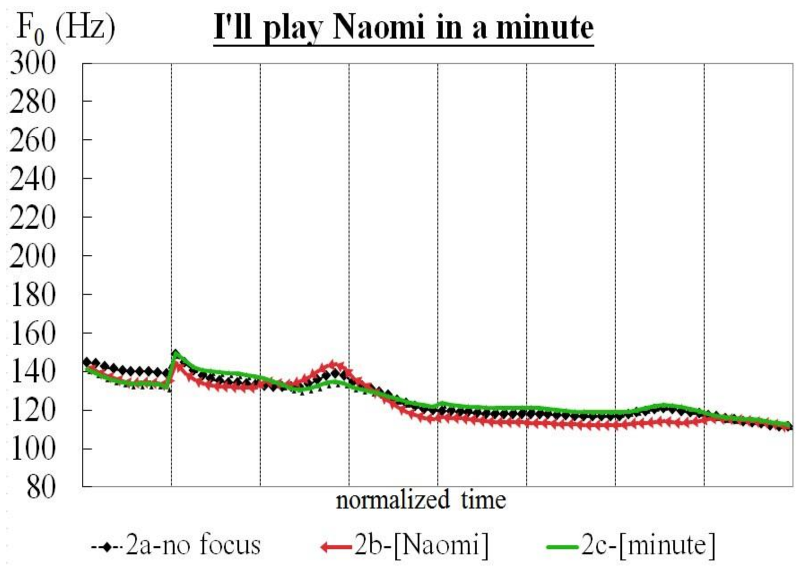
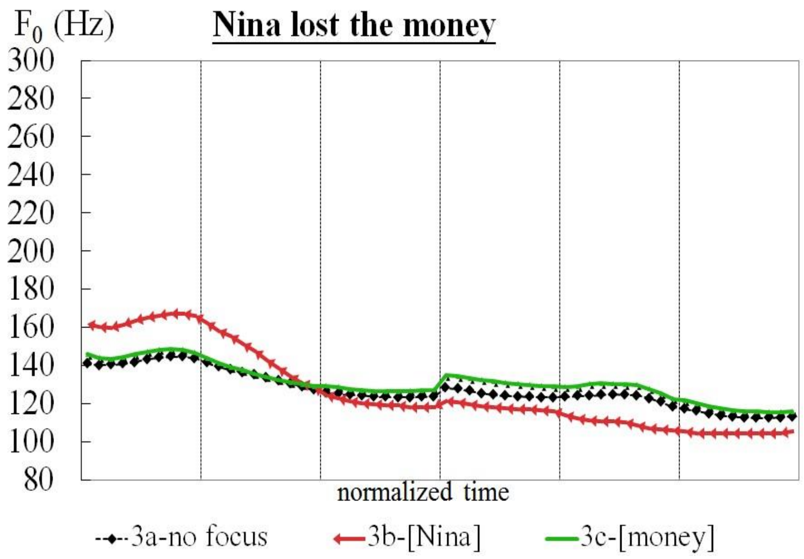

F0 (Hz)
Emma comes from Mali
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
4a-no focus
4b-[Emma]
4c-[Mali]

Figure 4.7 Time-normalized $F_0$ graphs of the 4 English test sentences (a) to (d), corresponding to Sentence Sets E1 to E4, averaged across the 5 bilingual speakers who show PFC in some of the test sentences in the graphical analyses (Table 4.5).

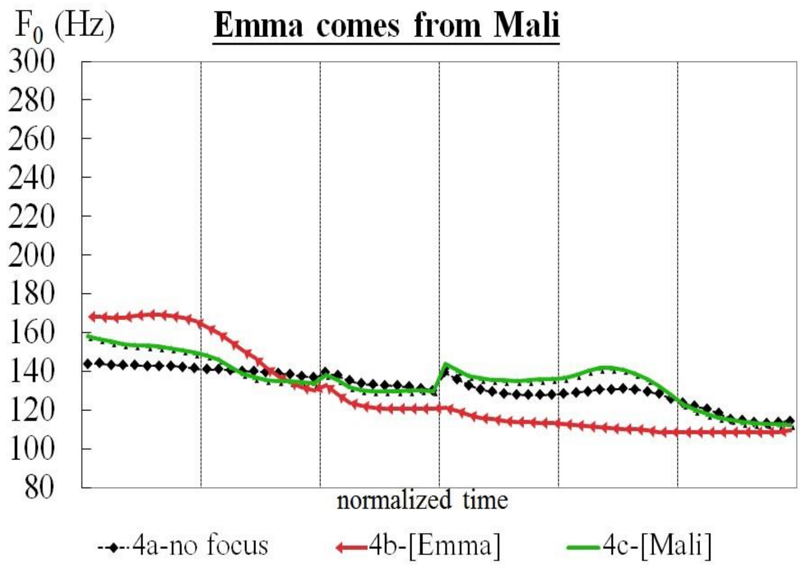
| (a) | (b) |
|---|---|
|   |   |

(c)



