INTONATION EFFECTS ON CANTONESE LEXICAL TONES

IN SPEAKING AND SINGING

LING ZHANG

Ph.D

The Hong Kong Polytechnic University

2013

THE HONG KONG POLYTECHINIC UNIVERSITY

Department of Chinese and Bilingual Studies

Intonation Effects on Cantonese Lexical Tones

in Speaking and Singing

Ling ZHANG

A thesis submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

December 2012

CERTIFICATE OF ORIGINALITY

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

________________________ (Signed)

Ling ZHANG (Name of student)

Abstract

Cantonese lexical tones are preserved in both speaking and singing (which can be regarded as a special kind of speech style). Through three main acoustic experiments, the present thesis is a systematic study of the intonation effects on Cantonese lexical tones in speaking and singing.

In Experiment I, three pairs of speech styles were compared: opera speaking versus opera singing, opera speaking versus normal speaking, and normal speaking versus normal singing. A perception test further examined the conclusions of the acoustic experiments from the perspective of listeners. The results of Experiment I showed that speaking and singing speech styles are different mainly in pitch slope. Speaking speech styles are associated with a declining intonation, while singing speech styles tend to be level or even show a slightly ascending tendency. The experimental data also indicate that a higher pitch register is usually employed when a speech type is considered more conspicuous, such as opera speaking (compared with normal speaking) and normal singing (compared with normal speaking).

Experiment II focused on the comparison between speaking and singing by layman speakers. Twelve native speakers of Cantonese were asked to read aloud and sing a set of lyrics taken from Cantonese songs. The data analysis was conducted in the framework of a matrix of the two dimensions of pitch configuration and pitch domain. At the syllable level and the utterance-body portion, speaking has a more pitch slope declination, a lower pitch register, and a wider pitch span, while singing has a more ascending pitch slope, a higher pitch register, and a narrower pitch span. Pitch slope at both the syllable and the utterance levels exhibits a downdrift pattern for speaking and an updrift pattern for singing. At the utterance-final portion, speaking shows a strong utterance-final declination effect while singing does not exhibit an obvious utterance-final effect.

Experiment III was conducted in the same framework as Experiment II, but with focus on the internal differences of normal speaking. Four sentence types were compared: declarative without a sentence-final particle (SFP, henceforth), declarative-derived question without an SFP, declarative with an SFP, and

declarative-derived question with an SFP. The results showed that there is a declination effect in the domains of the syllable level and the utterance-body portion. For utterances without an SFP, the utterance-final intonation is very different: in a declarative, the final syllable has a significantly lower pitch register and a slightly more declining pitch slope; in a question, the final syllable keeps the pitch register unchanged but has a sharply rising contour. For utterances with an SFP, the SFP carries the burden of utterance-final intonation. The utterance-final intonation can be regarded as a kind of segmentless SFP with floating pitch contour to be superimposed to the final syllable.

The present study enriches the knowledge of intonation typology and provides further insight into the relationship between lexical tones and intonation. In particular, the updrift intonation of singing fills the vacuum of non-declination in intonation types. The matrix of pitch configuration and pitch domain proposed here is a new model for experimental studies on intonation in tonal languages.

Publications Arising from the Thesis

1. Zhang, Ling. 2010. Downdrift and non-downdrift: Cantonese lexical tones in speaking and singing. Paper presented at IACL-18 & NACCL-22. Harvard University.

2. Zhang, Ling, and Kwan-Hin Cheung. 2011a. Cantonese lexical tones in speaking and singing. Proceedings of the Psycholinguistic Representation of Tone Conference: 32-35. The Chinese University of Hong Kong.

3. Zhang, Ling, and Kwan-Hin Cheung. 2011b. Experimental studies on Cantonese lexical tones in speaking and singing. Paper presented at the 16th International Conference on Cantonese and Yue Dialects, Hong Kong Polytechnic University.

4. Zhang, Ling. 2013. Declination and non-declination: Cantonese lexical tones in speaking and singing. Submitted. *Phonetica*.

Acknowledgements

First and foremost my sincere thanks go to my supervisor, Professor Kwan-Hin Cheung. He gave me much freedom to conduct research that interests me. He encouraged me to think and work independently. When I needed guidance, the discussions with him were always informative and inspiring. It is my great honor to have worked with him these years.

I would like to thank my committee members. Firstly I would like to thank the Chair, Dr. Roxana S. Y. Fung. Without her coordination, my oral examination would have to be arranged much later. I would also like to thank Prof. Marjorie K. M. Chan and Prof. K. K. Luke for their valuable advice and feedback.

I would also like to express my gratitude to the professors who have exerted a profound influence on me. I especially thank Prof. Robert Bauer for his warmly encouraging emails and useful suggestions. I also thank Dr. Sze-Wing Tang, Dr. Choi-Lan Tong, Dr. Chun-Fat Lau, Prof. Yun Mai, Prof. Chusheng Zhuang and Prof. Yu-en Gan for their kindly help and encouragement. I would also like to thank Dr. Xiaonong Zhu, whose class made me more interested in the studies of phonology and phonetics. It was a memorable experience to join the LSA 2009 Linguistic Institute held in UC Berkeley, and I learned a lot from the professors there: Prof. John Ohala, Prof. Ian Maddison, Prof. John Harris, Prof. Carlos Gussenhoven, Prof. Mary Beckman, Dr. Richard Wright, etc. I really enjoyed their wonderful classes and they also gave me valuable suggestions on my own research during our discussions after class. I would like to thank Prof. Thomas H. T. Lee, Prof. Patrick C. M. Wong, Dr. Lian-Hee Wee and Dr. Peggy Mok, for their useful comments during several conferences, which also helped me to improve my studies on my thesis.

Many friends gave me a hand when I was doing my thesis research. Dr. Murray Schellenberg suggested many useful references on singing in tonal languages and helped me to get access to the references I needed. He also helped me to proofread the final version of my thesis. Some friends volunteered to be the informants in my experiment, such as Tak-Sum Wong, Siu-Kei Cheng, Cherie Wong, April Liu, Suixin Zheng, etc. I faithfully appreciate their help.

I have to thank my friends who are also pursuing a linguistic dream. I thank all the guys of Room AG 511. I also thank Ying Chen, Yao Yao, Nian Liu, Shawn Chang, Becky Huang, Te-Hsin Liu, Yufuko Takashima, Hui Chen, Mingxing Li, Robert Xu, Zhen Qin, Lei Jiao, Ying Hong, Yuhang Xu, Jian Jin, etc. With them, I did not feel lonely on my way through these years.

Lastly, I am indebted to my family for their endless love. My parents raised me up and always stand by me. My husband Haiping Yin strongly supports my research and has persistent confidence in me. My mother-in-law and my parents helped us to look after our daughters Yue Xin and Yue Si, so that I could concentrate my mind on my own research. For my dear Yue Xin and Yue Si, you have been bringing us happiness and motivation to work harder and harder.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

DN = Declarative without sentence-final particle

DP = Declarative with sentence-final particle

f_0 = Fundamental frequency

ICF = Isolated citation form

NA = Not applicable

NSng = Normal Singing

NSpk = Normal Speaking

OSng = Opera Singing

OSpk = Opera Speaking

QN = Declarative-derived question without sentence-final particle

QP = Declarative-derived question with sentence-final particle

SNG = Singing

SFP = Sentence-final particle

SPK = Speaking

ST= Semitone

T1~6 = Tone 1~6

Chapter 1
Introduction and Literature Review

In tone languages, pitch (or its acoustic correlate fundamental frequency, f_0) is used to signify both lexical tones and intonation. The relationship between lexical tones and intonation has aroused increasing research interest. For a language with a large number of lexical tones like Cantonese, the competition between lexical tones and intonation for pitch manipulation is especially complicated.

It is very interesting and significant that Cantonese lexical tones are preserved not only in ordinary speech but also in authentic Cantonese singing (Chan, 1987a, 1987b; Chao, 1956/2006; Chen-Hafteck, 1999; Cheung, 2012a; Cheung, 2007, 2012b; Cheung & Wong, 2008; Ho & Ho, 2007; Ho, 2006, 2008; Schellenberg, 2011, 2013; Wong & Diehl, 2002). Thus, melody does not neutralize tonal contrasts in Cantonese and Cantonese speakers can identify the tones of lexical items in songs. As pitch is also the essential element for melody in singing, it would be very interesting to compare the relationship between tone and intonation in speaking and between tone and melody in singing. Actually Cantonese speakers can clearly judge whether an utterance is being spoken or sung, which indicates that the realization patterns of lexical tones between speaking and singing exhibit certain differences.

The present study explores the different intonation effects on Cantonese lexical tones in speaking and singing with the help of systematic experiments. The intonation effect of singing speech can serve as a reference, thus shedding new light on the research problem of tone-intonation relationship in speaking. Part of the results has been reported in several conferences (Zhang, 2010; Zhang & Cheung, 2011a, 2011b), while this thesis unifies those reports and provides a more systematic analysis with more data.

Speaking and singing are paired reference objects in our present study. The "speaking" speech style mentioned here is used in a broad sense: it is not only limited to the spontaneous narrative speech (the "speaking" referred in Esser and Polomski (1988)) but includes reading aloud as well. Actually in the controlled experiments, it is much more convenient and controllable to ask informants to "read aloud" and sing the same text.

Later in this chapter, we provide the necessary background information and a literature review, including tone studies, tone-intonation studies and tone-melody studies. In each area, we adopt a general-to-specific order for literature review: a general overview will be provided first, and then our concern will be more focused on the studies of Cantonese. At the end of this chapter, we briefly introduce the organization of the experiments conducted in the present study.

1.1 Tone

1.1.1 An overview of tone studies

In this section, we provide a brief account of the definition of tones, the phonological features of tones, tone studies in the Chinese tradition, and several contextual effects other than intonation on tones in continuous speech

Firstly we have to make it clear what tone is and what a tone language is. A classic definition was given by Pike (1976, p. 3):

> "A tone language may be defined as a language having lexically significant, contrastive, but relative pitch on each syllable. The languages of southeastern Asia (China, Burman, Indo-China, Siam) are largely tonal, as are the languages of Africa west of Ethiopia and south of the Sahara (Sudanic, Bantu, Bushman, and the Hottentot groups). In North America various tone languages are found in southwestern Mexico (Mixteco, Mazateco, Amuzgo, Chatino, Chinanteco, Chocho, Cuicateco, Otomi, Tlapaneco, Trique, and Zapoteco) and in the United States (Navaho, Apache, and others)."

Wang (1967, pp. 93-94) pointed out that the tone systems of the Sino-Tibetan family (together with many neighboring languages) of Southeast Asia are different from those of the American Indian languages as well as the African languages. He further demonstrated the differences between the tone systems of Sino-Tibetan languages and the other languages. Firstly, the tones of Sino-Tibetan languages are almost exclusively used lexically, with no correlation with the syntactic or morphological aspects of the language, while in many American and African languages, tones carry extensive load in the declensional and conjugational morphology. Secondly, the tone paradigms of Sino-Tibetan languages are typically more complex and have more distinct shapes. Thirdly, tone sandhi operates in different ways: tone sandhi in Sino-Tibetan languages is in the form of paradigmatic replacement, while in many Bantu languages and terrace-level tone languages in West Africa, the tone sandhi is essentially in the form of syntagmatic displacement.

For phonologists, a fundamental concern in tone studies is the phonological features of tones. Pitch or fundamental frequency (f_0) is the primary parameter of tone. Pitch is “a perceptual attribute which allows the ordering of sounds on a frequency-related scale extending from low to high. (Klapuri & Davy, 2006, p. 8)” Fundamental frequency (f_0) is the corresponding physical term. Compared with pitch, f_0 is objective and quantitatively rigorous. The main concern of this thesis is pitch because pitch deals with perception and (perceptually) self-monitored production. f_0 in the log scale is the best acoustical approximation of pitch and is therefore utilized to stand for pitch wherever applicable in the thesis.

The phonological features of tones can be divided into two main dimensions, pitch height (also known as pitch level, or pitch register) and pitch contour (also called pitch shape).

In the dimension of pitch height, from the definition of tone by Pike (1976, p. 3), we have noticed that “relative pitch” is emphasized. Ladefoged (1975/2006, p. 250) pointed out:

> “If you are working with a friend or with recordings of a speaker of a tone language, be careful not to imitate their exact pitches, unless they have just the same pitch range as you normally do. Contrastive tones must always be considered relative to the presumed mean pitch of the speaker.”

Thus, because of such interpersonal pitch range differences, the raw data are messy and conceal the linguistic patterns. Pitch normalization should be conducted when we carry out phonetic research on tones. In the present thesis, pitch normalization was achieved by converting the data to Z-Scores (details listed in Section 3.1.4, Page 90). The Z-Score reflects the relative distribution in the space of a certain individual. The absolute pitch variations of different individuals (typically of different genders) were normalized.

Acoustically absolute pitch can cover a broad range of values, but phonologically the number of distinctive pitch levels for tones is quite small. There have been many discussions on how many levels are needed to describe tones in all languages (Anderson, 1978; Chao, 1930/2006; Guo, 1993; Hulst &

Snider, 1993; Ian, 1978; Pike, 1976; Wang, 1967; Yip, 1990, 2002), but there is still no consensus. Duanmu (2000, p. 254) pointed out:

> "In particular, if downstep or downdrift belongs to phonetic implementation, which need not be represented with separate levels, far fewer phonemic levels are needed. In African languages, two phonemic levels, H and L, are often sufficient, although many more have been posited."

In Sino-Tibetan languages, it is common that there are three or four contrastive levels in a language. It is rare to have five contrastive levels – Gaoba Dong (Shi, Shi, & Liao, 1987), Black Miao and Tauhua Yao (Li, 1980) are reported to be such rare cases.

The most widely used tone-transcription method by Sino-Tibetan linguists is the five-point-scale method, which was proposed by Chao (1930/2006). This method was inspired by the staff notation for transcribing melody. According to this five-point-scale method, the tessitura of a speaker's voice is divided into four equal intervals, resulting in five pitch levels ˩, ˨, ˧, ˦, ˥. The integers 1 to 5 are assigned to the five levels in ascending order of pitch, respectively. From the literal meaning, Chao (1930/2006) posited five tone levels. However, Chao (1930/2006) also pointed out that "In order not to make distinctions too fine, points 2 and 4 are used either alone or with each other, but not in combination with 1, 3 or 5." Cheung (2012c) discussed this rule and found that this rule was not obeyed consistently in actual practice. Cheung (2012c) also noticed that in later works Chao himself violated this rule as well.

Wang (1967) was the first to discuss the distinctive features for tones. He posited seven binary features to represent all tones: [high], [central], [mid], [contour], [rising], [falling], and [convex]. Among them, [high], [central], [mid] are responsible for pitch levels of the system of Chao (1930/2006). These three binary features can also represent the five pitch levels, as listed in Table 1-1. The binary features in Wang (1967) can more flexibly represent the pitch levels in a certain language. Unlike the five-point-scale method, which is by design committed to five pitch levels, Wang (1967)'s method can accommodate five or fewer levels as [central] or [mid] can be suspended.

Table 1-1 Wang (1967)'s binary features and five tone levels

	˩	˨	˧	˦	˥
High	-	-	-	+	+
Central	-	+	+	+	-
Mid	-	-	+	-	-

Zhu (2012, p. 2) argued that

> "The five-point scale does not suffice when facing a more complicated tonal system, where not only pitch, but also phonation carries tonological information."

He introduced a Multi-Register and Four-Level model. To be more specific, he proposed that there are up to three tonological registers in terms of phonation types:

> "By these phonation types three tonological registers are defined. Not every phonation type can identify a register. Essentially, only three types, falsetto, voiceless, and breathy voice, if occurring in a tone language, can define an independent register. Falsetto defines the highest Register H. Voicelessness defines the commonest, default, Register M. Every tone language, every language indeed, has this register. Breathy voice defines the lowest Register L. (Zhu, 2012, p. 3)"

Zhu (2012) further suggested that the pitch range of each register can be divided into four levels, and each register will be one notch lower/higher in pitch than an adjacent one. Thus, there are six pitch levels that span three registers in Zhu (2012)'s model, as displayed in Figure 1-1.

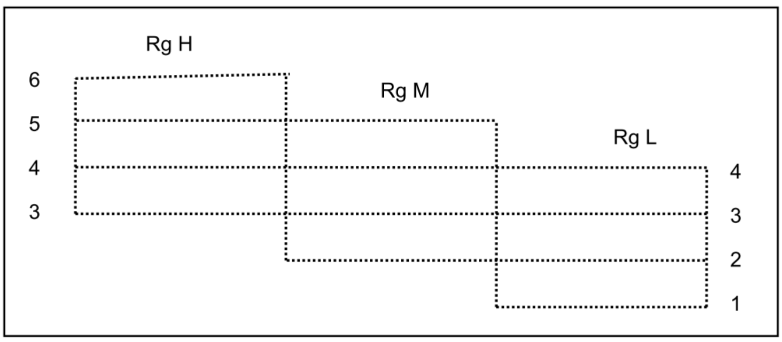

Figure 1-1 The Multi-Register and Four-Level tonal model, reproduced from Zhu (2012)

It should be noted that the term "register" has different referents in Zhu (2012) and in our present study. Zhu (2012) referred to the whole pitch space, which is closely related to the phonation types. In the present study, we do not discuss the factor of phonation types, and "register" refers to the relative height (represented by the median pitch point) of a syllable or an utterance.

