Cross-linguistic Perception of Intonation by Cantonese and Mandarin Listeners

XU, Bo

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy

in

Linguistics

The Chinese University of Hong Kong
September 2013

Thesis/Assessment Committee

Professor LEE Hun Tak Thomas (Chair)
Professor MOK Pik Ki Peggy (Thesis Supervisor)
Professor JIANG-KING Ping (Committee Member)
Professor SHIH Chilin (External Examiner)

論文評審委員會

李行德 教授 (主席)
莫碧琪 教授 (論文導師)
蔣平 教授 (委員)
石基琳 教授 (校外委員)

Abstract

This study investigates the interaction between psychoacoustic perception mechanism (the Frequency Code) and phonological knowledge, and the role of language background in the perception of intonation.

Cantonese and Mandarin have distinctive lexical tone and intonation systems. Previous studies showed that Mandarin questions were signaled by raising F0 contour of the whole utterance. In comparison to Mandarin, there are fewer studies on Cantonese intonation and each has their limitations. Two production experiments focusing on the final rising and the global F0 contour were thus conducted to confirm the intonation patterns of Cantonese. The results show that the final rising is the principal perceptual cue for Cantonese questions. It is resulted from a final high boundary tone that is strictly localized at the end of an utterance. The modification in global F0 contour is neither significant nor consistent. In questions ending with a sentence final particle (SFP), the final high boundary tone is located in the particle and does not influence pre-SFP part of the utterance.

To answer the central research question on cross-linguistic perception of intonation, 20 Mandarin listeners and 20 Cantonese listeners participated in perception experiments with different contexts (normal speech and low-pass filtered

speech) and conditions (complete utterances and utterances without the final tone). They completed forced choice tasks to identify whether the utterance they heard was a statement or a question. Identification accuracy, discriminability scores, bias scores and confidence ratings were included in the data analysis. The results show that when they listened to an unfiltered unfamiliar/less familiar language, they used both their phonological knowledge and the psychoacoustic code to process intonation. When they listened to low-pass filtered speech, they mainly relied on the psychoacoustic mechanism, even though perceptual patterns still show minor differences stemming from language background. The results also show complicated interactions between tone and intonation in perception.

Based on these findings, it is concluded that the interaction between psychoacoustic mechanism and phonological knowledge of tone and intonation are essential for intonation perception. Language background also plays an important part. When there is rich linguistic information in the speech signal, phonological knowledge may override psychoacoustic mechanism.

全文摘要

香港粵語和漢語普通話有著截然不同的韻律系統。粵語的聲調比普通話複雜，語調的型式也很不同。普通話疑問句的全句基頻都會提升，而粵語則依靠句末的聲調上揚來表示提問。這些差異提供了研究跨語言語調感知的機會，讓我們從中可以了解音系知識與跨語言普遍存在的心理聲學機制（基頻編碼）的互動。

因為在文獻中粵語疑問句全句基頻的變化仍存在爭議，所以本研究先從聲學的角度分析了粵語的語調。兩個產出實驗的結果顯示：粵語疑問句的信息承載在最後一個音節的後半部分，使這個音節的聲調，不論原有調形如何，都變為一個升調，而除去最後音節的整句基頻不會因為疑問句的影響而產生固然、顯著的變化。所以，粵語疑問句最重要的感知標記就是句末的升調。聲學分析也顯示：粵語的句末語氣詞“咩”承載了疑問句的信息。在有語氣詞的疑問句中，語氣詞以前的全句基頻並不會發生改變。

20 名香港粵語母語者和 20 名普通話母語者參加了感知實驗。在實驗中，他們聽到的雙語材料經過了條件的控制（有/無句子最後一個音節）及聲學的處理（有/無經低通濾波處理）。在聽完每一句句子後，他們必須決定這個句子是陳述句還是疑問句。回答的準確率、區別及偏向得分以及被試對自己答案的自信評分都納入了實驗結果的分析。結果顯示，在聽到一個不熟悉的語言時，他們會綜合使用基頻編碼和音系知識來進行判斷。對於經過低通濾波處理的句子，

他們主要依賴心理聲學機制進行判斷，但是仍會受到語言背景的影響。除此之外，實驗的結果也再次揭示了聲調和語調的複雜關系。

這項研究揭示了心理聲學機制和音系知識對於跨語言語調感知的互動和層次。當語言信息缺乏時，心理聲學機制會主導語調的感知，但語言背景仍對感知有一定的影響。當語言信息充盈時，各種因素同時起作用，但音系知識有可能凌駕於心理聲學機制，對感知產生更大的影響。

To my grandma Mary

Acknowledgements

Learning new things through writing a research paper is companied by learning new things about oneself on the path to becoming a better being. I have to ascribe my growth as a researcher and as a person to many people.

My guide on this path is Prof. Peggy Mok. She introduced me to the study of phonetics, showed me how to be a careful and enthusiastic experimentalist, and has left tiny, invaluable comments in every margin of every draft I have finished, many times during her vacation and leave. Meanwhile, she taught me to be humble and grateful. I could not ask for an advisor and mentor more generous and supportive.

I owe intellectual debt to my committee members. Thanks to the wisdom and patience of Prof. Thomas Lee and Prof. Ping Jiang, I took my first step in linguistics. His theory-oriented approach and her data-oriented approach to research have nurtured me tremendously and will continue to influence me in the future. Special thanks go to my external committee member, Prof. Shih Chilin. I have always learnt from her from her illuminating works, and to receive her insightful comments and suggestions was a wonderful reward.

I am deeply grateful to Prof. Virginia Yip, Prof. Candice Cheung, Prof. Donovan Grose, Prof. Foong Ha Yap, Dr. Baoning Fu, Prof. James Woodward, and Prof. Felix Sze, from whom I have learnt so much about different aspects of language, and who have given me decent training and immense encouragement. I owe many, many thanks to Prof. Gloria Zhang. Working with her widened my horizon, sharpened my skills; and this thesis could not be finished in time without her unreserved support.

I could not believe how lucky I am by choosing linguistics. This is such a tight, warm community. Many thanks to Prof. Alan Yu, Prof. Francis Nolan, Prof. Daniel

Hirst, Prof. Yi Xu, Prof. Patrick Wong, Prof. Jiahong Yuan, who I met briefly during their visits to my department or in conferences, but have showed me insights and inspirations that this thesis could not be completed without. I am more than fortunate to have met new mentors and friends during my trip to the LSA Linguistics Institute 2013. My gratitude also goes to Murray Schellenberg, Ling Zhang and Thomas Wang. They were working on related topics during the writing of this thesis, and have generously given me their advice and friendship.

There is a lovely bunch of cohorts and colleagues in the department: Vera, Patrick, Donghui, Hui, Haoze, Jess, Aishu, Margaret and more I cannot name here. Thanks for enlightening my mind in our discussions. I look forward to working with them in the future. Thanks also go to the staff in the general office, especially Yvonne, who never made any mistakes in taking good care of us.

I am spoiled to have Ethan and Krista standing by me in my highest and lowest. I am also indebted to two of my mentors back in my home college: Prof. Jianfei Zhang and Prof. Xianchu Wan. Their faith in me has motivated me to explore what I love – something they always wished me to do.

I must thank Mr. Steven Sondheim for his marvelous music and lyrics. Not a day has gone by did I not merrily roll along with his beautiful works. Thanks to him, I am still here, being alive, putting it together, finishing the hat and opening doors.

Finally, I am who I am because of my family. I cannot return the love, care and patience my parents have granted me. I dedicate this thesis to my late grandma Mary, who was an English teacher and taught me English IPA and phonics when I was a child. She urged me to pay attention to intonation when I started learning English. I hope this thesis is only one of the fruits you have seeded in me, grandma.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Defining intonation

As defined by Ladd (2008:4), intonation is "the use of suprasegmental features to convey 'postlexical' or sentence-level pragmatic meaning in a linguistically structured way". It is a universal linguistic component of the world's languages, but different languages may have specific intonational inventory and phonology (Bolinger 1964, 1979; Ladd 1981).

Intonation manipulates the pitch contour as the primary acoustic correlate (Liu 1924; Chao 1933), and makes use of duration and amplitude as secondary features as they are not so salient or consistent (Blicher et al. 1990; Gottfried and Suiter 1997; Liu and Samuel 2004). In many East Asian tone languages, the height and shape of pitch (F0) are also used to distinguish lexical meaning. Therefore, the change of pitch in these tone languages bears two linguistic functions: lexical tones on the lexical level and intonation on the postlexical level. This study examines the interaction

between lexical tones and intonation in two Chinese tone languages, Mandarin and Cantonese, with a perceptual perspective. Since pitch is the primary cue for both tone and intonation, pitch will be the primary acoustic correlate to be measured and discussed in this study.

In any language, intonation is used to convey multiple linguistic or paralinguistic meanings, such as mood, focus, emotion, attitude and social index. The distinction between statement and question is a basic use of intonation across languages, which has attracted much research attention. In Cantonese and Mandarin, this distinction is also the most studied aspect of intonation. This study will follow previous studies on this topic and focus on the intonational differences between statements and yes-no questions in Cantonese and Mandarin.

1.2 The Frequency Code

The cross-linguistic methodology in this study is motivated by a central research effort to investigate the interaction between the psychoacoustic mechanism and phonological knowledge in perceiving intonation. The most relevant psychoacoustic mechanism to the perception of statement and question is the Frequency Code proposed by Ohala (1983). It explains the cross-linguistic correspondence between

intonation contour and (para-)linguistic meaning, stemming from a biological basis (Gussenhoven 2004:80). Since the organ that produces a lower sound is usually larger, low-pitched individuals are associated with being dominant and aggressive, whereas individuals that produce a higher pitch are associated with being subordinate and submissive. Since puberty, the male larynx grows into almost twice the size of the female larynx, supplemented by a longer vocal tract than that of the female, allowing men to be able to produce lower voice pitch than women (Luchsinger and Arnold 1965; Fant 1975). The proposal is also supported by studies on animal behavior and evolutionary biology. It is shown that the length of vocal tract and formant dispersion resulting from it are positively correlated with body size in different species of birds (Fitch 1999), mammals (Fitch 1997; Riede & Fitch 1999) and human beings (Fitch & Giedd 1999). Some species of deer, among other animals, can lower their larynx to exaggerate the impression of their body size signaled by their voices (Fitch and Reby 2001). Puts et al. (2006) showed that men's physical and social ratings are increased by a masculine, low-pitched voice. Furthermore, they also showed that when addressing a competitor who is believed to be physically less dominant, men would lower their voice pitch; in contrast, men who believe they are less dominant would raise their voice pitch.

There are "affective interpretations" and "informational interpretations" of the

Frequency Code in the para-linguistic and linguistic domains respectively (Gussenhoven 2004). The different affective interpretations summarized by Gussenhoven are presented in pairs in Table 1.1. These para-linguistic contrasts between high and low voice pitch are supported by experimental findings (Rietveld et al. 2002; Uldall 1964). These studies showed that a high fundamental frequency, especially at the last quarter of the utterance contours, is correlated with meanings at the left side of Table 1.1, concurring with the prediction of the Frequency Code.

Table 1.1 Affective interpretations of the Frequency Code

(Adapted from Gussenhoven 2004)

High pitch	**Low pitch**
Feminine / Submissive	Masculine / Dominance
Friendly / Polite / Pleasant	Aggressive
Vulnerable	Confident / Protective

The "informational interpretations" of the Frequency Code relate high or rising contour to "uncertainty", and thus questioning, and relate low or falling contour to "certainty", and thus being assertive (Gussenhoven 2004; Ohala 1983). In languages, questions can be raised in different ways. Some questions are indicated by lexical markers such as wh-element or sentence final particles. Yes-no questions without

sentence final particles (SFPs), on the other hand, are usually raised with the sole modification of intonation, and would be assigned with a high or rising pitch contour according to the Frequency Code. Although not without exceptions (Gussenhoven 2002; Hirst 2013), most languages indeed use final rising or a raised contour to convey such questions. This study, as a result, will focus on the intonation of yes-no questions without SFP in Mandarin and Cantonese. A few studies have investigated the role of the Frequency Code in intonation perception, which will be reviewed in Section 2.5.3.

1.3 Outline of the thesis

This first chapter presents a general introduction of this thesis. Chapter 2 provides a detailed background of the current study. The chapter introduces lexical tones and intonation types in Cantonese and Mandarin in both production and perception. Phonetic patterns of question intonation in the two languages are compared, which provides the basis for predicting the results of the perception experiments. A review of studies on cross-linguistic perception gives rise to research questions in this study.

The following chapters provide details of the several experiments conducted in this study. Chapter 3 revisits the F0 of Cantonese question intonation by reporting

two production experiments. The first experiment looks at the final rising of Cantonese questions. The second experiment focuses on the global pitch contour of Cantonese intonation. Chapter 4 and Chapter 5 report two perception experiments that have native Mandarin and Cantonese listeners identify intonation type of both Mandarin and Cantonese. In Chapter 4, normal speech stimuli were used; while in Chapter 5, low-pass filtered speech stimuli were used. Chapter 6 summarizes the patterns found in the experiments and concludes the thesis.

CHAPTER 2

BACKGROUND

This chapter reviews previous studies that lead to the current study. The review can be divided into two parts: Sections 2.1-2.4 review the literature concerning lexical tone and intonation in Mandarin and Cantonese; Section 2.5 reviews studies that investigated cross-linguistic perception of various prosodic features. Section 2.6 offers a summary of the chapter.

2.1 Mandarin and Cantonese

Mandarin and Cantonese are two tonal Chinese languages spoken in China and many oversea Chinese communities. Mandarin (*guanhua*, 'speech of the officials') refers to different varieties and dialects spoken in northern and southeastern China. However, in the linguistic literature, it often refers to the standard form of Chinese (*Putonghua*, 'common speech') based on the Beijing dialect, promoted by the Chinese government nationally. It is the official language of China and is widely used in broadcasts, schools and formal communications. In this study, we follow

previous studies and use ‘Mandarin’ in the sense of Standard Chinese. Similarly, Cantonese can refer to either the different varieties and dialects spoken in southern China Lingnan area (*Yueyu* ‘Cantonese language’), or specifically refer to the dominant variety spoken in the Pearl River delta area like Hong Kong and Guangzhou (*Guangfu Hua*, ‘speech of Guangzhou’). This study investigates the Cantonese spoken in Hong Kong.

Although Mandarin and Cantonese are typologically related, they have many different prosodic features, among which two are relevant to this study. First, the lexical tone system in Mandarin mainly makes use of the direction of pitch contour to differentiate lexical meaning. In contrast, Cantonese has a more complicated system, utilizing both pitch height and pitch direction. Second, they use different mechanisms to signal yes-no questions: while Mandarin raises pitch level in a sentential domain, Cantonese adds a boundary tone at the very end of an utterance as the chief mechanism for questioning. The literature on lexical tones and intonation of Mandarin and Cantonese will be reviewed in Sections 2.2 and 2.3. Sentence-final particle (SFP), a phenomenon related to intonation in Mandarin and Cantonese, is discussed in Section 2.4.

2.2 Tone and intonation in Mandarin

2.2.1 Lexical tones in Mandarin

There are four lexical tones in Mandarin (Chao 1933), as seen in Table 2.1 and Figure 2.1. All of them differ in pitch shape. Tone 1 (T1) is a high level tone. T2 is a high rising tone. T4 is a falling tone. There are two allotones for T3. Produced as a citation form or before a boundary, it is a dipping tone (214 in Chao's tone letter); otherwise it is produced as a low tone (21 in Chao's tone letter, Duanmu 2007).

There is also a tone sandhi rule in Mandarin. In a T3-T3 sequence, the first T3 is produced as a high rising tone [35] (Shih 1986), perceptually indistinguishable from T2 (Peng 2000; Wang and Li 1967), especially when the linguistic context that implies the occurrence of tone sandhi is lacking (Speer et al. 1989).

Table 2.1 Lexical tones in Mandarin

	T1	T2	T3	T4
Tone Shape	Level	Rising	Dipping	Falling
Tone Letter	55	35	21/214	51

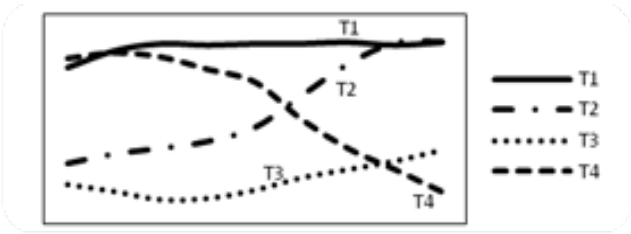

Figure 2.1 time-normalized pitch trajectories of Mandarin lexical tones

2.2.2 Production of Mandarin question intonation

Chao, in his pioneer works on Chinese intonation studies, described the relationship between the lexical tones and intonation in Mandarin as "small ripples riding on large waves", in which "small ripples" were lexical tones and intonation were "large waves" (Chao 1933, 1968). Chao (1968) stated that intonation could stretch or shrink the pitch range, or horizontally make the syllables longer or shorter. Besides looking at the global F0 contour over the utterance, Chao also described the modification of the final tone in a Mandarin question. He stated that questioning in Mandarin is a result of a rising ending added to the last syllable. For example, T1 (55 in Chao letter) became 56 after such an addition, T2 (35) became 36, T3 (214) became 216 and T4 (51) became 513 (Chao 1933, 1968).

This view of "successive addition" of a rising ending was supported by some subsequent studies. Ho (1977) found a trivial rising at the end of Mandarin questions, but the rising did not make the final lexical tone unrecognizable. Wu (1982) regarded intonation as the sum of lexical tone contours and tonal coarticulations, but he also

admitted that yes-no questions had an ending with a higher register than statements when a sentence final particle “ma” is not present. Lin (2004) observed various forms of high boundary tone at the end of Mandarin questions: higher F0 at the beginning of the final tone, higher F0 at the end of the final tone, or a combination of both. The Pan-Mandarin ToBI, a transcription system of Mandarin intonation, regards the question sentence final particle “ma” a high boundary tone (Peng et al. 2006). These studies suggest that questioning in Mandarin could be a local event.

However, other studies showed that the tone shape of the final tone remains very similar to the citation form acoustically (He and Jin 1992; Yuan et al. 2002; Yuan 2004). Yuan (2004) measured a large number of eight-syllable utterances that included statement-question minimal pairs, produced by eight native speakers. His finding showed that the global F0 curve of a question was raised and the whole sentence had a smaller declination slope compared to a statement. The final T2 of a question had a steeper rising and a higher ending than that of a statement. The final T4 of a question had a higher beginning and a higher ending than that of a statement, and hence had a flattened falling contour. Yuan concluded that a boundary tone is not necessary to model the distinction between statement and question in Mandarin. This claim is supported by Shi (2010), who stated that the boundary effect in Mandarin is

merely a break between prosodic units, instead of being an important cue for questioning.

Other studies also regarded that Mandarin questions were signaled by a global event, an event that requires the modification of the F0 contour of the whole utterance (Gårding 1984, 1987; O. J. Lee 2005; J. Shen 1985, 1994; X. S. Shen 1990; F. Shi 2010; P. Shi 1980; Yuan et al. 2002; Yuan 2006). Shi (1980) claimed that question intonation was a raised version of its statement counterpart starting from the beginning of the F0 contour. In Gårding's "grid" model (1984, 1987), intonation was represented by the "grid" consisted of a top line and a base line. The falling or rising of the "grid", primarily over the final phrase, was responsible for the distinction of statement and question in Mandarin. Shen (1985, 1994) also used a top line and a base line in his model to represent the pitch range in the tone-intonation contour. In his view, steep falling of both the top line and the base line indicate a statement, while a gradually falling top line and a slightly rising base line indicate a question. Shen (1990) proposed that intonation could be represented as "tunes". The "tune" of yes-no questions started and ended at a higher key than that of statements.

To sum up, Mandarin question intonation is different from statement intonation in both the global F0 contour and at the final position. However, the final lexical tone retains its canonical tone shape, and the modification of the final tone may be a

by-product of the global F0 modification. The difference between statement and question may become larger towards the end of the utterance, but the essential mechanism of Mandarin questioning is its raised F0 contour over the scope of the whole utterance.

2.2.3 Perception of Mandarin question intonation

A few studies have investigated the perception of Mandarin intonation. These studies revealed several asymmetries in their perception results. First of all, statements are easier to identify than questions. This bias towards statement shows that statement would be an unmarked sentence type. Listeners do not judge a sentence as a question unless they have captured the question feature in the sentence (Yuan 2004, 2011). Second, the sentence-final lexical tone does not affect the identification of statements, but influences the identification of intonational questions (yes-no questions that are signaled by intonation only, without any sentence final particles). Questions with a sentence-final falling tone (T4) are easiest to be identified in comparison to questions ending with the other tones. In contrast, a question ending with a final rising tone (T2) is the most difficult to identify and causes varied results among listeners (Yuan 2004). Yuan (2004, 2011) explained that this was due to the flattening of the falling tone at the end of a question. This was a language-and-tone-specific perceptual pattern; and

intonation perception was sensitive to the phonological tone identity at the end of an utterance. Third, question-final lexical tones are rarely affected by intonation perceptually. Each tone in statements or in questions can easily be recognized by listeners (Connell et al. 1983). In a study focusing on the boundary effect in Mandarin questioning, Lin (2004) suggested that manipulating the pitch register of the final tone (T1 and T4) of a given utterance could cause a categorical distinction between statement and question. However, his result should be taken with caution as there were only five listeners in his study. Jiang and Chen (2011), with a well-designed set of stimuli, showed that cutting off the final tone does not significantly influence perception. What is important to questioning is the last prosodic word of the utterance.

2.3 Tone and intonation in Cantonese

2.3.1 Lexical tones in Cantonese

There are six lexical tones in Cantonese (Vance 1977; Bauer and Benedict 1997; Kao 1971, Table 2.2 and Figure 2.2). They differentiate in pitch direction and pitch level. There are three level tones (T1: high level, T3: mid level, T6: low level), two rising tones (T2: high rising, T5: low rising) and one low falling tone (T4). Looking at the

pitch contour drawn from the production of a Cantonese speaker (Figure 2.2), one can observe that the lower pitch range of Cantonese is fairly crowded. There are short versions for each of the three level tones, also known as the checked tones or entering tones (T7, T8, T9). They are in effect syllables ending with unreleased oral stops and thus have an abrupt ending. A free variation of Tone 1 (producing the high level tone as high falling in some specific lexical items) can be found in the older generations, but this variation no longer exists among the younger speakers in Hong Kong (Bauer and Benedict 1997).

Table 2.2 Lexical tones in Cantonese

		T1	T2	T3	T4	T5	T6
Tone Shape		High level	High rising	Mid level	Low falling	Low rising	Low level
Tone Letter	Long	55	25	33	21	23	22
	Short	5		3			2

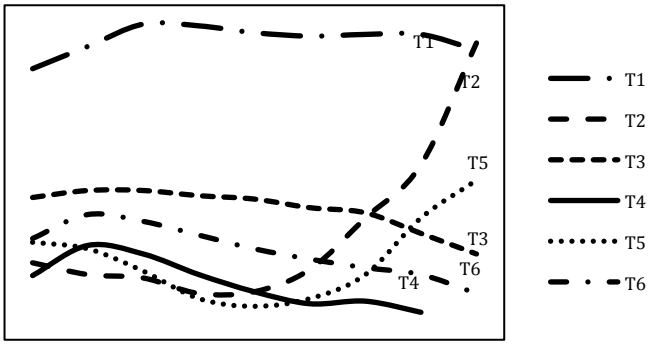

Figure 2.2 Time-normalized pitch trajectories of Cantonese lexical tones

2.3.2 Production of Cantonese question intonation

Yes-no questions are signaled differently in Cantonese than that in Mandarin, most notably for the realization of the final syllable. First, a significant final lengthening is observed in Cantonese questions, with the final syllable of a question much longer than that of a statement (Ma et al. 2006a). More importantly, a final rise at the last syllable of a question was observed. Mai (1998) impressionistically pointed out that the question intonation is successively added to the last lexical tone. His observations were in accordance with acoustic data in later studies (Lam 2002; Chow 2002; Gu et al. 2006; Lee 2008; Ma et al. 2006a; Ma et al. 2006b). The tone shapes of the lexical tone at the end of a question change to a rising tone and unanimously reach the pitch height of the end of canonical T2. However, T1, a high level tone, does not change much from its canonical form like the other tones do because of the upper F0 limit of speakers, i.e., a ceiling effect (Lee 2008; Ma et al. 2006a). Acoustically, even though the general tone shapes of the final tone are changed, the first halves of the pitch contours retain their original F0 level, followed by a rising contour. As a result, T3 (33) as a final tone would have a mid-level start, while T2 (25), T4 (21), T6 (22) share similar tone contour from start to end in the final position of questions.