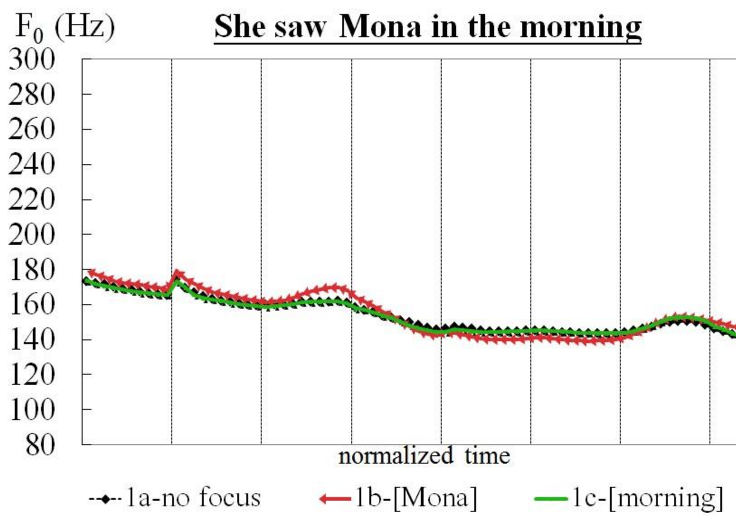
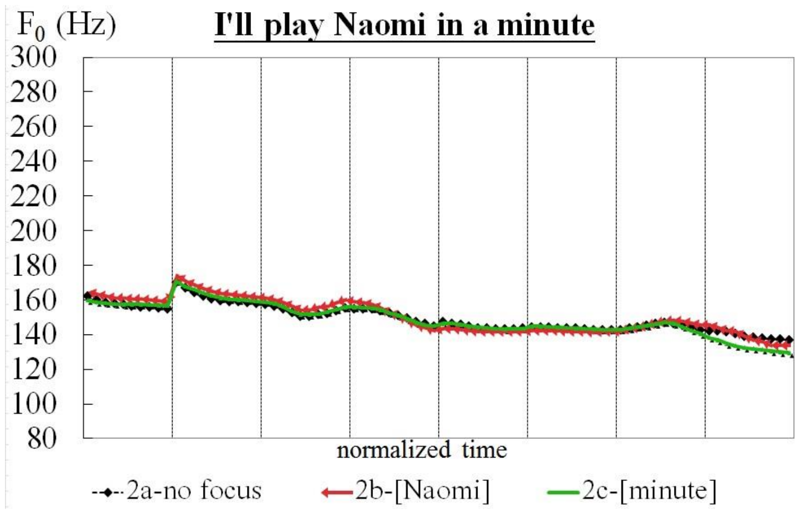
F0 (Hz)
Nina lost the money
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
3a-no focus
3b-[Nina]
3c-[money]


(d)

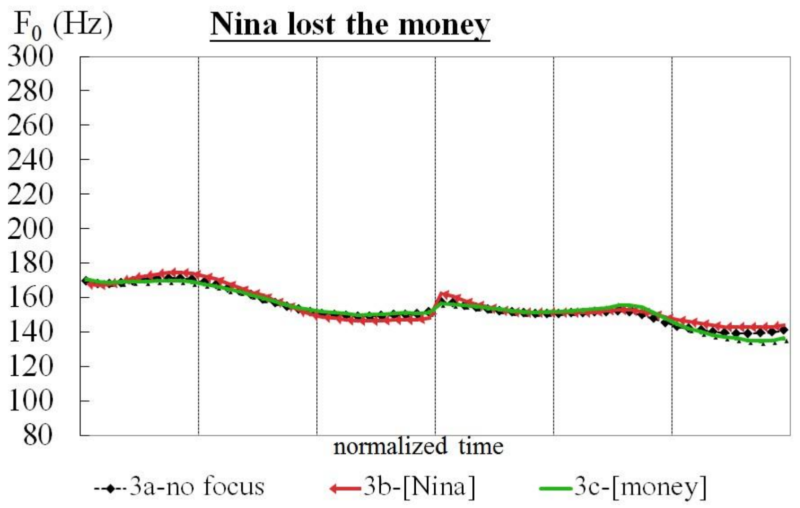


F0 (Hz)
Emma comes from Mali
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
4a-no focus
4b-[Emma]
4c-[Mali]

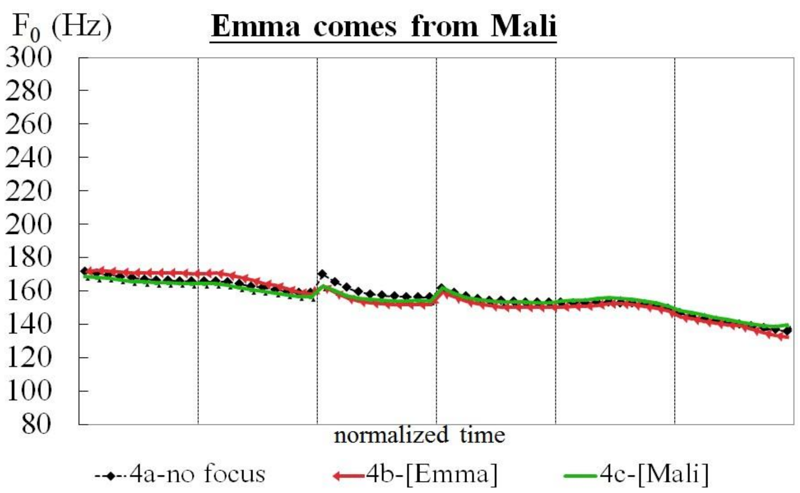
Figure 4.8 Time-normalized $F_0$ graphs of the 4 English test sentences (a) to (d), corresponding to Sentence Sets E1 to E4, averaged across the 5 bilingual speakers who do not show PFC in the graphical analyses (Table 4.5).