The other dimension of phonological features of tones is pitch contour. Among the binary distinctive features proposed by Wang (1967), [contour], [rising], [falling], and [convex] are responsible for pitch contour. Level tones can be represented as [-contour, -rising, -falling, -convex]; rising tones are [+contour, +rising, -falling, -convex]; falling tones are [+contour, -rising, +falling, -convex]; convex tones are [+contour, +rising, +falling, +convex]; and concave tones are [+contour, +rising, +falling, -convex].

Different from Wang (1967)'s method of positing binary features of pitch contour, Woo (1969) proposed that all contour tones are combinations of level tones. This theory can better connect the tone studies in Sino-Tibetan languages with other languages as it can convert various kinds of pitch contours into the simpler linear representation of successive level tones. Woo (1969) also proposed that tones must be realized on segments, and there can only be one tone per segment.

The above account of previous studies on the phonological representation of tones was presented along the two dimensions of pitch height and pitch contour. In our present study, when investigating the intonation effects of different speech styles on the Cantonese lexical tones, the dimensions of pitch height and pitch contour were also our main concerns in our analysis. The analysis on pitch register and pitch span was along the pitch height dimension while the analysis on pitch slope was along the pitch contour dimension.

As Cantonese is a dialect of Chinese, we would also like to provide a brief introduction to the tone studies in traditional Chinese linguistics. In Chinese philological and linguistic literature, tone studies have very different tradition and terms, and the historical tonal category is more emphasized.

Firstly, the tones are classified into four tone categories: Ping (平"Even"), Shang (上"Rising"), Qu (去"Departing") and Ru (入"Entering"), the names of which were first established in the fifth century A.D. (Pulleyblank, 1978). The

Ru tones are different from the other three tone categories in that the Ru-syllables contain a stop coda, while the syllables with other tones ended in a vowel or a nasal. It should be noted that in Chinese tradition, not only the pitch factor, but also the coda types are used as criteria for tone categorization.

Secondly, tones are classified into the Yin ("陰") and Yang ("陽") registers, which were correlated with the voicing qualities of the syllable onsets in Middle Chinese. The Yin tones were with higher or upper register, which occurred on syllables with voiceless onsets. The Yang tones were with lower register, which were occurred on syllables with voiced onsets.

As the four tone categories of Ping, Shang, Qu and Ru cross-cut with Yin-register versus Yang-register, there are eight tone categories in total: Yin-Ping, Yin-Shang, Yin-Qu, Yin-Ru, Yang-Ping, Yang-Shang, Yang-Qu, and Yang-Ru.

The names of tone categories in traditional Chinese linguistics might reflect certain phonetic properties in Middle Chinese, e.g., the names of Ping ("Even") and Shang ("Rising") indicated the tone shapes of these two tone categories in Middle Chinese. Besides, the tone categories in traditional Chinese linguistics can help us to identify the historical origin of tones in modern dialects in Chinese, and they can serve for cross-dialectal or diachronic comparison. However, it should be noted that the same tone category of different dialects may have very different phonetic values. Bao (1999, p. 11) pointed out that the historical classification by traditional Chinese linguistics –

> "does not give the phonetic pitch of the tones. The meanings of the tonal labels are obscure at best, and we have no clear idea how the tones in classical Chinese were realized phonetically. … The labels may reflect the phonetic properties of the tones at the time when they were coined, but, judging by tones of modern dialects, the names are hardly an indication of their pitch height or contour."

When describing the tone system in a modern Chinese dialect, we can still adopt the categorization and terms by traditional Chinese linguistics, while we are more concerned about the actual phonetic value of each tone.

For the investigation of the intonation effects of different speech styles on Cantonese lexical tones, we mainly adopt the Western linguistic practice and terminology, but when describing the tonal system in Cantonese in Section 1.1.2, the Chinese terms will also be presented.

Although the first step for tone studies in a certain language is to study the tones in isolated citation form, which is the easiest and the most common method in field studies on tones, in fact the most frequent use of tones is in continuous speech. Apart from the effect of intonation (which will be introduced below and further explored in later chapters), we would also like to provide a brief introduction to the contextual effects on tones in continuous speech, especially the coarticulation effect.

Norman (1988, p. 146) mentioned two kinds of contextual variants of tones. The first kind is defined as "phonetic sandhi", for which "the changes that take place may be no more than phonetic variants or allophones of the tone in question". The second kind is "phonemic sandhi", for which "the new shape that a tone assumes in a given context may actually coincide with the value of another distinct tone in the same context". For the phonemic sandhi, there is a most famous example in Chinese Mandarin: if there are two successive T3 syllables, the first syllable will become T2. In this thesis, we are more interested in "phonetic sandhi".

Tseng (1990) conducted both acoustic and perception studies on Chinese Mandarin tones, not only in isolation form, but also in continuous speech. She found that:

> "within a sentence, unless it is intended that every syllable be fully articulated, only a small number of syllables were produced with their lexical tones as the phonology predicts. (Tseng, 1990, p. 72)"

Her findings in the perception test were also consistent with the above production results:

> "The perception test demonstrates that without the support of context, native Mandarin speakers identified lexical tones extracted from spontaneous speech much more poorly than their judgment of tones produced in isolation . (Tseng, 1990, p. 91)"

Xu (1994) and Xu (1997) carried out systematic research on the contextual effects of adjacent tones (in Chinese Mandarin) in running speech. Through production and perception tests, Xu (1994) found that the nature of the tonal context influences the amount of deviation of a tone from its canonical form. In a "compatible" context, where adjacent tonal values agree, the deviation was relatively small. In a "conflicting" context, where adjacent tonal values disagree, the deviation was much greater, sometimes even to the extent of changing the direction of a dynamic tone. The perception test also showed that the identification of tones in a "compatible" context is much better than in a "conflicting" context.

Xu (1997) discussed the anticipatory and carry-over effects on the four Mandarin tones in bi-tonal sequences. In his experimental design, the bi-tonal sequences were nonsense words inserted in carrier sentences. The results showed that the carry-over effects and anticipatory effects are different in nature: the carry-over effects are mostly assimilatory, while the anticipatory effects are mostly dissimilatory. These two kinds of effects also have different magnitude: the carry-over effects are often large while the anticipatory effects are relatively small.

Apart from the coarticulation effects, other contextual effects, such as the effect of focus, may also influence the formation and alignment of f_0 contours. Xu (1999, p. 101) found that:

> "focus was found to significantly influence the global f0 shape of the entire sentences. A non-final focus substantially expands the pitch range, particularly the upper end, of the words directly under focus, and it suppressed the pitch range of post-focus words, while leaving that of pre-focus words largely intact."

From the above brief account on contextual effects, it can be inferred that lexical tones in continuous speech are affected by various factors of continuous speech and are in a dynamic situation. In static canonical form, lexical tones are intrinsic phonological categories of pitch features to contrast lexical items. In real life, these intrinsic categories go through a series of phonological, phonetic, syntactic and other rules to become their final phonetic output form in speech. Thus, the realization of lexical tones in connected speech deviates more or less from their canonical form. Our study here is not to challenge the previous phonological description of tones in Cantonese, but to call for attention to the fact that the tones obtained from actual speech (including isolated citation form) are not "pure tones" – they are affected by speech effects.

Ho and Ho (2007) compared the conventional description of Cantonese tones by Jones and Woo (1912), Chao (1947) and Hashimoto (1972), and called for a re-consideration of Chao (1930/2006)'s five-point-scale transcription method. They argued that the use of this “so-called short-hand of musical notation” is musically ill-defined because “a system of standardized tools for quantitative characterization like that in music is lacking (Ho & Ho, 2007, p. 64)”. Their denial of the method of five-point-scale tonal transcription may be excessive, but their argument that lexical tones should be dynamic rather than static is a valid point.

In addition to the contextual effects such as coarticulation, focus, etc., the intonation effects that we investigated in the present thesis, also clearly contribute to the dynamic status of lexical tones in continuous speech. In our quantitative analysis, as we have a large sample of data, the contextual effects of coarticulation and focus may be considered counterbalanced and can thus be neglected, and we only focus on the intonation effects on tones in different speech styles and different sentence types here.

1.1.2 Cantonese lexical tones

Cantonese has a rich lexical tone system. According to the western linguistic practice, where tone only focuses on pitch distinction, there are six lexical tones in Cantonese, as listed in Table 1-2. In traditional Chinese linguistics, which also takes coda type into account, it is generally accepted that there are nine tone categories or “sheng”, as shown in Table 1-3.

Table 1-2 Cantonese lexical tone categories in western linguistic practice

Category	Description	IPA Transcription	Numeral Transcription	Example	
T1	High-level /High-falling	˥/˥˧	55/53 (5)	夫 [fuː$^{55/53}$] ‘husband’	福 [fʊk^{5}] ‘happiness’
T2	High-rising	˨˥	25	苦 [fuː^{25}] ‘bitter’	鈪 [aːk^{25}] ‘bracelet’
T3	Mid-level	˧	33 (3)	富 [fuː^{33}] ‘wealthy’	霍 [fɔːk^{3}] ‘a family name’
T4	Low-level / Mid-low-falling	˨˩/˩	21/11	符 [fuː$^{21/11}$] ‘symbol’	
T5	Mid-rising	˨˧	23	婦 [fuː^{23}] ‘woman’	
T6	Mid-low-level	˨	22 (2)	父 [fuː^{22}] ‘father’	服 [fʊk^{2}] ‘obedience’

Table 1-3 Cantonese lexical tone categories in the Chinese tradition

Category		IPA Transcription	Numeral Transcription	Western category	Example
Yin-Ping	陰平	˥/˥˧	55/53	T1	夫 [fuː$^{55/53}$] ‘husband’
Yang-Ping	陽平	˩/˨˩	11/21	T4	符 [fuː$^{21/11}$] ‘symbol’
Yin-Shang	陰上	˨˥	25	T2	苦 [fuː^{25}] ‘bitter’
Yang-Shang	陽上	˨˧	23	T5	婦 [fuː^{23}] ‘woman’
Yin-Qu	陰去	˧	33	T3	富 [fuː^{33}] ‘wealthy’
Yang-Qu	陽去	˨	22	T6	父 [fuː^{22}] ‘father’
Gao-Yin-Ru	高陰入	˥	5	T1	福 [fʊk^{5}] ‘happiness’
Di-Yin-Ru	低陰入	˧	3	T3	霍 [fɔːk^{3}] ‘a family name’
Yang-Ru	陽入	˨	2	T6	服 [fʊk^{2}] ‘obedience’

The divergence of the categorization by Western linguistic practice and traditional Chinese linguistics is mainly related to the syllable coda types. In terms of tone or “sheng” category, the divergence lies in the Ru-tones. As discussed in Section 1.1.1, there were four tone categories – Ping, Shang, Qu, and Ru – in Middle Chinese, and the Ru-tones are different from the other three tones in that they have stop codas and sound shorter. In many modern dialects of Chinese, the stop codas of Ru-tones in Middle Chinese have been lost and the Ru-tones have merged with other tone categories segmentally, such that Ru-tones no longer exist in these dialects. In modern Cantonese, however, the Ru-tones are well preserved.

The syllables of Cantonese can be divided into two types according to the syllable codas: (1) syllables with a stop-coda of /p t k/, and (2) syllables with a sonorant-coda, i.e., ending in nasals, semi-vowels, or nuclear vowels. In our present study, the syllables of type (1) are referred to as “stopped syllables”, and those of type (2) are called as “non-stopped syllables”. The stopped syllables in modern Cantonese have Ru-tone origin in Middle Chinese. The Western linguistic practice treats the difference of stopped and non-stopped syllables as segmental, not tonal.

Here we only focus on the factor of pitch and analyze the tone system of Cantonese according to the Western practice. In terms of pitch height, the Cantonese tone system is rich in levels. According to the “Multi-Register and Four-Level” (RL) tonal model proposed by Zhu (2012), the tone system of Cantonese is located at the middle register. It does not use other phonation types to spread to the high and low registers. Among the tone systems described in Zhu (2012), Cantonese is the one that most fully utilize all the four levels within the middle register. The richness in pitch levels may be an important factor for tones to be preserved in singing. We will have more discussion on this topic in Section1.3.2.

In terms of pitch contour, there are two rising tones (the high-rising tone T2 and the mid-rising tone T5), two level tones without a falling variant (the mid-level tone T3 and the mid-low-level tone T6), and two tones with both level and falling variants (T1 can be either high-level or high-falling and T4 can be either low-level or mid-low-falling). (Chan, 1987a, 1987b; Chao, 1930/2006, 1947; Cheung, 2007; Hashimoto, 1972; Ho & Ho, 2007; Jones & Woo, 1912; Kam, 1977; Vance, 1977).

Actually it was not totally uncontroversial whether high-level and high-falling should be regarded as two variants of the same tone T1 or should be treated as two distinctive tones. In the 1960s, at least for some speakers in Guangzhou, these two tone shapes could distinguish lexical items (Zong, 1964), such as “孫 (grandson)[syːn^{55}]” versus “酸 (sour)[syːn^{53}]”. These speakers regarded syllables with high-level and high-falling pitch contours as distinctively different lexical items. Thus, for these speakers, high-level and high-falling are distinctive and there should be seven tones in their tonal system. During the same decade, it was reported that people in Hong Kong tended to use only the high-level variant for T1 (Cheung, 1969).

According to our own observation, most people in Guangzhou nowadays do not have phonemic distinction between high-level and high-falling any more: these two tone shapes are so called free variants of T1 in present-day Guangzhou Cantonese. To be more exact, these two variants are not totally free, i.e., they still conform to the only sandhi rule in Cantonese – the high-falling variant becomes high-level if it occurs immediately before another T1 (Chao, 1947; Gandour, 1981; Hashimoto, 1972; Kao, 1971). In present-day Hong Kong, the overwhelming realization of T1 is the high-level variant, similar to several decades ago. The usage of T1 variants is also one of the important differences between Guangzhou Cantonese and Hong Kong Cantonese.

The tone categories listed in Table 1-2 represent the standard pattern of Cantonese lexical tones. Several pairs of tones were reported to have a tendency to merge, especially T5 and T2, T5 and T3, T3 and T6, T6 and T4 (Bauer, Cheung, & Cheung, 2003; Cheung, 2002; Fok Chan, 1974; Fung & Wong, 2011; Jin, 2009, 2010; Kei et al., 2002; Killingley, 1993; Li, 2008; Mok & Wong, 2010a, 2010b, 2010c; Shao & Xian, 2004; So, 1996; Vance, 1977; Whitaker, 1955, 1956; Wong, 1982; Wong, 2008; Zhang, Peng, & Wang, 2011). Tone mergers in Cantonese are still on-going and the patterns and extent of the mergers for different individuals exhibit a complicated picture. None of the tone mergers reported in the previous literature has completed. In our present study, we still adopt the six-tone system. We also conducted a screening test of tone mergers before asking native informants to participate in our experiments. If an informant was found to have severe tone mergers, we did not continue the experiment on him/her.

Different from the Western linguistic tradition, in the Chinese linguistic tradition the features of coda and duration differences are taken into account for defining tone categories. The Western tone categories of T1, T3 and T6 are divided into two tone categories separately according to the type of syllable in Chinese traditional terms: non-stopped syllables belong to non-Ru tone, while stopped syllables are of Ru-tone. Accordingly, T1 is divided into Yin-Ping (陰平) and Gao-Yin-Ru (高陰入); T3 is divided to Yin-Qu (陰去) and Di-Yin-Ru (低陰入); and T6 is divided into Yang-Qu (陽去) and Yang-Ru (陽入). In the column of "Example" in Table 1-2, there are two sub-columns, with the first sub-column providing the examples of non-Ru-tones, and the second sub-column providing the examples of Ru-tones.

For the duration pattern of Ru-tones, as the stop coda turns the syllable into silence immediately, it follows that the f_0 trajectory of syllables with Ru-tones are much shorter than syllables with non-Ru-tones. Zhang (2006, 2008) further suggested that in continuous speech, after the stopped-syllables there is more acoustic blank and/or the onset of the following syllable is lengthened. If the extra acoustic blank and the lengthened part of the onset of the following syllable are reckoned as part of the duration for the stopped-syllables, the duration pattern of stopped-syllables is similar to the non-stopped-syllables. In this way, the prosodic feature of syllable isochrony[1] is not affected by the superficial shortness of Ru-tones in continuous speech of Cantonese.

To indicate the short duration of the actual pitch contour, a practice of the numeral transcription for Ru-tones with level contours is to use only one digit rather than two to represent the tone values, such as "福[fʊk^5]", "霍 [fɔːk^3]", "服[fʊk^2]". The one-digit representation is only applicable to the level tone shape. If the pitch contour shape is not level for the stopped syllables, we still use two numbers to represent the tone values, e.g., the high-rising stopped syllable "鈪(bracelet)" is transcribed as [aːk^{25}].

The example of "鈪(bracelet)" is important in that it leads to two controversies related to Cantonese tones: the first is about the changed tones or modified tones (變音, in some previous literature it was romanized as *pinjam*), and the second is about the exception to the nine-tone system in the Chinese linguistic tradition.

[1] According to different isochronous units of speech, languages can be classified into the rhythmic groups of either stress-timed or syllable-timed. Chinese is generally accepted as a kind of syllable-timed language, in which syllables recur at regular intervals.