While the final rising of Cantonese question intonation was widely acknowledged in the literature, neither the details of global F0 contour (pre-final F0

contour) nor its importance to Cantonese question relative to final rising has reached a consensus among researchers. Chang (2003) claimed that the final local event would not affect the global declination. However, he did not conduct a controlled experiment to investigate this issue. In Ma et al. (2006a), three different carrier phrases were used where the target syllables were embedded respectively in the initial, medial and final positions of the utterances. They measured the target syllables of all tones in different sentence positions. It was found that the F0 level in questions was significantly higher than that in statements for most tones in the initial position (except T4) and for T4 and T6 in the medial position. F0 values at specific time points along each tone's F0 contour showed that the F0 values are largely higher in questions than in statements. However, there was no significant difference for most tones in the initial and medial positions, except T5 (most time points) and T3 (all time points). In brief, Ma et al. (2006a) had mixed results concerning the global F0 contour of Cantonese questions. The global F0 difference between questions and statements remained unclear (significant in terms of F0 levels, not significant in terms of specific F0 values at time points), and is dependent on lexical tones. Besides, although their method illustrated the acoustic details in different positions of the utterance, the conclusion may be questionable because the results might be influenced by focus. Since all the tested syllables were target syllables

embedded in three carriers designed to test the three positions, the target syllables became the only new information in the carrier sentences. As a result, subjects might tend to emphasize the target syllables in each carrier phrase, resulting in a higher F0 level at all the positions.

Lee (2008) used a different technique to measure the global F0 contour. She linked the F0 values of two T1 (high level, 55) tones in the carrier phrase to draw an upper trend line, and linked the F0 values of two T6 (low level, 22) tones to draw a lower trend line. Average F0 level (the mean of upper line and lower line), declination (steps and gradients) and pitch span (pitch range in other studies) were calculated based on these trend lines. It was found that questions had a higher average F0 level and a more gentle declination than statements, whereas pitch span was not affected by intonation type. Her results showed that the global raise in Cantonese questions was consistent. Lee's utilization of the trend lines provided visual depiction of the pitch range change along the sentential contour, but a lot of the acoustic information of the individual lexical tones is lost along the intonation contour, especially for the non-level tones. Therefore, her study did not provide a complete picture of the global F0 contour of Cantonese questions.

Lee (2008) also suggested that the final rising might be a more important cue for Cantonese question than the global F0 pattern. She tested how sentence length affects

the global F0 contour and the final rising using utterances consisting of one, four and eight syllables embedded in a short story, each ending in each of the six lexical tones. She discovered that sentence length could affect pre-sentence-final intonation in both statements and questions, but not the final rising in questions. Based on this pattern, she argued that the final rising was the more consistent cue for Cantonese question.

Since the importance of the global F0 to Cantonese questions has not reached a consensus in the literature, two production experiments investigating the final rising and the global F0 contour respectively were conducted (see Chapter 3). The first experiment used a carrier phrase to control the non-final part of the utterances. The second experiment employed utterances consisting of one single lexical tone to measure the F0 variation of the whole non-final part of the utterances. In this way, the potential focus effect in Ma et al. (2006a) was avoided, while phonetic realization of different lexical tones in Cantonese questions, ignored by Lee (2008)'s method, can be observed.

2.3.3 Perception of Cantonese question intonation

Since tone shape gets altered by question intonation in Cantonese, the interaction between tone and intonation has been investigated from two main perceptual perspectives: the influence of question intonation on tone recognition, and the

identification of statement and question with different lexical tones.

Ma et al. (2006a) investigated tone identification under the influence of intonation. They found that the addition of question intonation on the final tone caused a large variability of tone recognition among listeners. The most common error was misidentifying the low tones (T4, T5, T6) as the high rising tone (T2). This is understandable because T2 has a low starting point, similar to the low tones. As a result, when the low tones have a rising tail at the end, the overall pitch contours are very similar to that of T2. In contrast, listeners made no mistakes in identifying question-final T1 and made few mistakes in identifying T3. This can also be explained by the production data (Ma et al. 2006a): T1 is distinct from T2 in questions in terms of both pitch level and pitch shape. T3 in a question final position, even though becomes rising at the end, starts from a middle pitch level, and thus is still different from the low starting point of T2.

In Ma et al. (2011), the perception of Cantonese intonation was tested. They examined three conditions: the whole utterance (carrier phrase + final syllable), isolated final syllable and isolated carrier phrase. Much like the data from Mandarin experiments (Yuan 2004), statement was shown to be the default intonation type in Cantonese. Statements are identified significantly better than questions, especially for the carrier phrase condition. Listeners showed equally high sensitivity in the

identification of statement and question in the whole utterance condition and isolated final syllable condition. In contrast, they failed to identify many questions in the carrier phrase condition. This result confirmed that the final syllable bears most information in Cantonese question intonation, in accordance to the production data (Ma et al. 2006a). A logistic regression model (R^2=86%) showed that the classification of statement and question could be predicted by F0 interval (the difference between the beginning and ending of the F0 in semitone) and F0 average level of the final syllable, but not by its F0 slope (the rate of F0 change) and duration. Statistical results also showed that average F0 in the carrier phrase could be used to predict statement and question. However, this acoustic cue did not seem to be salient to listeners, as Cantonese native speakers could not make good use of it to identify question in the carrier phrase condition (identification accuracy 48%, below chance level, Ma et al. 2011).

Lexical tone also played a role in intonation perception in Ma et al. (2011)'s results. Listeners achieved the best accuracy in discriminating question and statement when the utterances ended with T3. Even in the carrier phrase condition, listeners identified the T3 questions better than the other questions. In contrast, listeners performed the worst in distinguishing T1 and T2 statements and questions. Many errors were found in identifying T2 statements, which was a result of confusion

between the question rising ending and the rising lexical tone contour.

To sum up, different intonation patterns in Cantonese and Mandarin provide an opportunity to investigate cross-linguistic perception. Cantonese listeners look for the perceptual cue for questioning at the utterance ending, where the final rising takes place. Final lexical tone has an effect on intonation perception in Cantonese. In contrast, Mandarin listeners need to pay attention to the F0 contour of the whole utterance, although the difference between F0 contours of statements and questions is larger towards the end of the utterance. Mandarin listeners' identification of intonation is also sensitive to the identity of the final tone, as T4 (high falling) makes questions the easiest to identify, and T2 (high rising) the most difficult.

2.4 Sentence-final particles in Mandarin and Cantonese

Sentence final particles (SFP) are words that attach to the end of sentences. They function to indicate "speech act, evidentiary and emotional coloring in the language" (Matthews and Virginia 1994). SFPs share a similar function with intonation in non-tone languages (Kwok 1984; Wakefield 2010; Yau 1980). Yau (1980) suggested that the use of SFP might carry the function load of intonation and suppress the intonational contour of the sentence.

There are only a few sentence final particles in Mandarin, in comparison to the large number of SFPs in Cantonese. Exhaustive lists of SFPs in Cantonese have been provided in the literature (Law 1990; Fang 2003; Kwok 1984; Leung 2005; Matthews and Virginia 1994; Yau 1980). There has been no consensus on how many SFPs there are in Cantonese, counting from 30 (Kwok 1984) to 206 (Yau 1980), due to differences in definitions and individual intuitions.

Regarding the relationship between SFPs and intonation in Mandarin, the Pan-Model of ToBI (Peng et al. 2006) claimed that the Mandarin SFP "ma" is evidence of the boundary tone in Mandarin intonation, as a high "ma" and a low "ma" indicated question and statement respectively. However, they did not provide semantic, structural or psychological evidence to support the assumption that the two 'ma' in Mandarin are in effect the same lexical item underlyingly realized in different tones. Furthermore, no other minimal pair of sentence final particle with such functions can be found in Mandarin.

Other studies did not consider Mandarin SFPs to be carriers of the boundary tone. Shih (1997) showed that the SFP "le" in Mandarin did not influence the declination of Mandarin utterances. Lee (2005) suggested that the phonetic manipulation of pitch in Mandarin is relative to syntactic structure and pragmatic use. For instance, questions without 'ma' would have a more exaggerated pitch range

expansion compared with those with the question particle 'ma', and echo-questions with strong pragmatic meanings also have a significant expansion of pitch range. He and Jin (1992) also showed that some pragmatically strong questions and imperatives would have longer durations at the final syllable and are generally louder. More interestingly, Lee (2005) found that when the *"ma"* particle have a raising top line and expanded pitch range, the domain of such expansion is larger than a syllable unit (Lee 2005). The domain exceeded the SFP "ma" and covered the last noun phrase (NP) of the utterance. This suggests that it may not be appropriate to consider *"ma"* as a boundary tone in Mandarin.

Cantonese SFPs may play a more important role than their Mandarin counterparts in carrying intonation meaning. On one hand, Law (1990) proposed that some SFPs in Cantonese (i.e. "la", "a" "lo") did not have intrinsic lexical tone and their surface pitch contour were determined by intonation. This indicates that some SFPs in Cantonese might be intonational in nature. On the other hand, the SFPs in Cantonese seem to carry more informational load than those in Mandarin. Gu et al. (2006) investigated "*me1*", "*aa4*"and "*gwaa3*" at the end of Cantonese questions. It was found that the difference between F0 contours of particle questions (questions ending with SFPs) and statements is localized in the sentence-final part (essentially on the SFPs). Lengthened duration and enlarged amplitude of SFPs have also been

observed. The variation of duration and amplitude were found to convey or emphasize the intention of the speaker.

Wu (2008, 2009) compared statements and intonation questions (yes-no questions without SFPs) and found that their F0 patterns are similar except the final syllables. He also compared the intonation of sentences with and without SFPs. The F0 patterns of the utterance bodies for those sentences with SFPs are the same as those without SFPs. SFPs such as “me1” “ge2” display modification of pitch shape by intonation superimposition. Wu concluded that the pitch contours of SFPs in Cantonese were a combination of lexical tones and intonation. In other words, SFPs are lexical items intrinsically embracing intonational features.

In short, Cantonese has a larger amount of SFPs than Mandarin. SFPs in Cantonese may carrier more intonational meaning than those in Mandarin. However, the relationship between SFPs and the localized boundary tone, if any, in tone languages, needs further investigation.

2.5 Cross-linguistic perception of prosodic features

Both the Frequency Code and the phonological knowledge of lexical tone and intonation may affect listeners’ perception of intonation. Cross-linguistic perception provides a way to understand the interaction between the psychoacoustic and

phonological processings. Many studies have investigated cross-linguistic perception of prosodic features, especially lexical tones, revolving around the theme question concerning the roles of psychophysical characteristics and language background in speech perception. The following sections will review experiments on cross-linguistic perception of prominence, accent, stress, lexical tones and intonation, to serve as a background for the current study.

2.5.1 Cross-linguistic perception of prominence, accent and stress

Lehiste and Fox (1992) showed that English and Estonian listeners responded to different cues in identifying prominence. English-speaking listeners were more sensitive to amplitude cues, while Estonian listeners relied more on duration cues. Beaugendre et al. (2001) investigated the perception of accentuation boundary by Dutch, French and Swedish listeners. The results showed that the three groups of listeners utilized different cues to identify accentuation boundaries, resulting in significantly different responses. These studies showed that language background was important in perception, because language experience influence listeners' sensitivity to certain perceptual cues in the acoustic signal, while being insensitive to some others.

Dupoux et al. (1997, 2008) and Dupoux et al. (2001) recruited Spanish and

French speakers to identify contrastive stress. There is contrastive stress in Spanish but not in French. While Spanish speakers performed well in the task, French speakers could not perceive stress at a phonological level, despite their ability to detect the acoustic correlates of stress such as F0, duration and energy. When the participants of this experiment were expanded to speakers of other languages lacking contrastive stress, such as Finnish, Hungarian and Polish, it was found that speakers of French were the most insensitive to phonological stress, while Polish speakers performed better. Dupoux et al. (1997) argued that the degree of "deafness" to stress correlated with the ease with which infants could acquire stress regularity. To sum up, their works have shown that the inability to distinguish non-native phonological distinction was not restricted to the segmental domain. Language background had a strong influence on the perception of prosodic features such as stress.

Wang (2008) showed that Mandarin learners of English transferred their sensitivity in F0 to perceive English lexical stress. This result indicated that phonetic resources could be used in cross-linguistic perception of different prosodic features.

2.5.2 Cross-linguistic perception of lexical tones

Gandour (1981) was the pioneer to study the perception of linguistic pitch variation in tone languages. By studying the perception of the six tones in Cantonese, Gandour

proposed that the perception of lexical tone consisted of three dimensions: direction, contour and height. Since then, perception of lexical tones by native and non-native, non-tone language speakers attracted much research attention.

Studies in cross-linguistic perception of lexical tone showed that tone language speakers had a significant advantage over non-tone language speakers in perceiving tone categories and tonal contrasts (Hallé et al. 2004; Liang and van Heuven 2007; Peng et al. 2010; Xu et al. 2006), even after intensive training (A. L. Francis et al. 2008; Kaan et al. 2008). The experience in tone language and phonological tone categories in speakers' native language shaped the outcome of the categorical perception (Peng et al. 2010). Non-tone language speakers' perception of lexical tone was not categorical, but rather psychophysically based (Hallé et al. 2004). On the other hand, some studies found that non-native listeners showed better within-category pitch differences (DiCanio 2012; Stagray and Downs 1993). Furthermore, studies using PET scan showed that native and non-native speakers have different brain areas activated in perceiving lexical tones (Gandour et al. 2000; Klein et al. 2001).

However, studies have also shown that language background is not the only factor in tone perception. So and Best (2010) showed that the phonemic status and the similarity or dissimilarity between the native language and the target language

are important factors. Non-native tone perception was influenced by the phonetic features of the target tones. Moreover, contextual information is important in cross-linguistic perception of lexical tone information (Gottfried and Suiter 1997; Lee et al. 2010).

Finally, Burnham et al. (1996) showed that Thai tone pairs were perceived equally well by Thai and Cantonese listeners, but not as well by English listeners. More interestingly, they tested the perception of tonal contrasts in normal speech, low-pass filtered speech and musical sounds played by violin. For all three groups of listeners, tonal contrasts in music sounds were better discriminated than in low-pass filtered speech, which in turn was better discriminated than in normal speech. This indicates that listeners were more sensitive to acoustic differences in non-speech signal than in speech.

2.5.3 Cross-linguistic perception of intonation

Studies on cross-linguistic perception of prosodic features have shown that language background is very influential, whereas phonetic details of the target language may also be important in perception. While intonation patterns remain diverse in languages, cross-linguistic intonation perception has not been widely studied. Such studies usually have two main research questions: (a) What are the universal

tendencies that map intonation form and intonation meaning shared by speakers of different languages? (b) Having access to such universal features, how do language background and language experience shape the perception of intonation for speakers of a specific language? (Grabe et al. 2003; Vaissière 2005).

Handding-Koch and Studdert-Kennedy (1964) showed that Swedish and English listeners identified statements and questions according to the acoustic correlates of intonation type in their own language. Jiang and Chen (2012) examined perception of Mandarin intonation by Mandarin, Thai and English listeners. They found that the tone language speakers (Mandarin and Thai) used a "register contrast" strategy, and therefore made few mistakes in identifying statements ending with T1 (high level) and T2 (high rising) as questions. English speakers made mistakes in identifying these utterances because they relied on pitch shape and ending pitch height.

A cross-linguistic perception study conducted by Gussenhoven and Chen (2000) showed that three different groups of listeners (Dutch, Chinese and Hungarian) displayed similar association between question intonation and either a later or a higher F0 peak and higher end pitch in a made-up language. This concurs with the Frequency Code proposed by Ohala (1983, Gussenhoven 2004) that a high or rising pitch contour is associated with questioning. Grabe et al. (2003) focused on the discrimination between intonation contours rather than the statement-question

distinction. Their results showed that listeners of different language backgrounds (English, Spanish and Chinese) had access to a universal auditory mechanism that allowed them to distinguish rising from falling. They also indicated language experience played a role in the perception of intonation contour. They concluded that experience with a native language was added to the universal auditory mechanism in shaping speakers' perception of intonation.

Finally, Liang and van Heuven (2007) found that whether the listeners were a tone language speaker might have little to do with intonation perception, because the non-tone language speakers (Uygur) in their study was more sensitive to intonation than tone-language speakers (Beijing, Nantong and Changsha), when asked to identify Mandarin intonation. They explained that this was because the F0 contour carried less functional load for non-tone language speakers than tone language speakers.

To sum up, previous studies showed that the psychoacoustic mechanism and language background are both useful in cross-linguistic perception of intonation. When listening to a non-native language, listeners are influenced by the intonation patterns of their native language, with the Frequency Code being a possible resource to assist their perception.

2.6 Experiments in this thesis

Four experiments were conducted in this thesis in order to answer the central research question. The first two production experiments were designed to confirm that the question intonation of Cantonese is distinctive from that of Mandarin, so that the following perception experiments would stand on a solid acoustic grounds.

The first production experiment (Chapter 3, experiment 1) investigates the acoustic patterns at the final position of questions. The utterances used in this experiments are minimal pairs that differ only in intonation. With a clear pattern shown in the literature, it is expected that the results of this experiment will echo previous studies, showing a final rising in Cantonese questions. This rising contour may appear only at the second half of the final tone, thus keeping the canonical register of the first half of the tone. The rising contour would occur in every lexical tone except T1, because of the ceiling effect reported in previous studies (Ma et al. 2006a, Lee 2008).

The second production experiment (Chapter 3, experiment 2) aims to find out whether global F0 contour is an important part of Cantonese question intonation. The F0 of pre-final syllables will be measured. By testing utterances that consist of only one tone, the F0 of every tone at every position of the utterance are controlled and comparable. If the global F0 contour plays a role in Cantonese questioning, it is

expected that the pre-final F0 contour of intonation question will be different from those of statement and particle question. Intonation questions would have either a higher global pitch register, or a smaller declination slope, or both. On the other hand, if Cantonese question intonation is strictly localized in the final tone, the pre-final F0 contours of the utterances with the three intonation will not be significantly different from each other. In that case, listeners would not be able to recognize Cantonese questions without the final syllable.

The two perception experiments (Chapter 4 and Chapter 5) in this thesis enable us to observe the perception of intonation in three scenarios. First, Cantonese and Mandarin listeners will listen to utterances in their native language. The results should reduplicate previous studies that investigate Cantonese and Mandarin separately. Cantonese listeners will not be able to detect questions after the final syllable is cut, while Mandarin listeners should not have this problem. Meanwhile, language and tone dependent perceptual patterns are expected. The results of perceiving the listeners' respective native language will also serve as a baseline for further investigations, in which listeners will be asked to identify intonation types in a less familiar / unfamiliar language, or in unintelligible speech due to a low-pass filter.

When listening to an unfiltered less familiar / unfamiliar language (Chapter 4),

listeners may not have access to the phonological knowledge they used to have. For example, Cantonese listeners cannot find final rising in Mandarin utterances, and Mandarin listeners cannot rely on global F0 patterns to identify questions in Cantonese. Therefore, the effect of the Frequency Code may occur, and listeners may regard high or rising pitch contour as a signal for questioning. However, listeners may apply some native phonological knowledge to similar phonetic realizations in the other language. For example, Cantonese listeners may tend to interpret a rising tone (Mandarin T2 or sandhi-ed T3) at the end of an utterance as a signal for questioning. Therefore, the perceptual pattern of listening to a less familiar / unfamiliar language may be guided by both native phonological knowledge and the Frequency Code.

On the other hand, when listening to filtered speech in the second perception experiment (Chapter 5), where segmental and semantic information is absent, and individual lexical tones are difficult to recognize, the Frequency Code may dominate listeners' processing of pitch contour. Since the Frequency Code is assumed to be universal, there should be no substantial difference between the perceptual pattern between Mandarin and Cantonese listeners. Every utterance ending with a high or rising tone should be judged as a question, and those ending with a low or falling tone should be interpreted as a statement. The language and tone dependent

perception patterns described in previous studies (Yuan 2004; 2011; Ma et al. 2006a) should not be observed. If this expectation of perceptual results is not met, it may be implied that a strong effect of native language experience might be present, even when the speech is unintelligible.

2.7 Purpose of this study

Although both Cantonese and Mandarin utilize a high pitch in their intonation contour as a signal of questioning, the high pitch is realized differently. Mandarin employs pitch level over a sentential scale, whereas Cantonese uses a boundary tone that is localized in the sentence final position, which results in a high rising tone over the canonical lexical tones. The different scope and the different use of pitch (register raise vs. rising shape) cause differences in both the production and perception of intonation. In addition, as both lexical tones and intonation use F0 as the chief acoustic correlate, the knowledge about the interaction between tone and intonation in their native languages is very important for speakers of Mandarin and Cantonese in perceiving intonation (Yuan 2004, Ma et al. 2006a, 2010).

Furthermore, Mandarin and Cantonese presumably shared the psychoacoustic knowledge of the Frequency Code, which is supposed to be universal across languages. A study has showed that Mandarin listeners followed the Frequency Code

when perceiving artificial intonation contours (Gussenhoven and Chen 2000). What is unclear is how the phonological knowledge of lexical tone and intonation would interact with the psychoacoustic Frequency Code when listeners perceive intonation, and what would happen if they contradict each other. Since no previous studies have demonstrated this interaction, this study aims at filling this research gap.

Although some studies have explored cross-linguistic perception of intonation, none of them has examined the perception of materials stemming from two natural languages with distinct intonation patterns. Moreover, few studies have focused on and compared two groups of tonal language listeners, to explore how language experiences in two tone languages would influence the perception of intonation. This is particularly interesting considering the complicated interaction between the perception of lexical tone and intonation in Cantonese and Mandarin.

This study examined the process underlying intonation perception by changing the conditions listeners perceive intonation. Listeners were asked to identify intonation types (statement or question) in their native language, in an unfamiliar/less familiar language and in low-pass filtered speech. It is expected that both the Frequency Code and the phonological knowledge would help the listeners when they identify intonation types in the other language. However, the Frequency Code may be the sole mechanism they depend on when they listen to low pass-filtered speech.

The results will reveal the importance and interaction of different mechanisms in intonation perception, thereby demonstrate the interplay of language specified and universal factors in speech perception.

CHAPTER 3

REVISITING CANTONESE INTONATION

This chapter aims to investigate the production patterns of Cantonese intonation. The several studies focusing on Cantonese intonation each has their pros and cons, as summarized in Chapter 2. Particularly, there is a lack of sufficient direct evidence to support that global contour is not essential for Cantonese questions, in contrary to the importance of the raised global contour in Mandarin. Therefore, two experiments were conducted. Experiment 1, which controlled the non-final part of utterances, was designed to examine the final F0 rising. Experiment 2, on the other hand, intended to examine the global contour by controlling lexical tones of the whole utterances.

3.1 Participants and recording

Eight native Hong Kong Cantonese speakers (four males, four females) were paid to participate in the experiments. They were undergraduate students at The Chinese University of Hong Kong (CUHK) with no speech or hearing problems, aged

between 18 and 22. They had only limited trainings in linguistics and phonetics by the time they did the recording. They were not aware of the purpose of the experiment. Their speech recordings were screened to make sure that none of them were merging tones, which means that they could clearly produce the six distinct Cantonese lexical tones in citation form and in utterances.

The recording took place in a double-door sound-treated booth at CUHK. They were given the reading materials ten minutes before the recording to get familiar with the utterances. They were asked to read out the utterances as naturally as possible. The target utterances and filler utterances were randomized. The participants read the randomized list three times. Therefore, each target syllable has three tokens recorded. The participants could take a break whenever they like. The utterances they produced were recorded directly onto hard disk with a sampling rate of 44100 Hz. Materials of Experiment 1 and 2 were recorded in a single session that took 30 to 40 minutes to finish.

3.2 Experiment 1: the F0 patterns of the final rising

3.2.1 Reading materials

Since this experiment examined the final tone of Cantonese questions, the sentences

used in this experiment consisted of a common carrier phrase and a target syllable embedded at the end of the sentence or before the sentence final particle (SFP). Each carrier phrase was read in three conditions: as a statement, an intonation question (yes/no questions without an SFPs), and a particle question (questions with SFPs).