| (a) | (b) |
| --- | --- |
|  |  |

(c)


F0 (Hz)
Nina lost the money
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
3a-no focus
3b-[Nina]
3c-[money]

(d)


F0 (Hz)
Emma comes from Mali
300
280
260
240
220
200
180
160
140
120
100
80
normalized time
4a-no focus
4b-[Emma]
4c-[Mali]

For Group A speakers (Figure 4.6), the visual impression is that the mean $F_0$ tracings of the two longer sentences (Sentences E1 and E2, Figures 4.6a and 4.6b) are similar to each other, and so are the two shorter sentences (Sentences E3 and E4, Figures 4.6c and 4.6d). For Sentences E1 and E2 in Figures 4.6a and 4.6b, the average $F_0$ of the on-focus word is not much different from that of the neutral focus word. In contrast, when “Mona” or “Naomi” is in focus, the post-focus $F_0$ is clearly lower than in the neutral-focus counterpart. For Sentences E3 and E4 in Figure 4.6c and 4.6d, both on-focus $F_0$ increase and post-focus $F_0$ lowering can be observed. In summary, the longer sentences do not show much change in $F_0$ for the on-focus words, but post-focus $F_0$ lowering is observable, while the shorter sentences show both on-focus $F_0$ increase and post-focus $F_0$ lowering.

For Group B speakers in Figure 4.7, the two short sentences (Figures 4.7c and 4.7d) exhibit similar on-focus $F_0$ increase and post-focus $F_0$ lowering. However, there is little on-focus or post-focus $F_0$ change in the long sentences. Finally, Group C speakers in Figure 4.8 show virtually no variations in $F_0$ in either on-focus or post-focus words in any of the sentences from the corresponding sentences with neutral focus.

### 4.7.2 Statistical analyses

To support the graphical analysis, two-way repeated measures ANOVAs, with sentence length and focus as factors, were performed on max $F_0$ and mean intensity of the on-focus and post-focus words and their counterparts in the neutral focus condition. For post-focus max $F_0$, the value was computed based on only the last three syllables because the first syllable immediately after focus is subject to strong carryover influence from the preceding syllable, as can be seen in the $F_0$ graphs. The results are presented in Table 4.6.

Table 4.6 One-way repeated measures ANOVA results of the English data by three groups of speakers based on the time-normalized $F_0$ graphs, whether they seem to have PFC in all the test sentences (Group A), in some of them (Group B) or none of them (Group C). For each measurement, the comparison is between the on-focus or post-focus items with their neutral counterparts. In the tables, comparisons for sentence length and focus conditions are made, where L = longer sentences; S = shorter sentences; F = on-focus, PF = post-focus, N = neutral focus. ↓ specifies that in the pair of items compared, the former has a smaller value than the latter.

(a) max F0

(i) On-focus vs. Neutral focus

| | Group A | | Group B | | Group C | |
|---|---|---|---|---|---|---|
| Factors | F | *p* | F | *p* | F | *p* |
| Focus (F vs. N) | 2.495 | 0.189 | 1.079 | 0.358 | 0.426 | 0.550 |
| Sentence length (L vs. S) | 5.387 | 0.081 | 15.048 | ***0.018*** ↓ | 137.2 | ***<0.001*** ↓ |
| Focus x Sentence length | 1.641 | 0.269 | 3.116 | 0.152 | 0.002 | 0.964 |

(ii) Post-focus vs. Neutral focus

| | Group A | | Group B | | Group C | |
|---|---|---|---|---|---|---|
| Factors | F | *p* | F | *p* | F | *p* |
| Focus (PF vs. N) | 11.458 | ***0.028*** ↓ | 3.250 | 0.146 | 1.066 | 0.360 |
| Sentence length (L vs. S) | 8.098 | ***0.047*** ↓ | 1.057 | 0.362 | 12.74 | ***0.023*** ↓ |
| Focus x Sentence length | 0.122 | 0.744 | 1.585 | 0.277 | 0.105 | 0.762 |

For Group A speakers, max $F_0$ of on-focus words is not significantly different from that of the neutral-focus words, and there is no difference between long and short sentences. Similarly, there is no statistical difference in mean intensity between on-focus and neutral words, but sentence length does show a significant difference. In post-focus words, both max $F_0$ and mean intensity are lower than in neutral focus words, confirming what can be observed in Figure 4.6. There is also a significant difference in post-focus max $F_0$ due to sentence length. But it is the longer sentences that have lower post-focus max $F_0$, apparently because the early focus is on the first word in the short sentence, but on the third word in the longer sentences. This gives more post-focus words relative to the entire sentence in the short sentences than in the long sentences.