There is a lot of literature that mentions the phenomenon of changed tones in Cantonese (Bauer & Benedict, 1997; Chao, 1947; Chen, 2000; Cheung, 1986; Cheung, 1969; Hashimoto, 1972; Kam, 1977; Matthews & Yip, 1994; Yip, 1990; Yu, 2007; Zong, 1964). The changed tones in Cantonese are morphologically derived. They usually signal familiarity or diminutiveness, or denote the nominalization of a verb, or indicate various other complicated semantic nuances. There are two kinds of phonological output of the changed tone: one is high-level (with tone value of 55), and the other is high-rising (with tone value transcribed as 25 or 35 in the literature).

According to Yu (2007), the morphologically derived high-rising tone and the lexical high-rising tone are in the status of near merger, i.e., the pitch-measuring-points of changed tone are higher than the lexical tone in production, but they have no distinction in perception. However, among the literature on Cantonese changed tones, it is a norm that the difference between the morphologically-derived and the lexical high-rising tones are neglected. In our present study, the morphologically derived high-rising tone is treated as T2.

It was generally assumed that the lexical high-rising tone only apply to non-stopped syllables and they all belonged to the Yin-Shang tone in the Chinese linguistic tradition. However, the stopped syllables can also acquire the high-rising pitch contour for morphological reasons, e.g.: "鴨 (duck) [aːp^{3}] -> [aːp^{25}]", "盒(box) [hɐp^{2}] -> [hɐp^{25}]". For lexical items such as "鈪(bracelet) [aːk^{25}]", historically it should have got the status that the high-rising pitch contour must have been derived by way of the morphological change, but it has gone through lexicalization and nowadays most Cantonese speakers cannot tell the origin tone of it, i.e., the high-rising pitch contour of "鈪(bracelet)" is no longer a morphologically changed tone but a real lexical tone now. In Western linguistic practice, there is no problem at all that "鈪(bracelet)" belongs to T2, but in terms of Chinese linguistic tradition, it is an exception to the nine-tone system and we cannot find a tone category for it. Cases like "鈪(bracelet)" show that the six-tone system of the Western linguistic practice is more appropriate for present-day Cantonese.

Apart from the canonical form, some of the literature also mentions contextual effects on Cantonese lexical tones in continuous speech. Hashimoto (1972, p. 93) observed that:

> "While in normal speech, the chain of individual tones is observed to be linked together by regions of transitions, where the onset and coda of the tones are modified, each according to its immediately preceding and following tones. For example, the rising tones in context tend to have a greater fall at the onset if preceded by higher frequency tones and a more prominent rise if followed by the same, and the height of the fall of both the low falling and the low level tones is in direct variation with the frequency of the preceding tone, moreover, they acquire a slight rising coda if followed by a higher frequency tone."

Chang (2003) further studied the coarticulation effects on Cantonese lexical tones. The experimental results showed that:

> "Anticipatory and carryover effects on slope are both assimilatory in nature but on height are dissimilatory in nature. Both anticipatory and carryover effects on height encompass the whole of the duration of the target tones. An anticipatory effect on slope is minimal and extends for only the last quarter of the target tone's duration, whereas a carryover effect on slope extends for not more than the first half of the target tone's duration. The carryover effect is greater than the anticipatory effect. The contrastive pattern of the six individual tones remains despite the extensive phonetic variation in height and slope (Chang, 2003, p. 29)."

In our present study, we will focus on the intonation effects on Cantonese lexical tones and will not go into the coarticulation effects. However, we should keep in mind that the coarticulation effects may interact with the intonation effects on the final realization patterns of lexical tones.

1.2 Tone and intonation

1.2.1 An overview of intonation, especially declination

The definition of intonation has been full of ambiguity. Bolinger (1978) characterized intonation as a 'half-tamed savage'. From previous works on intonation, we can summarize several consensus points for a definition of intonation:

Firstly, intonation involves suprasegmental phonetic features. Pitch is universally acknowledged to be the primary parameter of intonation (Cruttenden, 1986; Gussenhoven, 2004; Hirst & Cristo, 1998; Ladd, 1996; Pierrehumbert, 1981; Pike, 1945; Tench, 1996; Wells, 2006; Yuan, Shih, & Kochanki, 2002). Some other terms may also relate to suprasegmental phonetic features, such as prosody. Hirst and Cristo (1998, p. 3) wrote, “The term intonation has often been used interchangeably in the literature with that of prosody.” It is generally accepted that prosody is a broader concept and intonation is a component of prosody (Ma, 2007). Apart from the parameter of fundamental frequency, other parameters such as intensity and duration are included as the physical correlates of intonation according to some linguists (Beckman, 1986; Hirst & Cristo, 1998; Ladd, 1996). In our present study, we adopt the narrower definition of intonation and only conduct investigation on the parameter of fundamental frequency.

Secondly, intonation refers to the pitch variation at the utterance level above the levels of syllable and word. This criterion is used to differentiate intonation from tone. Some authors refer to this feature as “postlexical”(Hirst & Cristo, 1998; Ladd, 1996), “supralexical” or simply “non-lexical”(Hirst & Cristo, 1998). Intonation and tone are different in domain and in the type of meanings conveyed. Tone applies to lexical items while intonation applies to “phrases or utterances as a whole, such as sentence type or speech act, or focus and information structure (Ladd, 1996, p. 7)”. The meanings conveyed by intonation are often less concrete (Cruttenden, 1986, p. 10).

Thirdly, pitch variation at the utterance level may convey linguistic or paralinguistic information. Wells (2006, p. 11) summarized the functions of intonation, including the attitudinal function, the grammatical function, the focusing (also called accentual or informational) function, the discourse (or cohesive) function, the psychological function, the indexical function (i.e., intonation may act as a marker of personal or social identity). Among these linguistic and paralinguistic functions, the linguistic functions in terms of structured features are the main concern of linguists. Ladd (1996, p. 8) emphasized that:

> “intonational features are organized in terms of categorically distinct entities (e.g. low tone or boundary rise) and relations (e.g. stronger than / weaker than). They exclude ‘paralinguistic’ features, in which continuously variable physical parameters (e.g. tempo and loudness) directly signal continuously variable states of the speaker (e.g. degree of involvement or arousal).”

One point that should be noted in the definition by Tench (1996, p. 2) is that he excluded singing from linguistic uses:

> "intonation is the linguistic use of pitch in utterances. By saying linguistic, we hope to avoid reference to other uses of pitch such as singing, and to subjective, aesthetic evaluations about how 'nice' and 'pleasant' (or 'ugly' and 'unpleasant') an accent's intonation is."

In our present study, "linguistic" is taken in a broader sense, which not only refers to properties of ordinary speaking, but also includes properties of singing. Singing is regarded as a special kind of speech, and it can be juxtaposed with speaking for comparison. Our investigation revealed that singing and speaking in Cantonese have structured differences which are non-lexical. According to Gussenhoven (2007, p. 253), intonation refers to "the structured variation in pitch which is not determined by lexical distinctions". It can be inferred that the speech styles of speaking and singing have different intonation effects, which are linguistically structured.

In previous literature on intonation studies, there have been various approaches and models. In traditional auditory methods, the "impressionistic" or "proto-phonological" approach was usually adopted by linguists and language teachers who were interested in describing intonation either for pedagogical purpose or for the general development of intonational phonology. In this approach, intonation is presented as impressionistic pitch curves. The disadvantage of this approach is the lack of agreement on the inventory of categorically distinct elements. There is no objective instrumental evidence for settling such disagreement. In this sense, the instrumental approach is more scientific, or, at least, more objective.

Among the intonation models proposed in previous studies, there are two basic types called the "Contour Interaction" type and the "Tone Sequence" type (Hirst & Cristo, 1998; Ladd, 1983). In the Contour Interaction type, the intonation contour is seen as "the result of superposing on each other pitch configurations of different sizes (Hirst & Cristo, 1998, p. 12)". In the Tone Sequence type, "the pitch movements associated with accented syllables are themselves what make up sentence intonation (Ladd, 1983, p. 40)".

It can be found from an exploration of previous literature that the Contour Interaction type is used more often in the intonation studies of tonal languages, because the pitch curves in tonal languages have to be decomposed into the tone layer and the intonation layer. Actually some famous studies on Chinese intonation are in the model of Contour Interaction type. In all the studies of Chao (1932/2006, p. 123), Chao (1956/2006, p. 598) and Chao (1968, p. 39), it was mentioned that the actual pitch movement of Chinese speech is the algebraic sum of tone and intonation. Chao (1968, p. 39) compared the relationship between tones and intonation as ripples riding on top of waves. Chao (1933/2006) suggested that tone and intonation can be combined through "addition" in two ways. The first is "successive addition", where a rise or a fall is added to the end of an utterance; this is similar to a boundary H or L in English. The second is "simultaneous addition", where intonation is superimposed on lexical tones, while the lexical tones remain distinctive. The simultaneous addition is also observed in the phonetic studies of Shen (1989) and He and Jing (1992). As far as our present study is concerned, the proposed model is also of the Contour Interaction type – the fundamental frequency curves were factored out as interacting components of lexical tones and intonation.

Meanwhile, the Tone Sequence type applies more to non-tonal languages, as pitch in non-tonal languages primarily signifies intonation. Although occasionally the Tone Sequence model has been seen in the intonation studies of tonal languages, these intonation models are borrowed from the studies of non-tonal languages. There are two important intonation models of the Tone Sequence type: the IPO approach and the autosegmental-metrical (AM) approach. The IPO approach first appeared in the mid-1960s at the Institute for Perception Research (IPO) in Eindhoven. It was originally motivated by the search for a model of the intonation of Dutch for use in speech synthesis, and later developed into a general theory of intonational structure. In the IPO approach, which is a phonetically accountable approach to intonation, contours are idealized as sequences of pitch movements and connecting line segments. The IPO approach has its roots in the work on speech perception, which assumes that certain pitch movements are interpreted as relevant by the listener (Ladd, 1996). Unlike the IPO approach, the AM approach (since the late 1970s) grew out of theoretical problems in phonology. Liberman (1975), Bruce (1977) and Pierrehumbert (1980) were the representative works in the AM approach. Ladd (1996, p. 42) summarized the four basic tenets of the AM approach to intonation: (1) linearity of tonal structure; (2) distinction between pitch accent and stress; (3) analysis of pitch accents in terms of level tones; (4) local sources for global trends.

Based on the AM approach, a framework called ToBI (Silverman et al., 1992) has been developed for labeling intonation. In addition to an inventory of tones, restricted for English to H(igh) and L(ow), ToBI makes use of diacritic symbols, e.g., "H%"represents a boundary tone, "H*" represents the "strong" tone of a pitch accent, and "H-" represents a phrase accent. In addition to English, ToBI has also been applied to other non-tonal languages such as German (Grice, Baumann, & Benzmüller, 2005) and Greek (Arvaniti & Baltazani, 2005), and even tonal languages such as Mandarin (Peng et al., 2005) and Cantonese (Wong, Chan, & Beckman, 2005). Although ToBI is a general framework for developing intonation models in different languages, the details of each ToBI system are language-specific as every language has its own characteristics of intonation.

As the model we propose in the present study is a matrix composed of two dimensions of pitch domain and pitch configuration, below we will also have a brief review on these two aspects.

For pitch domain in non-tonal languages such as English, tonality, tonicity, and tone are three important concepts. Tonality is concerned with how to break the material up into chunks (Wells, 2006, p. 6), i.e., the units of intonation (Tench, 1996, p. 31). Tonicity is concerned with what is to be accented within the intonation units (Wells, 2006, p. 6), i.e., the focal point of intonation units (Tench, 1996, p. 53). Tone is the pitch contour which is realized on the accented point.

For tonal languages, intonation is mainly of the superimposing kind. Cruttenden (1986, p. 10) mentioned two different domains: the first is when "the pitch level of the whole utterance may be raised or lowered"; and the second is when "the final tone of the utterance may be modified in various ways". When describing the characteristics of unemphatic yes/no questions, Hirst and Cristo (1998, p. 25) summarized that they have global, local or recurrent features compared with the declarative utterances. The global feature refers to the raising of the pitch of all or part of the utterance, while the local feature is a high final pitch. Fox, Luke, and Nancarrow (2008) explicitly divided the domain of intonation into the utterance-body intonation and utterance-final intonation. As

Fox, Luke, and Nancarrow (2008) mainly focused on Cantonese intonation, we will have further discussion on their work in Section 1.2.2. In the model we propose in the present study, intonation operates on two levels – the syllable level and the utterance level. An utterance divides into two parts – the utterance-body portion and the utterance-final portion. Utterance-level intonation operating at both parts of has been extensively investigated in previous studies as introduced above. The intonation effect on the syllable level has been neglected in previous studies but is emphasized in our present study. Details of our framework will be shown in Section 3.1. If an utterance is monosyllabic, the syllable level and the utterance level coincide with each other.

Pitch configuration in our framework is analyzed in terms of three factors – pitch slope, pitch register and pitch span. Pitch slope involves the direction of the pitch contour (whether it is falling/rising/level) and the gradient of the direction (whether the fall/rise tendency is significant). The most studied pitch slope is the universal declination intonation, which will be illustrated in more details later. Another extensively studied pitch direction is the rising tendency of interrogatives. Hirst and Cristo (1998, p. 25) reviewed the intonation investigations of different languages in their edited volume, and found that in US English, Swedish, Brazilian Portuguese, Finnish, Hungarian, Western Arabic, Vietnamese and Thai, the interrogative utterances have a raising of the pitch of all or part of the utterance. Bolinger (1978) reported that about 70% of a sample of nearly 250 languages were said to use a final-rising tone to signify questions and that the remaining 30% used a higher over-all pitch for questions than for non-questions.

While pitch slope corresponds to the property of pitch contours, pitch register and pitch span are used to characterize pitch levels. Ladd (1996, p. 260) pointed out that:

> "'pitch range' is not a single variable. Inspection of comparable F_0 data from different speakers reveals two partially independent dimensions of variation: differences of overall level, and differences in what I propose to call span (the range of frequencies used)."

The term "pitch register" used in the present study corresponds to the dimension of "overall level" in Ladd (1996, p. 260). It is the relative pitch height of a syllable or an utterance. We avoid using the term of "pitch level" here because "level" may be easily misunderstood as the tone shape of level tones. Meanwhile, the term "pitch span" by Ladd (1996, p. 260) is adopted in our present study, which refers to the range of the frequencies spanned.

An important feature of pitch register and pitch span is their relative nature. Ladd (1996, p. 252) wrote:

> "There are two obvious ways of thinking about the relative nature of pitch features: they may be relative to the speaker's voice, or they may be relative to other parts of an utterance. For reasons that will become clear presently, I will refer to these two approaches as normalising and initialising respectively."

We also pay attention to the relative nature of pitch. In Chapter 3 and Chapter 4, the calculation of pitch slope, pitch register and pitch span is based on the Z-Score, which has the individual differences normalized. As each of the pairs of speech styles or sentence types being compared are with the same text, each pair can be regarded a kind of minimal pair, we therefore did not carry out the initializing procedure.

Below we will focus on the declination intonation. Declination is the most widespread and widely studied intonation (Beckman & Pierrehumbert, 1986; Duanmu, 2000; Gussenhoven & Rietveld, 1988; Huang, Yang, & Lü, 2007; Ladd, 1984; Liberman & Pierrehumbert, 1984; Lindau, 1986; Pierrehumbert, 1979; Prieto, Shih, & Nibert, 1996; Shih, 1997; Strik & Boves, 1995; Ting, 1971/1984; Wang, Chen, & Lu, 2004). It is known as the general pitch lowering over a syntactic unit such as a sentence (Ladefoged, 1975/2006, p. 252). Most languages are reported to have declination intonation, whether it is a tonal or a non-tonal language.

Perception studies showed that listeners unconsciously expect this declination intonation in daily speech and automatically normalize it: if two peaks are perceived as equal, the second one actually have lower f_0 value; or for two peaks of equal f_0 value, the later one is interpreted as having higher prominence than the first one (Collier, 1990; Gussenhoven & Rietveld, 1988; Pierrehumbert, 1979). Thus, listeners consider their speech kept at the same pitch level, whereas the fact is that it is declining.

The physiological causes of declination have been controversial (Cruttenden, 1986; Ladd, 1984; Liberman, 1967; Ohala, 1978; Strik & Boves, 1995). It has not been answered definitively whether it is automatic and caused by reduced subglottal pressure that accompanies expiration, or whether it is controlled by laryngeal muscles and not by subglottal pressure. The "automatic reduced subglottal pressure" view as suggested by Cruttenden (1986, p. 168) has been more popular:

> "the explanation for declination has often been related to the decline in transglottal pressure as the speaker uses up the breath in his lungs. A more recent explanation suggests that an upward change of pitch involves a physical adjustment which is more difficult than a downward change of pitch, the evidence being that a rise takes longer to achieve than a fall of a similar interval in fundamental frequency."

The "actively controlled" view has also found experimental support. Ohala (1978, p. 32) argued that the declination effect is due to active laryngeally caused changes in vocal cord tension, and it is a purposeful rather than an "automatic" effect, because the gradual pitch decrement in utterances serves a useful linguistic purpose in signaling clause and sentence boundaries.

There are several terms related to pitch downtrend of utterances: declination, downstep, downdrift, and final-lowering.