The Carrier phrase[1]:

呢個字讀___。(statement)

ne1 go3 zi6 duk6 ___.

This character is ___.

呢個字讀___? (intonation question)

ne1 go3 zi6 duk6 ___?

This character is ___?

呢個字讀___咩? (particle question)

ne1 go3 zi6 duk6 ___ me1?

This character is ___SFP?

Two sets of target syllables were used. All six lexical tones were included for these syllables. They consisted of only sonorant segments so that the F0 pattern throughout the whole syllable can be seen. Due to the variant pronunciations among young Cantonese speakers, they pronounced the first syllable ("aa" or "ngaa") with or without a velar nasal onset. In either case, they were asked to read the

[1] In this thesis, Cantonese words are transcribed in Jyutping, a romanization system invented and promoted by the Linguistic Society of Hong Kong (LSHK).

pronunciations they felt most comfortable with, and be consistent across all the lexical tones.

Target syllables: (in the order of the lexical tones)

(1) aa / ngaa: 鴉 啞 阿 牙 啞 訝

(2) iau: 優 柚 幼 游 有 又

3.2.2 Data analysis

The target syllables were labeled manually in Praat (Boersma and Weenink 2013). Because the chosen target syllables were all-sonorant syllables, the F0 curves of the whole syllables were used. The duration of each syllable was extracted and automatically divided into ten equidistant points by a Praat script. The F0 value at each point was tracked by Praat. Manual checking of all the obtained F0 data was performed.

As the results of the current production study will contribute to our understanding of intonation perception, and since psycho-acoustic scales (semitone or ERB-rate) could better reflect listeners' intuition regarding equivalent interval of intonational span across speakers (Nolan 2003), F0 values were first converted to semitone (ST) for further analysis. The use of semitone instead of Hz also helps to normalize the difference in pitch range between individuals. The semitone scale is

transformed logarithmically from the physical Hertz scale. In this study, the F0 values in Hz tracked by Praat was converted to semitone using the following formula.

$$ST = [\lg(Hz)-\lg(50)] / \lg(1.059463) \quad \text{(Lee 2008)}$$

After averaging the F0 values for each vowel for every speaker, the data was analyzied statistically. Creakiness of various speakers, especially when they produced T2, T4 and T5, resulted in some missing F0 values. To avoid dropping a whole data set of a participant due to some missing values, linear mixed model analysis was adopted to analyze the data statistically. F0 values of the first and the last measurement points of each tone were excluded to avoid frequent creakiness at the beginning and ending of the syllables. The F0 values in semitone from measurement point 2 to 9 were the dependent variables. Lexical tone (6 levels) and intonation (3 levels: intonation question, statement, particle question) were within-subject factors. Bonferroni correction was used in the post hoc tests.

3.2.3 Results

Figures 3.1 to 3.6 show the F0 contours (from point 2 to point 9) of the last syllable of each utterance in Experiment 1. In the legend of each figure, “I-Question” refers

to an intonation question (a yes-no question without any sentence final particles), and "P-Question" stands for a particle question (a question ending with a sentence final particle "me1"). The F0 contours of female (on the left) and male (on the right) speakers were plotted separately.

In Figure 3.1, while the statement and particle question contours remain similar, the final T1 of the question intonation sentence has a rising shape and is constantly higher than the other two contours. Moreover, the average F0 value of the end of the rising T1 (about 32 semitone for female speakers and 26 semitone for male speakers) is higher than the average F0 value of the question rising ending of the other tones (Figures 3.2-3.6). This pattern does not concur with reports in the previous study (Ma et al. 2006a), which stated that T1 at the end of an intonation question does not rise up like the other tones do. According to Ma et al. (2006a), this is because T1 is at the top of a speaker's pitch range, and cannot rise higher than its canonical form, i.e., a ceiling effect. The results in this study may be different from the previous findings for two possible reasons. First, it may be due to individual difference. Some speakers may produce it as a rising tone like the other question-final tones, and some may produce it as its canonical form without rising. It is also possible that the production pattern resulted from the experiment design, which asked participants to read out statement-question minimal pairs. The speakers may produce the question-final T1

with a rising contour to emphasize the distinction between questions and statements.

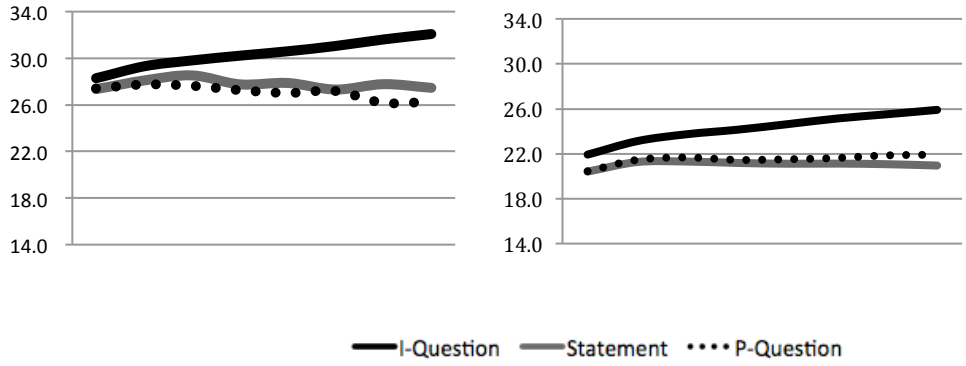

Figure 3.1 F0 contours (in semitone) of final T1 by female (L) and male (R) speakers

Figure 3.2 shows the final T2 contours of the three intonation utterances. As T2 is already a high rising tone, the three contours are expected to be quite similar in terms of height and shape. This is true for the female speakers, whose contours almost overlap, except the end of the particle question, which is lower than the other two contours. However, for the male speakers, while the first half of the three contours overlap, the ending of the question intonation rises above the other two contours, with the ending of the particle question slight lower than the statement. It is possible that in comparison to female speakers, male speakers may not make full use of their upper pitch range when they produce statements, and therefore has more room for rising for questions ending with T2.

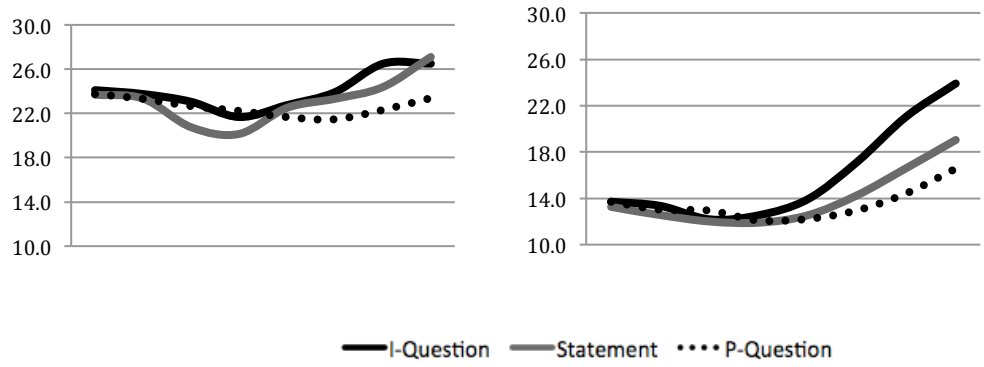

Figure 3.2 F0 contours (in semitone) of final T2 by female (L) and male (R) speakers

T3 to T6 are mid and low tones in Cantonese. Final rising at the end of intonation question is expected to be observable for all these tones. Figures 3.3 to 3.6 clearly give such an impression. Regardless of tone and gender, the beginning of the final tone contour of different utterance types is very similar. However, from the middle of the contour, the intonation question contour rises up, while the statement and particle question contours remain very similar, even entirely overlapping in some cases (T5, both female and male). The results indicate that high rising in intonation question is consistent for all the mid and low tones in Cantonese, and only takes place at the second half of the final tone. The ending tone of a particle question does not show this pattern.

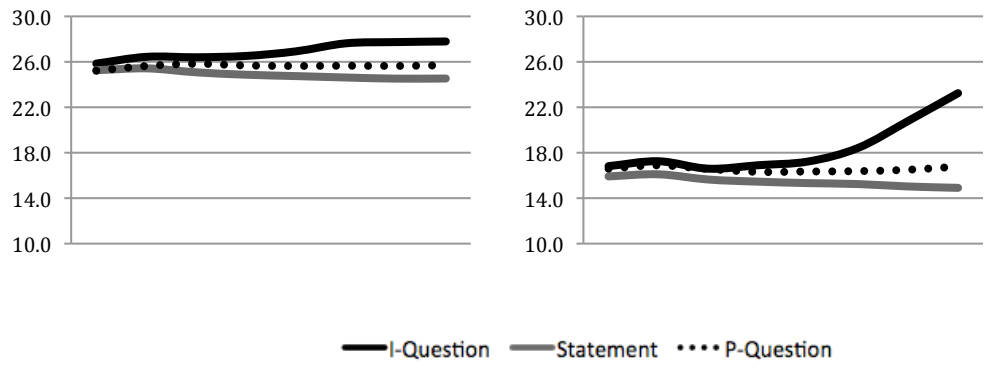

Figure 3.3 F0 contours (in semitone) of final T3 by female (L) and male (R) speakers

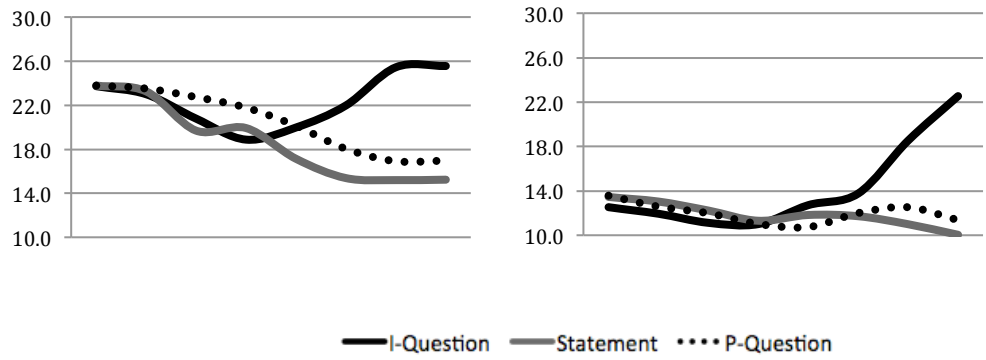

Figure 3.4 F0 contours (in semitone) of final T4 by female (L) and male (R) speakers

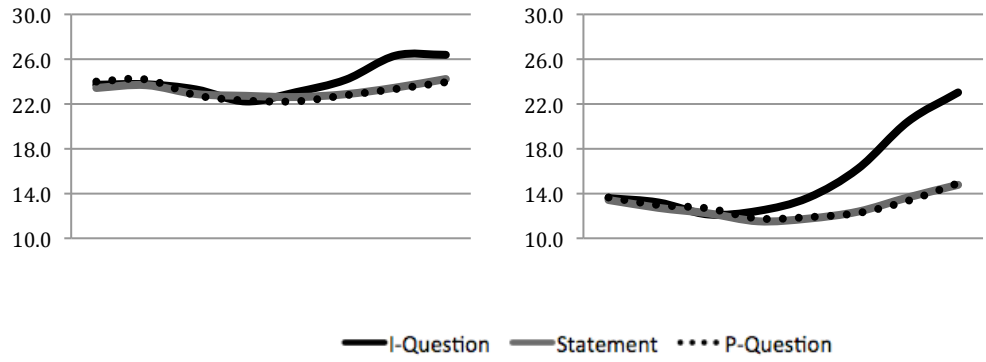

Figure 3.5 F0 contours (in semitone) of final T5 by female (L) and male (R) speakers

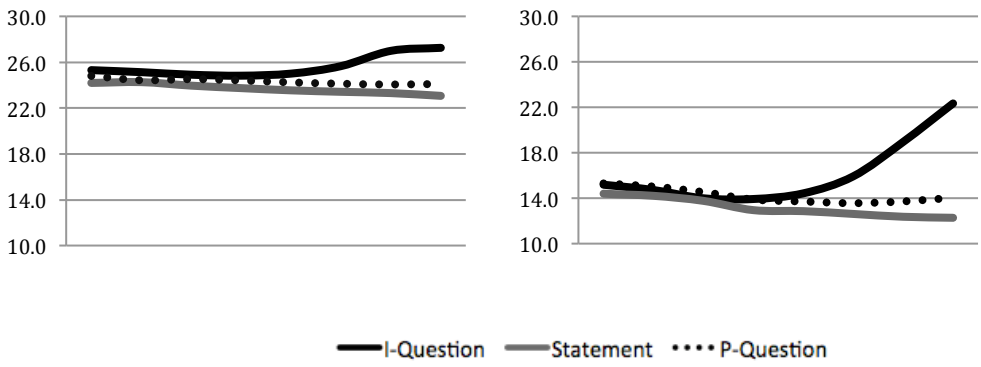

Figure 3.6 F0 contours (in semitone) of final T6 by female (L) and male (R) speakers

Results of two-way ANOVA (Tone × Intonation) test using the Linear mixed model analysis are shown in Table 3.2. It is found that a main effect of Tone exists throughout the contours ($p<0.001$ in every time point, see Table 3.1), a natural result of six different lexical tones. The main effect of Intonation occurs starting from point 6 [$F(2,285)=3.17$, $p<0.05$] and remains at point 7 [$F(2,281)=11.69$, $p<0.001$], point 8 [$F(2,281)=39.87$, $p<0.001$] and point 9 [$F(2,275)=57.29$, $p<0.001$]. In these final four points, the F0 of intonation questions is significantly higher than those of statements and particle questions, while there is no difference between statements and particle questions (statistic details see Table 3.1). The Tone × Intonation interaction becomes only marginal significant until the 9th time point [$F(10,275)=1.83$, $p=0.055$]. The statistic results confirm that the first half of the ending tones of intonation question retains its canonical form, and the final rising only occur at the second half of the tones (from point 6 onwards). The pre-particle syllable of particle questions does not

rise up and is statistically the same as the final tone in statements, indicating that the information load of a particle question rely entirely on the particle.

Table 3.1 The results of ANOVA test of the final-rising experiment data

Time point	Tone	Intonation	Interaction
2	F(5,286)=7.99	n.s.	n.s.
3	F(5,284)=13.39	n.s.	n.s.
4	F(5,281)=18.68	n.s.	n.s.
5	F(5,286)=23.08	n.s.	n.s.
6	F(5,285)=23.59	F(2,285)=3.17	n.s.
7	F(5,281)=23.15	F(2,281)=11.69	n.s.
8	F(5,281)=19.54	F(2,281)=39.87	n.s.
9	F(5,275)=18.74	F(2,275)=57.29	n.s
Significance at	p<0.001	p<0.05 (6) P<0.001 (7-9)	

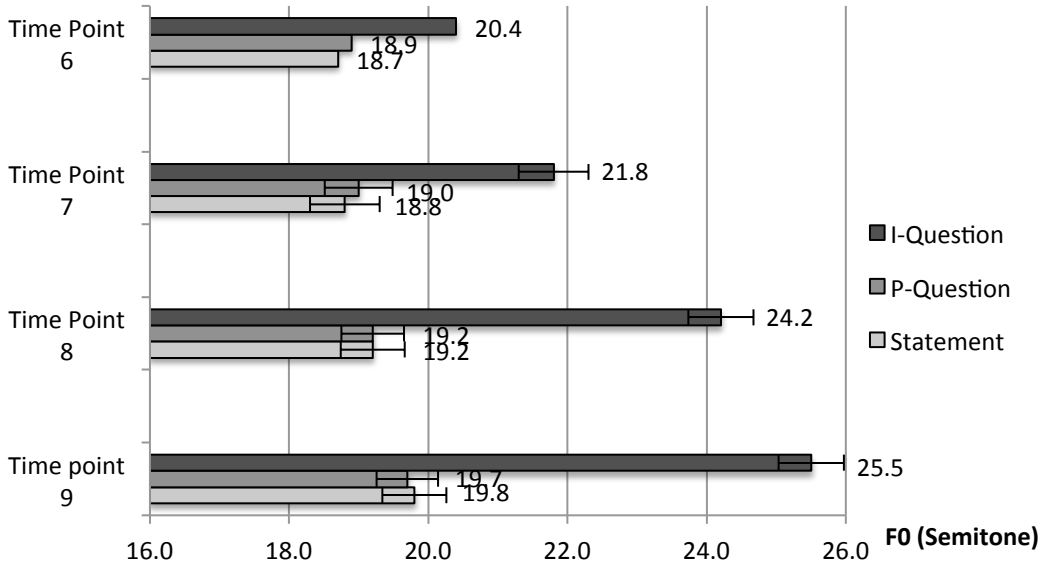

Figure 3.7 Average F0 (semitone) of the final tone contour at the last four time points of the three Intonation types

3.3 Experiment 2: the F0 of the global contour

3.3.1 Reading Materials

The second experiment in the production study investigates the global contour of the three intonation types. In order to achieve this, the lexical tones of the whole utterances need to be controlled. Six sentences as shown in Table 3.1, each consisted of only one single lexical tone, were designed. Every sentence is made up of five syllables, which can be read as a statement and as an intonation question. When read as a particle question, one syllable in the sentence was deleted and was replaced by the additional sentence final particle for question “me1” (the word in bracket in

Table 3.2), so as to control the length of all the utterances. The deletion did not change the general meaning of the whole sentence. These sentences were given to the participants randomly and were mixed with an equal number of filler sentences, which also appeared in three different intonation types. The participants were encouraged to read them naturally without placing emphasis on any specific part in the sentences.

Table 3.2 Stimuli of the Global F0 production test

(。- statement; ？- intonation question; 咩？- particle question)

T1	今晚煲(雞)湯 。/ ？ / 咩? gam1 maan1 bou1 (gai1) tong1 . / ? / me1? Making (chicken) soup tonight . / ? / SFP?
T2	小姐(想)飲酒 。/ ？ / 咩? siu2 ze2 (soeng2) jam2 zau2 . / ? / me1? The girl (wants to) drink . / ? / SFP?
T3	阿霸(去)訓覺 。/ ？ / 咩? aa3 baa3 (heoi3) fan3 gaau3 . / ? / me1? Mr. Ba (goes to) sleep . / ? / SFP?
T4	人群(來)遊行 。/ ？ / 咩? jan4 kwan4 (loi4) jau4 haang4 . / ? / me1? The crowd (comes to) parade . / ? / SFP?
T5	李理買蟹(柳) 。/ ？ / 咩? lei5 lei5 maai5 haai5 (lau5) . / ? / me1? Lily buys crab (stick) . / ? / SFP?
T6	學校做(壞)事 。/ ？ / 咩? hok6 haau6 zou6 (waai6) si6 . / ? / me1? The school does something (bad) . / ? / SFP?

3.3.2 Data analysis

All syllables in the target sentences were manually labeled in Praat (Boersma and Weenink 2013). The boundaries of the syllables were determined by the periodicity of the waveforms and the starting/ending of the formant. When two assimilated nasal sounds (as in the T1 sentence "tonight") or two neighboring vowels (as in the T6 sentence "do-bad") have no clear visual distinctions in formant and waveform, the amplitude curve was used to assist the segmentation. The first four syllables in the sentences were targets for investigation. F0 values of ten equidistant points along each tone contour of the target syllables were tracked by Praat. Following the procedure of Experiment 1, manual checking of the automatically tracked F0 values was performed before all the F0 values in Hertz were transferred into semitone.

Linear mixed model analysis (ANOVA) was employed to analyze the data. F0 values of the first and the last measurement point of each tone were excluded in the statistical analysis, due to similar consideration in Experiment 1. The F0 values in semitone from measurement points 2 to 9 were the dependent variables, whereas lexical tone (6 levels), intonation (3 levels: intonation question, statement, particle question), and position in the sentence (4 levels, i.e., the four syllable positions) were within-subject factors. Bonferroni correction was used in the post hoc tests.

3.3.3 Results

Figure 3.8 shows the average F0 contour of all participants for all six sentences. It can be observed that the differences among intonation question, statement and particle question are quite minor. Moreover, the differences in the the three contours are not consistent. For example, some seemingly consistent patterns may be observed in the current data, such as intonation question contour is higher than the other two in T6 utterance; or intonation and particle question contours are very similar and are both consistantly higher than statement contour in T3 utterance. However, mixed patterns are also found: intonation question contour is higher than the other two in the second and third syllable of T1 utterance; but particle question becomes higher in the fourth word in the same utterance. While relatively lower in some cases, statement contour is higher than the other two at many points in the third and fourth words in T4 utterance. The F0 contours of the three intonation types even entirely overlap at the second word in T2 utterance. The complicated tone-intonation-position interaction would make the listeners very difficult to decode intonation information based on such confusing patterns.

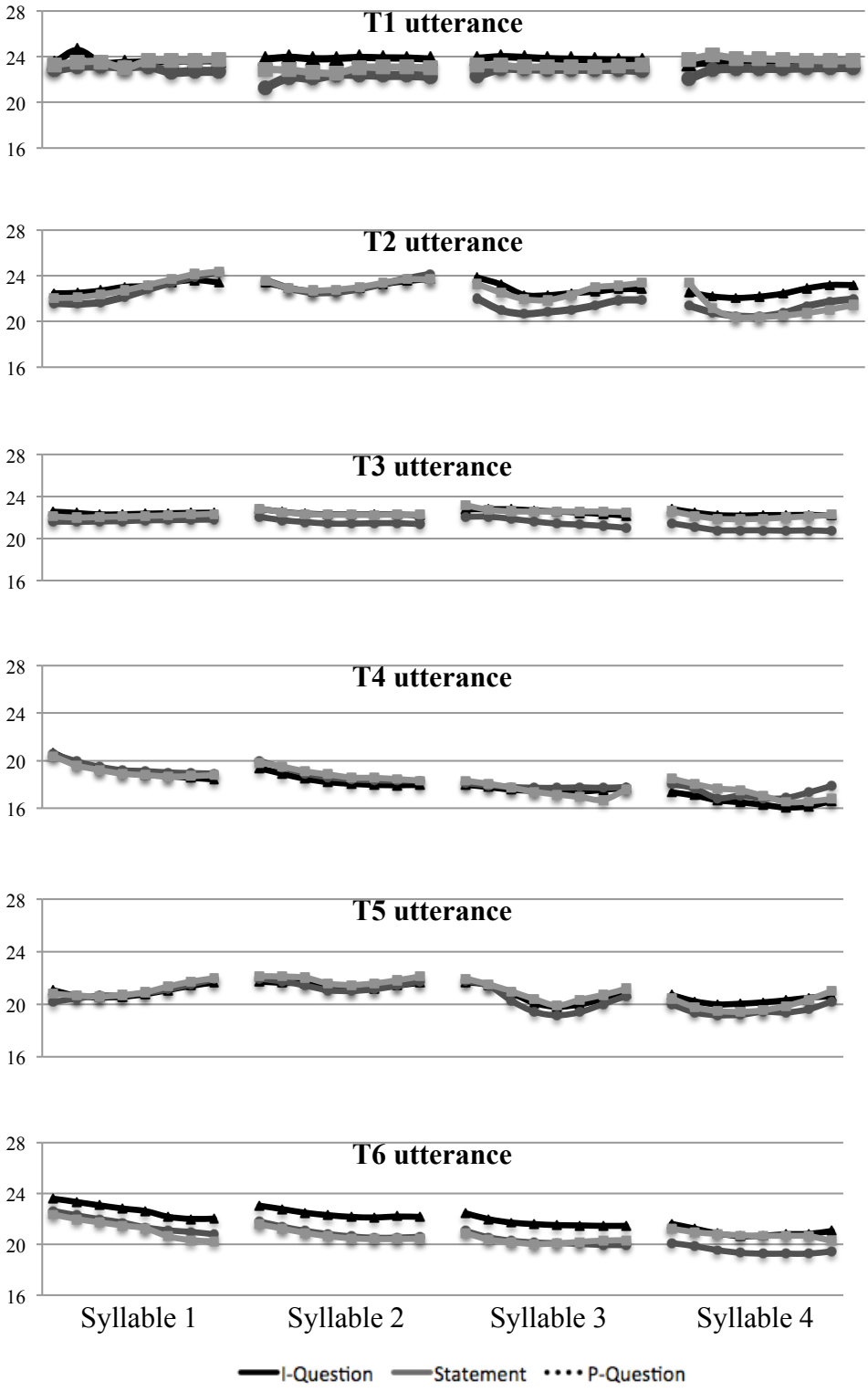

Figure 3.8 Average F0 contour in semitone of all participants

Linear mixed model analysis (ANOVA) shows that the main fact of tone is significant at every point ($p<0.001$ in every analysis), reflecting the difference of the six lexical tones. No significant main effect was found for intonation and position at any point throughout the data. Nor is there any interaction effect among tone, intonation and position. This result shows that Cantonese intonation does not influence F0 value at any point in the pre-final positions. It clearly shows that that Cantonese global contour does not play any role in carrying intonation information.