For Group B speakers (Figure 4.7), on-focus and neutral-focus words do not show significant differences in either max $F_0$ or mean intensity. There is a significant lowering of mean intensity in post-focus words relative to neutral-focus words, but this is not matched by a significant difference in max $F_0$. Max $F_0$ of on-focus words is significantly lower in longer than in shorter sentences, apparently because of lack of on-focus $F_0$ raising in the former.

For Group C speakers, there is no significant difference in max $F_0$ or mean intensity due to focus. But there is a significant difference in post-focus max $F_0$ due to sentence length, which, judging from Figure 4.8, has little to do with focus, but is likely due to greater declination in a longer than in a shorter sentence.

## 4.8 <u>Discussion</u>

In this chapter, we have tested two hypotheses: (a) PFC can be transferred from English to Cantonese through simultaneous bilingualism, and (b) PFC is unaffected in the English of bilingual speakers. For the Cantonese sentences produced by both Hong Kong speakers and bilingual speakers in England, we have compared the on-focus and post-focus words to the neutral-focus words in terms of $F_0$, intensity, excursion size and duration. Both Hong Kong and bilingual speakers show similar on-focus and post-focus $F_0$ variations in that the on-focus change in $F_0$ is minimal for the static tones (Tones 1, 3, 4 and 6). The on-focus expansion in $F_0$ is more obvious for the dynamic tones (Tones 2 and 5) of the Hong Kong speakers than the bilingual speakers, but there is considerable on-focus increase in both duration and intensity for both groups of speakers. More importantly, post-focus lowering in $F_0$ or intensity is absent in both groups of speakers.

For English, based on graphical inspection, three groups of bilingual speakers were identified. The first (Group A) consistently lowered post-focus $F_0$ in all sentences; the second (Group B) lowered post-focus $F_0$ in some sentences; and the third (Group C) did not lower post-focus $F_0$ in any of the test sentences.

The aim of the production experiments is to provide evidence as to whether PFC is likely to be transferred from a language with PFC to another without PFC, and our findings hint that the answer is negative. We have seen that for Group A bilingual speakers, who show PFC in their English, their Cantonese does not show PFC, and for Group C speakers, they do not show PFC in both their English and Cantonese. Based on this observation, we believe that PFC can be lost but not gained across languages. Also, the different behaviours of post-focus $F_0$ variations of the bilingual speakers suggest that there is a continuum of the realization of PFC in the bilinguals, where some speakers retain this feature, while there is a partial loss of this feature in some of the speakers and total loss in the rest of the speakers.

Although it is possible that focus realization of the bilingual subjects in this study is not representative of all bilinguals, and that the groupings above may reflect different levels of proficiency in English or Cantonese, there is little doubt that these speakers' spoken English is of native or near-native proficiency, since all these speakers were born and raised in the UK and received education up to at least the undergraduate level. Further studies are required to ascertain the representativeness of the present findings, but the results are nevertheless of important value in showing the possible cross-linguistic influence on focus realization in bilinguals.

As cross-linguistic research on the existence of PFC increases, a language atlas based on its existence will show interesting links between previously seemingly unrelated languages. Our present study provides evidence for the non-transferability of PFC across languages, and if more upcoming researches have similar conclusions, then our currently held positions towards language families may have to be revised.

# Chapter 5 - Conclusions

## 5.1 About Cantonese SFPs

Cantonese SFPs have previously been described as having tones similar to the ordinary lexical tones. The present study, for the first time, has provided systematic experimental data in support of this long-held impression. The SFPs are referred to by different phonetic segments and lexical tones, and this now seems to be an appropriate notation given the close resemblance of the $F_0$ of the SFPs with lexical tones. On the other hand, the traditional view has ignored the interaction of tone and intonation on the SFPs, and the contribution of sentence intonation when discussing the meanings and functions of SFPs. From the $F_0$ contours and acoustic measurements obtained in this study, we found that despite the resemblance of SFPs to the lexical tones in contour shape, differences in final $F_0$, $F_0$ velocity and duration exist in some cases. These differences are attributable to the pragmatic functions performed by the SFPs. In particular, sentences containing question SFPs display an intonation pattern similar to that of the intonation questions, indicating that the $F_0$ of SFPs is a mixture of both intonation and lexical tones. More interestingly, we have made two surprising findings. The first is that some of the SFPs, which have been suggested to be heard as lively or animated, show much shorter duration than the corresponding Tones-in-statement. The second one is that the SFP [wɔ˧˥˩] (Figure 2.2) seems to contain an additional intonation component that is sequentially attached to the end of a tonal target, which is reminiscent of the utterance-final edge tone in Mandarin (Chao, 1968; Li et al., 2011; Mueller-Liu, 2006). Both of these findings are therefore worth further examination in future research.

The current study is the first to examine $F_0$ contours and duration of Cantonese SFPs in fine detail. As a result, it has raised more questions than definitive answers. Although we have demonstrated with phonetic data that SFPs are likely to be a combination of lexical tone and intonation, it is not possible for us to definitively identify the reason behind the specific deviation of each SFP from its tone counterpart. For example, why is an SFP signalling politeness shorter in duration? Why should the reminder SFP [wɔ˨˥] (Figure 2.2) have longer duration with a likely extra sequential tone? We have tried to suggest possible links based on existing theories and findings, such as Ohala (1984)'s theory of frequency code and Chao (1968)'s observed sentence-final extra tone. But only future studies specifically designed to look into these issues can generate more definitive answers. Our findings here nevertheless have taken a first step in establishing the phonetic patterns for further research.