Crystal (2008, pp. 157-158) discussed the terms of downstep, downdrift, and declination under the entry of "downstep". He described "downstep" as:

> "a term used in the PHONOLOGY of TONE languages, referring to a lowering process which applies to the second of two high-tone SYLLABLES. A downstepped high tone would be slightly lower than the preceding high tone, but not so low as to be equivalent to a low tone. The process has been widely observed in African languages. …Downstep is phonologically CONTRASTIVE, and is usually distinguished from downdrift, a sequential process whereby high tones after low tones become progressively less high throughout an intonational unit. These effects have been described more generally as 'register lowering' or 'key lowering'. Declination is often used as an equivalent for downdrift, but this term also has a more general phonetic use ('F_0 declination'), referring to a gradual descent of pitch level and narrowing of pitch range throughout an utterance, partly as a result of reduction in subglottal air pressure, as speakers use up the breath in their lungs. Such effects, of course, are not restricted to tone languages."

The definitions of declination, downdrift and downstep are similar in Connell and Ladd (1990) and Connell (2002): declination refers to the phonetic feature of continuous lowering from the beginning to the end of the intonation unit; downdrift is an iterative lowering of successive high pitches within an intonation unit separated by intermediate low points; downstep is an iterative lowering of successive high pitches within an intonation unit without intervening low tones. Connell and Ladd (1990) and Connell (2002) also mentioned another term of "final lowering", which is a rapid lowering occurring before the final boundary of an intonation unit.

In our present study, we use the term "declination" for the global pitch downtrend across an utterance as well as the local falling tendency of an individual syllable; it is therefore a broader concept than usual. The term "downdrift" is also a kind of declination but is more specified for the overall downward terracing pattern of intonation. We prefer the term "downdrift" rather than "downstep" in our present study because the overall pattern is not like level steps but like steps each with declination of its own (Fox, Luke, & Nancarrow, 2008)

Among the extensive literature on declination intonation, researchers have taken it for granted that the declination of utterances is natural and right and have never considered whether there exists a non-declination counterpart. Declination was alone without a reference object. If we limit our research only to normal speech, as usual, declination will probably continue to be alone. However, if we broaden our horizons and take other speech styles such as singing into our research field, we can immediately find a reference object for declination and thus make the previous research problems on declination much easier to understand. This is what we have done in our present study.

1.2.2 Tone and intonation in Cantonese

From the introduction in Section 1.1.2, it can be seen that Cantonese has a rich lexical tone system. How pitch is manipulated between tone and intonation in Cantonese has aroused the interest of many linguists. For the intonation patterns in Cantonese, the most studied problems include the declination intonation in declaratives and the final-rising intonation of interrogatives. Cantonese also has a rich system of sentence-final particles (**SFPs**, henceforth), which can also convey many kinds of subtle sentential connotations and thus share some functions of intonation. The presence or absence of SFP may also influence the final realization pattern of intonation. Thus, the relationship between SFP and intonation is also worth discussing.

Fox, Luke, and Nancarrow (2008) explicitly put forward the theory that the system of intonation for Cantonese comprises two main components: utterance-body intonation and utterance-final intonation. These two components will also be referred to in the following discussion on the intonation patterns of declaratives and questions.

As pitch in connected speech has to signify both lexical tones and intonation in Cantonese, a convenient method to study intonation pattern is to compare the pitch of same-tone syllables located at different positions of an utterance. It is observed that the pitch for the same lexical tone is successively lowered across an utterance in Cantonese (Bauer & Benedict, 1997; Fox, Luke, & Nancarrow, 2008; Lam, 2002; Li, Lee, & Qian, 2002; Vance, 1976; Wong, 1999; Wong & Diehl, 2003), which conforms to the universal declination intonation of most languages.

Apart from the overall trend of declination over the utterance, another important feature of the utterance-body intonation is the pitch upstep at syntactic boundaries. This kind of pitch reset can exert its syntactic functions for parsing. It can help us identify grammatical structures in speech, i.e., the pitch upstep marks the beginning of new grammatical units such as phrases and clauses. To study the pitch reset in Cantonese utterances, Chow (2006) designed a set of stimuli, for which two different interpretations can be derived by conveying prosodic boundaries at different syntactically motivated locations. He found that there was obvious upward adjustment at syntactic boundaries.

The overall declination trend and the upstep at phrase boundaries together make a typical downdrift terracing pattern for the declarative utterances of the same-tone-sequences (Fox, Luke, & Nancarrow, 2008; Li, Lee, & Qian, 2002): pitch keeps going downwards within a phrase, then goes upwards to reset at the syntactic left-boundary (though the pitch peak is still lower than the preceding phrase), and continues falling down until a new phrase starts. Figure 1-2 provided a depiction of the above description.

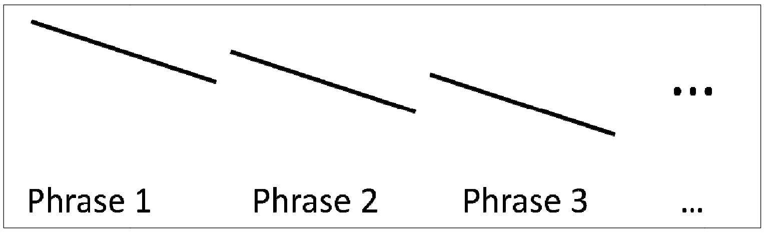

Figure 1-2 Schematic diagram depicting utterance-body downdrift

Utterance-final intonation (or boundary tone) mainly affects the pitch of the last syllable in an utterance. In Cantonese, it mainly exerts an effect on utterances without sentence final particles. Like most languages, Cantonese has a final lowering effect at the end of declarative sentences (Fox, Luke, & Nancarrow, 2008; Lam, 2002; Vance, 1976). The experimental investigation by Vance (1976) found that the sentence-final tone lowering in Cantonese makes the intelligibility somewhat weaker sentence-finally than sentence-medially. However, he also showed that the tonal distinctions are not neutralized in final position. Fox, Luke, and Nancarrow (2008) observed that the "level tones (55, 33, 22)" are in fact slightly falling at final position in "neutral" utterances. For "assertive" utterances there is a stronger falling tendency, which widens the fall of falling and level tones, and levels out rising tones or narrows their range. We will also further investigate the final lowering effect in our present study.

For the interrogatives in Cantonese, the most outstanding feature is the final rising pitch contour (Lam, 2002; Ma, 2007; Ma, Ciocca, & Whitehill, 2004, 2006a, 2006b, 2011; Mai, 1998, 2000; Xu & Mok, 2011). This final rising intonation may even neutralize the tonal differences. The acoustic investigations in previous studies (Lam, 2002; Ma, 2007; Ma, Ciocca, & Whitehill, 2004, 2006a, 2006b; Xu & Mok, 2011) reveal that the final syllable of a question was modified by sentence intonation to a rising contour regardless of the original contour. In the perception test carried out by Mai (2000), he posited minimal pairs of T2~T3, T3~T6, T5~T6, T2~T5 and T2~T6 in the final position of several authentic sentences. These sentences were presented to the participants as stimuli. The participants were asked to judge which lexical item they heard. With the follow-up statistical analysis, Mai (2000) concluded that T2~T3, T3~T6 and T5~T3 could be at least partially differentiated, and T2~T5, T2~T6 and T5~T6 tended to be confused.

Previous studies on the interrogatives in Cantonese were mainly concerned with utterance-final intonation. Seldom did they include a study of the utterance-body intonation. In fact, we should also be curious about the utterance-body intonation of the interrogatives – whether they are also affected by the declination effect. In our present study, this problem will also be tackled through acoustic experiments.

There are a large number of sentence-final particles in Cantonese. According to Luke (1990, p. 1), previous work has identified about thirty utterance particles in Cantonese. As these particles can be used in combination, the actual number of particles (simple and compound) currently in use in Cantonese should be in the region of a hundred. Luke (1990, p. 3) also described the distinctive features of this word class, including: (1) they have no semantic content; (2) they serve to indicate the mood of a sentence; (3) they are used to express attitudes and emotions; (4) they are attached (as boundary forms) to the end of sentences.

As sentence-final particles can also convey many kinds of subtle intonation-type meaning, they are closely related to intonation. Wakefield (2010) examined the English intonational equivalents of four Cantonese SFPs. He found that each SFP translates consistently into English as a specifically-shaped pitch contour by Cantonese/English native-bilingual participants, regardless of the discourse context or the syntactic structure of the sentence. This study confirmed that the segmental SFPs in Cantonese and the suprasegmental intonation in English share the same function and meaning.

Cheung (1986, p. 251) pointed out the interrelationship among tone, intonation, and SFPs: "Tone shares the form and SP[2] shares the content of intonation." Both tone and intonation make use of pitch as the main phonetic manifestation, while SFP and intonation fulfill similar function to express sentential connotations. Cheung (1986, p. 256) further put forth two exploratory views. Firstly, "the intrinsic pitch shape of S[F]P is not the same thing as tone: it is intonational." This observation builds up a bridge between SFP and intonation such that SFP and intonation are no longer separated. Secondly, "S[F]P may be segmentless." This may initially seem hypothetical, as traditionally SFP are

[2] The "SP" in Cheung (1986) is the abbreviation for "sentence final particle", henceforth.

regarded as a kind of function word, which should be composed of a segment or a sequence of segments. However, we find the concept of "segmentless SP" reasonable as it helps us to connect SFP with intonation. The present study also intends to provide some phonetic evidence to the two viewpoints in Cheung (1986, p. 256). Furthermore, from the experiment described in Chapter 4, we found that the existence of an SFP in the utterance will obviously affect the utterance-final intonation. It is worthwhile to conduct systematic comparison between the utterances with and without an SFP.

1.3 Tone and melody

1.3.1 An overview of singing in tonal languages

Singing can be regarded as a special style of speech. The "singing speech" we discuss in the present study is limited to that with concrete lexical items rather than lyric-less hums or exclamations. Compared with speaking speech, singing speech has its own characteristics, such as a particular voice quality (Sundberg, 1977, 2001). Meanwhile, singing speech also preserves the linguistic features and constraints of a certain language. Apart from the segmental features of vowels and consonants, the prosodic features of a certain language can also be reflected, such as the pitch patterns and duration patterns. For example, singing reflects patterns of durational contrast between successive vowels in spoken sentences, as well as patterns of pitch interval variability in speech of British English and French (Patel, Iversen, & Rosenberg, 2006).

In terms of pitch, it is self-evident that in singing, people should follow a certain kind of musical scale; while in speaking people need not to do so. Whether or not to follow a musical scale is a universal distinction between singing speech and speaking speech. This is applicable to all the languages in the world, including tonal languages and non-tonal languages.

For non-tonal languages, the situation is simple: the pitch shape in singing just needs to conform to the requirement of the musical tune; and the singing speech does not make phonemic changes to the lexical items. In most circumstances, the only thing lost in the singing speech is the speaking tune, or the speaking intonation. Since the musical tune can vary freely, it is meaningless to compare the singing tune and the speaking intonation.

For tonal languages, the lexical tones can be either neutralized or partly/fully preserved in the singing speech. In the cases that tones are totally neutralized in singing, such as modern pop-songs singing in Chinese Mandarin (Chan, 1987a, 1987b; Chao, 1956/2006; Ho, 2006; Pian-Chao, 2000; Schellenberg, 2012b, 2013; Wee, 2007), syllables are only required to obey the melody in singing and people cannot identify the original tone categories of the lexical items. Apart from the loss of tonemic distinction for lexical items, the tone-neutralized singing in tonal languages is not different from singing in non-tonal languages. It is also pointless to compare the intonation effects of speaking and singing without regard of tones.

In our present study, we are more interested in cases where tones are partly or fully preserved in singing. From our introduction in Section 1.1.1, we have seen that the tone systems of the Sino-Tibetan language family of Southeast Asia have different structures and features from American Indian languages and African languages. Our exploration of previous literature shows that all those language families include some tonal languages preserving their lexical tones in singing to some extent. Among these tonal languages, many just partly preserve or mimic their lexical tones in singing, and only a few tonal languages were reported to have high correspondence between speaking tones and song melody. In a certain tonal language, different types of singing tones may also have different degrees of tone-preservation. Below we will briefly introduce what kinds of tonal languages were reported to have tones preserved in singing, and to what extent the tones are preserved. In addition to the overview of the world's tonal languages, we will provide a brief introduction to singing in Chinese dialects.

Schellenberg (2012a) made an overview of singing in tonal languages. He sorted out the percentage of correspondence in songs of nine different languages for which statistical results are available or calculable. Here we reproduce this table (Schellenberg, 2012a, p. 270) in Table 1-4, which provided very useful information to us. The descriptions about this table (Schellenberg, 2012a, pp. 269-270) are also summarized below.

The asterisks in the column of "Language" indicate that the results for those languages have been calculated post hoc from data published in the papers cited. There are three possible transitions from one syllable to the next—up, down or the same. The “Parallel” column in Table 1-4 gives the proportion of instances in which the transition is parallel between the speech melody and song melody – e.g. the speech melody goes up, and so does the song melody. From 48% for Kalami to 92% for both Cantonese and Zulu, in all cases, this is well above the 33.3% chance (using chi-square tests, all have $p \leqslant 0.01$). The column labeled as “Not opposing” in Table 1-4 gives the percentage when transitions that are not directly opposite (i.e., one transition goes up or down and the other remains the same). This strategy seems common in some of the languages.

Table 1-4 Proportions of correspondence between speech melody and sung melody. (Table reproduced from Schellenberg (2012a, p. 270))

Language	Paper	Number of Artifacts	Number of transitions	Parallel	Not opposing
Cantonese Chinese	Wong and Diehl (2002)	4	281	92%	98%
Ewe	Jones (1959)	1	105	68%	95%
Ewe*	Hornbostel (1928)	1	35	49%	89%
Hausa	Richards (1972)	1	380	53%	96%
Kalami (Gawri)	Baart (2004)	14	434	48%	89%
Shona	Schellenberg (2009)	3	140	53%	67%
Thai	List (1961)	8	no data	76%	no data
Wu-Ming Tai	Mark and Li (1966)	6	(320 syll)	63%	no data
Xhosa*	Starke (1930)	25	281	67%	95%
Zulu*	Rycroft (1979)	2	36	92%	97%

Among the languages listed in Table 1-4, some have a considerable number of mismatches between linguistic tones and musical melody. For example, in *rō*, the most popular style of song and poetry in Kalam Kohistani (a tonal language of the northwestern corner of the Indo-Aryan language territory), the investigation by Baart (2004) concluded that:

> "Even though a rising sequence of tones more often than not corresponds to an ascending sequence of musical notes in a song, and a falling sequence of tones more often than not corresponds to a descending sequence of musical notes, yet the number of instances where tones and sung pitches do not match is considerable. (Baart, 2004, p. 14)"

Apart from the languages listed in Table 1-4, there are also other tonal languages reported to have their lexical tones preserved in singing, such as the Igbo language (an African language) studied by Ekwueme (1974a, 1974b), the Chewa language (a language in Central East Africa) and the Navaho language (an American Indian language) studied by Herzog (1934).

Some special types of music in tonal languages reproduce the features of the lexical tones, not only singing. Herzog (1934) reported that in the horn music in the Jabo language (spoken in Western Liberia), the musical style employs a four-tone scale corresponding to the four linguistic tones (or registers):

> "On the horn, as representative of this group, the four language registers are actually reproduced by an equal number of tones. These are the only tones the horn commands, and are produced by leaving it open, stopping it at either end, and half-stopping the larger end. In the following signals, the tones noted are those given by the horn from which they were recorded. … The horn-signals follow patterns of speech too closely to be regarded as musical forms in the proper sense.(Herzog, 1934, pp. 454-456)"

Another example was the sung speech whistled with a leaf in the Akha language (a three-tone language in South East Asia). According to (Meyer, 2007, p. 7):

> "… whistled speech transposes primarily the F0. It has a very rich oral repertoire of songs. …As the aim is to transmit a poetic message, the melody of the songs follows the pitch of their lyrics…. The whistlers also say that whistling is more melodic (aesthetical justification)."

In some tonal languages, in different types of singing performance, there are different degrees of the correspondence between linguistic tones and musical melody. Below are examples from Bantu languages, Thai and the Shona language.

Rycroft (1979, p. 306) put forward that:

> “(in Bantu languages) it seems that exact melodic imitation of the pitch contours of speech throughout an entire song is of very rare occurrence. Also, the degree to which linguistic features influence the music tends to vary considerably, both within one and the same speech-community, and beyond this, even among different items from the same category of song.”

He further provided many examples to illustrate the continua of the tone-melody correspondence. The high-correspondence examples include (in the order from high to low): (1) the personal solo songs in Zulu – “the melodic line matches the total speech contour almost exactly (Rycroft, 1979, p. 307)”; (2) the war-cries (*izaga*, singular, *isaga*), and some items used for the *isigekle* recreational dance – “pitch movement consists entirely of stylized and exaggerated speech contours, rendered in a ‘singsong’ manner like choral recitation, without any fixed musical notes (Rycroft, 1979, p. 308)”; (3) the Zulu *izibongo* (the praise-poetry) – “four recurrent levels of pitch resembling Sol-fa notes *doh’*, *te*, *soh* and *Doh* appear to predominate and serve as a basic tonality in the musical sense (Rycroft, 1979, p. 308)”; (4) “choral speech” of the war-cry type –“melodic contours in place of sentence intonation (Rycroft, 1979, p. 308)”.

Meanwhile, there are also examples of lower-correspondence due to various factors: (1) Zulu and Swazi antiphonal dance-songs – alternative essential-tone patterns are reflected (in speech, many words take one tone pattern when occurring finally but a different one when non-final; while in song, either of these patterns can be adopted) (Rycroft, 1979, p. 310); (2) Zulu communal songs and dance-songs – influenced by the effects of syllable elision and length distortion; (3) Venda children’s songs, songs of Ewe language, Southern Bantu music, children’s game-songs and jingles, Zulu and Swazi *imilolozelo* jingles, ancient Nguni anthems and sance-songs, Zulu *uhawu* victory song – violation of speech tones to different extent.