3.4. Summary

The current data show that final rising only modifies the second half of the question-final lexical tone. Out of ten equidistant points along the pitch contour, only starting from the sixth point does question intonation have a significant impact on the final tone. This result echoes Ma et al. (2006a), whose experiment investigated the influence of intonation on tonal perception in Cantonese. Their results showed that question-final T4, T5 and T6 had a higher risk than the other tones to be mistaken as T2, because T2 shared a low start with those tones. The highly localized rising could be regarded as a result of an alignment to a boundary high target (H%). That is why Cantonese T3 can remain a mid start, instead of being replaced by the contour of T2 (low start and high rising) entirely. The final rising is a result of tone alignment with

this high target.

The current data also show that the lexical tone before a question SFP is not affected by the final rising, suggesting that the question intonation is also localized in the sentence final position in a particle question. The pattern suggests that intonation question and particle question share the same intonation pattern in Cantonese. However, this result alone cannot make a claim that SFPs in Cantonese are used as intonation holders for two reasons. First, many SFPs in Cantonese are associated with certain pragmatic meanings that are not directly associated with their realized pitch contour. Second, as Wu (2008) has shown that the SFP "me" at the end of yes-no questions have a rising contour instead of a high level pitch contour, equivalent to the final rising pattern of the last tone of questions without SFPs. It was implied that the SFP "me" undergoes tonal modification of final rising like other question-final lexical items do. Therefore "me" has an intrinsic lexical tone and should not be regarded simply as an intonation holder. With the presence of a complicated SFPs system, the phonology and pragmatic functions of SFPs in Cantonese utterances need more systematic investigation in the future.

The results of Experiment 1 suggest that the ceiling effect of T1 question ending may be subject to individual differences. While previous studies found no evidence to show that T1 question in Cantonese was also signaled by final rising like questions

ending with the other tones were, the current findings show that there exists a high boundary target in a question already ending with a high tone. Whether the high boundary target is realized as rising for T1 needs further investigation.

Contrary to Lee (2008), the results in this study also show that there is no global F0 event for question intonation in Cantonese. Different intonations do not have a significant influence on pre-final tones in both intonation questions and particle questions. In comparison to the inconsistency and insignificance of the global raising, the final rising is a reliable and strong cue for Cantonese questions. Because no global effect for question intonation was detected, it is necessary to model Cantonese question intonation with a boundary high target, realized as final rising in intonation questions or carried by SFPs in particle questions.

CHAPTER 4

CROSS-LINGUISTIC PERCEPTION USING NORMAL SPEECH

Both Mandarin and Cantonese use a high pitch to signal question, but in distinctive ways. Mandarin uses a global pitch raise and Cantonese uses a localized rising at the end of the last tone. This study examines how both groups of listeners perform when they listen to both languages in an identification task. Two experiments were conducted. The first one investigated listeners' identification in normal speech, and examined how they identified intonation type of the two languages respectively. It is expected that listeners would identify the intonation of their native language by using their phonological knowledge; while they might need the Frequency Code when interpreting the intonation of the less familiar / unfamiliar language. The procedure and result are presented in this chapter. The second experiment, further looking at their reaction when the segmental information and identity of the language were concealed by a low-pass filter, is discussed in Chapter 5.

4.1. Method

4.1.1 Stimuli Design

The materials in both languages were designed to compare the effects of final rising and global contour on the perception of question intonation, and the interaction between lexical tones and intonation. Accordingly, two sets of nine-syllable sentences in Mandarin and Cantonese shown in Tables 4.1 and 4.2 were included in the experiment.

The final two syllables of each sentence share the same lexical tone. With the final syllables cut off, the utterances still remained meaningful, and the ending tone remained the same (with the exception of T3 in Mandarin because of the unavoidable tone sandhi, in which case the second last T3 dipping tone would be realized as a high rising contour).

Table 4.1 Mandarin utterances used in the experiment[1]

Finals	Sentences in Chinese and pinyin with English translation
T1	妈妈今晚炖的是鸡(汤)。 ma1 ma jin1 wan3 dun4 de shi4 ji1 (tang1). 'Mommy cooked chicken (soup) for tonight'
T2	亚马逊是最长的河(流)。 ya3 ma3 xun4 shi4 zui4 chang2 de he2 (liu2).

[1] The Mandarin words used in this study are transcribed in Pinyin, designed and promoted by the Chinese government. Mandarin tones are specified in tone category numbers instead of diacritics for easiness of reference and typography.

	‘Amazon is the longest river.’
T3	他最大的缺点是懶(散)。 ta1 zui4 da4 de que1 dian3 shi4 lan3 (san3). ‘His biggest shortcoming is laziness.’
T4	工人在修公园的路(面)。 gong1 ren2 zai4 xiu1 gong1 yuan2 de lu4 (mian4). ‘The workers are repairing the road in the park.’

Table 4.2 Cantonese utterances used in the experiment

Finals	Sentences in Chinese and Jyutping with English translation
T1	媽媽今晚煮嘅係雞(湯)。 maa1 maa1 gam1 maan1 zyu2 ge3 hai6 gai1(tong1). ‘Mommy cooked chicken (soup) for tonight’
T2	琴日老黃佢地好早(走)。 kam4 jat6 lou5 wong2 keoi5 dei6 hou2 zou2 (zau2). ‘Yesterday Mr. Wong and his friends leave early.’
T3	佢琴晚三點先去瞓(覺)。 keoi5 kam4 maan5 saam1 dim2 sin1 heoi3 fan3 (gaau3). ‘He went to bed at three last night.’
T4	亞馬遜係最長嘅河(流)。 aa3 maa5 seon3 hai6 zeoi3 zeong4 ge3 ho4 (lau4). ‘Amazon is the longest river.’
T5	阿婆琴日買嘅係蟹(柳)。 aa3 po4 kam4 yat6 maai5 ge3 hai6 haai5 (lau5). ‘Granny bought crab (stick) yesterday.’
T6	工人鋪緊黃大仙路(段)。 gung1 jan4 pou1 gan2 wong4 daai6 sin1 lou6 (dyun6). ‘The workers are paving the Wong Daai Sin Road.’

4.1.2 Recording and Editing

Two native Hong Kong Cantonese speaker (1 male, 1 female) and two native

Mandarin speakers (1 male, 1 female) were recorded reading the sentences for the experiment. The Cantonese speakers were undergraduate students born and raised in Hong Kong, aged 22. Their pronunciations were screened before the recording to make sure that no tone merge was present in their pronunciation. The Mandarin speakers came from Shaanxi (male) and Hubei (female), both Mandarin-speaking districts. Both of them were qualified and experienced Mandarin instructors at the Chinese University of Hong Kong, aged 30 (male) and 26 (female). The recording took place in a sound-treated room. The target sentences appeared on a computer screen in a random order. With clear instruction, each subject read the sentences for multiple times in two forms: as a question and as a statement. To eliminate the potential discrepancy caused by the different focus patterns in the two languages, the speakers were instructed to read the sentences focus-neutrally.

After screening the naturalness of the utterances produced by two linguistically trained native speakers of each language, one statement and one question of every final tone from each recorded speaker with the best quality were chosen. Four presentation sentence conditions were prepared for the perception test: complete statements and statements with the final syllable cut off, complete questions and questions with the final syllable cut off. All cutting points were at the zero crossing following the offset of the penultimate syllables of the utterances. The average

amplitude of all the utterances was also normalized using PRAAT.

4.1.3 Listeners and Procedures

Twenty Cantonese (5 male and 15 female) and twenty Mandarin listeners (2 male and 18 female) participated in the perception experiment. They were MA or undergraduate students at the Chinese University of Hong Kong, between 18 and 24 years old. All the Cantonese listeners were native Hong Kong Cantonese speakers, and could speak Mandarin with varying proficiency. All the Mandarin listeners came from Mainland China, speaking Standard Mandarin in their daily life. Eighteen of them had been in Hong Kong for less than two months and found Cantonese difficult to understand. Two of them have arrived in Hong Kong for a year and could understand some basic Cantonese, but did not speak it in daily life. All the listeners spoke English as a foreign language. The subjects were naive listeners since none have received any systematic training in linguistics or phonetics by the time they participated in the experiment. None of them had a reported history of speech or hearing disorder.

The perception experiment was carried out in a sound-attenuated room. The materials were presented to them in blocks on a computer screen using the software E-Prime. A practice session preceded the testing blocks to familiarize the subjects

with the experiment task. In this session, they heard 8 utterances different from the target sentences in the experiment. Four of them were spoken by a female Cantonese speaker and four were spoken by a male Mandarin speaker. The first four utterances (2 from each language) were low-pass filtered (for data in Chapter 5) and could only be heard as humming, the next four were normal speech (2 from each language). After the practice session, the listeners had to finish four blocks of low-pass filtered speech identification (Chapter 5) before they proceeded to the blocks with normal speech, to prevent them from guessing the content of the filtered materials.

Before each block, subjects were told that the utterances within that block were either all in Mandarin or Cantonese. The blocks and the tokens within each block were randomized. Each of the 80 stimuli (4 sentence conditions × 2 genders × 4 tones in Mandarin and 6 tones in Cantonese) was repeated twice, resulting in 160 stimuli in total. Task instructions were given visually on screen and verbally by the author. Listeners listened to each trial only once before they decided whether the utterance they just heard was a statement or a question. They input the answer by pressing a button on the keyboard. Then they rated their confidence of their answer in a 1-7 scale (1 being the least confident and 7 the most confident) before moving on to the next utterance.

4.1.4 Data analysis

Average identification accuracy and confidence rating were calculated. Three-way (Intonation type × Condition × Final tone) repeated measures ANOVAs were conducted on the identification scores and confidence rating for each language group, Corrections for violations of sphericity were made, where appropriate, using the Greenhouse–Geisser estimates of sphericity. Bonferroni correction was applied to make pairwise comparison. Analysis of correlation between identification accuracy and confidence rating measured by Person's r was also conducted. In each figure below, the numbers on the horizontal axis stand for the lexical tones; L stands for complete utterances; C stands for cut-off utterances.

As pointed out by previous studies using native listeners (Ma et al. 2011, Yuan 2011), the perception of intonation was biased towards statement. Statement might be the default choice when listeners were not sure but forced to choose an answer. To account for this bias, the accuracy data was also analyzed in the theoretical framework of signal detection theory (Macmillan and Creelman 2005). Discriminability and bias measures were computed using the following formula (Yuan 2011).

$$A' = 0.5 + [(H-FA)*(1+H-FA)]/[4*H*(1-FA)]$$

$$B'' = [(1-H)*(1-FA)-H*FA]/[(1-H)*(1-FA)+H*FA]$$

In the formula, H (hits) is the percentage of correctly identified questions. FA (false alarm) is the percentage of statements that are mistaken as questions. The score of A' (0~1) measures discriminability between statement and question. The closer to 1, the higher discriminability it indicates; whereas a score of 0.5 indicates chance level. The score of B'' (-1~1) measures bias. Positive scores show favor to statements; while negative scores show favor to questions; and no bias is detected if the score equals 0.

4.2 Results

4.2.1 Cantonese listeners listening to Cantonese utterances

Figure 4.1 shows the identification accuracy of Cantonese intonation by Cantonese listeners. The listeners performed well except in identifying the cut-off questions. Statistical analysis indicates that the main effect of Intonation type (statement or question) is significant [$F(1,19)=277.32$, $p<0.001$]. Statements (M=99%, SD=0.06) were better identified than questions (M=57%, SD=0.43). The main effect of Condition (complete or cut-off) is also significant [$F(1,19)=336.65$, $p<0.001$], with complete utterances (M=97%, SD=0.10) better identified than cut-off utterances (M=59%, SD=0.44). The interaction between Intonation type and Condition is also

significant [F(1,19)=306.78, p<0.001], indicating that the difference between statement and question is sensitive to Condition. Figure 4.1 shows that only in the cut-off condition, identification accuracy of statements is clearly higher than questions.

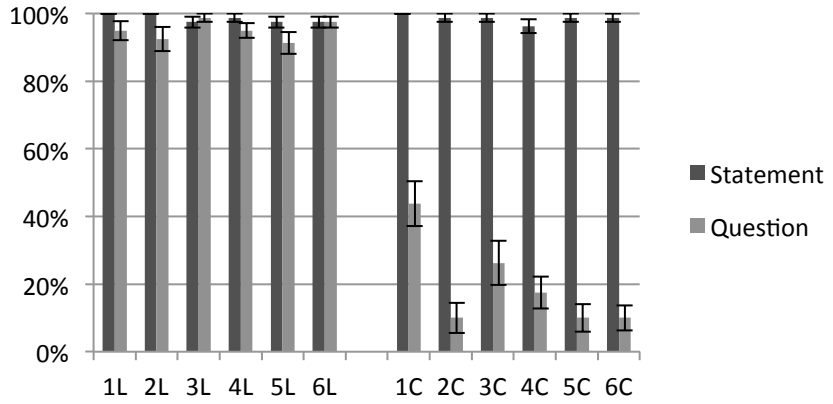

Figure 4.1 Average percentage of correct identification of Cantonese normal utterances by Cantonese listeners
(L: complete utterances; C: cut-off utterances)

Final tone also has a significant impact over the identification of intonation [F(5,95)=13.48, p<0.05]. The Tone × Condition [F(3.35, 63.73)=6.66, p<0.001] and Tone × Intonation type [F(5,95)=10.29, p<0.001] interactions are also significant, indicating that the differences between tones are subject to Condition and Intonation type. In pairwise comparisons corrected by Bonferroni adjustment, no difference was found between different ending tones in the complete condition. However, within the cut-off condition, while the identification of statement remained

similar to the performance in the complete utterance condition, questions were poorly identified, all falling below the chance level (50%). Among these cut-off questions, listeners performed better in identifying T1 question than questions with the other tones (significant at $p<0.05$ in every comparison except the one between T1 and T3). This shows that the final syllable of Cantonese has a strong impact over question identification for Cantonese listeners. In addition, a question ending with the high level tone was most likely to be identified as question than questions ending with other tones for native listeners.

Table 4.3 Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'') of Cantonese complete (upper panel) and cut-off (lower panel) utterances by Cantonese listeners (L: complete utterances; C: cut-off utterances)

	1L	2L	3L	4L	5L	6L
H	0.95	0.93	0.99	0.95	0.91	0.98
FA	0.00	0.00	0.03	0.01	0.03	0.03
A'	0.99	0.98	0.99	0.98	0.97	0.99
B''	1.00	1.00	-0.34	0.61	0.58	0.00

	1C	2C	3C	4C	5C	6C
H	0.44	0.10	0.26	0.18	0.10	0.10
FA	0.00	0.01	0.01	0.04	0.01	0.01
A'	0.86	0.74	0.80	0.73	0.74	0.74
B''	1.00	1.00	0.99	0.98	1.00	1.00

The discriminability scores in Table 4.3 show that complete utterances have

strong discriminability across all final tones. However, when the final syllable is cut off, the discriminability scores drop for all utterances. Cut-off utterances ending with T1 have the highest discriminability score in comparison to utterances ending with the other tones. This indicates that the intonation perception of T1 utterances is the least affected by eliminating the final syllable. The bias scores show that Cantonese listeners display a bias towards statement in utterances with most tones, except for utterances ending with T3 (bias towards question) and T6 (no bias). However, in the cut-off condition, the identification of intonation displays a strong bias towards statement for every final tone, probably because of the lack of the final tone that Cantonese listeners depend on to identify questions.

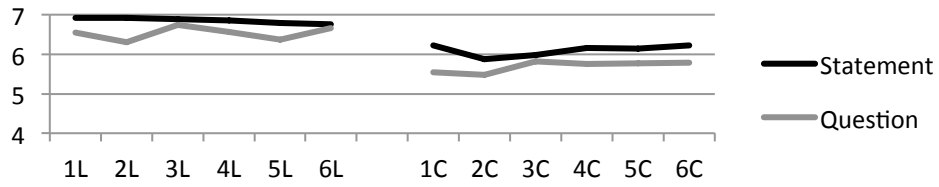

Figure 4.2 Average confidence rating of Cantonese normal utterances by Cantonese listeners (L: complete utterances; C: cut-off utterances)

Generally, there is a strong and significant correlation between confidence rating and identification accuracy [Pearson's $r(22)=0.748$, $p<0.01$]. Figure 4.2 shows the average self-rated confidence level of Cantonese listeners' judgment. Main effects of

Intonation type [F(1,19)=14.80, p<0.05] and Condition [F(1,19)=12.32, p<0.05] were found. Listeners were significantly more confident in identifying statements (M=6.48, SD=0.80) than questions (M=6.11, SD=1.20); and they were significantly more confident with complete utterances (M=6.70, SD=0.75) than cut-off utterances (M=5.90, SD=1.12). The result shows no main effect of Final tone [F(5,95)=2.06, p>0.05]; nor are there any interactions of tone with other factors (p>0.05 in each interaction). This indicates that Cantonese native listeners were more diffident when they came across questions and cut-off utterances, no matter what tone the utterances ended with.

4.2.2 Mandarin listeners listening to Cantonese utterances

Figure 4.3 shows the identification accuracy of Cantonese intonation by Mandarin listeners. Similar to Cantonese speakers, Mandarin listeners performed better in identifying statements (M=92%, SD=0.16) than questions (M=59%, SD=0.42), with a significant main effect of Intonation type [F(1,19)=109.10, p<0.001]. They achieved a better identification accuracy in complete utterances (M=95%, SD=0.13) than in cut-off utterances (M=56%, SD=0.41), with a significant main effect of Condition [F(1,19)=607.43, p<0.001]. The significant interaction Condition × Intonation type [F(1,19)=258.30, p<0.001] shows that the difference between

statement and question is significant only in the cut-off condition.

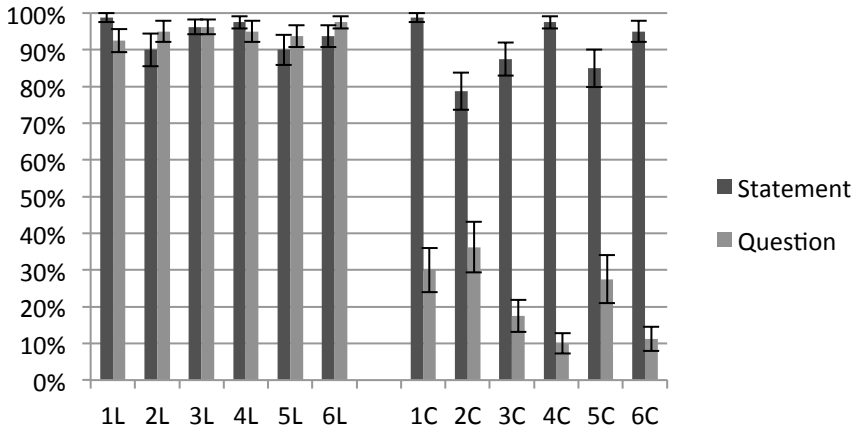

Figure 4.3 Average percentage of correct identification of Cantonese normal utterances by Mandarin listeners
(L: complete utterances; C: cut-off utterances)

The main effect of Final tone is not significant [$F(1,19)=2.14$, $p>0.05$]. However, the interactions between tone and the other factors are all significant: Final tone × Intonation type [$F(2.67,50.68)=5.27$, $p<0.05$], Final tone × Condition [$F(2.67,50.68)=5.27$, $p<0.05$]. The results suggest that the perceptual differences of intonation caused by tones are dependent on intonation type and condition for Mandarin listeners. Specifically, there is no significant difference in identification accuracy among complete utterances with different tones, in either statements or questions. The general identification accuracy of cut-off questions is quite low. All utterances were identified below the chance level (50%), showing that Mandarin

listeners also strongly relied on the last syllable to identify questions in Cantonese. More specifically, in the cut-off condition, Mandarin listeners identified T2 statements poorly, with accuracy significantly lower than that of T1 and T4 statements ($p<0.05$ in each comparison). This suggests that Mandarin listeners might have misinterpreted the high rising T2 of cut-off T2 statements as a signal for questions. Meanwhile, T2 question was best identified among all cut-off questions, significantly higher than the worst identified T4 question ($p<0.05$).

Table 4.4 Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'') of Cantonese complete (upper panel) and cut-off (lower panel) utterances by Mandarin listeners (L: complete utterances; C: cut-off utterances)

	1L	2L	3L	4L	5L	6L
H	0.93	0.95	0.96	0.95	0.94	0.98
FA	0.01	0.10	0.04	0.03	0.10	0.06
A'	0.98	0.96	0.98	0.98	0.96	0.98
B''	0.73	-0.36	0.00	0.34	-0.25	-0.44

	1C	2C	3C	4C	5C	6C
H	0.30	0.36	0.18	0.10	0.28	0.11
FA	0.01	0.21	0.13	0.03	0.15	0.05
A'	0.81	0.65	0.59	0.71	0.65	0.66
B''	0.99	0.73	0.94	0.99	0.87	0.99

Table 4.4 shows the discriminability and bias scores of the identification of Cantonese intonation by Mandarin listeners. High discriminability scores across all

final tones are shown in the complete condition. Similar to Cantonese listeners, the discriminability score drop in the cut-off condition, with T1 utterances remaining a higher score than the other ending tones. Listeners show bias towards question when identifying utterances ending with T2, T5 and T6. However, in the cut-off condition, a uniformed bias towards statement across all ending tones is shown.

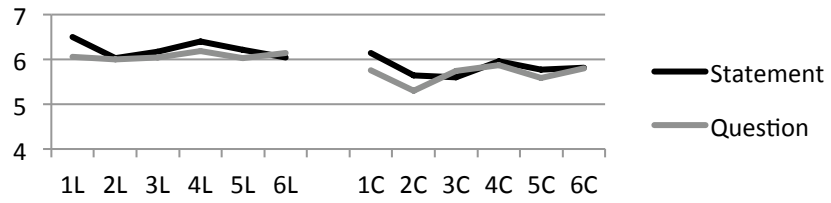

Figure 4.4 Average confidence rating of Cantonese normal utterances by Mandarin listeners (L: complete utterances; C: cut-off utterances)

Figure 4.4 shows the average confidence rating by Mandarin listeners. Similar to Cantonese listeners, the self-rated confidence of Mandarin listeners is significantly correlated with their identification accuracy [Pearson's $r(22)=0.603$, $p<0.01$]. ANOVA shows that Mandarin listeners did not differ in their confidence ratings for statements and questions, as the main effect of Intonation type is not significant [$F(1,19)=3.25$, $p>0.05$]. On the other hand, listeners gave significantly different confidence ratings to complete and cut-off utterances, as the main effect of Condition [$F(1,19)=10.02$, $p<0.05$] is significant, with complete utterances ($M=6.15$, $SD=1.43$) having higher confidence rating than cut-off utterances ($M=5.75$, $SD=1.20$). The

main effect of Final tone is also significant [$F(3.11,59.11)=3.87$, $p<0.05$], so is the Intonation type × Final tone interaction [$F(5,95)=4.20$, $p<0.05$]. These results show that Mandarin listeners had different levels of confidence when they listened to questions and statements, depending on the ending tone of the utterance.

When Mandarin listeners listened to complete utterances, they showed no difference in conference ratings across different Final tones ($p>0.05$ in every comparison). However, in the cut-off condition, they were most confident in identifying T1 statements ($M=6.15$, $SD=1.23$), significantly higher than T2 statements ($M=5.65$, $SD=1.22$, $p<0.05$), which echoes the identification accuracy result. Surprisingly, they had the lowest confidence for T2 cut-off questions ($M=5.30$, $SD=1.34$), which they did relatively well in identification. The confidence of T2 cut-off question is significantly lower than T3 and T6 questions ($p<0.05$ in each comparison). This inconsistence in identification accuracy and confidence suggests that even though listeners might be able to categorize T2 cut-off questions correctly, they were unsure about their answers. Cantonese cut-off utterances ending with T2 might be confusing for Mandarin listeners.