## 5.2 About prosodic focus

We have tested three hypotheses using two production experiments (Chapters 3 and 4 respectively): (1) Prosodic focus is realized in Hong Kong Cantonese without PFC; (2) PFC can be transferred from English to Cantonese through simultaneous bilingualism, and (3) PFC is unaffected in the English of bilingual speakers. For the Cantonese data, we have compared the on-focus and post-focus words to the neutral-focus words in terms of mean $F_0$, intensity, duration and excursion size. Both Hong Kong and bilingual speakers show similar on-focus and post-focus $F_0$ variations in that the on-focus change in $F_0$ is minimal for the static tones (Tones 1, 3, 4 and 6). The on-focus expansion in $F_0$ is more obvious for the dynamic tones (Tones 2 and 5) of the Hong Kong speakers than the bilingual speakers, but there is considerable on-focus increase in both duration and intensity for both groups of speakers. More importantly, post-focus lowering in $F_0$ or intensity is absent in both groups of speakers.

For the English data, based on graphical inspection, three groups of bilingual speakers were identified. The first (Group A) consistently lowered post-focus $F_0$ in all sentences; the second (Group B) lowered post-focus $F_0$ in some sentences; and the third (Group C) did not lower post-focus $F_0$ in any of the sentences.

One aim of the present experimental study on the speech of English-Cantonese bilinguals is to provide evidence as to whether PFC is likely to be transferred from a language with PFC to another without PFC through contact, and our findings hint that the answer is negative. We have seen that for Group A bilingual speakers, who show PFC in their English, their Cantonese does not show PFC, and for Group C speakers, they do not show PFC in both their English and Cantonese. Based on this observation, we believe that PFC can be lost but not gained across languages. Also, the different behaviours of post-focus $F_0$ variations of the bilingual speakers suggest that there is a continuum of the realization of PFC in the bilinguals, where some speakers retain this feature, while there is partial loss of this feature in some of the speakers and total loss in the rest of the speakers.

As cross-linguistic research on the existence of PFC increases, a language atlas based on its existence will show interesting links between previously seemingly unrelated languages. Our present study provides evidence for the non-transferability of PFC across languages, and if more upcoming researches have similar conclusions, then our currently held positions towards language families may have to be revised.

## 5.3 Contributions of this dissertation

Although much research has been done to identify the tonal and intonational features of Cantonese and their interaction, comprehensive and well-controlled experimental studies are still lacking, especially those concerning the tonal nature of SFPs. With recent advances in acoustic analyses; we see a good opportunity for re-thinking some of the old questions and exploring new ones regarding tone-intonation interaction in general and Cantonese prosody in particular. In addition, owing to the recent finding that there are different focus realizations in closely related languages, it has become interesting to compare post-focus compression cross-linguistically with a similar experimental protocol. The present dissertation aimed at providing an important research framework under which important questions relating to Cantonese prosody can be answered.

The overarching theme of the experimental studies in this dissertation is the surface $F_0$ of Cantonese being an interplay between the systems of lexical tone and intonation, and the two important factors in communication which influence it, namely the SFPs and prosodic focus. The first concern was whether the surface $F_0$ of Cantonese SFPs is derived directly from lexical tones, purely intonational, or due to a combination of tone and intonation (Chapter 2). Previous researchers on Cantonese prosody did not look at what subtle differences there might be between SFPs and ordinary lexical tones. Through a robust experimental design, detailed graphical and statistical analyses, we have demonstrated the nuances previously underexplored. The results demonstrated by the comparisons between the SFPs and ordinary syllables with respect to several acoustic measurements (final $F_0$, final $F_0$ velocity) enable us to say with better confidence that some SFPs are clearly a combination of tone and intonation.

The second area of exploration in this dissertation (Chapters 3 and 4) was prosodic focus in Cantonese by monolingual and bilingual speakers. For the monolingual speakers, the acoustic correlates of focus in words with different lexical tones were analyzed in detail, and a perception test was carried out. The production and perception tests together provided a fuller description of Cantonese prosodic focus than by previous research. Not only was the present account important to Cantonese research per se, it also added to our knowledge about the transferability of PFC across languages. The latter information will be of value to cross-linguistic study of PFC (Xu et al., 2012).

As we studied the surface $F_0$ of Cantonese SFPs, some potential future research has been identified, namely, the perception tests involving pre-final syllables as stimuli for determining the significance of sentence-body $F_0$ in the overall function of the system of SFPs (p. 71). Admittedly, what has been explored in the present dissertation is just the preliminary step in uncovering an intricate story behind the system of SFPs, but nonetheless a significant one along the right direction towards a solid and thorough description of tone-intonation interaction in SFPs. It is hoped that the research method demonstrated in this dissertation can serve as a worthwhile reference for future prosodic studies in general and Cantonese phonetic research in particular.

# References

Abramson, A. S., & Svastikula, K. (1983). Intersections of tone and intonation in Thai. *Status Report on Speech Research, Haskins Laboratories, SR-74/75,* 143-167.