List (1961, p. 30) also observed that “not all genres of Thai song or recitation show the same degree of coordination of speech melody and song melody.” The coordination of the speech melody and song melody also shows variation in songs of central Thai: there is high coordination in recitations or chants used in the public schools (either traditional or improvised), and in folk songs such as lullabies; there is less coordination in some classical song; there is least coordination in the present day popular song.

The above examples seem to suggest that there is less correspondence as the musical style is more “contemporary” and influenced by Western composition. However, according to the studies on three songs in Shona by Schellenberg (2009, p. 143):

> “both the hymn and the national anthem are written in a more ‘modern’ musical idiom yet they both show higher rates of parallelism than the very traditional story-song.”

It can be concluded that whether the singing genre is modern or traditional is not the determining factor affecting the degrees of correspondence between the speech tune and the singing melody.

It should be noted that for some of the tonal languages, singing not only realizes the tonal distinction, but also preserves some of the prosodic features of the speaking intonation in those languages, especially the declination intonation (which has been introduced in Section 1.2.1).

Leben (1983) offered a reinterpretation of the study on a Hausa song by Richards (1972), as he found that the correspondence shown in Richards (1972)’s data was statistically significant but admitted many deviations. Leben (1983, p. 148) suggested that:

> “for this particular Hausa song, a comparison should be made not between the musical melody and the phonemic tone levels but rather between the musical melody and a normalized intonational realization of the lyrics.”

Ekwueme (1974a, 1974b) described the downward terracing pitch pattern of singing in the Igbo language: the musical phrase –

> “starts at the highest point and gently works its way down to the lowest point. This is not only the shape of Igbo tunes, it is the common shape of melodies of all other Africans south of the Sahara.(Ekwueme, 1974b, p. 339)”.

Starke (1930, p. 46) also observed a similar pattern in Xòsa singing:

> “Xòsa song, like Xòsa speech, has a unit of division. In speech we call this unit a ‘breath-group’; in song, have called it a ‘phrase’. Within this unit there is, in both song and speech, a tendency for the intonation level to be lowered, so that a final high tone (for example) is always lower than an initial high tone.”

Through the observation on a song of the Chewa tribe in Central East Africa, Herzog (1934, p. 457) found that "working side by side with linguistic influences, there is evident a strong musical force: the downward trend of melody". He further pointed out this kind of trend is "so common in primitive music".

The declination shown in the singing of these tonal languages is not consistent with what we found for Cantonese singing intonation. We will have further discussion on this problem in Chapter 5.

As Cantonese is a dialect of Chinese, apart from the general overview of singing in world's tonal languages, here we will especially provide a brief introduction to tones and singing in Chinese dialects.

Chao (1956/2006) described several types of melodic speech styles in Chinese, which partly or fully preserve the tones: (1) singsong (the examples include Mandarin children's songs and Cantonese venders' cries), which is based largely on the phonemic tones of the words, spoken in a stereotyped manner; (2) chanting (the examples are Tang poems chanting in classical education, in Changchow dialect, one of the Wu dialects), which is largely, but not unambiguously (a limited range of variations are allowed) determined by tone; (3) recitative, which is in the traditional form of Chinese drama, and follows largely the tones of a special artificial dialect known as the Zhong-zhou pronunciation or Central pronunciation; (4) tonal composition, such as kunqu (崑曲) style of plays and songs, in which the dependence on word tones was not so close as to fix the melody unambiguously, but suggested and limited the range of possibilities, resulting in somewhat stereotyped forms of melodies; (5) atonal composition[3] (the classical style is still somewhat followed), which is rare but still exists in contemporary Chinese song composition, while "most of contemporary Chinese song composition is independent of tone (Chao, 1956/2006, p. 610)".

Pian-Chao (2000) elaborated Chao (1956/2006) on the various types of relationship between the word tone in Chinese language and the musical tone. In her description of Peking Opera, Pian-Chao (2000, p. 186) mentioned that:

> "when there is an attempt to make the melodic adjustments to the word tones of a few selected words, we often find that syllables in the Even Tones tend to be set to higher, and slower melodic segments, while the

[3] Nowadays atonal music is a cover term used to define music that seems to lack a clear tonal centre. It is evident that Chao's term of "atonal composition" was totally different from the present-day "atonal music".

words in the Oblique Tone category tend to be set to lower and relatively faster moving melodic turns".

She provided an example from "霸王別姬 The King's Farewell", in which she found the upward scooping at the word 愁 (T2 in Mandarin, a rising tone) and the downward sliding at the word 地 (T4 in Mandarin, a falling tone).

On the relationship between speech tones and melody of Beijing Opera, Stock (1999) also conducted detailed investigation. Stock (1999, p. 183) suggested that:

> "the subordination of musical elements to the sounds of the spoken language in Chinese opera is neither automatic nor complete. Even in a genre where language is of unquestioned importance, music-structural considerations may, sometimes, challenge the dictates of speech-tone and lyric structure in the production of a finished musical text. A singer's melodic style may take precedence over his or her setting of speech tone."

Like Pian-Chao (2000), Stock (1999) also adopted the dichotomy of Even Tones versus Oblique Tones.

For the older classical opera Kunqu Opera, the dichotomy of Even Tones versus Oblique Tones is also applicable. Pian-Chao (2000, p. 187) mentioned that:

> "Even Tone words Pyng Sheng tzyh 平聲字 use a generally lower and descending melodic line, while the Oblique Tone words Tzeh Sheng tzyh 仄聲字 are more often given faster and higher notes."

Pian-Chao (2000, pp. 187-188) further pointed out that:

> "if the word tones are indeed taken into account, it still does not necessarily mean that every single word in a line is treated according to some rules. Usually this applies only to some key words, and more often to the ending words in a line."

Liu (1974, p. 65) also noticed the importance of the distinction between the Even Tones and Oblique Tones in Kunqu Opera:

> "These ch'ü p'ái are in reality tone tunes that translate the *p'íng tsè* (even and oblique) tonal sequences of prosody into similar melodic contours of song."

Liu (1974, pp. 73-74) provided a more detailed illustration of the four tones of Chinese Mandarin and the melodic tone contours in Kunqu Opera:

> "It appears that tones 1 and 4 show similar features, the main difference being the addition of one pitch for the latter. The most common version of tone 3 is identical with tone 2 as far as the shape is concerned. However, a relatively higher pitch register for tone 2 distinguishes it from tone 3 in speech. …melody contours in K'unch'ü need not always adhere to the tonal contours of speech (except perhaps initially), but that they have a certain amount of freedom within the concept of p'ing tsê to follow the dictates of musical expression."

According to our own observation, in many other Chinese dialects, lexical tones are also partly preserved in classic opera singing, like Yuju opera in Henan Province and Lüju opera in Shandong Province. Apart from the classical opera, we also explore the folk singing art. We found that there is a kind of Mandarin singing that preserves the lexical tones. It is a ballad singing art called Jingyun Dagu, or "Big Drum" Recitative. In a Jingyun Dagu performance, one singer stands to perform the story-singing and beat the drum, while other accompanists playing the trichord and dulcimer. The four lexical tones of Mandarin are discernible from the singing narratives. Pian-Chao (2000, p. 188) also referred to this "Peking Drum Song Jing Yunn Dah Guu 京韻大鼓", and she observed that the word tones are constantly imitated closely in this kind of performance.

From the above introduction to singing in the world's tonal languages and especially in Chinese dialects, we can see that either lexical tones are not fully preserved in singing, or the resource of singing performance is limited. The case of Cantonese is of strategic importance here, as lexical tones are fully preserved in many kinds of Cantonese songs singing, not only the traditional opera songs, but also innumerable pop songs, children's songs, advertisements songs, etc. In addition to setting Cantonese lyrics to a fixed melody, sometimes the lyrics come out first and the melody is generated somewhat accordingly, such as in some aria types in Cantonese opera (Cheung & Wong, 2008; Yung, 1989). The corpus for studying Cantonese singing is vast and limitless.

There are also some songs sung in Cantonese that ignores the lexical tones of the language. A song without regard to the tones in Cantonese is not regarded as a typical Cantonese song. These songs usually have a foreign or Mandarin origin. As all words in Chinese have Cantonese pronunciation, a song that is written for Mandarin Chinese can be sung segmentally in Cantonese. Hymns offer one example of songs falling into this category. Many hymns were written for singing in Mandarin in the first place, and when they were directly sung in

Cantonese, the tones were neglected. The need to retain the religious content is one reason for such practice. However, nowadays many hymns that used to be sung without tone-melody match have had their lyrics amended to achieve tone-melody match, and many even have more than one version of lyrics. For Cantonese speakers, if a song is without regard to tones, the song sounds unnatural and odd to them. Only when tone-melody match is achieved will the song sound right to native speakers.

1.3.2 Tone-melody matching in Cantonese singing

The studies on Cantonese singing have covered a broad range of research topics, e.g.: the development of Cantonese opera song singing (Wong, 1981) and pop song singing (Wong, 2003), the features and rules of Cantonese opera singing (Cheung, 2012b; Wong, 1981; Yung, 1989), the rhymes in Cantonese singing (Chan, 2010; Cheung, 2012b; Cheung, 1996), etc. It is especially interesting that Cantonese singing does not neutralize the tonal contrast, and Cantonese speakers can identify the original tone categories of lexical items in singing.

The perfect match of the lexical tones and musical tones in singing is an outstanding characteristic of Cantonese. The multi-register structure of the tonal system in Cantonese may be an important factor, which make it easier for tones to fit the melody movement.

It was observed that when T1 and T4 are in songs, they tend to be produced as their level variants 55 and 11 (Chan, 1987a, 1987b; Cheung, 2007). Thus, the six lexical tones can be divided into four tone registers according to the tone value of the ending point: T1 and T2 are high (~5), T3 and T5 are middle (~3), T6 is low (~2), and T4 is bottom (~1), as shown in Table 1-5. Wong and Diehl (2002) was a little different from the classification listed here in that T6 and T4 were grouped to a register, but the classification in Table 1-5 is more generally accepted. Cheung and Wong (2008) explored the aria types and narrative songs in Cantonese opera, where the melody is not fixed but dependent on the tones of the lyrics. They found that at the default notional key of C, the default mapping between the tone note and musical note is: high – D^4; mid-high – C^4; mid-low – A^3; low – G^3.

Table 1-5 Four tone registers during singing

Tone register	Ending point	Pitch-image label[4]	Typical tonic sol-fa correspondence	Tone(s) subsumed	Numbers as examples
High	5	Sharp 尖[tsiːm^{55}]	2/re	T1	1;3;7
				T2	9
Mid-high	3	High 亢[kɔːŋ^{33}]	1/dol	T3	4;8
				T5	5
Mid-low	2	Low 下[haː^{22}]	6̣/la	T6	2;6;10
Low	1	Deep 沉[ts^{h}ɐm^{11}]	5̣/so	T4	0

One may still wonder how this four-register-scale can be matched to an array of a much wider range of musical notes. Cheung (2012a) stated and demonstrated that the elasticity and transposability of lexical tones are the gist of the answer to this question. He pointed out that both the entire pitch range and the pitch level gradations are elastic for lexical tones. An obvious example is that people of different genders have quite different pitch range and pitch register. Even the same speaker in different contexts also exhibits variation in absolute pitch. Pitch range can be compressed or widened. Also, the pitch register can be transposed upwards or downwards. For example, the pitch range is G^3 to D^4 at one instance but is transposed to the range D^4 to A^5 at another instance. With the interplay of elasticity and transposability, melodies can be translated into different permutations of the four registers.

According to the above mechanism, the six lexical tones can be categorized to four registers, but it does not mean that tones within the same tone register cannot be differentiated in singing. T1 and T2 are of the same tone register (High) and regarded as equal when setting lyrics to a melody, but they have different contour when singing (Cheung, 2012a; Cheung, 2007, 2012b; Ho, 2008; Schellenberg, 2011, 2013; Yung, 1989). T1 keeps flat while T2 has an ornamental upward glide at the beginning or adopts two successive notes in ascending order to make a rising contour (Cheung, 2012a; Cheung, 2007, 2012b; Ho, 2008; Yung, 1989). Similar situation also happens to the Mid-high tone

[4] The terms listed in the column of "Pitch-image label" are cited from Cheung (2012b).

register comprising of T3 and T5. Therefore, the implementation of singing keeps the original tone shapes and makes tones of the same register differentiated from each other.

Previously, research on the relationship between lexical tones and melody was mainly conducted by comparing different lyrics of the same melody and the correspondence between musical notes and tone letters (Chan, 1987a, 1987b; Cheung, 2012a; Cheung, 2007; Cheung & Wong, 2008; Chow, 2012; Ho, 2006, 2008; Lau, 2010; Wong & Diehl, 2002). These studies helped to explain tone-melody matching in Cantonese songs. From the perspective of linguistics, it is more important to identify the different phonetic realization patterns of lexical tones in speaking and singing. Wong, Chan, and Lau (2001) applied signal technique of acoustic measurement to the singing of Cantonese opera, but their sample was small and they did not include normal speech for comparison. To answer the question how the lexical tones can be sung, we have to tell how the lexical tones can be both spoken and sung while keeping these two styles of speech undoubtedly distinct. Our present experimental study provide a clear answer.

The differences between the speaking and singing speech styles in Cantonese have seldom been mentioned. The exception is Yung (1989, p. 13):

> "the distinction between 'song' and 'speech' appears to be based upon the presence or absence of musical tones, which are defined here as ones with stable pitches and with consistent intervallic relationships among them such that, theoretically, a scale can be abstracted. Although speech types do not involve musical tones, they often exhibit identifiable metrical and rhythmic patterns, and verse structures that involve patterns in phrase lengths, rhymes, and linguistic tones."

According to Yung's categorization of Cantonese Opera, there are seven kinds of "speech types" and three kinds of "songs", which are collectively called "oral delivery types". The "speech types" include "Baak (Plain Speech)", "Tokbaak (Supported Speech)", "Eanbaak (Comic Rhymed Speech)", "Logubaak (Percussion Speech)", "Sibaak (Poetic Speech)", "Haugu (Rhymed Speech)", "Baaklaam (Patter Speech)" (Yung, 1989, p. 57). The three types of "songs" are "Aria Types", "Fixed Tunes", and "Narrative Songs" (Yung, 1989, p. 14). He found that occasionally the distinction between speech and song can be blurred.

Indeed, the "presence or absence of musical tones" is the underlying distinction between singing and speaking. However, Yung (1989) did not further explore the tonal-realization differences in singing and speaking. Yung (1989) did not clearly illustrate what the lexical tones are like with the presence of musical tones in singing, or what the lexical tones are like in the absence of musical tones in speaking. Our present study will go further on this problem and provide concrete analysis on the realization patterns of Cantonese lexical tones in speaking and singing.

1.4 An overview of the experiments in the present study

There were three experiments carried out in the present study.

Experiment I was a pilot study which compared three pairs of speech styles: opera speaking versus opera singing, opera speaking versus normal speaking, and normal speaking versus normal singing. The first two pairs were obtained from existing recordings of authentic materials produced by top-notch opera professionals and the last pair was produced by a lay person. The use of authentic materials can make our study more objective as they are not influenced by our experimental purpose. Cantonese opera was adopted here also because it is convenient to find both speaking and singing utterances produced by the same speaker. The first and the third pairs of speech styles are both comparison between speaking and singing. The first pair is in a classic opera setting, and the third pair is in a normal setting. The two sub-types of speaking (opera speaking and normal speaking) in the second pair of speech styles help us to bridge the comparison between the first and the third pairs and find out the most essential factor to differentiate speaking and singing.

On the basis of the findings from the acoustic data, we modified a speaking utterance to the singing pattern (Speaking-to-Singing) and also transformed a singing utterance to the speaking pattern (Singing-to-Speaking) as stimuli for a small-scale perception test. Details of these adjustments are described in Chapter 2. This perception test can help us to interpret the results of the acoustic experiments from the perspective of listeners.

Experiment I only used data from one speaker. It may be challenged that the findings are due to that speaker's idiosyncrasies. Thus, we moved on to conduct

Experiment II and Experiment III, for which we recruited twelve native speakers of Cantonese.

In Experiment II, we had more informants performing normal speaking and normal singing. These two speech styles were abbreviated as “speaking and singing” in Experiment II. Twelve native informants of Cantonese were asked to read aloud a newly designed recording questionnaire, which is reproduced in Appendix F. This questionnaire consists of a set of lyrics containing all six lexical tones as well as lyrics comprising a series of reduplicative utterances. This questionnaire can help us get an equal number of syllables for each tone, and can help us observe pitch configurations at the utterance level. Syllables in isolated citation form are also included in this questionnaire.

In addition to the intonation effects of speech styles, we also want to further observe intonation effects of different sentence types in speaking, for which we carried out Experiment III. The sentence types investigated here actually make up a 2*2 matrix. One dimension is declarative versus question, and the other dimension is with a sentence-final particle versus without a sentence-final particle.

Experiment II and Experiment III were carried out at one go with the same informants. Similar measurements and statistical methods were applied to both experiments. The analysis on the acoustic data of these two experiments was conducted in the framework of the matrix composed of pitch configuration and pitch domain, as introduced in Section 3.1.1. We will observe whether different speech styles and different sentence types will yield different results in the matrix. More details about Experiment II will be introduced in Chapter 3 and Experiment III will be analyzed in Chapter 4.