4.2.3 Mandarin listeners listening to Mandarin utterances

Figure 4.5 shows the identification accuracy for Mandarin listeners listening to

Mandarin normal utterances. The main effect of Intonation type is significant [$F(1,19)=23.02$, $p<0.001$], as statements (M=95%, SD=0.14) are better identified than questions (M=73%, SD=0.33). The main effect of Condition is also significant [$F(1,19)=148.52$, $p<0.001$], with complete utterances (M=94%, SD=0.15) better identified than cut-off utterances (M=74%, SD=0.33). The interaction Intonation type × Condition is significant [$F(1,19)=53.32$, $p<0.001$], indicating that the difference between statement and question depends on the conditions.

There is also a significant main effect for Final tone [$F(3,57)=18.65$, $p<0.001$] and the interactions Final tone × Condition [$F(3,57)=7.19$, $p<0.001$], and between Final tone × Intonation type [$F(1.99,37.78)=8.032$, $p<0.001$] are also significant. In the complete utterance condition, statements were well identified across different tones. T3 (M=78%, SD=0.26) complete question has the lowest identification accuracy, significantly lower than T2 and T4 ($p<0.05$ in each comparison). There is no significant difference among T1, T2 and T4 complete questions. In the cut-off condition, while T1, T2 and T4 statements remain well identified, T3 statement has the lowest accuracy (M=74%, SD=0.29), probably due to the penultimate high rising pitch contour induced by tone sandhi. The identification accuracy of T3 cut-off statement is significantly lower than T2 and T4 statements ($p<0.05$ in each comparison). Among the cut-off questions, T4 question remains almost unaffected

by the elimination of the last syllable, while the identification accuracy for T1 and T2 questions become much lower than their complete counterparts. The identification accuracy for T4 is significantly higher than all the other tones ($p<0.05$ in every comparison). The results suggest that T3 is a difficult tone for the identification in Mandarin questions in the compete condition. In contrast, T4 may be the easiest ending tone for Mandarin questions. Cutting off the final syllable influences the perception of intonation, but the exact influence depends on different tones.

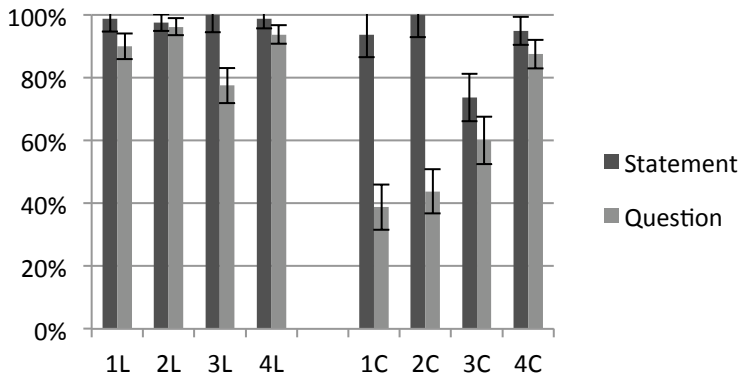

Figure 4.5 Average percentage of correct identification of Mandarin normal utterances by Mandarin listeners
(L: complete utterances; C: cut-off utterances)

Table 4.5 shows the discriminability and bias score of Mandarin listeners listening to Mandarin utterances. They had high scores of discriminability over all final tones in the complete condition, with T3 slightly lower than the others. In the cut-off condition, the discriminability scores of T1, T2 and T3 dropped, with T3

lowest among all the final tones. However, the discriminability score of T4 remains high, indicating Mandarin listeners grouped T4 questions and statements correctly even when the final syllable was cut off. All the bias scores show that Mandarin listeners have a bias towards statement in their judgment. Interestingly, in the complete condition, T3 has the highest bias score; while in the cut-off condition, it has the lowest. This clearly is a result of the rising penultimate tone in T3 utterances.

Table 4.5 Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'') of Mandarin complete (upper panel) and cut-off (lower panel) utterances by Mandarin listeners (L: complete utterances; C: cut-off utterances)

	1L	2L	3L	4L
H	0.90	0.96	0.78	0.94
FA	0.01	0.03	0.00	0.01
A'	0.97	0.98	0.94	0.98
B''	0.80	0.21	1.00	0.68

	1C	2C	3C	4C
H	0.39	0.44	0.60	0.88
FA	0.06	0.00	0.26	0.05
A'	0.80	0.86	0.76	0.95
B''	0.92	1.00	0.30	0.46

Figure 4.6 shows the self-rated confidence level of Mandarin listeners. The confidence given by this group of listeners significantly and strongly correlates with their identification accuracy [Pearson's $r(14)=0.862$, $p<0.01$]. Intonation type has a

significant main effect [F(1,19)=11.33, p<0.05], with statements (M=6.35, SD=1.29) being rated more confidently than questions (M=6.03, SD=1.42). There is also a significant main effect for condition [F(1,19)=4.53, p<0.05]. Listeners were more confident with complete utterances (M=6.41, SD=1.46) than cut-off utterances (M=5.98, SD=1.22). However, there is neither a significant main effect nor any significant two-way interactions for Final tones. Native Mandarin listeners found questions and cut-off utterances more difficult than statements and complete utterances. But they were confident about their own language and therefore did not differenciate their confidence across tones within the same condition and intonation type.

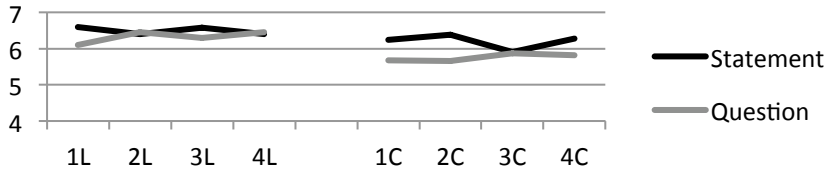

Figure 4.6 Average confidence rating of Mandarin normal utterances by Mandarin listeners (L: complete utterances; C: cut-off utterances)

4.2.4 Cantonese listeners listening to Mandarin utterances

Figure 4.7 shows that Cantonese listeners had lower identification accuracy than native Mandarin listeners, especially for questions. The main effect of Intonation

type is significant [$F(1,19)=71.31$, $p<0.05$], with statements ($M=95\%$, $SD=0.17$) higher than questions ($M=56\%$, $SD=0.37$). A main effect of Condition is also found [$F(1,19)=55.50$, $p<0.05$], as complete utterances ($M=85\%$, $SD=0.27$) are better identified than cut-off utterances ($M=66\%$, $SD=0.38$). The difference between the identification accuracy of statements and questions is subject to conditions, as the interaction Intonation type × Condition [$F(1,19)=12.18$, $p<0.01$] is significant.

There is a main effect for Final tones [$F(3,57)=13.10$, $p<0.05$], as well as significant interactions (Final tone × Intonation type [$F(2.11,40.03)=5.86$, $p<0.01$] and Final tone × Condition [$F(3,57)=26.00$, $p<0.001$]). For the complete utterances, listeners performed well over statements with all ending tones. This is not the case for questions, as T3 complete question is poorly identified ($M=43\%$, $SD=0.20$, below chance level). The identification accuracy is significantly lower than complete questions with the other tones ($p<0.05$ in every comparison). For the cut-off statements, T3 statements is the worst identified ($M=71\%$, $SD=0.34$), with identification accuracy significantly lower than the other tones ($p<0.05$ in every comparison). For the cut-off questions, T1 and T2 questions (both identified below chance level) are identified significantly more poorly than T3 and T4 questions ($p>0.05$ in every comparison). There are no significant differences between the identification accuracy of T1 and T2 ($p>0.05$) or between that of T3 and T4 ($p>0.05$).

The result shows that Cantonese listeners also found T3 to be a difficult ending tone. They also found questions ending with T1 and T2 in the cut-off condition confusing.

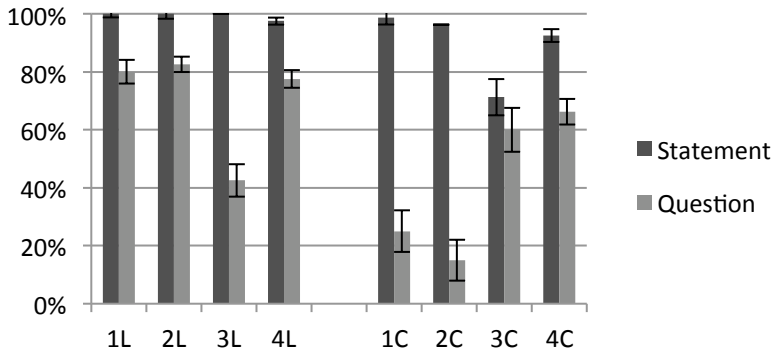

Figure 4.7 Average percentage of correct identification of Mandarin normal utterances by Cantonese listeners
(L: complete utterances; C: cut-off utterances)

Table 4.6 Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'') of Mandarin complete (upper panel) and cut-off (lower panel) utterances by Cantonese listeners
(L: complete utterances; C: cut-off utterances)

	1L	2L	3L	4L
H	0.80	0.83	0.43	0.78
FA	0.00	0.00	0.00	0.03
A'	0.95	0.96	0.86	0.93
B''	1.00	1.00	1.00	0.84

	1C	2C	3C	4C
H	0.25	0.15	0.60	0.66
FA	0.01	0.04	0.29	0.08
A'	0.80	0.72	0.74	0.88
B''	0.99	0.99	0.25	0.73

Table 4.6 shows that in the complete condition, Cantonese listeners have a high discriminability score in T1, T2 and T4, with T3 the lowest. The discriminability scores decreased in the cut-off condition, with T4 remaining the highest. The bias score in the complete condition is very high across tones, indicating that Cantonese listeners have a strong bias toward statement in this condition. In the cut-off condition, the bias score of T1 and T2 remain very high, a result of their poor performance of identifying T1 and T2 questions in this condition. The bias score of T3 cut-off utterances is the lowest among all the tones, indicating a relatively week bias towards statement.

Figure 4.8 shows the self-rated confidence level by Cantonese listeners to their answers in Mandarin utterances. The confidence rating has a significant correlation to the identification accuracy [Pearson's $r(14)=0.695$, $p<0.01$]. The main effect of intonation type is significant [$F(1,19)=21.34$, $p<0.05$]. Listeners shows more confidence in statements ($M=6.43$, $SD=0.87$) than in questions ($M=5.66$, $SD=1.31$). There is a main effect for condition [$F(1,19)=12.80$, $p<0.05$], as more confidence is shown in complete utterances ($M=6.34$, $SD=1.04$) than in cut-off utterances ($M=5.74$, $SD=1.23$). There is also a significant main effect for tone [$F(3,57)=8.58$, $p<0.05$]. Within the complete utterance condition, no significant difference between every two tones is found ($p>0.05$ in every comparison), except a marginal difference between

T3 and T2 complete questions (p=0.06). In the cut-off utterance condition, T3 cut-off statement has lower confidence level than all the others. However, there is no significant difference between every comparison (p>0.05). Apart from the accuracy data, T4 cut-off question receives the least confidence from the listeners, significantly lower than T2 question (p<0.05). This suggests that even though Cantonese listeners could figure out the identity of T4 cut-off questions, they still found it difficult to recognize.

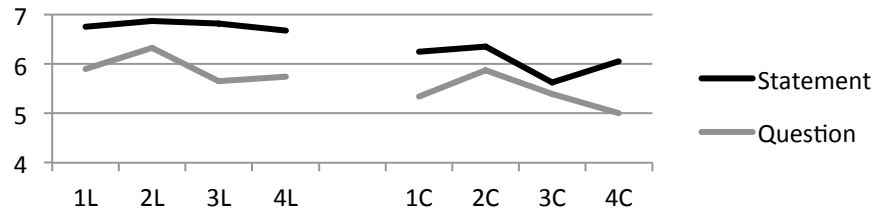

Figure 4.8 Average confidence rating of Mandarin normal utterances by Cantonese listeners (L: complete utterances; C: cut-off utterances)

4.3 Discussion

In general, the listeners from both groups performed well in the complete utterance condition. When listening to their native language, they displayed a ceiling performance: every utterance was successfully identified for above 90% of time (except T3 question in Mandarin, 74% in accuracy). There is no significant difference in identification accuracy among different tones in complete utterances,

except that the accuracy of Mandarin T3 complete question was significantly lower than the other tones. Native listeners were also very confident about their judgment in their own language. There is no significant difference in self-rated confidence among tones when listening to complete or cut-off utterances in their native language, showing that they were comfortable with the tasks concerning their native language, regardless of the tones. They had an overall higher confidence towards complete utterances than cut-off utterances, probably due to the unnaturalness of the cutting utterances.

On the other hand, listeners performed more poorly in a language that was not familiar to them. Mandarin listeners who do not understand Cantonese did relatively well when they listened to Cantonese complete utterances, probably owing to two reasons. First, the final rising in Cantonese is very prominent and easy to detect. Second, they were English learners for a long time and were familiar with the concept that questions could be signaled at the boundary position, even though boundary tone was not a necessary component in Mandarin question intonation. Therefore, while they could not understand most Cantonese utterances, the final rising might be an important cue that they relied on. In comparison, the Mandarin-speaking Cantonese listeners did not do as well in identifying Mandarin complete questions (around 80% for T1, T2 and T4 and 43% for T3), showing that

they could, to some degree, recognize the global intonation contour but could not make full use of it. The confidence level varies with tones when listeners listened to the unfamiliar/less familiar language. Sometimes utterances with high identification accuracy were given a low confidence level, e.g., T2 cut-off question in Cantonese by Mandarin listeners and T4 cut-off question in Mandarin by Cantonese listeners. This suggests that listeners still found those utterances confusing when they intuitively followed some cues and got it right.

Since the identification of complete utterances has been successful in most cases, the significant differences exhibited in the cut-off utterance condition worth more attention, and reveal aspects of intonation perception such as question-statement asymmetry, language-tone specific mechanism, and universal intonation code.

The results further the insights into question-statement bias previously observed in individual languages (Yuan 2004, Ma et al. 2010). For both groups of listeners in both languages, the question identification accuracy was constantly lower than statement, especially in cut-off utterances. The bias scores are positive for all the Mandarin utterances, all the Cantonese cut-off utterances, and most Cantonese complete utterances. The exceptions cannot be explained by the current the study. It could be the characteristics of the several Cantonese final tones, or it could be due to the design of the current study. But the general positive bias scores show that

statement is a preferred intonation type for both languages, particularly when identifying questions becomes more difficult because of the elimination of the final tone. Moreover, this experiment recorded the self-rated confidence of the listeners, giving a direct and new perspective to the listeners' attitude. In most cases, regardless of the language, the listener groups, the tones or the condition, listeners gave a higher confidence score to statements than questions. This indicates that statement was the preferred choice when strong acoustic cues for questions were not present. The fact that both groups of listeners showed the same preference towards statements suggests that the bias towards statement being an unmarked sentence type may be universal.

The results confirm the intonation patterns of Cantonese investigated in previous studies and in this study. Without the final syllable, the identification accuracy of questions in Cantonese decreased dramatically. This result is compatible with acoustic studies (Ma et al. 2006a) claiming that a boundary tone located at the very end of the final tone is the critical cue for Cantonese questions. The only cut-off question identified relatively well by Cantonese listeners was T1. This is probably because questions ending with T1 in Cantonese do not rely so much on the final rising, as T1 is already at the top of speakers' pitch range. Besides, as predicted by the Frequency Code, a high level tone at the end may have facilitated the

identification of questions.

Eliminating the last syllable did not influence the perception of Mandarin questions as much as it did for Cantonese questions. The overall average identification accuracy of Mandarin cut-off utterances (M=74%) by Mandarin listeners is higher than that of Cantonese cut-off utterances by Cantonese listeners (M=59%). When Cantonese listeners listening to Cantonese, none of the cut-off questions achieved the chance level (50%), with the average identification accuracy of T2-6 questions below 30%. However, in Mandarin, native listeners were able to successfully recognize cut-off questions ending with T3 and T4. Even though they did not perform as well in the cut-off questions with T1 and T2, the average accuracy (around 40%) is still higher than the Cantonese cut-off questions ending with T2-T5. This difference in the result indicates that Mandarin question intonation does not rely heavily on the final syllable like Cantonese questions do.

Agreeing with Yuan (2004, 2011), T4 seems to be an easy tone for question identification. In the complete condition, there is no significant difference between the identification accuracy of complete T1, T2, and T4 questions for Mandarin listeners, as a ceiling performance is shown. However, in the cut-off condition, T4 question was hardly influenced by the loss of the final tone, and was significantly better identified than questions with all the other tones.

Nonetheless, Yuan's explanation cannot account for the current data. What he examined was the effect of the final tone, while the T4 question with an identification advantage was deprived of the final tone in this experiment. Figure 4.9 shows the F0 contour of the Mandarin T4 question. The arrows point at the penultimate tone in the complete utterance, which is also the last tone in the cut-off condition. Due to anticipatory tonal coarticulation, the penultimate T4 has a rising tail to prepare for the next high falling tone. This rising tail may give away the identity of a question. However, this rising tail alone cannot explain the high accuracy of T4 questions, as T2 and T3 cut-off questions that actually ended with a rising tone was not identified as accurately. This indicates that the tone-specific mechanism Yuan (2006) used to explain the identification advantage of questions ending with T4 could be extended to the pre-final tones. This idea is reasonable because the flattening of the final T4 (Yuan 2004) may take place at the pre-final T4 as a result of the raised sentential F0 contour. Figure 4.9 shows that the penultimate T4 had a gentle falling contour. This might benefit a tone-specific perception of question intonation. Although the perceptual result shows that Cantonese listeners did well in identifying T4 cut-off question, they showed the least confidence in judging T4 questions. This suggests that even if they were attending to the same cues Mandarin listeners used, they did not find the rising tail prominent and/or they were

unsure of the tone-specific intonation mechanism in Mandarin.

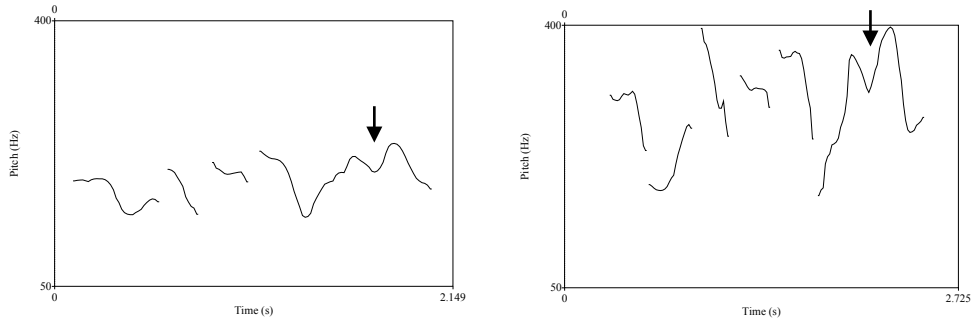

Figure 4.9 The F0 contour of the Mandarin T4 question (left: M; right: F), with the arrows pointing at the penultimate tone in the utterances

On the other hand, cutting off the final tone has brought down the identification accuracy for Mandarin questions ending with other tones, especially T1 and T2 for both listener groups. This might be caused by tonal coarticulation. As shown in Figure 4.10, in order to prepare for the final rising tone, the penultimate T2 produced by both Mandarin speakers did not rise to a high position in their pitch range, so the pitch contour might be regarded as a low rising and could not provide sufficient cue for question intonation. On the other hand, T1 (high level) cut-off questions were identified poorly by Mandarin listeners. Cut-off T1 questions ended with a perceivably shorter syllable than a complete utterance (see Figure 4.11), where final lengthening should have occurred. Table 4.4 showed the duration of each syllable of the Mandarin T1 questions. The penultimate syllable is approximately the duration of

the mean of all the syllables in the utterances, much shorter than the final syllable. Listeners might find the ending rather abrupt and did not receive a strong cue from the shortened final tone, and therefore regarded the cut-off questions as statements. Furthermore, Mandarin listeners were doing much better in the cut-off questions with T1 (M=39%, SD=0.33) and T2 (M=44%; SD=0.32) than Cantonese listeners (T1: M=25%, SD=0.27; T2: M=15%, SD=0.21). This contrast shows that Mandarin listeners were still using some cues that were not familiar to the Cantonese listeners, namely the global intonation contour specific to Mandarin.

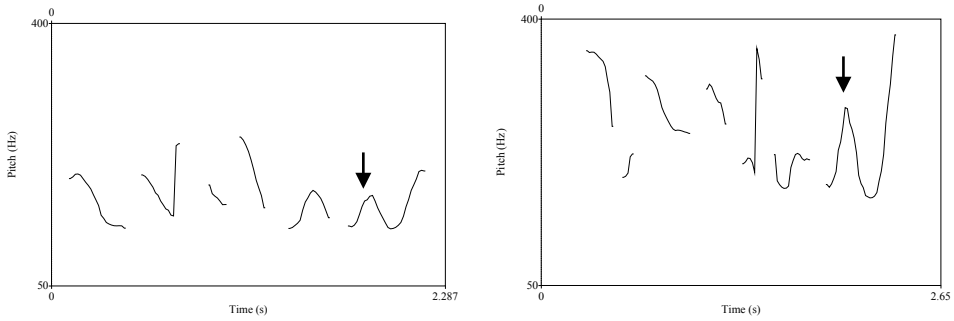

Figure 4.10 The F0 contour of the Mandarin T2 question (left: M; right: F), with the arrows pointing at the penultimate tone in the utterances

Table 4.7 Duration (ms) of each syllable of complete Mandarin T1 question

Syllable No.	1	2	3	4	5	6	7	8	9	
Pinyin	ma1	ma	jin1	wan3	dun4	de	shi4	ji1	tang1	
Character	妈	妈	今	晚	炖	的	是	鸡	汤	Mean
Female	181	241	294	287	274	9	186	224	327	225
Male	177	239	260	221	326	8	162	208	358	218

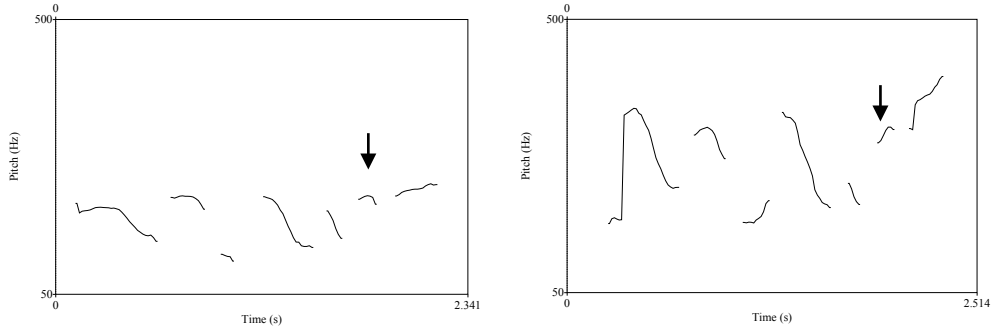

Figure 4.11 The F0 contour of the Mandarin T1 question (left: M; right: F), with the arrows pointing at the penultimate tone in the utterances

Interestingly, when listening to the T3 cut-off question in Mandarin, Cantonese listeners had a relatively high identification accuracy. The identification accuracy of T3 cut-off questions is significantly higher than those of T1 and T2 questions, and is not significantly different than that of T4. Cantonese listeners were used to having the final tone change to a rising contour. This pattern shows that they applied their linguistic knowledge of question intonation here, and correctly identify questions ending with a sandhied rising tone. On the other hand, Mandarin listeners' identification accuracy of T3 question cut off question is significantly lower than T4 questions, but is not significantly different from those of T1 and T2. They had a similar performance in questions ending with sandhied rising T3 and T2 cut-off questions.

The cross-linguistic results also lend some support to Ohala's proposal of the

Frequency Code (Ohala 1983), which stated that smaller vocal cords and faster vibrations rates were used for expression of power relations: a high and high-ending pitch corresponds to appealing and questioning, a low and low-ending pitch corresponds to authority and assertiveness. According to the Frequency Code, one may predict that high-rising or high register pitch contours would be more often identified with questions, whereas low or falling pitch contours would be recognized as statements. This seems to be the case in the current data.

For example, the Mandarin T3 (low tone) ending has caused troubles for both native and non-native listeners. In the complete utterance condition, while statements ending with T3 were all successfully identified (100%), T3 questions were poorly identified by both groups of listeners. The identification accuracy of T3 complete questions is significantly lower than questions with the other three tones for both groups, especially for Cantonese listeners, whose identification accuracy dropped below chance level (M=42.5%). This shows that a low tone is not associated with a question by listeners. Since Cantonese listeners rely more on the final tone in judging questions, they performed particularly poorly in identifying a question ending with a low tone. In addition, when the final tone was cut off, the T3 utterances were left with a final tone of the sandhi-ed T3 (original penultimate tone in the complete utterances), which is a high rising tone. It can be observed that statements ending

with this rising version of T3 received significantly lower identification accuracy by both groups of listeners. Obviously, listeners did not associate a rising tone with a statement this time.