Bahill, A. T., Kallman, J. S., & Lieberman, J. E. (1982). Frequency limitations of the two-point central difference differentiation algorithm. *Biological Cybernetics, 45,* 1-4.

Bamgbose, A. (1966). *A Grammar of Yoruba*. Cambridge: Cambridge University Press.

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

Bauer, R. S., Cheung, K. H., Cheung, P. M., & Ng, L. (2004). Acoustic correlates of focus-stress in Hong Kong Cantonese. In *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001*, Tempe, Arizona State University Program for Southeast Asian Studies Monograph Series, 29-49.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology, 3,* 255-309.

Bhatia, T. K., & Ritchie, W. C. (Eds.) (2004.) *The Handbook of Bilingualism.* Cambridge, MA: Blackwell.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5* (9/10), 341-345.

Bodman, N. C. (1980), Proto-Chinese and Sino-Tibetan: data towards establishing the nature of the relationship", in F. van Coetsem, L., & R. Waugh, (Eds), *Contributions to Historical Linguistics: Issues and Materials* (pp. 34-199). Leiden: E. J. Brill.

Bolinger, D. (1978). Intonation across languages. In J. Greenberg (Ed.), *Universals of Human Language, Volume 2: Phonology* (pp. 471-524). Palo Alto, California: Stanford University Press.

Botinis, A., Fourakis, M., & Gawronska, B. (1999). Focus identification in English, Greek and Swedish. In *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, USA, Vol. 2, 1557-1560.

Bruce, G. (1982). Textual aspects of prosody in Swedish. *Phonetica, 39,* 274-287.

Chang, N. T. (1958). Tones and intonation in the Chengtu dialect (Szechuan, China). *Phonetica, 2*, 59-85.

Chao, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique, 45,* 24-27.

Chao, Y. R. (1933). Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology, 4,* 121-134

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

Chen, S-w, Wang, B., & Xu, Y. (2009). Closely related languages, different ways of realizing focus. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009),* Brighton, UK, 1007-1010.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics, 36,* 724-746.

Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). A study of sentence stress production in Mandarin speakers of American English. *Journal of the Acoustical Society of America, 109* (4), 1681-1690.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech – Evidence from $F_0$ patterns of neutral tone in standard Chinese. *Phonetica, 63,* 47-75.

Cheung, H. N. 張洪年 (2007). *香港粵語語法的研究* [A Grammar of Cantonese Spoken in Hong Kong] (Revised edition). Hong Kong: The Chinese University Press.

Cheung, H. N. 張洪年 (2009). *Cantonese Made Easy: 早期粵語中的語氣助詞* [Cantonese Made Easy: Sentence-final particles in early Cantonese]. *Bulletin of Chinese Linguistics, 3* (2), 131-169.

Chow, D. (2010). Tone, intonation and tonogenesis – the case of utterance final particles. Paper presented at *The Montreal-Ottawa-Toronto Phonology Workshop*. Carleton University, Canada. March, 2010.

Chow, S. (2001). The intonation patterns in Hong Kong Cantonese. In L.-H. Cai, et al. (Eds.), *Proceedings of the 5th National Conference on Modern Phonetics* (pp. 137-141). Beijing: Tsinghua University Press.

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding emotions in speech with the size code — A perceptual investigation. *Phonetica, 65,* 210-230.

Chung, L. (2010). *Prosodic Focus Production by English-Cantonese Bilinguals.* Unpublished linguistic project in IBSC Speech Science and Communication. London, UK: University College London.

Connell, B. A., Hogan J. T., & Rozsypal A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics, 11,* 337-351.

Cooper, W. E., Eady, S. J. & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America, 77* (6), 2142-2156.

Crystal, D. (1969). *Prosodic Systems and Intonation in English.* Cambridge: Cambridge University Press.

de Jong, K. J., & Zawaydeh, B. A. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30,* 53-75.

Dohen, M., & Lœvenbruck, H. (2004). Pre-focal rephrasing, focal enhancement and post-focal deaccentuation in French. In *Proceedings of The 8th International Conference on Spoken Language Processing*, Jeju, Korea, 1313-1316.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, *80* (2), 402-416.

Fang, X. 方小燕 (2003). *廣州方言句末語氣助詞* [Sentence-final Mood Helping-words in Guangzhou Dialect]. Guangzhou, China: Jinan University Press.

Fernald, A. & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27* (2), 209-221.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics, 36* (4), 680-703.

Flege, J. (1995). The phonetic study of bilingualism. In G. Bloothooft, V. Hazan, D. Huber & J. Llisterri (Eds.) *European Studies in Phonetics and Speech Communication* (pp. 98-103). Utrecht: OTS Publications.

Fok-Chan, Y. Y. (1974). *A Perceptual Study of Tones in Cantonese*. Hong Kong: University of Hong Kong Press.

Fox, A., Luke, K. K., & Nancarrow, O. T. (2008). Aspects of intonation in Cantonese. *Journal of Chinese Linguistics*, *36* (2), 321-367.

Fry, D. B. (1968). Prosodic phenomena. In B. Malmberg (Ed.), *Manual of Phonetics* (pp. 365-410). Amsterdam: North Holland.