From the brief introduction of the experiments above, it can be seen that the scope of the experiments is from broad to narrow. Firstly we conducted studies on three pairs of speech styles to get a preliminary idea as to what differentiates speaking and singing (whether they are in an opera or a normal setting), and what are the minor differences between sub-categories of speaking. Then we concentrated on only one pair of speech styles – normal speaking and normal singing, which represent the more general universal status and for which it is easier to recruit more informants to participate in our experiment. More data were obtained in this second experiment and the experimental method as well as the framework was improved to further study the different intonation effects that speaking and singing exert on the lexical tones in continuous speech. Finally we only focused on the normal speaking speech and explored intonation patterns of several sentence types in our last experiment. Through studying the internal differences of speaking, we can better understand the coverage of speaking and the boundary between speaking and singing. The three main experiments were interrelated and our quest got more focused as our study moved on.

Chapter 2
Experiment I: Three Pairs of Speech Styles

In Experiment I, we compared three pairs of speech styles: opera speaking versus opera singing, opera speaking versus normal speaking, and normal speaking versus normal singing. Cantonese opera was adopted as part of the experimental materials here, for the reason that it is convenient to find both speaking and singing utterances without the interference of background music. As well, the materials for the first two pairs of speech styles are authentic speech, which are not influenced by our experimental purposes, and are, thus, more objective. As linguists may be more concerned about the performance of layman speakers, comparison of the third pair is also conducted, which also serves as a pilot study for the experiments in Chapter 3.

Comparison of these three pairs of speech styles reveals that speaking and singing speech styles differ mainly in pitch slope, whether it is opera speaking versus opera singing or normal speaking versus normal singing. At the utterance level, speaking exhibits downward tendency. Singing does not have this tendency and even shows a slightly upward trend. At the syllable level, the falling/level tones have a much more declined pitch slope, while rising tones have weakened slope in speaking. In singing, the pitch contours of falling/level tones tend to be flattened, and the rising tones have more tilted pitch slope than in speaking. A small-scale perception test further confirmed that pitch slope is the key factor to differentiate speaking and singing.

For normal speaking versus normal singing, there is another difference, namely that of pitch register – normal singing usually has higher register than normal speaking. However, opera speaking and opera singing have similar overall pitch register, which shows that the pitch register of singing is not necessarily higher than that of speaking. Meanwhile, the main difference between opera speaking and normal speaking is that opera speaking has significantly higher pitch register than normal speaking. Thus, register differences alone cannot signify whether a speech style is speaking or singing. The actual function of higher register is that it usually indicates a conspicuous kind of speech style.

2.1 Method of Experiment I

2.1.1 Experimental materials

The resources of the experimental materials used in Experiment I are listed in Appendix A.

For the first pair of speech styles, opera speaking (**OSpk** henceforth) versus opera singing (**OSng** henceforth), we only focused on the opera recordings of one famous Cantonese opera actor, Mun Cheen-shuih (MCS, 文千歲). Four of his high-quality CDs were selected and those utterances without background music were used for our acoustic analysis. The detailed information of these utterances is listed in Appendix B, including the code for our later reference, the type of melody, the Chinese text with IPA transcription, and the resource of the utterance. For the OSpk utterances, most of them belong to the "白 (Plain Speech)" and "詩白 (Poetic Speech)" types. For the OSng utterances, some of them are not full sentences because we wished to avoid the interference of background music. In certain instances where the background music started, the affected part of the utterance was cut off. All the selected utterances were segmented to syllables for further measurement. As the intrinsic pitch shape of sentence final particles is intonational rather than a real lexical tone (Cheung, 1986, p. 256), we excluded the sentence final particles in our analysis. There are 262 speaking syllables and 346 singing syllables in all.

For the second pair of speech styles, opera speaking versus normal speaking (**NSpk** henceforth), we adopted materials from a Hong Kong SAR government website on Cantonese opera, which is also listed in Appendix A. These recordings were made by another famous Cantonese opera actor, Yuen Siu-fai (YSF, 阮兆輝). YSF demonstrated the opera speaking and normal speaking versions of the same text, which can be regarded as a minimal pair of these two styles of speech. These utterances were transcribed in Appendix C in both orthography and IPA. In some utterances, there is a sentence final particle in OSpk that is missing in NSpk. We segmented these sample utterances into syllables. The sentence final particles were also excluded in this experiment. There are 59 pairs of opera speaking and normal speaking syllables.

For the third pair of speech styles, normal speaking versus normal singing (**NSng** henceforth), we asked a layman informant ZSX to read aloud and sing the same lyrics we selected. The recordings were made with a Marantz PMD620 Professional Handheld Digital Audio Recorder. The microphone was set to mono channel, the sampling rate was 44,100 Hz, and the resolution was 16-bit. ZSX was a 26-year-old male born and raised in Guangzhou, China. ZSX was a reliable informant in that his performance was stable and he kept singing in tune. The lyrics in the questionnaire include classic opera songs, pop songs and children's songs. To make sure that the informant is familiar with the tunes of the songs, the songs used in the questionnaire are all very popular ones. There are 724 pairs of speaking and singing syllables for our analysis.

We carried out a small-scale perception test as part of this experiment. This perception test was conducted on the basis of the findings in the above acoustic experiments. The original materials come from some of the recordings by MCS and ZSX. Details of this perception test will be described in Section 2.5.

2.1.2 Measurement and calculation methods

All the pitch measurement in this experiment was carried out using the Praat software program (Boersma & Weenink, 2009). For every syllable, measurement was applied to the vocalic portion, and ended at the end of the vowel. As a pilot study, for the convenience of comparing a large number of syllables, we only used two pitch points for every collected syllable in our analysis – the maximum pitch point (f_{Max} henceforth) and the minimum pitch point (f_{Min} henceforth). For level or falling tones, f_{Max} coincides with the beginning pitch point ($f_{\text{Beginning}}$ henceforth), and the f_{Min} is the ending pitch point (f_{Ending} henceforth) in most circumstances. For rising tones, correspondence is the other way round, i.e., f_{Max} coincides with f_{Ending}, while f_{Min} coincides with $f_{\text{Beginning}}$.

When we wanted to know more details of the pitch configuration, the data of two pitch points per syllable seemed insufficient. In these cases, we wanted to have more measuring points to more accurately approximate the pitch contour. In these circumstances we measured pitch at every 5% of the syllable duration and connected these 21 evenly-distributed pitch points. There is no doubt that pitch contours drawn by this method are more detailed and accurate than the schematic straight lines defined by the two extreme pitch points. In this chapter, we will use the more detailed method to draw pitch contours in the following circumstances: (1) the lengthened syllables by MCS; (2) the comparison between the minimal pairs of OSpk versus NSpk by YSF.

Some parameters are defined below to help us study the characteristics of pitch contours in different speech styles.

To show the direction and slope of pitch contour, we define *r* as follows:

$$r = \frac{f_{\text{Ending}}}{f_{\text{Beginning}}}$$ **Formula 2-1**

where f_{Ending} represents the f_0 value of the ending point, and $f_{\text{Beginning}}$ stands for the f_0 value of the beginning point. By the definition of *r*, if *r* is less than 1, the pitch contour is falling; if *r* is greater than 1, the pitch contour is rising. The pitch contour is level if *r* approximates 1. In contrast, the pitch contour tends to be steep if *r* is more deviated from 1.

Pitch in Cantonese has to signify both tone and intonation. We cannot simply compare the pitch of successive syllables in an utterance for its global direction of pitch shape, because these syllables probably belong to different tone categories and have their intrinsic tone registers. The method of comparing successive syllables for the pattern of utterance is only applicable to the situation of same-tone syllables. To avoid the interferences of remote factors (for instance, if two syllables are in different syntactic units, there is pitch reset effect at the boundaries), only the pitch of adjacent same-tone syllables is compared here. Furthermore, syllables are limited to those within a phrase to exclude the effect of upstep at the phrase boundary. Let

$$u = f_i - f_{i+1}$$ **Formula 2-2**

where f_i represents the f_0 value of syllable i, and f_{i+1} stands for the f_0 value of syllable i+1. There are three possibilities of u, namely $u > 0$, $u < 0$ and $u = 0$, respectively representing downdrift, updrift, and same register, as depicted in Figure 2-1.

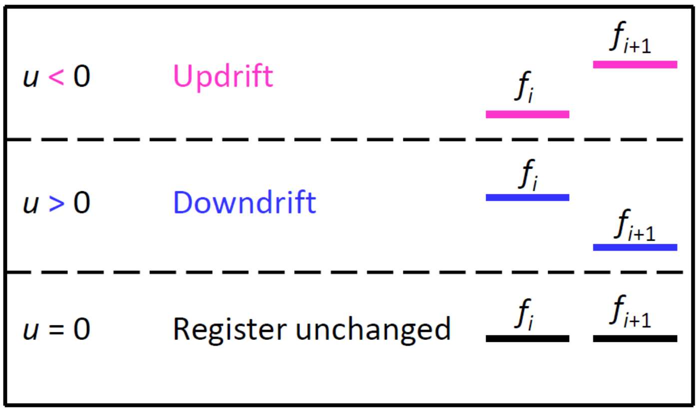

Figure 2-1 Three possibilities of *u*: updrift, down drift and register unchanged

2.2 Results and discussions on OSpk versus OSng

The experiment for a comparison between OSpk and OSng was our first step to study the intonation effects on Cantonese lexical tones in speaking and singing. Thus, we hope to have a more detailed and overall analysis in this section (Section 2.2). First we would like to make a close observation of the raw data of the OSpk and OSng utterances in Section 2.2.1, which can provide us with a basic impression of how speaking differs from singing, and lead us on the right track for further research. Then we move on to carry out calculations and analysis at the syllable level and at the utterance level, which will be discussed in Section 2.2.2 and Section 2.2.3, respectively. As many of the OSng utterances are not full sentences (the end of which was cut off to avoid the interference of background music) and many of the OSpk utterances end with SFP rather than ordinary syllables, there are not enough data to compare the utterance-final syllables with the non-final syllables in OSpk and OSng. The utterance-final intonation effect will not be discussed for the data of OSpk and OSng.

2.2.1 Preliminary observations

Table 2-1 lists the tessitura of MCS in OSpk and OSng. We notice that the tessitura is higher and wider in opera performance than in our daily conversation. This is probably because the speech and emotion in opera performance are exaggerated and more dramatic than in our daily normal speech. The data in Table 2-1 reveal that the maximum and minimum pitch values of OSng are a little higher than OSpk, while the tessitura in OSng is a little narrower than OSpk. All these differences are not significant. OSpk and OSng, while they are clearly identified as different types of speech, occupy similar tessitura, from which it can be inferred that tessitura is not the key contrastive factor for the distinction between speaking and singing.

Table 2-1 Tessitura of MCS in OSpk and OSng (Unit: Hz)

Pitch point \ Speech	OSpk	OSng
f_{Max}	532.67	537.17
f_{Min}	115.85	126.13
Tessitura (=f_{Max} –f_{Min})	416.82	411.04

Figure 2-2 displays the distribution of the two extreme pitch points, f_{Max} and f_{Min} of every syllable by MCS. The upper half shows the pitch distribution in OSpk and the lower half exhibits the counterpart in OSng. f_{Max} is represented by unfilled symbols and f_{Min} is denoted by solid symbols within each tone category. For level/falling tones, f_{Max} is on the left and f_{Min} is on the right, while for rising tones, the direction is reversed. Here the tone categories are arranged according to the four tone registers in the order from high to low (as introduced in Section 1.3.2, T1 & T2 > T3 & T5 > T6 >T4), and the level and falling tones precede the rising tones within the same tone register. In this order, the pitch distribution of different tone registers can be easily seen.

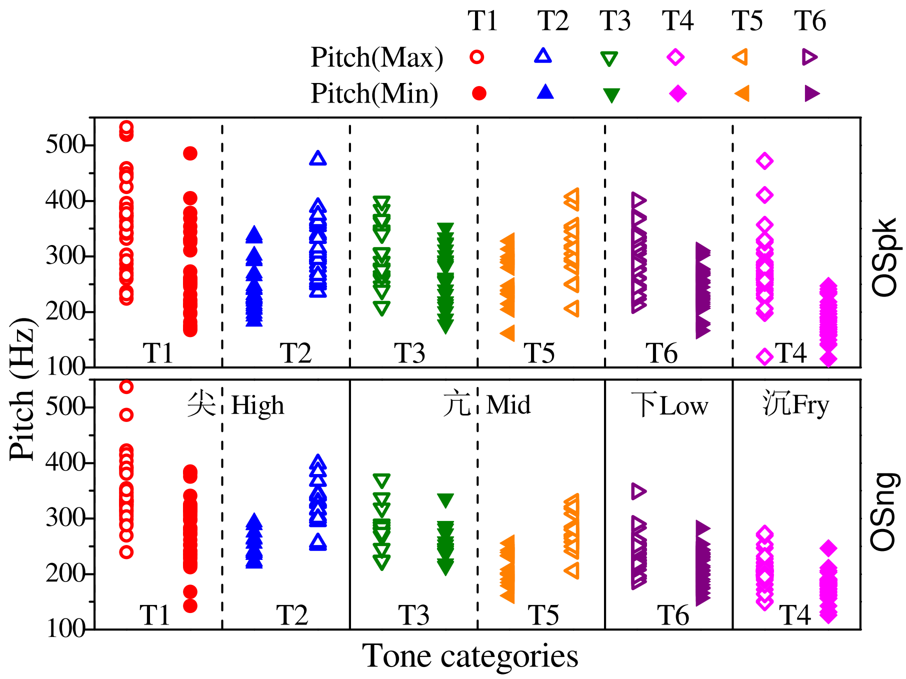

Figure 2-2 Pitch distribution of beginning and ending points in OSpk and OSng

It can be observed from Figure 2-2 that there is great overlap among different tone categories in the graphs of both OSpk and OSng. It can be inferred that the absolute pitch of an isolated syllable may not be able to indicate its tone category. For instance, almost all the tone categories have at least one pitch point at about 300 Hz in OSpk. If a pitch line is at 300 Hz, people cannot tell which tone category it belongs to. The large overlap of pitch distribution among tones may be the acoustic reason for the regularity found in Peng et al. (2012), Wong (1999) and Varley and So (1995) that the accuracy of judgment on Cantonese level tones drops dramatically in isolation.

Comparing the two graphs of OSpk and OSng, we found that the overlap across different tones is greater in OSpk than in OSng. For the pitch distribution within the same tone category, the pitch points in OSpk are more dispersed while the distribution range in OSng is more concentrated. Therefore, compared with OSpk, OSng is more conforming to the sequence from high to low.

According to our observation, there are some extra long syllables (which are more than 500 milliseconds) in the opera performance. Traditionally when a syllable in OSng needs to be lengthened, the actors/actresses are required to keep the articulation of a particular segment sustained but change the frequency of their vocal fold vibration. This explains why the lengthened syllables in OSng have curving pitch contours. In OSpk, it is less common to have lengthened curving syllables. Occasionally there are some such syllables in "詩白 (Poetic Speech)", and they are usually used to express strong emotion.

For the curving pitch contours of the lengthened syllables in OSpk and OSng, our concern is whether lexical tones can be realized and which part of the curving contour expresses the lexical tone. We found that the part realizing lexical tones in lengthened curving syllables is different in OSpk and OSng. Below we will illustrate the different realization patterns with some concrete examples.

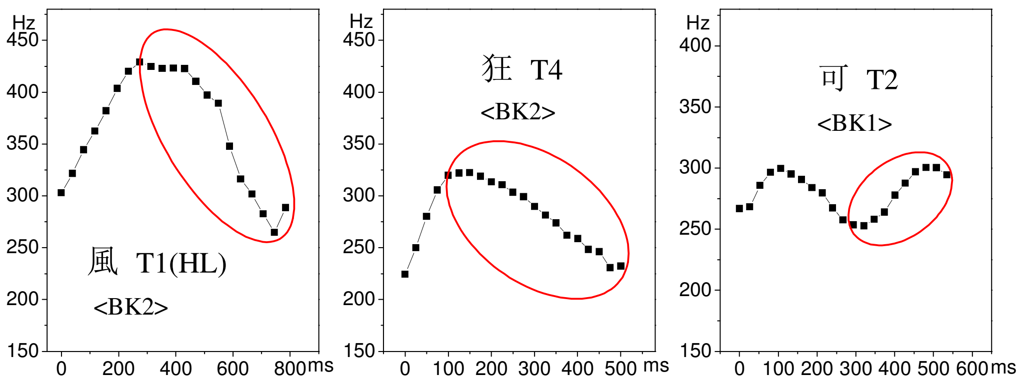

Figure 2-3 Examples of curving pitch contours of lengthened syllables in OSpk

Figure 2-3 provides three examples of curving pitch contours of lengthened syllables in OSpk. These examples have 21 evenly-distributed pitch points per syllable, which can accurately reflect the detailed pitch contours. The ellipses are used to denote where the lexical tones are realized. For the two falling tones T1 (HL) and T4 shown in the first two graphs, the beginning part of the curving pitch contour rises abruptly to the peak before starting the real falling tone contour. For the rising tone T2 displayed in the last graph, there is a convex bend before the real rising tone contour commences. These three examples of lengthened syllables in OSpk share a common pattern that the lexical tone features are realized from the middle part of the pitch contour, with the beginning part moving in the opposite direction as a preparation.