Evidence supporting perception based on the Frequency Code can also be found in the Cantonese perception result. When the final tone was cut off, Mandarin listeners had problems recognizing Cantonese T2 (high rising) statements. This suggests that they may have confused the final rising tone as a signal to questioning. Furthermore, they had a much better identification accuracy in T1 (high level), T2 (high rising), T5 (low rising) questions than T4 (low falling) and T6 (low level) question. The accuracy of T2 cut-off question is significantly higher than T4 (low falling) question, which scored the worst. This indicates that Mandarin listeners related a high and/or rising pitch contour at the end of an utterance with questions; whereas they associated an ending low pitch to statements, agreeing with the prediction of the Frequency Code. However, it should be noted that since T2 (high rising) is easily confused with the ordinary rising at the end of Cantonese questions, Mandarin listeners found the identification of the cut-off T2 utterances difficult as a whole, as their confidence ratings for both statements and questions ending with T2 are the lowest among all the ending tones.

4.4 Summary

The combined results of the identification accuracy and confidence rating show that statement is the unmarked intonation type for both groups of listeners. The results also confirmed the production patterns in the two languages. The final tone is indispensible for Cantonese questions. Eliminating the final tone had some impact on the identification of Mandarin question, depending on the final lexical tone. Tonal coarticulation and tone sandhi had a great influence over the perception of intonation.

The knowledge about their intonation and lexical tone system helps the listeners to perceive intonation. Cantonese listeners expected T1 to have a different pattern in question ending and did not rely on final rising that appeared in the other tones. The flattening effect caused by the raised global contour became a strong cue for Mandarin listeners. The language-tone specific mechanisms suggest that the phonological knowledge about the language's intonation system and lexical tone system plays an essential part in the perception of intonation.

The results also suggest that the Frequency Code may provide another key to understand the perception of intonation in general. Due to unfamiliarity of the stimuli or editing of the speech material, phonological knowledge was not always applicable. In those cases, listener showed a preference to associate questions with high and/or rising ending pitch and statements with low and/or falling pitch.

However, in this experiment, listeners were well aware of the languages they were exposed to and in most cases used their pre-existent phonological knowledge to process the data. Therefore, a follow-up experiment reported in the next chapter was conducted with low-pass filtered speech materials. The goal is to examine whether listeners would rely on the Frequency Code entirely to perceive intonation when no segmental and semantic information can be used.

CHAPTER 5

CROSS-LINGUISTIC PERCEPTION USING FILTERED SPEECH

The previous perception experiment used normal speech stimuli to show the interaction between universal Frequency Code and language-tone specific perception. However, it is unclear how tonal language listeners would interpret intonation contours when they are given unintelligible speech by which they could not employ the tone-intonation interrelationship existing in their own language. This experiment aims to investigate whether listeners would rely on the universal Frequency Code entirely to perceive intonation when they could not recognize the semantic context and thus without the top-down phonological information by using low-pass filtered materials.

5.1 Method

The stimuli are the same set of utterances used in the previous experiment. Low-pass filters were applied to every sound file with 100 Hz bandsmoothing. The cutoff frequency varied according to individual pitch range, determined by

the highest pitch each speaker produced in their questions ending with T2 (a high rising tone in both languages). The cutoff frequency was 230 Hz for the Cantonese male speakers, 300 Hz for the Cantonese female speaker, 250 Hz for the Mandarin male speaker, and 400 Hz for the Mandarin female speaker. Informal tests showed that native speakers of both languages could not understand the content of the low-pass filtered sentences. They reported only hearing some low frequency humming.

The same groups of listeners in the last experiment (See Chapter 4) participated in this experiment, following the procedure and requirement of the normal speech experiment. However, they listened to the filtered speech blocks for this experiment before the normal speech blocks for the last experiment, so that the segmental and semantic information would be unknown to them when they listened to the low-pass filtered speech.

Before each block, subjects were told that the utterances within that block were either all in Mandarin or Cantonese. However, they could not tell which was which because of the low-pass filter. The blocks and tokens within each block were randomized. Each of the 80 stimuli (4 sentence conditions × 2 genders × 4 tones in Mandarin and 6 tones in Cantonese) was repeated twice, resulting in 160 stimuli in total.

5.2 Results

5.2.1 Cantonese listeners listening to Cantonese filtered utterances

Figure 5.1 shows the identification accuracy of Cantonese intonation by Cantonese listeners. The listeners performed well except in identifying the cut-off questions. Statistical analysis indicates that the main effect of Intonation type is significant [$F(1,19)=76.16$, $p<0.001$]. Statements (M=89%, SD=0.18) are better identified than questions (M=53%, SD=0.42). The main effect of Condition (complete or cut-off) is also significant [$F(1,19)=102.30$, $p<0.001$], with complete utterances (M=86%, SD=0.24) better identified than cut-off utterances (M=56%, SD=0.41). The difference between statements and questions mainly occur in the cut-off condition, as the interaction Intonation type × Condition is significant [$F(1,19)=148.43$, $p<0.001$].

The different ending tones also have a significant impact over the identification of intonation [main effect of Final tone: $F(5,95)=3.06$, $p<0.05$]. And this impact is also subject to intonation types and conditions with significant interactions: Final tone × Intonation type [$F(5,95)=2.69$, $p<0.05$] and Final tone × Condition [$F(5,95)=6.41$, $p<0.001$]. Within the cut-off condition, while the identification of statement remained similar to the performance in the complete utterance condition,

listeners performed significantly better in identifying T1 question than cut-off questions in the other tones ($p<0.05$ in every comparison). T1 question was the only tone that reached the chance level (50%) in the cut-off condition, while questions in the other tones were identified very poorly. This shows that the final syllable of Cantonese has a strong impact over question identification for Cantonese listeners. In addition, utterances ending with a high level tone (T1) were more likely to be identified as a question.

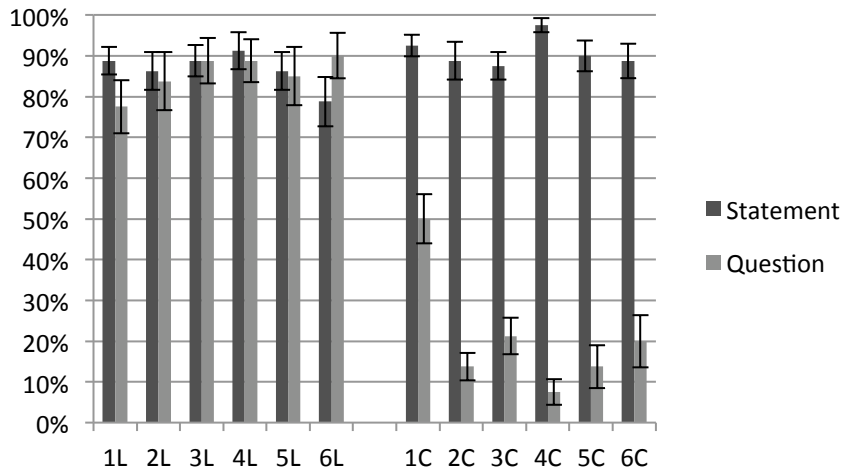

Figure 5.1 Average percentage of correct identification of Cantonese filtered utterances by Cantonese listeners
(L: complete utterances; C: cut-off utterances)

The discriminability and bias scores shown in Table 5.1 indicate that complete utterances have strong discriminability across all ending tones. However, when the final syllable is absent, the discriminability scores drop sharply, among which T1

cut-off utterances have the highest score and T2 and T5 utterances have the lowest. The bias scores show that in the complete condition, the identification of all utterances has a slight bias towards statement except T3 (0, no bias) and T6 (negative, bias towards question). Nevertheless, in the cut-off condition, all the utterances have a high positive bias score, indicating that listeners had a strong bias towards statement with all ending tones. This shows that listeners tended to choose statement as the answer when the strong cue at the end of the questions is removed.

Table 5.1 Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'') of Cantonese complete (upper panel) and cut-off (lower panel) utterances by Cantonese listeners (L: complete utterances; C: cut-off utterances)

	1L	2L	3L	4L	5L	6L
H	0.78	0.84	0.89	0.89	0.85	0.90
FA	0.11	0.14	0.11	0.09	0.14	0.21
A'	0.90	0.91	0.94	0.94	0.92	0.91
B''	0.39	0.10	0.00	0.14	0.05	-0.42

	1C	2C	3C	4C	5C	6C
H	0.50	0.14	0.21	0.08	0.14	0.20
FA	0.08	0.11	0.13	0.03	0.10	0.11
A'	0.83	0.55	0.63	0.68	0.58	0.63
B''	0.85	0.96	0.93	1.00	0.97	0.94

Figure 5.2 shows the confidence rating of Cantonese listeners. The confidence rating has a significant correlation to the identification accuracy [Pearson's $r(22)=0.544$, $p<0.01$]. Generally, there is a main effect of Intonation type

[F(1,19)=5.45, p<0.05]. In the complete condition, listeners were more confident in the questions than the statement. In the cut-off condition, the difference in condition is not as obvious. Listeners were more confident in the complete condition (M=5.31, SD=1.12) than in the cut-off condition (M=4.72, SD=1.21) with a significant main effect of Condition [F(1,19)=24.56, p<0.001]. There is also a significant main effect of Final tone [F(5,95)=4.29, P<0.05]. However, Bonferroni adjustment does not reveal any significant pairwise comparison, except that the difference between T6 and T2 complete utterance is marginally significant (p=0.05).

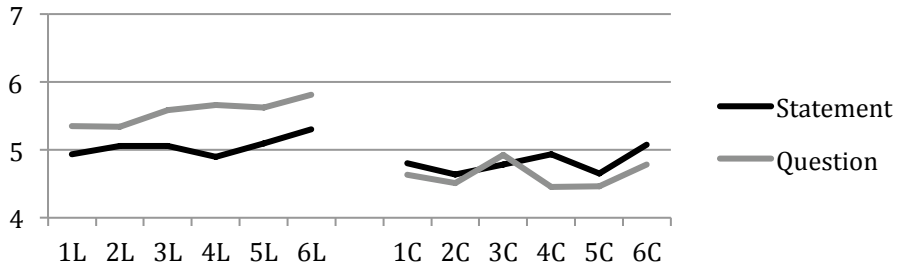

Figure 5.2 Average confidence rating of Cantonese filtered utterances by Cantonese listeners (L: complete utterances; C: cut-off utterances)

5.2.2 Mandarin listeners listening to Cantonese filtered utterances

As shown in Figure 5.3, Mandarin listeners performed very similarly to Cantonese listeners in identifying the low-pass filtered Cantonese utterances. Like Cantonese listeners, Mandarin listeners achieved a better identification accuracy in statement

(M=87%, SD=0.21) than questions (M=60%, SD=0.40), with a significant main effect of Intonation type [$F(1,19)=66.57$, $p<0.001$]. Complete utterances (M=92%, SD=0.17) are easier to identify than cut-off utterances (M=54%, SD=0.38), as the main effect of Condition is significant [$F(1,19)=474.44$, $p<0.001$]. Interaction between Intonation type and Condition is also significant [$F(1,19)=256.97$, $p<0.001$], showing that the difference between question and statement is dependent on conditions.

Final tone also has a significant main effect on the identification [$F(5,95)=2.635$, $p<0.05$], which is subject to intonation types (Final tone × Intonation type [$F(5,95)=5.21$, $p<0.001$]) and conditions (Final tone × Condition [$F(5,95)=7.98$, $p<0.001$]). Specifically, complete statement ending with T2 (a high rising tone) appeared to be identified worse than the other tones, probably because of the confusion caused by the high rising pitch at the end. However, this difference is not significant ($p>0.05$ in every comparison). In the cut-off condition, statements remained well identified in comparison to questions. For cut-off questions, T1 is significantly better identified than T3-T6; T2 and T5 are significantly better identified than T4 ($p<0.05$ in each comparison). T1 is also the only tone that reached the chance level for Mandarin listeners. The statistical results of Mandarin listeners show that in the cut-off condition, questions ending with a high tone (T1) are

identified the best while a rising ending tone (T2 and T5) seem to be easier for Mandarin listeners to identify than questions ending with a low falling tone (T4).

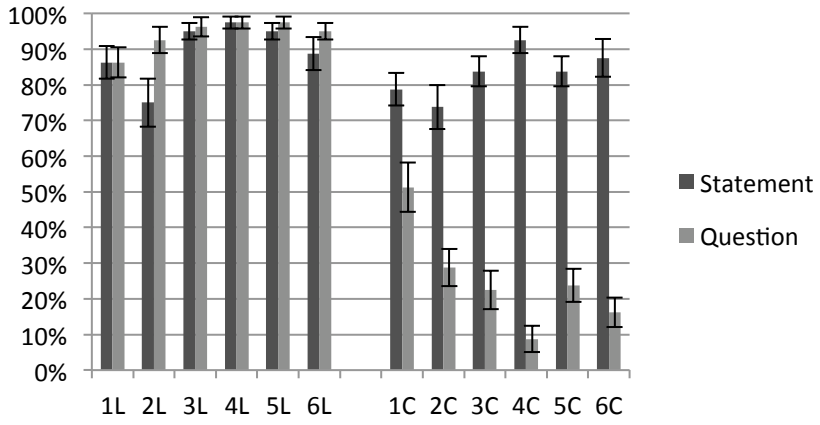

Figure 5.3 Average percentage of correct identification of Cantonese filtered utterances by Mandarin listeners
(L: complete utterances; C: cut-off utterances)

Table 5.2 Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'') of Cantonese complete (upper panel) and cut-off (lower panel) utterances by Mandarin listeners
(L: complete utterances; C: cut-off utterances)

	1L	2L	3L	4L	5L	6L
H	0.86	0.93	0.96	0.98	0.98	0.95
FA	0.14	0.25	0.05	0.03	0.05	0.11
A'	0.92	0.91	0.98	0.99	0.98	0.96
B''	0.00	-0.61	-0.15	0.00	-0.34	-0.41

	1C	2C	3C	4C	5C	6C
H	0.51	0.29	0.23	0.09	0.24	0.16
FA	0.21	0.26	0.16	0.08	0.16	0.13
A'	0.74	0.53	0.59	0.54	0.60	0.57
B''	0.56	0.75	0.89	0.98	0.89	0.95

The scores in Table 4 show that discriminability by Mandarin listeners was hindered by cutting off the final syllable of Cantonese utterances, as discriminability scores decreased considerably in the cut-off condition. T1 cut-off utterances have a higher score than cut-off utterances with the other tones, which corresponds to the accuracy data. Surprisingly, Mandarin listeners seem to have a bias towards question except T1 and T4 utterances (in both cases, B''=0, no bias found), when they listened to complete Cantonese utterances. The bias is strongest for T2 utterances. However, the bias is reversed to statement when the final tone is cut-off, showing that listeners had lost the information they need to identify question, and chose statement as a default answer, as predicted by the production pattern of Cantonese question.

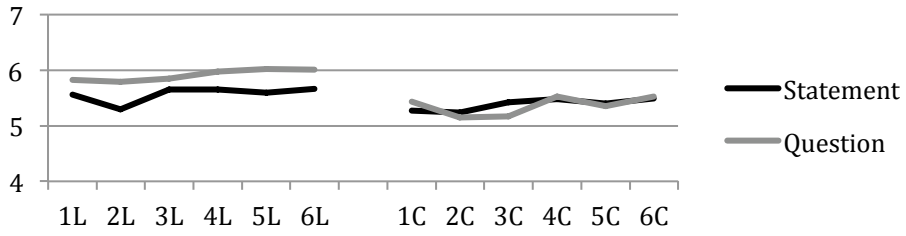

Figure 5.4 Average confidence rating of Cantonese filtered utterances by Mandarin listeners (L: complete utterances; C: cut-off utterances)

Figure 5.4 shows the confidence rating of Mandarin listeners listening to filtered Cantonese utterances. A significant correlation between confidence rating and identification accuracy can be detected [Pearson's $r(22)=0.566$, $p<0.01$]. There is a

main effect of Intonation type [$F(1,19)=11.13$, $p<0.05$]. Listeners were more confident with question than statement in the complete condition, while the confidence rating of statement and question largely overlap with each other in the cut-off condition. A main effect of Condition [$F(1,19)=15.57$, $p<0.05$] indicates that listeners were significantly more confident in the complete condition ($M=5.74$, $SD=1.30$) than in the cut-off condition ($M=5.37$, $SD=1.22$). There is also a significant main effect of Final tone [$F(5,95)=3.31$, $p<0.05$] and a significant Final tone × Condition interaction [$F(1,19)= 6.27$, $p<0.05$]. However, no significnt pairwise comparison is detected by Bonferroni adjustment.

5.2.3 Mandarin listeners listening to Mandarin filtered utterances

Figure 5.5 shows the identification accuracy for Mandarin listeners listening to Mandarin low-pass filtered utterances. The main effect of Intonation type is significant [$F(1,19)=57.39$, $p<0.001$], as statements ($M=91\%$, $SD=0.21$) are better identified than questions ($M=69\%$, $SD=0.35$). The main effect of Condition is also significant [$F(1,19)=173.57$, $p<0.001$], with complete utterances ($M=93\%$, $SD=0.15$) better identified than cut-off utterances ($M=68\%$, $SD=0.37$). The interaction Intonation type × Condition is also significant [$F(1,19)=34.41$, $p<0.001$], as the statement-question difference is dependent on conditions.

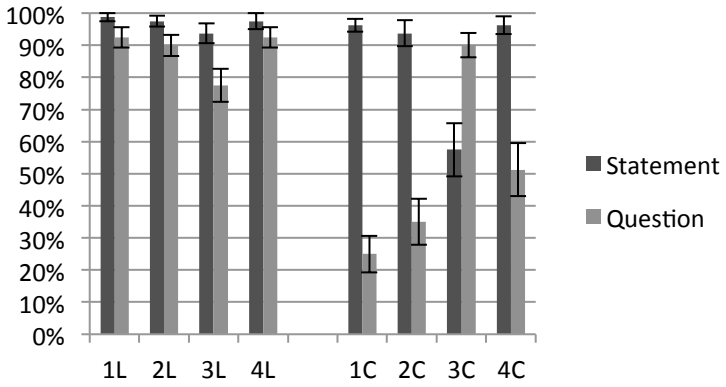

Figure 5.5 Average percentage of correct identification of Mandarin filtered utterances by Mandarin listeners

(L: complete utterances; C: cut-off utterances)

Although the main effect of Final tone is not significant [$F(3,57)=1.73$, $p>0.05$], the significant interactions Final tone × Intonation type [$F(2.29,43.59)=15.94$, $p<0.001$] and Final tone × Condition [$F(3,57)=6.74$, $p<0.01$] suggest that tones affect listeners' identification depending on the intonation type and condition. For example, complete question ending with T3 ($M=76\%$, $SD=0.23$) is the least well identified by listeners among the complete questions, significantly poorer than T1 ($M=93\%$, $SD=0.14$, $p<0.05$). Among the cut-off utterances, T3 statement is significantly more poorly identified than statements in all other tones, whereas T3 question is significantly better identified than the other tones ($p<0.05$ in all comparisons), both owing to the rising penultimate tone due to tone sandhi. Besides, within the cut-off condition, T4 question ($M=51\%$, $SD=0.37$) is

significantly better identified than T1 question (M=25%, SD=0.26, $p<0.05$), suggesting that for Mandarin listeners, questions ending with T4 are still easier to identify than some other tones, even though the original final tone is cut off.

Table 5.3 Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'') of Mandarin complete (upper panel) and cut-off (lower panel) utterances by Mandarin listeners (L: complete utterances; C: cut-off utterances)

	1L	2L	3L	4L
H	0.93	0.90	0.78	0.93
FA	0.01	0.03	0.06	0.03
A'	0.98	0.97	0.92	0.97
B''	0.73	0.63	0.63	0.52

	1C	2C	3C	4C
H	0.25	0.35	0.90	0.51
FA	0.04	0.06	0.43	0.04
A'	0.77	0.78	0.84	0.86
B''	0.97	0.93	-0.74	0.92

The scores in Table 5 suggest that in the complete condition, utterances ending with T3 have the least discriminability between statement and question, while utterances ending with the other tones have a fairly good discrimination. In the cut-off condition, utterances generally have lower discriminability scores than their complete counterparts. Cut-off utterances ending with T3 and T4 have a slightly higher discrimination score than those with T1 and T2. There is a bias towards

question in complete utterances identification. However, when the final tone is cut off, the listeners show a strong bias towards statements with utterances ending with T1, T2 and T4, but a bias towards question with those with T3, echoing the accuracy result.

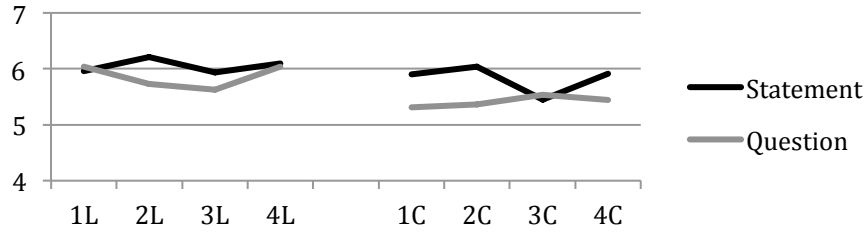

Figure 5.6 Average confidence rating of Mandarin filtered utterances by Mandarin listeners (L: complete utterances; C: cut-off utterances)

The confidence rating as shown in Figure 5.6 has a significant correlation to the identification accuracy [Pearson's $r(12)=0.867$, $p<0.01$]. There is a main effect of Intonation type [$F(1,19)=11.04$, $p<0.05$] in the confidence rating. Listerners were more confident about statements (M=5.94, SD=1.36) than questions (M=5.64, SD=1.32). A main effect of Condition [$F(1,19)=6.56$, $p<0.05$] indicates that the listeners were more confident in the complete condition (M=5.96, SD=1.43) than in the cut-off condition (M=5.62, SD=1.25). No effect of Final tone was found. However, Bonferroni adjustment reveals that T3 cut-off statement (M=5.45, SD=1.30) has significant lower confidence rating than T2 (M=6.04, SD=1.31) and

T4 (M=5.91, SD=1.27). This is probably because the penultimate sandhi-ed T3 was interfering with the listeners' judgement.

5.2.4 Cantonese listeners listening to Mandarin filtered utterances

Figure 5.7 shows the general pattern of Cantonese listeners' identification accuracy for Mandarin utterances. Intonation type has a significant main effect [$F(1,19)=40.69$, $p<0.001$], with statements (M=88%, SD=0.20) better identified than questions (M=54%, SD=0.35). The main effect of Condition is also significant [$F(1,19)=69.82$, $p<0.001$], as complete utterances (M=81%, SD=0.26) are more accurately identified than cut-off utterances (M=62%, SD=0.37). The Intonation type × Condition interaction is also significant [$F(1,19)=9.89$, $p<0.001$], showing that the differences caused by Intonation type and Condition are dependent with each other.