Fung, R. S-Y. (2000). *Final Particles in Standard Cantonese: Semantic Extension and Pragmatic Inference*. PhD Dissertation, the Ohio State University.

Gandour, J., Siripong, P., & Sumalee, D. (1994). Tonal coarticulation in Thai.

*Journal of Phonetics, 22,* 477-492.
Gibbons, J. (1980). A tentative framework for speech act description of the utterance particles in conversational Cantonese. *Linguistics, 18,* 763-775.
Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21,* 205-228.
Gu, W., & Lee, T. (2007). Effects of focus on prosody of Cantonese speech – A comparison of surface feature analysis and model-based analysis. In *Proceedings of the International Workshop Paralinguistic Speech '07,* Saarbrücken, Germany, 59-64.
Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
Ḥa, K. P. (2010). Prosody of Vietnamese from an interactional perspective: ờ, ừ and vâng in backchannels and requests for information. *Journal of the Southeast Asian Linguistics Society, 3* (1), 56-76.
Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica, 33*, 353-367.
Ho, A. T. (1977). Intonation variations in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative. *Phonetica*, *34,* 446-456.
House, D. (2005). Phrase-final rises as a prosodic feature in wh-questions in Swedish human–machine dialogue. *Speech Communication*, *46,* 268-283.
Huang, B. H., & Jun, S-A. (2011). The effect of age on the acquisition of second language prosody. *Language and Speech*, *54* (3), 387-414.
Ipek, C. (2011). Phonetic realization of focus with no on-focus pitch range expansion in Turkish. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong.
Ishihara, S. (2001). Stress, focus and scrambling. In E. Guerzoni, & O. Matushansky, (Eds.). *A Few from Building E-39, MIT Working Papers in Linguistics 39* (pp. 151-185). Cambridge: MIT Department of Linguistics and Philosophy.
Ishihara, S. (2002). Syntax-phonology interface of Wh-constructions in Japanese. In *Proceedings of Tokyo Conference on Psycholinguistics 2002 (TCP 2002)*, Tokyo, 165-189.
Jin, S. (1996). *An Acoustic Study of Sentence Stress in Mandarin Chinese*. Ph.D. Dissertation, the Ohio State University.

Kohler, K. (2005). Timing and communicative functions of pitch contours. *Phonetica, 62,* 88-105.

Kormos, J. (2006). *Speech Production and Second Language Acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica, 55,* 243-276.

Kwok, H. (1984). *Sentence Particles in Cantonese.* Hong Kong: Centre of Asian Studies.

Lam, K. P. 林建平 (2002). *香港粵語句調研究* [A Study of Intonation in Hong Kong Cantonese]. PhD Dissertation, Chinese University of Hong Kong.

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

Law, S. P. (1990). *The Syntax and Phonology of Cantonese Sentence-final Particles*. PhD Dissertation, Boston University.

Lee, B, Guion, S. G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition, 28,* 487-513.

Lee, W. S. (2004). The effect of intonation on the citation tones in Cantonese. In *Proceedings of the First International Symposium on the Tonal Aspects of Languages*, Beijing, China, 107-110.

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

Leung, C. S. 梁仲森 (1992). 香港粵語助詞的研究 [A Study of the Utterance Particles in Cantonese as Spoken in Hong Kong]. MPhil Dissertation, Hong Kong Polytechnic.

Leung, W. M. (2010). *On the Synchrony and Diachrony of Sentence-Final Particles: The Case of WO in Cantonese*. Saarbrücken, Germany: Lambert Academic Publishing.

Lewis, M. P. (Ed.). (2009). *Ethnologue: Languages of the World* (16th ed.). Dallas, Texas: SIL International.

Li, A., Fang, Q., & Dang, J. (2011). Emotional intonation in a tone language: Experimental evidence from Chinese. In *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong, 1198-1201.

Li, B. (2006). *Chinese Final Particles and the Syntax of the Periphery*. PhD

Dissertation, Leiden University.

Lindau, M. (1986). Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America, 80,* 757-764.

Lindsay, D., & Ainsworth, W. A. (1985). Two models of nuclear intonation. *Journal of Phonetics, 13,* 163-173.

Linguistic Society of Hong Kong 香港語言學學會 (2002). *粵語拼音字表* [*Guide to LSHK Cantonese Romanization of Chinese Characters*]. (2nd Ed.). Hong Kong: Linguistic Society of Hong Kong.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*, 70-87.

Liu, F., & Xu, Y. (2007a). Question intonation as affected by word stress and focus in English. In *Proceedings of The 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1189-1192.

Liu, F., & Xu, Y. (2007b). The neutral tone in question intonation in Mandarin. In *Proceedings of Interspeech 2007*, Antwerp, Belgium, 630-633.

Luke, K. K. (1990). *Utterance Particles in Cantonese Conversation.* Amsterdam, Philadelphia: John Benjamins Publishing Company.

Ma, J. K-Y., Ciocca, V., & Whitehill, T. L. (2004). The effects of intonation patterns on lexical tone production in Cantonese. Paper presented at *the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, Beijing, China.

Ma, J. K-Y., Ciocca, V. & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *Journal of the Acoustical Society of America*, *120* (6), 3978-3987.