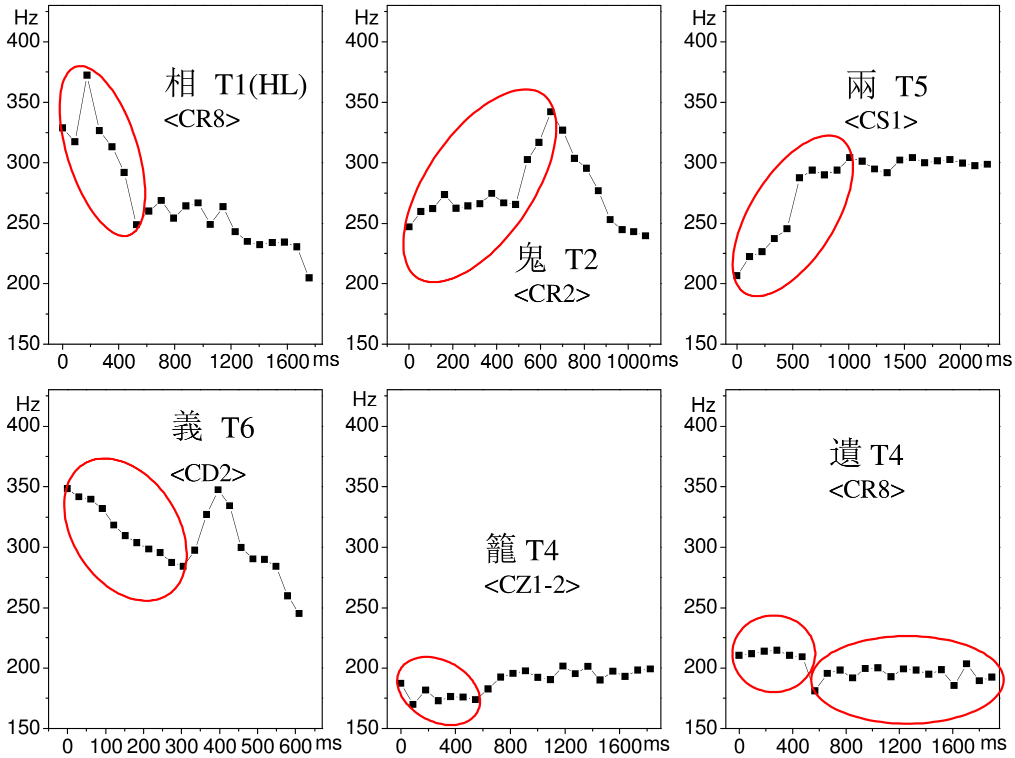

Figure 2-4 Examples of curving pitch contours of lengthened syllables in OSng

For the lengthened curving syllables in OSng, there are six examples shown in Figure 2-4. It can be observed that the realization portion of the lexical tones is quite different from that of the lengthened syllables in OSpk. Actually there are two types of realization. The first type covers the examples of “相”, “鬼”, “兩”, “義” and “籠”. In these five graphs, the original lexical tones are realized at the very beginning of the curving pitch contours. After realizing the lexical tones, the remaining pitch contours have more freedom to move: it can move along the same direction of the lexical tone but become less sloping, like the one shown in the first graph; it can also move on the opposite track, as the examples shown in the second and the fifth graph; it can keep the pitch value of the final point of the tone-realization part to make a nearly constant level contour, as shown in the third graph; it can change the moving direction as a zigzag curve, like the example displayed in the fourth graph. The final graph of Figure 2-4 displays an example of the second type of tone-realization pattern, i.e., tone features are approximated by two successive level contours, denoted by the two ellipses.

Therefore, the realization portions of lexical tones for lengthened curving syllables are different in OSpk and OSng. In OSpk, there may be a preparation part that moves in the opposite direction at the very beginning and the lexical tones are realized after this preparation. In OSng, the lexical tones can be realized by two methods. The first method is that the lexical tones are realized at the beginning part and the remaining pitch contour is free to change its movement trajectory. The second method is that the lexical tone is approximated by two successive level pitch contours. For the measurement of the f_{Max} and f_{Min} in our current experiment, as well as the measurement of the whole pitch contour in Experiment II (occasionally some singing syllables in Experiment II also have curving pitch contours), only the tone-realization part is measured if there is a lengthened curving pitch contour.

2.2.2 Calculation and analysis at the syllable level

The above was a preliminary observation on the raw pitch data of MCS. Below we will have further calculations and analysis to reveal the differences between OSpk and OSng.

Firstly we would like to conduct a detailed analysis at the syllable level. Figure 2-5 shows the mean pitch distribution of different tones in OSpk and OSng. The beginning point and the ending point of each tone are connected to approximate the mean pitch contour. The standard deviation of each point is indicated by an error bar. Pitch contours in OSpk and OSng are marked by solid lines and short dashed lines respectively. The order of tone categories is the same as in Figure 2-2 (Page 50), which arranged the tones according to the four registers from high to low, with the level/falling tones preceding the rising tones within the same tone register.

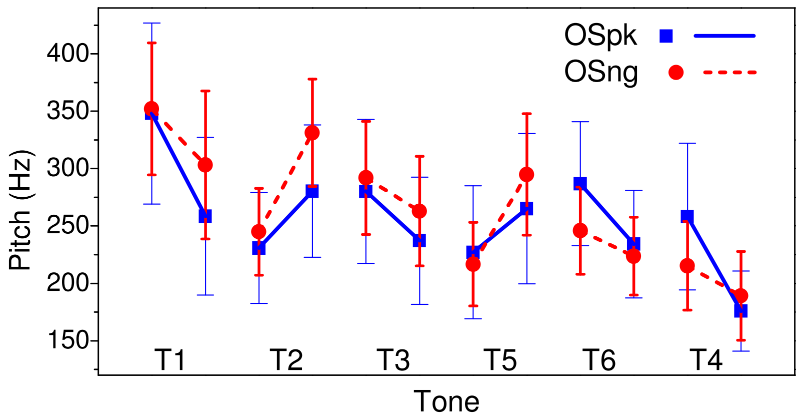

Figure 2-5 Schematic mean pitch contours in OSpk and OSng, with error-bars indicating ±1 standard deviation.

The mean r values of the different lexical tones are listed in Table 2-2, along with the mean pitch of beginning and ending points. Table 2-2 is a useful supplement to Figure 2-5 as it provides more accurate data. It should be noted that r as listed in Table 2-2 is the mean of the r values of all the syllables within a certain tone category, not the ratio of the mean ending pitch to the mean beginning pitch.

Table 2-2 Mean values of pitch and r in OSpk and OSng

Tone	Speech Styles	Pitch (Hz)		r
		$f_{Beginning}$	f_{Ending}	
T1	OSpk	348.16	258.41	0.7592
	OSng	352.06	303.23	0.8648
T2	OSpk	230.76	280.45	1.220
	OSng	245.11	331.22	1.369
T3	OSpk	280.27	237.19	0.8484
	OSng	292.00	262.92	0.8999
T4	OSpk	258.19	175.86	0.7027
	OSng	215.28	189.07	0.8776
T5	OSpk	227.04	265.07	1.173
	OSng	216.70	294.92	1.367
T6	OSpk	286.82	234.22	0.8230
	OSng	245.93	223.74	0.9116

Independent-sample *t*-tests were carried out to examine whether these factors are significant to differentiate OSpk and OSng. The results are listed in Table 2-3 to Table 2-5. In the tables listing statistical results, "**0.000**" represents $p < 0.001$, henceforth. The results in Table 2-3 reveal that the *p*-values of *r* are very consistent among all tones – they are all much smaller than 0.05. These results show that *r* of OSpk and OSng are significantly different, i.e. pitch contour sloping is an important factor to distinguish these two styles of speech.

Table 2-3 Independent-sample *t*-test results on *r* in OSpk and OSng

Tone	*t*	*p*	*df*
T1	-4.188	0.000	152
T2	-3.631	0.001	74
T3	-3.027	0.004	61
T4	-9.879	0.000	135
T5	-7.008	0.000	61
T6	-5.941	0.000	113

The comparison of pitch register is a little more complicated. We notice that most of the pitch points show no significant difference between OSpk and OSng, but the *p*-values of the ending points of T1, T2, T4 and the beginning points of T4 and T6 are below 0.05, which indicates the pitch distribution of these points are significantly different in these two styles of speech.

Table 2-4 Independent-sample *t*-test results on $f_{Beginning}$ in OSpk and OSng

Tone	*t*	*p*	*df*
T1	-0.352	0.747	152
T2	-1.429	0.157	74
T3	-0.815	0.418	61
T4	4.892	0.000	135
T5	0.813	0.390	61
T6	4.744	0.000	113

Table 2-5 Independent-sample *t*-test results on f_{Ending} in OSpk and OSng

Tone	*t*	*p*	*df*
T1	-4.045	0.000	152
T2	-4.178	0.000	74
T3	-1.952	0.056	61
T4	-2.026	0.045	135
T5	-1.940	0.057	61
T6	1.383	0.209	113

Let's go back to Figure 2-5 (Page 54) for a better understanding. The ending points of T1, T2 and T4 in OSng are obviously higher than in OSpk. The beginning points of T4 and T6 in OSng are much lower than in OSpk. The combined effect is that the falling tendency of T1, T4 and T6 is weakened and the rising trend of T2 is strengthened in OSng.

To compare the different pitch distribution of OSpk and OSng, we can observe the relative location of the solid lines and the dashed lines. The following patterns are displayed in Figure 2-5 (Page 54): 1) for high tones T1 and T2, OSng uses higher register; 2) for mid-high tones T3 and T5, OSng uses the same register as OSpk; 3) for mid-low tone T6 and low tone T4, OSng uses lower register. The pattern shows that the mechanism at work is pitch span elasticity, narrower for OSpk and wider for OSng.

In sum, at the syllable level, an obvious difference between OSpk and OSng is that they have different pitch contour sloping tendency. Compared with OSng, the falling and level tones in OSpk are more declining, while the rising tones are less ascending. Pitch register and span are only a little different in these two styles of speech, and the mechanism at work is pitch span elasticity.

2.2.3 Calculation and analysis at the utterance level

As mentioned in Section 2.1.2, to observe the pitch trend at the utterance level, our method was to compare the adjacent same-tone syllables which are within a syntactic phrase. These pairs of syllables are randomly distributed in an utterance. If the comparison of u shows an overwhelmingly regular pattern, it can be regarded as the direction of the whole utterance rather than a specific location of the utterance.

Firstly we would like to provide the overall statistical analysis on the adjacent same-tone syllable pairs in OSpk and OSng, which is listed in Table 2-6. The data shows that for OSpk, the overwhelming pattern is $u > 0$ (f_{Max}: 23 versus 3; f_{Min}: 21 versus 5), which indicates OSpk has strong tendency of declination. For OSng, f_{Max} tend to be elevated across the utterance, as there are much more $u < 0$ pairs (25 versus 9). However, the comparison of f_{Min} does not show a strong preference (16 versus18) of a certain direction, as the number of $u > 0$ pairs is similar to the number of $u < 0$ pairs.

Table 2-6 Statistical analysis on the adjacent same-tone syllable pairs in OSpk and OSng

u	Indication of u	OSpk		OSng	
		f_{Max}	f_{Min}	f_{Max}	f_{Min}
> 0	$f_i > f_{i+1}$	23	21	9	18
= 0	$f_i = f_{i+1}$	0	0	0	0
< 0	$f_i < f_{i+1}$	3	5	25	16

To have a closer observation of the adjacent same-tone syllables in OSpk and OSng, we sorted them into several categories according to the u values of the successive f_{Max} and f_{Min}. Table 2-7 lists the data of OSpk, and the data of OSng are shown in Table 2-8. In the column for "Category" in Table 2-7 and Table 2-8, the letter **T** represents the top value – f_{Max}, and the letter **B** represents the bottom value – f_{Min}. If $u > 0$, we add a "+" before the letter T or B; if $u < 0$, the symbol "-" is used instead.

Table 2-7 Data of the adjacent same-tone syllables in OSpk (Pitch unit: Hz)

Category	Syllable	Tone	f_{Max}	f_{Min}	Utterance Code
[+T, +B]	之	1	231.57	181.6	BR1
	詩	1	224.36	173.01	
	一	1	353.44	343.92	BR5
	喲	1	331.63	325.47	
	小	2	231.83	197.46	BX3
	姐	2	214.69	163.41	
	小	2	226.86	207.36	BX4-1
	姐	2	216.51	193.69	
	點	2	298.83	231.33	BD2
	樣	2	287.74	222.42	
	過	3	228.64	222.49	BH1
	去	3	222.49	207.27	
	答	3	273.61	204.69	BX4
	應	3	213.9	165.5	
	確	3	183.55	169.01	BX8
	確	3	179.3	166.95	
	紅	4	180.01	137.66	BX2-1
	娘	4	153.77	135.95	
	紅	4	277.04	199.8	BX7
	娘	4	203.6	163.46	
	樓	4	283.44	196.76	BS1
	臺	4	259.38	160.34	
	而	4	273.48	212.36	BR5

Category	Syllable	Tone	f_{Max}	f_{Min}	Utterance Code
	來	4	264.27	208.07	
	蒙	4	410.77	234.72	BK2
	塵	4	330.03	211.93	
	有	5	160.55	150.58	BX3
	禮	5	155.57	124.34	
	墓	6	315.09	234.35	BS2
	穴	6	278.09	177.37	
	杜	6	318.29	277.37	BR6-1
	麗	6	290.81	253.76	
	就	6	335.99	310.63	BR6
	是	6	328.27	309.41	
	杜	6	367.73	276.02	BR6-2
	麗	6	309.5	266.06	
	的	1	288.93	243.5	BX8
	的	1	281.57	267.83	
	夫	1	340.86	213.34	BS2
	妻	1	303.27	214.99	
	小	2	191.26	154.11	BX8
[+T, -B]	姐	2	177.1	159.4	
	紅	4	183.99	135.67	BX2-2
	娘	4	157.95	136.58	
	無	4	238.28	156.96	BS1
	緣	4	233.59	163.39	
	三	1	417.14	273.23	BH2
	生	1	422.4	247.62	
	詩	1	359.47	330.55	BX8
[-T, +B]	中	1	362.62	295.58	
	小	2	212.11	195.19	BX4-2
	姐	2	218.05	192.23	

In Table 2-7, there is not a category of [-T, -B]. This indicates that either or both of f_{Max} and f_{Min} shows the $u > 0$ pattern in OSpk, i.e., either or both of f_{Max} and f_{Min} is declining for successive same-tone syllables in OSpk.

Among the three categories shown in Table 2-7, there is no doubt that Category [+T, +B] is the dominant pattern as it applies to 18 pairs out of the total 26 pairs, while there are only 5 pairs of syllables in Category [+T, -B] and 3 pairs in Category [-T, +B]. Therefore, for successive same-tone syllable pairs in OSpk, the overwhelming pattern is that both f_{Max} and f_{Min} are declining forwards across the utterance. Occasionally one value of f_{Max} and f_{Min} is not declining, but there is never a case of both values increasing. It can be concluded that OSpk shows an overwhelmingly downward trend for the utterance-level pitch shape.

Table 2-8 Data of the adjacent same-tone syllables in OSng (Pitch unit: Hz)

Category	Syllable	Musical Note	Tone	f_{Max}	f_{Min}	Utterance Code
[-T, -B]	龜	$\dot{1}$	1	328.26	299.85	CH1
	山	$\dot{1}$	1	332.69	327.12	
	依	$\dot{1}$	1	404.67	380.33	CH3
	稀	$\dot{1}$	1	422.13	387.7	
	西	3	1	300.92	270.15	CX2
	廂	3	1	309.89	292.7	
	冰	3	1	325.05	304.17	CD3
	釋	3	1	346.36	308.77	
	迷	$\underset{\cdot}{5}$	4	173.82	142.17	CH1
	人	$\underset{\cdot}{6}$	4	216.64	173.82	
	漁	$\underset{\cdot}{5}$	4	157.95	135.82	CH3
	娘	$\underset{\cdot}{6}$	4	204.22	179.95	
	由	$\underset{\cdot}{5}$	4	201.29	189.26	CX3
	來	$\underset{\cdot}{6}$	4	203.7	190.36	
	為	$\underset{\cdot}{5}$	4	189.06	161.85	CD1
	媒	$\underset{\cdot}{6}$	4	204.18	186.85	
	難	1	4	269.39	205.69	CD2

Category	Syllable	Musical Note	Tone	f_{Max}	f_{Min}	Utterance Code
	忘	2	4	272.76	246.81	
	前	5̣	4	172.13	126.13	CD3
	嫌	5̣	4	203.35	186.92	
	團	6̣	4	180.95	156.35	CD4
	圓	1	4	209.9	160.78	
	儼[5]	5̣	4	183.48	155.87	CR1
	然	6̣	4	199.6	188.67	
	全	6̣	4	163.74	143.13	CR6
	無	1	4	207.8	180.09	
	暫	6̣	6	338.89	289.19	CH2
	避	6̣	6	343.91	324.24	
	夜	6̣	6	224.48	214.16	CR3
	靜	6̣	6	224.58	215.07	
[-T, +B]	江	1̇	1	471.94	462.75	CH2
	邊	1̇	1	492	435.86	
	今	6	1	324.78	301.76	CX5
	宵	6	1	328.5	294.09	
	呑	1̇	1	413.59	384.16	CS1
	聲	1̇6	1	426.19	381.48	
	飛	3	1	285.5	279.23	CR1
	仙	3	1	290.82	186.58	
	分	3	1	333.34	325.04	CD3
	清	53	1	388.17	216.23	
	濤	4	1	334.29	315.99	CZ3
	濤	43	1	380.46	271.88	
	惡	3	3	403.93	384.68	CH6
	霸	3	3	421.97	370.35	
	若	7̣	6	237.75	232.52	CR2

[5] The correct pronunciation of "儼" should be T5, but MCS mispronounced it as T4.