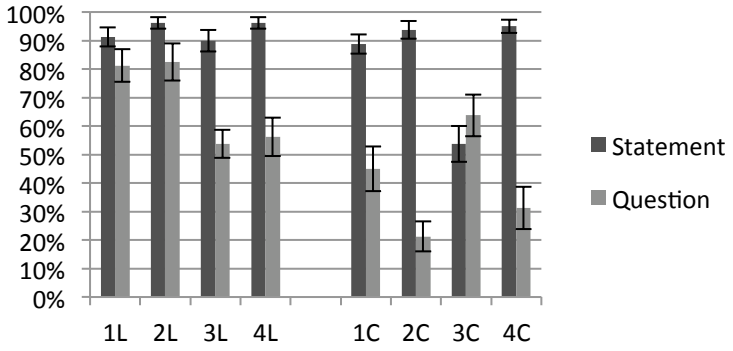

Figure 5.7 Average percentage of correct identification of Mandarin utterances by Cantonese listeners

(L: complete utterances; C: cut-off utterances)

There is also a significant main effect for Final tone [$F(3,57)=5.28$, $p<0.005$]. And the interactions Final tone × Intonation type [$F(3,57)=10.36$, $p<0.001$] and Final tone × Condition [$F(3,57)=6.51$, $p<0.01$] are both significant. In the complete condition, the identification accuracy of T3 (M=54%, SD=0.22) question is significantly lower than T1 (M=81%, SD=0.25) and T2 (M=83%, SD=0.29) questions ($p<0.05$ in each comparison). In the cut-off condition, the T3 utterance ends with the sandhi-ed T3 (high rising). As a result, the T3 cut-off question (M=64%, SD=0.33) was better identified than the other tones (comparisons to T2 (M=21%, SD=0.23) and T4 (M=31%, SD=0.33) questions were significant, $p<0.05$). The T3 cut-off statement (M=54%, SD=0.28) was identified most poorly among all the tones (significant compared to every other tone, $p<0.05$). Furthermore, unlike the pattern by Mandarin listeners, the T4 complete question (M=56%, SD=0.30) was

poorly identified by Cantonese listeners, with the accuracy significantly lower than that of T1 and T2 questions ($p<0.05$ in each comparison). The identification of T2 cut-off question (M=21%, SD=0.23) was also not successful, significantly poorer than T1 and T3 ($p<0.05$ in each comparison). The perceptual pattern suggests that Cantonese speakers mainly use the final tone in their judgment, much like the way they identify intonation type in their own language. Final high level and high rising pitch contours (in the case of complete T1, T2 utterances and cut-off sandhi-ed T3 utterances) seem to be a strong indicator for question, while a dipping tone and a falling tone (complete T3 and T4 utterances) are associated with statements.

Table 5.4 Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'') of Mandarin complete (upper panel) and cut-off (lower panel) utterances by Cantonese listeners (L: complete utterances; C: cut-off utterances)

	1L	2L	3L	4L
H	0.81	0.83	0.54	0.56
FA	0.09	0.04	0.10	0.04
A'	0.92	0.94	0.83	0.87
B''	0.41	0.69	0.77	0.90

	1C	2C	3C	4C
H	0.45	0.21	0.64	0.31
FA	0.11	0.06	0.46	0.05
A'	0.78	0.72	0.65	0.78
B''	0.81	0.96	-0.20	0.95

The discriminability scores in Table 5.4 show that complete utterances ending with T1 and T2 are better discriminated than those with T3 and T4, concurring with the accuracy data. In both the complete and cut-off conditions, T3 utterances have the lowest discriminability scores, indicating that Cantonese listeners found intonations with an ending T3 most difficult to recognize. Similar to the results of Mandarin listeners, Cantonese speakers have a bias towards statement in complete utterances ending with all tones. When the final syllable is absent, this bias was strengthened in T1, T2 and T4 utterances, but a bias towards question was shown in T3 utterances, due to the sandhi-ed T3.

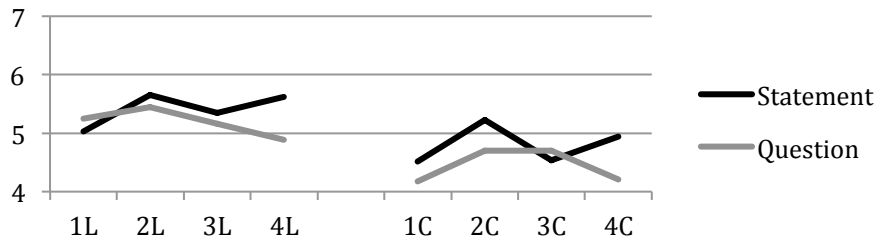

Figure 5.8 Average confidence rating of Mandarin filtered utterances by Cantonese listeners (L: complete utterances; C: cut-off utterances)

Figure 5.8 shows the confidence rating by Cantonese listeners when they listened to Mandarin filtered utterances. A significant correlation was found between confidence rating and identification accuracy [Pearson's $r(14)=0.689$, $p<0.01$]. There is a main effect for Intonation type in the confidence data [$F(1,19)=8.73$, $p<0.05$].

Staments have a higher confidence rating (M=5.11, SD=1.13) than questions (M=4.82, SD=1.20). A main effect of Condition [F(1,19)=45.04, p<0.001] indicates that listeners were more confident in the complete condition (M=5.30, SD=1.11) than the cut-off condition (M=4.63, SD=1.14). The main effect of Final tone [F(3,57)=9.09, p<0.001] and the interaction of Final tone × Intonation type [F(3,57)=7.33, p<0.001] are both significant. In the complete condition, the confidence rating of T1 statements (M=5.03, SD=1.07) is significantly lower than T2 (M=5.65, SD=0.96) and T4 (M=5.63, SD=1.04) statements (p<0.01 in each comparison). The confidence rating of T2 complete question (M=5.45, SD=1.04) is significantly higher than T4 question (M=4.89, SD=1.13, p<0.05). In the cut-off condition, T2 statements (M=5.23, SD=0.96) have significant higher confidence ratings than T1 (M=4.51, SD=1.13) and T3 (M=4.94, SD=1.20) statements (p<0.05 in each comparison). The confidence rating of T2 cut-off question (M=4.70, SD=1.06) is significantly higher than T1 (M=4.18, SD=1.09, p<0.05).

5.3 Discussion

The perceptual results of this experiment further support the acoustic patterns of Cantonese intonation found in previous production studies (Ma et al. 2006b, Gu et al.

2006, Chapter 3 of this thesis). In Cantonese, where questions are signaled by a boundary tone located at the very end of the last syllable, cutting off the final tones of questions significantly affects the perceptual accuracy of questions in all tones. The identification accuracy by Cantonese listeners is below the chance level (50.0%) in all tones except T1 (just reaching 50.0%). Mandarin listeners also did poorly with cut-off questions in Cantonese, with only T1 reaching the chance level (51.3%). T1, the high level tone in Cantonese, probably was the easiest to be associated with questioning. Among the other tones, Mandarin listeners did relatively better in T2 (28.8%) and T5 (23.8%), probably because these two rising lexical tones are more easily associated with questions. The worst identified cut-off question is the one with T4, the low falling tone, which might be least likely to be associated with question intonation.

Cutting off the final tone of Mandarin filtered utterances also affects listeners' judgment. With the exception of sandhi-ed T3, the identification accuracy of the cut-off questions is lower than that of complete questions. The accuracy of T4 question, which was perceived well by both groups of listeners in the previous experiment using unfiltered speech, decreased in this experiment. With the listeners unaware of the identity of the language and the lexical tone, the penultimate T4 might not present a cue strong enough for successful identification of questions.

Regarding the specific tones in the Mandarin perceptual data, T3 provides an interesting case. The identification accuracy of complete T3 questions by Mandarin listeners (77.5%) is significantly lower than all the other tones; and that of complete T3 questions by Cantonese listeners (53.8%) is significantly lower than T1 and T2. While T3 is only a low target under most circumstances in connected speech, it does have a rising tail at the end of a question (Yuan 2004a). Figure 5.9 shows the F0 contours of the Mandarin T3 complete questions played to the participants. The Mandarin T3 rising tail (the second arrows in Figure 5.9) did not reach the top of the speakers' pitch range, and was much lower than the end of the penultimate high rising tone (the first arrows in Figure 5.9, caused by the T3 tone sandhi). For the male speaker, the final rising end was at 148 Hz while the penultimate one was at 218 Hz; for the female speaker, the final rising end was at 295 Hz while the penultimate one was at 350 Hz. Interestingly, the results showed that listeners, especially Cantonese listeners, disregarded this rising tail after the low target in their perception by only correctly identifying 54% of the T3 complete questions. They might consider this rising tail as a low rising in comparison to the penultimate tone and thus did not associate it with questions.

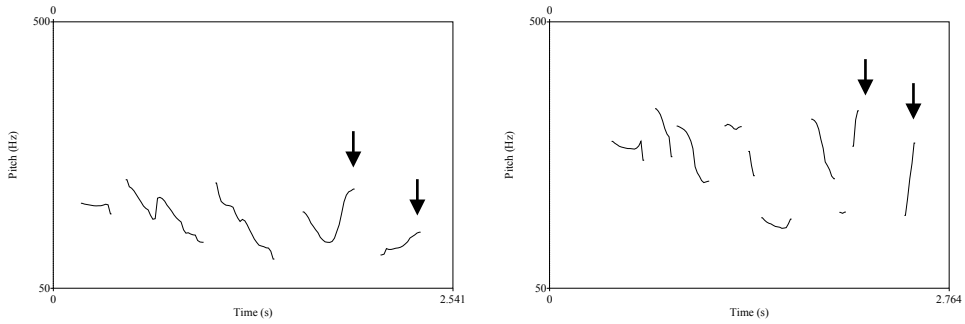

Figure 5.9 The F0 contour of the Mandarin T3 question (left: M; right: F), with the arrows pointing at the penultimate and final tones in the utterances

In the cut-off condition, the Mandarin T3 utterances ended with a high rising tone because of tone sandhi, and are more easily identified as a question than a statement. Consequently, the cut-off T3 statement was identified significantly less successfully than the other tones; while the cut-off T3 question was identified more accurately than the other tones and more successfully than complete T3 questions. Compared to the successful identification of T3 cut-off questions resulting from a high rising tone sandhi form, T2 (high rising) cut-off questions and T1 (high level) cut-off questions were identified poorly by both groups of listeners because of tonal coarticulation and modification of duration in connected speech as discussed in Chapter 4.

To sum up the accuracy results, under the circumstance that top-down information of language identity and lexical tone is absent, Cantonese and Mandarin listeners show generally similar perceptual patterns. When listening to

Cantonese utterances, both groups performed well in complete utterances, but did poorly in recognizing questions in the cut-off condition, where they displayed a strong bias towards statements. Among the cut-off questions, T1 questions were best identified. This pattern shows that both groups found the final rising in Cantonese very useful for question identification. When this cue is removed, they could no longer find a strong cue in the filtered Cantonese utterances. When listening to Mandarin utterances, both groups did well in perceiving T1 and T2 complete utterances. T3 complete questions were identified less accurately than T1 and T2 by both groups. In the cut-off condition, both groups were tricked by the sandhi-ed T3, resulting in high accuracy in identifying T3 question and low accuracy in T3 statement. Since the low-pass filter made individual tones and language identity difficult to recognize, the language-tone specific effects found in previous studies (Yuan 2004, Xu and Mok 2012) were weakened in the current study. For example, the perceptual asymmetry that Mandarin T4 questions were better identified than T2 questions, as shown in Yuan (2004), Yuan (2011) and also the normal speech experiment discussed in Chapter 4, was not found in this experiment. Mandarin listeners did not show a significantly different accuracy in identifying T2 questions and T4 questions in both complete and cut-off conditions. The effect of a universal mechanism, namely the Frequency Code, shared by both groups of listeners,

was expected to dominate their decision regardless of the native language they spoke and the language they heard. The general similar results from the two groups seem to support this prediction.

However, differences between the two groups can also be observed in the results. Cantonese listeners performed poorly in identifying Mandarin T4 complete questions, in which Mandarin listeners did well. In addition, the identification accuracy of Mandarin T1 cut-off questions by Cantonese listeners has no significant difference from that of T3 cut-off questions. In contrast, Mandarin listeners did poorly in identifying Mandarin T1 cut-off questions, with accuracy significantly lower than that of T3. These differences show that language background plays a role in cross-linguistic intonation perception. Mandarin listeners did well in recognizing T4 complete questions because they were familiar with question intonation ending with a falling tone. They had better perception accuracy with questions ending with T4 than questions ending with the other tones in previous studies (Yuan 2004, Yuan 2011). This shows that the interaction between final tone and intonation in their native language influences their perception of filtered speech. For Cantonese listeners, who were used to looking for rising cues for question at the end of an utterance, a falling tone was naturally not associated with a question according to the Frequency Code.

Furthermore, Mandarin listeners could not identify cut-off T1 questions probably because of the shortened duration of the penultimate T1 syllable, as mentioned in Chapter 4. However, Cantonese listeners seem to relate the ending high pitch to a question regardless of its duration. This may be because of their sensitivity towards the F0 height at the end of question, which led them to consider the shortened T1 as an H% that is important for Cantonese question. These different perception patterns suggest that language background still influence the perception of unintelligible speech, even with the presence of a universal F0 mechanism.

5.4 Summary

This experiment examines cross-linguistic perception of intonation of low-pass filtered speech by Mandarin and Cantonese listeners. When lexical and semantic information is absent, as well as when lexical tone identity is difficult to recognize, listeners relied mainly on the information provided by the sentential F0 contour. The results confirm the acoustic patterns of Cantonese intonation found by previous studies. The fact that both Mandarin and Cantonese listeners performed below chance level in identifying Cantonese questions without the final tone clearly indicate that the boundary tone in Cantonese is critical for question signaling. The

final tone in Mandarin also seemed to carry some reliable information of the intonation type. The results show that for both groups of tonal language listeners, their interpretation of intonation is dominated by a universal mechanism, i.e., the Frequency Code. The high register and a rising contour at the end of an utterance are associated with question signaling, whereas the low register and a falling contour at the end of an utterance are the signal for statements. However, the different perceptual patterns between the two listener groups also indicate that language background still plays a role in cross-linguistic perception of intonation even when the speech is unintelligible. The interaction between final tone and intonation seems to have an influence on the perception results. The results further show that the realization of F0 contours resulting from language-specific reasons such as tone sandhi and tonal coarticulation also contribute to intonation perception, especially when top-down phonological knowledge cannot be used.

CHAPTER 6

GENERAL DISCUSSION

This thesis investigated the perception of Mandarin and Cantonese intonation by both groups of listeners. In addition, the acoustics of Cantonese intonation has been revisited. The results covered several related topics and have addressed different research questions. First, both the production and perception of Cantonese intonation were investigated. The results support that Cantonese heavily relies on the final high target instead of the global F0 contour to signal questions. Second, the cross-linguistic perceptual results support the previous claim of statement-question bias in intonation perception. Third, the results answer the central research question of this study and reveal the interaction between psychoacoustic mechanism and phonological knowledge in intonation perception through different conditions of experiments. Section 6.1 – 6.3 will discuss those results respectively, followed by a discussion of the limitation of this study. The chapter is ended by an outlook of future research directions and a brief conclusion.

6.1 Cantonese intonation

The production experiments show that final rising is the principal perceptual cue for Cantonese question. The results of the first experiment concur with previous studies (Ma et al. 2006b, Lee 2008) and show that the final rising only occurs at the second half of the final syllable. Ma et al. (2006b) and Lee (2008) found that T1 (high level) did not undergo final rising because its canonical pitch height is already at the top of speakers' pitch range, i.e., a ceiling effect. However, speakers in this study produced Cantonese questions ending with T1 with a high rising like the other tones. Whether the rising of question-final T1 is due to individual difference or experiment design needs further investigation.

In the second production experiment, utterances consisted of only one tone were used to test the pre-final F0 contour of Cantonese question intonation. This method focuses on the phonetic details over the course of the whole utterances, while lexical tone and position in the utterance are controlled. Our results find no consistent or significant difference between the pre-final contours of statement, intonation question and particle question. This indicates that the global F0 contour is not an important perceptual cue for Cantonese questions.

We also looked at the role of sentence final particle in Cantonese question. Lee (2008) found that in Mandarin the rising contour at the end of particle questions were

not restricted to the final syllable, but rather covered the final NP. Our study shows that this is not the case in Cantonese. In the final rising experiment, the results show that the penultimate tones of particle questions have no significant difference from the last tones of statements at any time points. In the global F0 contour experiment, no difference was found between the F0 contours of the pre-final part of particle questions, intonation questions and statements. These results clearly indicate that the high boundary target is carried by the sentence final particle alone and does not influence the tones before the particle.

The production patterns were echoed by the perception patterns. The identification accuracy of Cantonese listeners listening to both normal and filtered speech is integrated into Figure 6.1. When listening to the complete utterances, Cantonese listeners performed generally well, with identification of normal speech slightly better than that of filtered speech. No significant difference between the identification of statement and question was detected. However, when the final tone was cut off, while the identification accuracy of statements remained high, the accuracy for questions decreased significantly for both normal and filtered speech. With the identification accuracy of questions ending with T1 reaching the chance level of 50% (filtered speech) or nearly 50% (normal speech), the accuracy with the other tones were far below the chance level. The results clearly demonstrate that

Cantonese listeners were not able to detect questioning when the final tone was absent. The reason for the slightly better performance on T1 was that T1 was the only final tone that did not necessarily become rising at the end of questions in Cantonese. Therefore listeners may be able to detect other cues in questions ending with T1. It is also possible that, as predicted by the Frequency Code, a high level tone at the end was naturally more easily associated with questions.

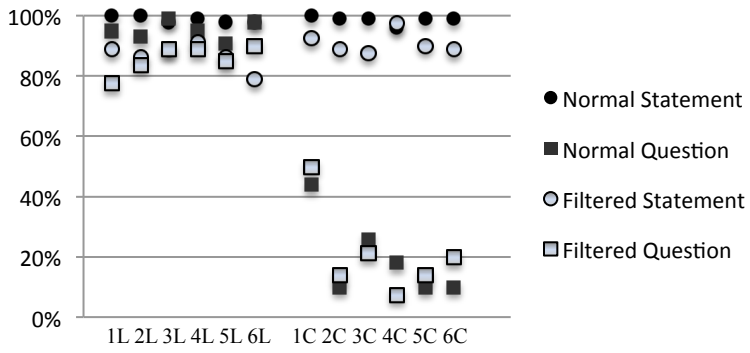

Figure 6.1 Identification accuracy of Cantonese listeners listening to Cantonese normal and filtered speech

To sum up, the production experiments show that the second half of the final rising in the final syllable is the essential and reliable signal for Cantonese questions, while the global F0 contour does not play a role in Cantonese questions. This conclusion was supported by the perception experiments, where questions without the final syllable could not be identified.

6.2 Statement-question bias

Previous studies have often noticed the imbalance between the identification of statement and question in intonation identification (Yuan 2004, 2001, Ma et al. 2006). They have shown that listeners had a bias towards statements. Therefore statement was regarded as an unmarked intonation type. The results of this study generally support this view. The identification accuracy of statements across conditions and listener groups is generally quite high, while identification accuracy of questions is more easily influenced by final tones and conditions. Furthermore, in most cases, listeners of both groups appeared to be significantly more confident in identifying statements than identifying questions. Regardless of language and listener groups, bias scores in most cases are positive, indicating a bias towards statement, especially for utterances in the cut-off condition. This shows that listeners tended to label a sentence as a statement when strong cues for question intonation were absent. This perceptual bias towards statement was reported in previous studies with individual languages (Mandarin, Yuan 2004; 2001; and Cantonese, Ma et al. 2011). Yuan (2004) suggested that the bias may be a universal tendency and statement is an unmarked intonation type. Our results lend support to this proposal by presenting further evidence from cross-linguistic and confidence data. However, results from more languages are needed to corroborate the universality of statement being an unmarked

intonation type.

An unexpected result is that for complete Cantonese utterances, both groups of listeners displayed bias towards questions in utterances with some final tones in both the normal speech and the filtered speech experiments. It is possible that the bias towards question in utterances ending with T2 and T5 is the result of confusion between the rising lexical tone and final rising intonation. However, it is unclear why complete utterances ending with T3 and T6 also led to a bias towards question. This pattern may be accidental, as the differences in identification accuracy between statements and questions in those utterances are not significant.

The confidence data in filtered speech perception may give another possible explanation for the unexpected pattern. Both groups of listeners gave higher confidence ratings to questions than to statements in complete condition in the filtered speech experiment. This confidence pattern suggests that listeners may be using a strategy in the task. In the Cantonese blocks, half of the utterances were questions with high rising ending. When listeners heard a prominent rising, they may immediately treat it as a question; when they did not hear such a rise, they would then decide whether it was a statement and question. The bias towards questions may be result from such a strategy. However, given our limited data, we cannot provide a definitive answer to this pattern, which needs further investigation.

6.3 Cross-linguistic perception

Tone language speakers have three tools to assist their perception of intonation: the psychoacoustic Frequency Code that may be universal across languages, the phonological intonation knowledge of their own language, and the knowledge about lexical tone contours that would interact with intonation patterns. Studies have shown that the Frequency Code plays an important part in intonation perception (Gussenhoven and Chen 2000), but intonation perception is also shaped by listeners' first language phonology of intonation (Handding-Koch and Studdert-Kennedy 1964, Jiang and Chen 2012). However, no previous studies have systematically demonstrated the interactions between these three types of knowledge tone language speakers have in perceiving intonation. This study has investigated these interactions through different conditions.

6.3.1 Interaction of different perceptual knowledge

First, listeners were asked to identify intonation of an unfamiliar/less familiar language in unfiltered normal speech. In this context, listeners were tested which knowledge was used primarily in their perception. The results show that the listeners

were using all three tools in their perception.

Mandarin listeners' performance in identifying Cantonese utterances was influenced by either the Frequency Code or their language experience. Mandarin listeners showed a ceiling performance across all the tones in the complete condition. In other word, they performed well in identifying both statements and questions, and there is no significant difference among different final tones. In the cut-off condition, Mandarin listeners did poorly with questions except those with T1, because no perceptual cue was present in the signal. The results show that the final rising in Cantonese is a useful cue for Mandarin listeners to identify questions as well. Because the Mandarin listeners in this study did not understand Cantonese, we propose two possibilities why Mandarin listeners were doing so well in identifying complete Cantonese utterances. The first is that they may have relied on the Frequency Code, and hence recognized the salient final rising as a cue for Cantonese questions. It is also possible that Mandarin listeners obtained the knowledge of questions signaled by final rising from learning English which has a similar pattern.

Cantonese listeners' identification of Mandarin intonation was facilitated by different tools that interacted with each other. In most cases, they used the phonological knowledge of intonation in Cantonese by locating the perceptual cue for question at the end of the utterances, as this is where the cue for their native

language is placed. Then they made use of the Frequency Code to associate a high rising tone with questions and a low tone with statements. For example, they performed well in T3 cut-off questions but poorly for T3 complete questions. Finally, the experience in knowing Mandarin also helped them to identify questions with T4.

6.3.2 The role of the Frequency Code and language background

In the experiment with low-pass filtered speech, segmental information was eliminated and F0 became the main resource of information left in the stimuli. The low-pass filter restricted the listeners' use of native phonological knowledge of tone and intonation. Therefore, the Frequency Code might have become the most available and useful tool for intonation perception.

The listeners performed generally well in the low-pass filtered speech experiment. In the complete conditions of both languages, listeners reached high identification accuracy around 80% in utterances with most of the final tones. This indicates that the impact of the low pass filter on the perception of intonation is not severe. The F0 contours provide most information needed for identifying intonation types. The listeners even rated high confidence in identifying questions in this condition. The perceptual results of filtered Cantonese speech show that listeners associated final high or rising pitch with questions, consistent with the prediction by

the Frequency Code.

For the perception of filtered Mandarin utterances, the Frequency Code also seems to play an important role. The results show that a low tone (canonical T3) is associated with statements and a high rising tone (sandhi-ed T3) is associated with questions. The unsuccessful identification of T4 questions was because the low-pass filter has concealed the identity of language and lexical tone from the listeners. They were unable to use the knowledge of tone-intonation interaction to identify questions and therefore had to follow the Frequency Code to associate the final falling tone with statements.

However, although the Frequency Code is shown to have dominated the perception processing of intonation in the filtered context, the perceptual patterns of the two groups of listeners were not identical. For example, the identification accuracy of T1 cut-off question in Mandarin is not significantly different from T3 question for Cantonese listeners, but is significantly lower than T3 questions for Mandarin listeners. T4 complete question in Mandarin was well identified by Mandarin listeners but not so by Cantonese listeners. These differences stem from the different language backgrounds the two groups of listeners have. Language background seems to be an influential factor for the perception of intonation, even when top-down linguistic knowledge is inhibited by the low-pass filter.

In summary, the results show that two interactions are important in intonation perception. First, the interaction between tone and intonation: no matter how intonation modifies F0 contour, at the final tone or on a sentential level, listeners are sensitive to the identity of lexical tone at the end of the utterance. Without knowing the identity of the lexical tones, listeners might find it difficult to identify the correct intonation types. Second, the interaction between tone-intonation knowledge and psychoacoustic code: They co-direct listeners' perception of intonation, especially when listeners are given an unfamiliar/less familiar language or filtered speech. Listeners follow the Frequency Code when phonological knowledge is not applicable. When the linguistic resources are rich in the speech signal, and when phonological knowledge and the psychoacoustic code conflict with each other, phonological knowledge would override the Frequency Code. As a result, listeners could perceive intonation patterns that seemingly contradict the Frequency Code in their native/familiar language. Finally, language background is an important and strong factor for intonation perception even with filtered speech materials, as different patterns arose from the two groups of listeners because of their language background.