Mai, Y. 麥耘 (1998). 廣州話語調說略 [A preliminary discussion of Cantonese intonation]. *Cantonese Research and Teaching, 3*, 269-280.

Matthews, S., & Yip, V. (2011). *Cantonese: A Comprehensive Grammar* (2nd ed.). London and New York: Routledge.

Mixdorff, H. (2004). Quantitative tone and intonation modeling across languages. In *Proceedings of International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, Beijing, 137-142.

Mueller-Liu, P. (2006). Signalling affect in Mandarin Chinese – the role of utterance-final non-lexical edge tones. In *Proceedings of Speech Prosody 2006,* Dresden, Germany: PS6-3-0048.

Niebuhr, O. (2007). The signalling of German rising-falling intonation categories – the interplay of synchronization, shape, and height. *Phonetica, 64,* 174-193.

O'Connor, J. D. & Arnold, G. F. (1961). *Intonation of Colloquial English.* London: Longman.

O'Shaughnessy, D. (1979). Linguistic features in fundamental frequency pattern. *Journal of Phonetics, 7,* 119-145.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of $F_0$ of voice. *Phonetica, 41,* 1-16.

Palmer, H. E. (1922). *English Intonation with Systematic Exercises*. Cambridge, UK: Heffer.

Pell, M. D. (2001). Influence of emotion and focus on prosody in matched statements and questions. *Journal of the Acoustical Society of America, 109,* 1668-1680.

Pike, K. L. (1945). *The Intonation of American English.* Ann Arbor: University of Michigan Press.

Pike, K. L. (1948). *Tone Languages.* Ann Arbor: University of Michigan Press.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29,* 191-215.

Pittayaporn, J. (2007). Prosody of final particles in Thai: Interaction between lexical tones and intonation. Paper presented at *Workshop on "Intonational Phonology: Understudied or Fieldwork Languages."* Saarland University. Saarbrücken, Germany.

Potisuk, S., Gandour, J., & Harper, M. P. (1997). Contextual variations in trisyllabic sequences of Thai tones. *Phonetica, 53,* 200-220.

Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, *24* (4), 445–473.

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

Sadock, J.M., & Zwicky, A.M. (1985). Speech act distinctions in syntax. In T. Shopen (Ed.), *Language Typology and Syntactic Description: Clause structure.* Volume 1 (pp. 155-196). Cambridge, England: Cambridge University Press.

Shen, X-n. S. (1990). *The Prosody of Mandarin Chinese*. Berkeley: University of California Press.

Sluijter, A., & van Heuven, V. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the 4th*

*International Conference of Spoken Language Processing ICSLP 96*, Philadelphia, 630-633.
Tan, Y. (2011). The origin and nature of high rising diminutive tone change in Siyi dialect. In *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)*, University of Oregon, 190-207.
Turk, A. E. & Sawusch, J. R. (1997) The domain of accentual lengthening in American English, *Journal of Phonetics*, *25*, 25-41.
University of Hong Kong (n.d.). Hong Kong Cantonese Corpus. Retrieved from http://www0.hku.hk/hkcancor/
Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, *33,* 368-392.
Wang, B. & Xu, Y. (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics*, *37*, 502-520.
Wang, B., Wang, L., & Qadir, T. (2011). Prosodic realization of focus in six languages/dialects in China. In *Proceedings of the International Congress of Phonetic Sciences*, Hong Kong.
Wong, Y. W. (2006a). Contextual tonal variations and pitch targets in Cantonese. In *Proceedings of Speech Prosody 2006*, Dresden, Germany, PS3-13-199.
Wong, Y. W. (2006b). Realization of Cantonese rising tones under different speaking rates. In *Proceedings of Speech Prosody 2006*, Dresden, Germany, PS3-14-198.
Wong, Y. W., & Xu, Y. (2007) Consonantal perturbation of f0 contours of Cantonese tones. In *Proceedings of The 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1293-1296.
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61-83.
Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*, 179-203.
Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*, 55-105.
Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica*, *58*, 26-52.
Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, *46* (3-4), 220-251.
Xu, Y. (2005-2012). ProsodyPro. Retrieved from

http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/.

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics, 38,* 329-336

Xu, Y. (2011). Speech prosody: A methodological review. *Journal of Speech Sciences, 1,* 85-115.

Xu, Y., & Wang, M. (2009). Organizing syllables into groups – Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics, 37,* 502-520.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication, 33,* 319-337.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33,* 159-197.

Xu, Y., Chen, S.-w., & Wang, B. (2012). Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family? *The Linguistic Review, 29,* 131-147.

Yip, V., & Matthews, S. (2000). *Intermediate Cantonese: A Grammar and Workbook.* London: Routledge.

Zee, E. (1999). Chinese (Hong Kong Cantonese). In The International Phonetic Association, *Handbook of the International Phonetic Association* (pp. 58-60). Cambridge, U.K.: Cambridge University Press.

Zerbian, S, Downing, L.J., & Kügler, F. (2009). Introduction: tone and intonation from a typological perspective. *Lingua, 119* (6), 817-826.

Zerbian, S., Genzel, S., & Kügler, F., (2010). Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). In *Proceedings of Speech Prosody 2010*, Chicago.