6.4 Theoretical significance

The results of this study indicate that language experience shapes prosodic

perception. It is shown that native language influenced the perception even in unintelligible filtered speech. In addition, by comparing the results of the two perception experiments, three perceptual patterns were observed. Listeners primarily use their phonological knowledge in perceiving intonation. When they listened to a less familiar / unfamiliar unfiltered speech, they made use of both phonological knowledge and the Frequency Code. Finally, they largely relied on the Frequency Code when listening to the filtered speech. These patterns imply that listeners may process different materials differently. While the universal and biologically-based Frequency Code lays the foundation for intonation perception, phonological knowledge comes into play when speech, especially familiar or native speech materials are exposed to listeners. When phonological knowledge and the Frequency Code contradicted with each other, the former takes advantage in the perception of native language.

Some implications regarding the Frequency Code arise from the results. First, the effect of the Frequency Code seems more prominent in the final position of utterances than in the global register. When listening to the other language, Mandarin listeners performed very well in identifying Cantonese intonation; while Cantonese listeners failed to do so in identifying Mandarin intonation. This comparison is quite surprising given the fact that the Mandarin listeners in the experiments did not

understand Cantonese, while the Cantonese listeners could speak Mandarin with varying proficiency. Assuming that the listeners were assisted by the Frequency Code when perceiving an unfamiliar / less familiar language, the Frequency Code is more helpful for the perception of Cantonese intonation which uses final rising to signal questions, than Mandarin intonation which uses global raise of F0. Furthermore, the listeners may have some awareness about the Frequency Code, since the confidence rating in every context and condition significantly and positively correlated with the identification accuracy.

Finally, this study focused on the perception of intonation by tone language speakers. In addition to the complicated interaction between final tone and intonation echoing previous studies (Yuan 2004; 2011; Ma et al. 2006a; 2011), it is also found that tone sandhi and tonal coarticulation also influence intonation perception. The results suggest that intonation cannot be processed independent of lexical tones. In tone languages, lexical level and postlexical level are closely related.

6.5 Limitations of the current study

While this study investigated the cross-linguistic perception of intonation by tone languages speakers, a topic that has not been well explored, a number of limitations

should also be acknowledged for future improvement. First, as a small-scale study, the speech materials given to the listeners were produced by one male and one female speaker from each language. Only one statement and one question from each speaker were adopted as repeated stimuli. Although this method produced controlled perceptual patterns, the results may be influenced by individual features of the chosen speakers. The perceptual materials lack inter-speaker and intra-speaker variations that are common features of everyday speech. In future study on the same topic, diversity in the speech stimuli can be used in the perception experiments.

Another limitation lies in the listeners participating in this study. It would be ideal to recruit native listeners of Mandarin and Cantonese who knew nothing about the intonation patterns of the other language. However, it was very difficult to find such subjects. For the Mandarin listeners, although we made sure most of them arrived in Hong Kong within two or three months and had never learned Cantonese, they had learned English for many years. As a result, although they could not understand Cantonese, they may be familiar with the notion of high rising terminal from their experience in learning English. In addition, due to the close connection between Chinese mainland and Hong Kong, it became very difficult to find Hong Kong Cantonese listeners, especially those in the younger generation, who had not learned Mandarin Chinese at some point in their life. All the Cantonese listeners in

this study could speak Mandarin. However, the proficiency and the frequency of their Mandarin use were difficult to control. If we could recruit listeners without knowledge of the other language at all, the perceptual differences between the two groups would be more pronounced.

6.6 Future directions

This study has investigated the acoustics of Cantonese yes-no question. More related research questions can be asked. For example, how intonation type (statement and question) in Cantonese interacts with other use of intonation such as emotion and attitude. Would they alternate the final rising of Cantonese questions, or would they modify F0 contour in a larger scope?

This study has examined the effect of the sentence final particle “me” on Cantonese question intonation. It is found that “me” carries the high boundary of Cantonese question and the F0 contour before “me” is not influenced by the intonation question. Further studies can be conducted to find out if this conclusion can be extended to other sentence final particles at the end of questions, especially for those without an explicit question meaning or those without a high tone. It is also unknown how sentence final particle clusters in Cantonese, such as “aa-maa”, are

influenced by question intonation, whether the high boundary target would be localized in the final syllable or extended to the whole cluster (more than one syllable). Furthermore, the acoustics of Cantonese questions should be examined with regard to the syntactic construction of questions.

The perception experiments had Mandarin and Cantonese listeners listen to the others' language, in order to evaluate the interactions between their phonological knowledge and the universal Frequency Code in their perception of intonation. However, to generalize the conclusion of this study beyond Chinese speakers, cross-linguistic tests between other language pairs and by other language speakers should be conducted. It would be interesting to have non-tone language speakers to identify intonation types in tone languages. Although Liang and van Heuven (2007) showed that Uygur speakers are sensitive to intonation in tone languages, their conclusion was weakened by the lack of understanding of the Uygur intonation system. As a result, it is difficult to figure out how Uygur speakers were actually processing the pitch contour in their experiment. In fact, as Jiang and Chen (2012) shows, English speakers may confuse the ending lexical tone with boundary tone for intonation. Furthermore, some varieties of non-tone languages may use a high rising contour systematically at the end of statements (Hirst 1998, 2013, Alcoba and Murillo 1998, Botinis 1998). In the American native language Chickasaw, rising

pitch is associated with statements while falling pitch is associated with questions (Gordon 1999, 2005). Cross-linguistic studies with listeners of those languages would demonstrate how phonological knowledge overrides or gives way to the psychoacoustic code. Future studies in cross-linguistic perception of intonation will continue to reveal the general and language-specific mechanisms in prosodic perception, and contribute to a better understanding of speech perception theories in general.

6.7 Conclusion

This study investigated cross-linguistic perception of intonation by Cantonese and Mandarin speakers. First, the results confirmed intonation pattern in both languages. Cantonese questions rely on the final rising in production and perception. Mandarin intonation is not influenced by eliminating the final tone, but is affected by the identity of the final tone. Second, the study shows that statement is an unmarked intonation type, and cross-linguistically there is a bias towards statements in intonation perception. Third, there is complicated tone-intonation interaction in the perception of both Cantonese and Mandarin. Besides the identity of tone, tonal coarticulation and tone sandhi also influence the interpretation of intonation. In addition, the results of the filtered speech perception experiment show that F0 is the

principal cue for intonation perception. Most importantly, the study demonstrates the interaction between the psychoacoustic Frequency Code and tone-intonation phonological knowledge. When listening to an unfamiliar/less familiar language, listeners make use of both mechanisms in interpreting intonation. Phonological knowledge seems to have an advantage over the Frequency Code in perceiving intonation. When rich linguistic information is present, phonological knowledge may override the Frequency Code. Whenever phonological knowledge is not applicable, the Frequency Code will direct the listeners' identification. Finally, the results also show that language background plays an important role in intonation perception.

BIBLIOGRAPHY

Alcoba, Santiago, and Murillo, Julio. 1998. "Intonation in Spanish." Pp. 152-166 in *Intonation Systems: A Survey of Twenty Languages*, edited by Hirst, Daniel, and Di Cristo, Albert. Cambridge: Cambridge University Press.

Bauer, Robert S., and Paul K. Benedict. 1997. *Modern Cantonese phonology*. Mouton de Gruyter.

Beaugendre, Frédéric, David House, and Dik J. Hermes. 2001. "Accentuation boundaries in Dutch, French and Swedish." *Speech Communication* 33(4):305–318.

Blicher, Deborah L., Randy L. Diehl, and Leslie B. Cohen. 1990. "Effects of syllable duration on the perception of the mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement." *Journal of Phonetics* 18:37–49.

Bolinger, Dwight L. 1964. "Intonation as a universal." Pp 833-834 in *Proceedings of the Nineth International Congress of Linguists*. Cambridge, U.S.

Bolinger, Dwight L. 1979. "Intonation across languages." in *Unversals of Language Volume 2: Phonology*, edited by Joseph Harold Greenberg, Charles Albert Ferguson, and Edith A. Moravcsik. Stanford: Standford University Press.

Botinis, Antonis. 1998. "Intonation in Greek." Pp 291-313 in *Intonation Systems: A Survey of Twenty Languages*, edited by Hirst, Daniel, and Di Cristo, Albert. Cambridge: Cambridge University Press.

Burnham, Dennis et al. 1996. "Perception of lexical tone across languages: Evidence for alinguistic mode of processing." Pp. 2514–2517 in *Proceedings of the Fourth International Conference on Spoken Language processing*. Philadelphia, U.S.

Chang, Choi-Yeung. 2003. Flynn. Intonation in Cantonese. LINCOM Europa.

Chao, Yuen Ren. 1933. "Tone and intonation in Chinese." *Bulletin of the Institute of History and Philosophy* 4:121–134.

Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. University of California

Press.

Chow, Shuen. 2002. “A phonetic study of the intonation in Hong Kong Cantonese.” M.Phil. thesis. City University of Hong Kong. Hong Kong.

Connell, Bruce A., John T. Hogan, and Anton J. Rozsypal. 1983. “Experimental evidence of interaction between tone and intonation in Mandarin Chinese.pdf.” *Journal of Phonetics* 11:337–351.

DiCanio, Christian T. 2012. “Cross-linguistic perception of Itunyoso Trique tone.” *Journal of Phonetics* 40(5):672–688.

Duanmu, San. 2007. *The Phonology of Standard Chinese*. Oxford University Press.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian, and Hacques Mehler. 1997. “A Destressing ‘Deafness’ in French?” *Journal of Memory and Language* 36(3):406–421.

Dupoux, Emmanuel, Sharon Peperkamp, and Núria Sebastián-Gallés. 2001. “A robust method to study stress ‘deafness’.” *The Journal of the Acoustical Society of America* 110(3):1606–1618.

Dupoux, Emmanuel, Núria Sebastián-Gallés, Eduardo Navarrete, and Sharon Peperkamp. 2008. “Persistent stress ‘deafness’: the case of French learners of Spanish.” *Cognition* 106(2):682–706.

Fang, Xiao-yan. 2003. *Sentence-final Particles in Guangzhou Dialect*. Guangzhou: Jinan University Press.

Fant, Gunnar. 1975. “Non-uniform Vowel Normalization.” Speech Transmission Laboratory Quarterly Progress and Status Report 16 (2-3): 1–19.

Fitch, W. Tecumseh. 1997. “Vocal Tract Length and Formant Frequency Dispersion Correlate with Body Size in Rhesus Macaques.” The Journal of the Acoustical Society of America 102 (2): 1213–22.

Fitch, W. Tecumseh. 1999. “Acoustic Exaggeration of Size in Birds via Tracheal Elongation: Comparative and Theoretical Analyses.” Journal of Zoology 248 (1): 31–48.

Fitch, W. Tecumseh, and Jay Giedd. 1999. “Morphology and Development of the Human Vocal Tract : A Study Using Magnetic Resonance Imaging.” Journal of

Acoustical Society of America 106: 1511–1522.

Fitch, W. Tecumseh, and D. Reby. 2001. "The Descended Larynx Is Not Uniquely Human." Proceedings of the The Royal Society 268:1669–1675.

Francis, Alexander L., Valter Ciocca, Lian Ma, and Kimberly Fenn. 2008. "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers." *Journal of Phonetics* 36(2):268–294.

Gandour, Jackson T. 1981. "Perceptual dimensions of tone - Evidence from Cantonese." *Journal of Chinese Linguistics* 9:20–36.

Gandour, Jackson T. et al. 2000. "A crosslinguistic PET study of tone perception." *Journal of cognitive neuroscience* 12(1):207–22.

Gordon, Matthew. 1999. "The intonational structure of Chickasaw." Pp. 1993-1996 in Proceedings of the Sixteenth International Conference of Phonetic Sciences. Saarbrücken, Germany.

Gordon, Matthew. 2005. "Intonational phonology of Chickasaw." Pp. 301-330 in *Prosodic Models and Transcription: Towards Prosodic Typology*, edited by Jun, Sun-Ah. Oxford University Press.

Gottfried, Terry L., and Traci L. Suiter. 1997. "Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones." *Journal of Phonetics* 25:207–231.

Grabe, E., B. S. Rosner, and Jose E. Garcia-Albea. 2003. "Perception of English Intonation by English, Spanish, and Chinese Listeners." *Language and Speech* 46(4):375–401.

Gu, Wentao, Keikichi Hirose, and Hiroya Fujisaki. 2006. "Modeling the Effects of Emphasis and Question on Fundamental Frequency Contours of Cantonese Utterances." *IEEE Transactions on audio, speech and language processing* 14(4):1155–1170.

Gussenhoven, Carlos. 2002. "Intonation and interpretation : phonetics and phonology." Pp. 47–57 in *Proceedings of the First International Conference on Speech Prosody*. Aix-en-provence, France.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge University Press.

Gussenhoven, Carlos, and Aoju Chen. 2000. “Universal and language-specific effects in the perception of question intonation.” in *Proceedings of the Sixth International Conference on Spoken Language Processing*. Beijing, China.

Gårding, Eva. 1984. “Chinese and Swedish in a Generative Model of Intonation. In Eds. by Elert, Claes-Christian, Johansson, and Eva Strangert.” Pp. 79– 91. in *Nordic Prosody III*, edited by Claes-Christian Elert and Iréne u. Eva Strangert Johansson. Stockholm: University of Umeå.

Gårding, Eva. 1987. “Speech act and tonal pattern in Standard Chinese: constancy and variation.” *Phonetica 44: 13-29* 44:13–29.

Hallé, Pierre a., Yueh-Chin Chang, and Catherine T. Best. 2004. “Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners.” *Journal of Phonetics* 32(3):395–421.

Handding-Koch, K., and M. Studdert-Kennedy. 1964. “An experimental study of some intonation contour.” *Phonetica* 11:175–185.

He, Yang, and Song Jin. 1992. “Experimental investigation of intonation in Beijing Mandarin.” *Language Teaching and Research* 2:71–96.

Hirst, Daniel. 1998. “Intonation in British English.” Pp. 56-77 in *Intonation Systems. A Survey of Twenty Languages*, edited by Hirst, Daniel, and Di Cristo, Albert. Cambridge: Cambridge University Press.

Hirst, Daniel. 2013. “UK declarative rises and the frequency code.” Pp. 149–160 in *Prosody and Iconicity*, edited by Sylvie Hancil and Daniel Hirst. John Benjamin Publishing Company.

Ho, Aichen T. 1977. “Intonation variation in a Mandarin sentence for three expressions interrogative, exclamatory and declarative.” *Phonetica* 34:446–457.

Jiang, Ping, and Aishu Chen. 2011. “Representation of Mandarin intonations : Boundary tone revisited.” Pp. 97–109 in *Proceedings of the Twenty-third North American Conference on Chinese Linguistics*, vol. 1. Eugene, U.S.

Jiang, Ping, and Aishu Chen. 2012. “Tonal effect on Mandarin intonation perception : A comparative study of listeners with different L1 backgrounds Research questions.” in *International Symposium on Bilingualism and Comparative Linguistics*. Hong Kong.

Kaan, Edith, Christopher M. Barkley, Mingzhen Bao, and Ratree Wayland. 2008. "Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study." *BMC neuroscience* 9:53-70.

Kao, Diana L. 1971. *Structure of the Syllable in Cantonese*. Paris: The Hague, Mouton.

Klein, D., R. J. Zatorre, B. Milner, and V. Zhao. 2001. "A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers." *NeuroImage* 13(4):646–53.

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

Ladd, D. Robert. 1981. "On intonational universals." Pp. 389–397 in *The Cognitive Representation of Speech*, edited by T. Myers, J. Laver, and J. Anderson. North-Holland Publishing Company.

Ladd, D. Robert. 2008. *Intonational Phonology*. 2nd ed. Cambridge University Press.

Lam, Kin-Ping. 2002. "A Study of Intonation in Hong Kong Cantonese". Ph.D. dissertation, The Chinese University of Hong Kong, Hong Kong.

Law, Sam-Po. 1990. "The syntax and phonology of Cantonese sentence-final particles". Ph.D. dissertation. Boston University.

Lee, Chao-Yang, Liang Tao, and Z. S. Bond. 2010. "Identification of acoustically Modified Mandarin tones by Non-native Listeners." *Language and Speech* 53(2):217–243.

Lee, Hoi-Yee Diana. 2008. "The Intonation of Cantonese Declarative Questions." M.Phil. thesis. Cambridge University. Cambridge, U.K.

Lee, Ok Joo. 2005. "The Prosody of Questions in Beijing Mandarin." Ph.D. dissertation. The Ohio State University.

Lehiste, Ilse, and Robert Allen Fox. 1992. "Perception of Prominence by Estonian and English Listeners." *Language and Speech* 35(4):419–434.

Leung, C. 2005. *A Study of the Utterance Particles in Cantonese as Spoken in Hong Kong. Monograph Series Computational and Linguistic Analysis of Asian*

Languages, No. 2. Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.

Liang, Jie, and Vincent J. J. P. van Heuven. 2007. "Chinese tone and intonation perceived by L1 and L2 listeners." in *Tones and Tunes, Volumn 2: Experimental studies in word and sentece prosody*, edited by Carlos Gussenhoven and Riad Tomas.

Lin, Maocan. 2004. "On Production and Perception of Boundary Tone in Chinese Intonation." in *Proceedings of the First International Symposium on Tonal Aspects of Languages*. Beijing, China.

Liu, Fu. 1924. A report of experiments on the four tones in Chinese. Qunyi Book Company.

Liu, Siyun, and Arthur G. Samuel. 2004. "Perception of Mandarin Lexical Tones when F0 Information is Neutralized." *Language and Speech* 47(2):109–138.

Luchsinger, R. L., and G. E. Arnold. 1965. *Voice, Speech and Language*. Constable and Co. Ltd.

Ma, Joan K-Y., Valter Ciocca, and Tara L. Whitehill. 2006a. "Effect of intonation on Cantonese lexical tones." *Journal of the Acoustic Society of America* 120(6):3978–3987.

Ma, Joan K-Y., Valter Ciocca, and Tara L. Whitehill. 2006b. "Quantitative analysis of intonation patterns in statements and questions in Cantonese." in *Proceedings of the Third International Conference on Speech Prosody*. Dresden, Germany.

Ma, Joan K-Y., Valter Ciocca, and Tara L. Whitehill. 2011. "The perception of intonation questions and statements in Cantonese." *Journal of the Acoustic Society of America* 129(2):1012–1023.

Mai, Yun. 1998. "Intonatin of Guangzhou Cantonese." *Guangzhou Cantonese Research and Teaching* 3:269–279.

Matthews, Steven, and Yip Virginia. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.

Ohala, John J. 1983. "Cross-language use of pitch: an ethological view." *Phonetica* 40:1–18.

Peng, Gang et al. 2010. “The influence of language experience on categorical perception of pitch contours.” *Journal of Phonetics* 38(4):616–624.

Peng, Shu-hui. 2000. “Lexical versus ‘phonological’ representations of Mandarin sandhi tones.”Pp. 152–167 in *Acquisition and the lexicon: Papers in Laboratory Phonology V*, edited by M. B. Broe and J. Pierrehumbert. Cambridge: Cambridge University Press.

Peng, Shu-hui et al. 2006. “Towards a Pan-Mandarin system for prosodic transcription.” Pp. 1–63 in *Prosody Typology: The Phonology of Intonation and Phrasing*, edited by Sun-Ah Jun. Oxford University Press.

Puts, David Andrew, Steven J.C. Gaulin, and Katherine Verdolini. 2006. “Dominance and the evolution of sexual dimorphism in human voice pitch.” *Evolution and Human Behavior* 27(4):283–296.

Riede, Tobias, and Tecumseh Fitch. 1999. “Vocal Tract Length and Acoustics of Vocalization in the Domestic Dog (Canis Familiaris).” The Journal of Experimental Biology 202: 2859–2867.

Rietveld, Toni, Judith Haan, Linda Heijmans, and Carlos Gussenhoven. 2002. “Explaining attitudinal ratings of Dutch rising contours: morphological structure vs. the frequency code.” *Phonetica* 59(2-3):180–94.

Shen, Jiong. 1985. “The pitch range of lexical tone and intonation in Beijing Mandarin.” Pp. 73–130 in *Exprimental Phonetic Studies of Beijing Mandarin*, edited by Tao Lin and Lijia Wang. Beijing University Press.

Shen, Jiong. 1994. “Construction and types of Chinese intonation.” *Dialect* 3:221–228.

Shen, Xiao-nan Susan. 1990. *The prosody of Mandarin Chinese*. University of California Press.

Shi, Feng. 2010. *Boundary Tone and Focus Tone--on the hierarchy of Chinese intonation system*. Public Talk at The Chinese University of Hong Kong.

Shi, Peiwen. 1980. “Intonation variations in four types of Mandarin sentences.” Language Teaching and Research 2:71–81.

Shih, Chilin. 1986. “The Prosodic Domain of Tone Sandhi in Chinese”. Ph.D dissertation. University of California, San Diego.

Shih, Chilin. 1997. “Declination in Mandarin.” In Proceedings of an ESCA Workshop, 293–296. Athens, Greece.

So, Connie K., and Catherine T. Best. 2010. “Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences.” Language and Speech 53(2):273–293.

Speer, Shari R., Chilin Shih, and Maria L. Slowiaczek. 1989. “Prosodic Structure in Language Understanding: Evidence from Tone Sandhi in Mandarin.” Language and Speech 32 (4): 337–354.

Stagray, J., and D. Downs. 1993. “Differential sensitivity for frequency among speakers of a tone and a nontone language.” *Journal of Chinese Linguistics* 21:143–163.

Uldall, Elizabeth. 1964. “Dimensions of meaning in intonation.” Pp. 271–279 in *In Honour of Daniel Jones: Papers Contributed on the Occasion of his Eightieth Birthday*, edited by D. Abercrombie Trim, D.B. Fry, P.A.C. MacCarthy, N.C. Scott, and J.L.Trim. London: Longman.

Vaissiere, Jacqueline. 2005. “Perception of intonation.” Pp. 236–263 in *The Handbook of Speech Perception*, edited by David B. Pisoni and Robert E. Remez. Blackwell Publishing.

Vance, Timothy J. 1977. “Tonal distinction in Cantonese.” Phonetica 34: 93-107.

Wakefield, John C. 2010. “The English equivalents of Cantonese sentence-final particles: A contrastive analysis.” Ph.D dissertation. The Hong Kong Polytechnic University. Hong Kong.

Wang, Qian. 2008. “L2 Stress Perception : The reliance on different acoustic cues.” in *Proceedings of the Fourth International Conference on Speech Prosody*. Campinas, Brazil.

Wang, William S-Y., and Kungpu Li. 1967. “Tone 3 in Pekinese.” *Journal of Speech and Hearing Research* 10(3):629–636.

Wu, Wing-li. 2008. “An acoustic phonetic study of the intonation of sentence-final particles in Hong Kong Cantonese.” *Asian Social Science* 4(2):23–29.

Wu, Wing-li. 2009. “Sentence-final particles in Hong Kong Cantonese : Are they tonal or intonational?” in *Proceedings of Interspeech*.

Wu, Zongji. 1982. “Variation of lexical tones in Mandarin intonation.” *Studies of The Chinese Language* 6:439–449.

Xu, Yisheng, Jackson T. Gandour, and Alexander L. Francis. 2006. “Effects of language experience and stimulus complexity on the categorical perception of pitch direction.” *The Journal of the Acoustical Society of America* 120(2):1063-1074.

Yau, S.-C. 1980. “Sentential connotations in Cantonese.” *Dialect* 1:35–52.

Yuan, Jiahong. 2004. “Intonation in Mandarin Chinese: Acoustics, perception, and computational modeling.” Ph.D. dissertation. Cornell University.

Yuan, Jiahong. 2006. “Mechanisms of question intonation in Mandarin.” Pp. 19–30 in *Chinese Spoken Language Processing*, edited by Q. Huo, B.Q. Ma, E-S. Chng, and H Li. Springer.

Yuan, Jiahong. 2011. “Perception of intonation in Mandarin Chinese.” *The Journal of the Acoustical Society of America* 130(6):4063–4069.

Yuan, Jiahong, Chilin Shih, and Greg P. Kochanski. 2002. “Comparison of declarative and interrogative intonation in Chinese.” Pp. 711–714 in *Proceedings of the First Internation Conference on Speech Prosody*. Aix-en-Provence, France.