Category	Syllable	Musical Note	Tone	f_{Max}	f_{Min}	Utterance Code
	是	7̣	6	245.24	232.13	
	熟	3	6	217.76	205.37	CR3
	睡	3	6	235.26	197.71	
	夜	1	6	196.23	180.89	CZ2
	望	1	6	196.77	173.98	
[+T, -B]	莫	2	3	262.29	188.48	CD1
	雁	2	3	251.13	229.69	
[+T, +B]	潘	5	1	401.64	378.18	CX4
	安	4	1	332.95	320.05	
	真	4	1	364.77	341.35	CX4
	心	3	1	333.6	305.1	
	癡	4	1	384.31	375.08	CR6
	心	3	1	329.57	320.44	
	想	$_{3}4$	2	295.41	242.59	CH5
	起	$_{3}4$	2	287.9	213.26	
	就	3	6	220.64	208.88	CH3
	是	3	6	219.77	204.88	
	待	3	6	265.07	238.68	CX2
	月	3	6	254.76	225.67	
	妙	3	6	224.69	207.27	CX5
	藥	3	6	222.95	199.54	
	自	1	6	302.26	267.95	CR1
	在	1	6	301.04	235.63	

Among the four categories listed in Table 2-8, the distribution of syllable pairs is as follows: [-T, -B]: 15 pairs; [-T, +B]: 10 pairs; [+T, -B]: 1 pair; [+T, +B]: 8 pairs. Category [-T, -B] takes the highest proportion, i.e. it is more often to have both f_{Max} and f_{Min} of successive same-tone syllables moving upwards in OSng. Category [+T, +B], which is the majority cases in OSpk, only occupies a small proportion of all the syllable pairs in OSng. This reflects that the utterance shape of OSng is quite different from OSpk. OSng has more freedom to break the downwards moving trend of utterance.

In Table 2-8 for the data of OSng, there is one more column than Table 2-7, which lists the musical notes of the syllables. The musical note in smaller font represents that it is an ornamental note. All the musical notes listed were transcribed by an informant with professional musical training. This informant does not know Cantonese and only transcribed the OSng utterances into musical notes according to her perception of the pitch. The transcribing of the musical notes here can act as a reference, and we can compare them with the actual pitch values.

In our expectation, a pair of adjacent same-tone syllables should have similar pitch and thus should be transcribed as the same musical note. If the musical transcription showed the second syllable higher than the first one, we marked the musical notes with boldface letters. If the second syllable was lower than the first one, the musical notes were marked with italic letters.

From the comparison of the musical notes of successive same-tone syllables, it can be observed that more than half of the syllables were transcribed as the same musical note within a pair, which met our expectation. These cases are distributed amongst all the four categories listed in Table 2-8. It can be inferred that if there is only a slight difference in the actual pitch values (f_{Max} and f_{Min}) between the first and second syllables within a pair, perceptually they still belong to the same pitch register. Thus, more than half of the adjacent same-tone syllables in OSng are perceptually regarded as keeping the same pitch register. It is less frequent to have the second syllable transcribed as a lower note than the first syllable – there are only 3 pairs.

Sometimes (there were 9 pairs) the second syllable even have a higher musical note than the first syllable. Most of these cases happen in Category [-T, -B], with one case in Category [-T, +B]. To explain the special case in Category [-T, +B], we found that the former syllable "分" is the HH variant of T1 and the latter syllable "清" is the HL variant of T1 – no wonder $u > 0$ for f_{Min}. In fact the second syllable uses an obviously higher pitch register and is transcribed with a higher musical note than the first syllable.

In Category [-T, +B], there are two more pairs of syllables in which the first syllable is the HH variant of T1 and the second syllable is the HL variant of T1 – "吞聲" and "濤濤". For the second syllable of the falling variant, the main note is lower than the preceding ornamental note. According to the explanation of an example in Cheung (2007, p. 14), in Cantonese singing, the high-falling tone contour can be realized by adding a higher ornamental note before the main note. Here "聲" and the second "濤" also use similar mechanism to make a falling tone contour.

To sum up the behavior of the adjacent same-tone syllables in OSpk and OSng, the majority of syllables in OSpk have the pattern of [+T, +B], and there is never a pair showing the pattern of [-T, -B]. On the contrary, the most frequent cases in OSng are [-T, -B]. Only a few pairs of the adjacent same-tone syllables in OSng exhibit the pattern of [+T, +B]. Therefore, it can be concluded that the utterance trend of OSpk is undoubtedly downward. This utterance downtrend is dramatically weakened in OSng, and sometimes OSng even shows a tendency to uptrend.

2.3 Results and discussions on OSpk versus NSpk

There were not sufficient pairs of adjacent same-tone syllables within one phrase – there were only 4 pairs in YSF's data. The data were not enough for carrying out statistical analysis similar to that in Section 2.2.3. Thus, we only drew the pitch contours of the whole utterances of OSpk and NSpk for observation in Section 2.3.1, and carried out statistical analysis at the syllable level by methods similar to Section 2.2.2, the details of which will be illustrated in Section 2.3.2.

2.3.1 Comparison of the pitch contours between OSpk and NSpk

In Figure 2-6 to Figure 2-8, each pair of OSpk and NSpk pitch contours for the same-text utterances are juxtaposed in the same graph for comparison.

In each graph, the vertical axis displays the pitch values (in Hz) and the horizontal axis represents the time axis. The OSpk pitch points are denoted by solid triangles while the NSpk pitch points are represented by unfilled circles. The corresponding Chinese characters and tone categories are listed above or below the pitch contours. There are nine intervals (corresponding to the nine syllables, available for filling pitch contours) separated by dashed vertical lines in each graph. The ideal cases are one utterance per graph, but there are some utterances which only consist of two to five syllables. To save space, all the graphs of Figure 2-6 and the middle graph of Figure 2-8 display two short utterances. Between the two utterances in these three graphs, there is an empty interval flanked by solid lines.

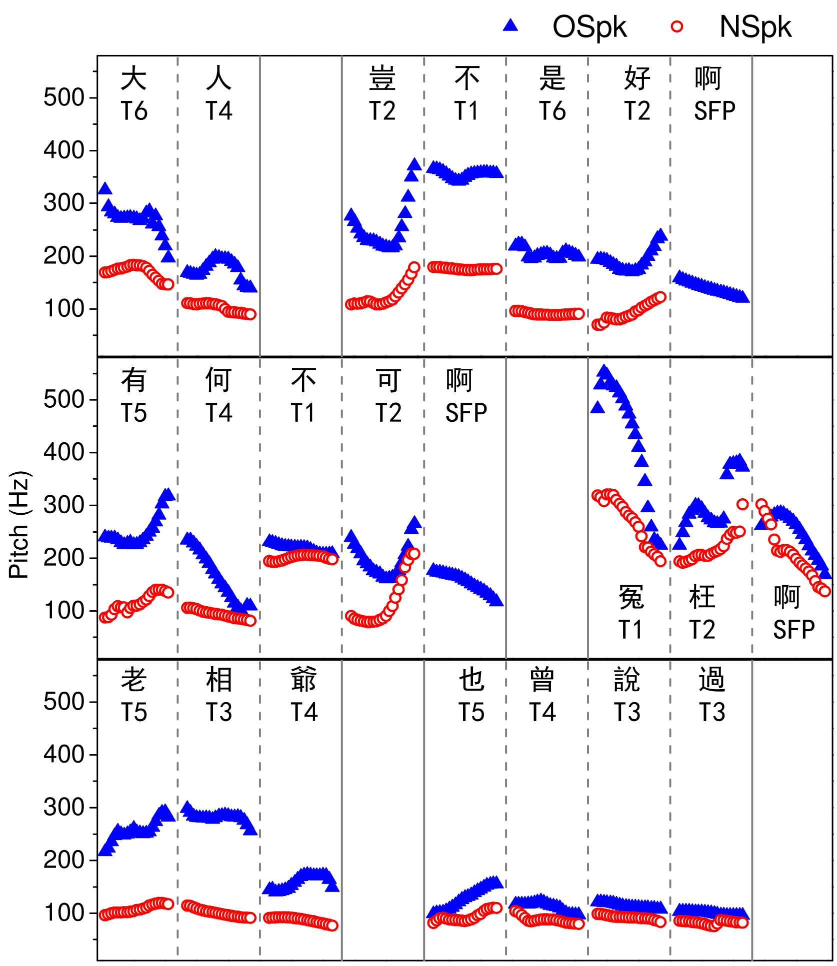

Figure 2-6 Mean pitch contours of OSpk and NSpk utterances (1)

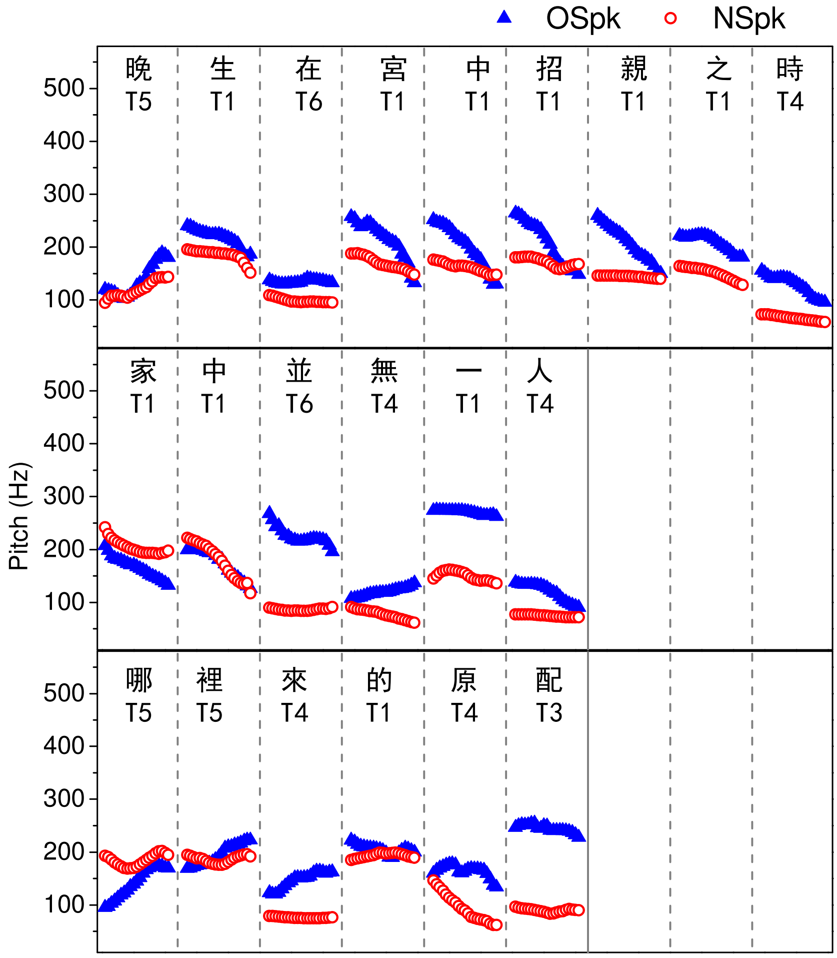

Figure 2-7 Mean pitch contours of OSpk and NSpk utterances (2)

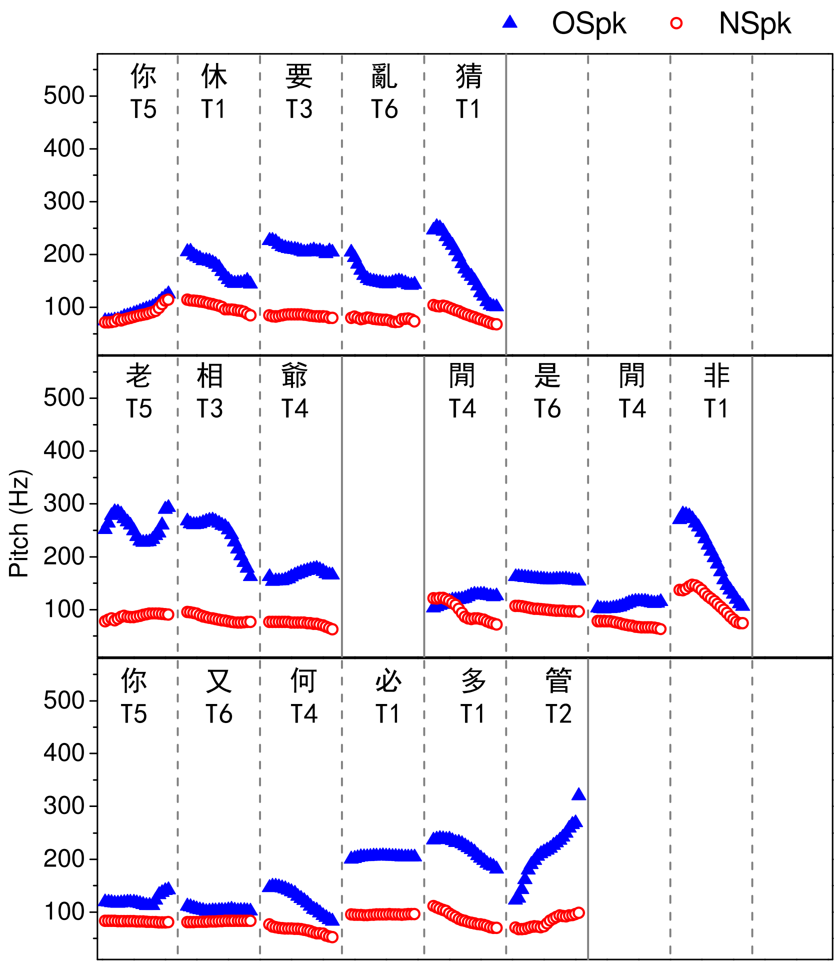

Figure 2-8 Mean pitch contours of OSpk and NSpk utterances (3)

By comparing the OSpk and NSpk pitch contours in Figure 2-6 through Figure 2-8, it can be observed that OSpk pitch contours are usually much higher than the corresponding NSpk pitch contours. Only some of the syllables in the last two graph of Figure 2-7 have higher NSpk pitch contours than OSpk.

In terms of pitch slope, some falling pitch contours in OSpk decline more dramatically than in NSpk, such as "冤" in the second graph of Figure 2-6, "時" in the first graph of Figure 2-7, "猜" and "非" in the first and second graphs of Figure 2-8. Sometimes the rising tones also have a steeper ascending slope in OSpk, like "哪" in the last graph of Figure 2-7, "管" in the final graph of Figure 2-8. Making the pitch slope much more tilted (i.e., making the fall or rise more conspicuous), is probably an effective method to emphasize that the syllables or the utterances are stressed, as the OSpk utterances carry more exaggerated intonation. There are also some cases showing the opposite behavior, e.g., "原"

in the last graph of Figure 2-7 has a more sharply falling pitch slope in NSpk, “可” in the second graph of Figure 2-6 has a more sharply rising pitch slope in NSpk. The above are only our preliminary observations on the pitch register and pitch slope of syllables in OSpk and NSpk and we are not sure whether these differences are significant or not. Further analysis will be conducted in Section 2.3.2.

At the utterance level, we would like to discuss two things applicable to some examples of successive same-tone syllables, namely the behavior of final syllables and the addition of sentence-final particles (**SFP**).

Though we cannot conduct statistical analysis at the utterance level using methods similar to Section 2.2.3 because of the data limitations, we can still make preliminary observations of the successive same-tone syllables in the graphs. Here are two examples of successive same-tone syllables within a phrase: for the two T3 syllables (“說過”) in the last graph of Figure 2-6, and the two T1 syllables (“家中”) in the middle graph of Figure 2-7, both OSpk and NSpk have lower pitch contours for the second syllable. It can be inferred that both OSpk and NSpk show a declining utterance trend.

From our preliminary observations, utterance-final syllables usually increase their pitch slope in OSpk compared with NSpk, i.e., pitch contours of level/falling tones are more declining, and rising tones are more ascending in OSpk. Probably a more tilted pitch slope for the final syllable of an utterance can enhance the expression of dramatic emotion in OSpk.

As mentioned in Section 2.1.1, in some utterances, there is an SFP in OSpk which is missing in NSpk. There are two examples of such cases displayed in Figure 2-6: the second short utterance in the first graph and the first short utterance in the second graph. YSF was supposed to pronounce the OSpk and NSpk versions of the same text but he did not strictly stick to “the same text” and added an extra SFP at the end of the OSpk version. Probably this was caused by the more dramatic emotion in the opera performance. If the emotion is too strong, a better way to express it is to append an SFP at the end of the utterance in addition to showing it in the pitch configuration. This is why YSF could not help adding SFP in OSpk.

2.3.2 Calculation and analysis at the syllable level

Similar to Section 2.2.2, here Figure 2-9 was drawn to directly reflect the mean pitch distribution of different tones in OSpk and NSpk, with Table 2-9 as a follow-up supplement for more accurate data of the pitch values at the beginning points, the pitch values at the ending points, as well as the data of the *r* values.

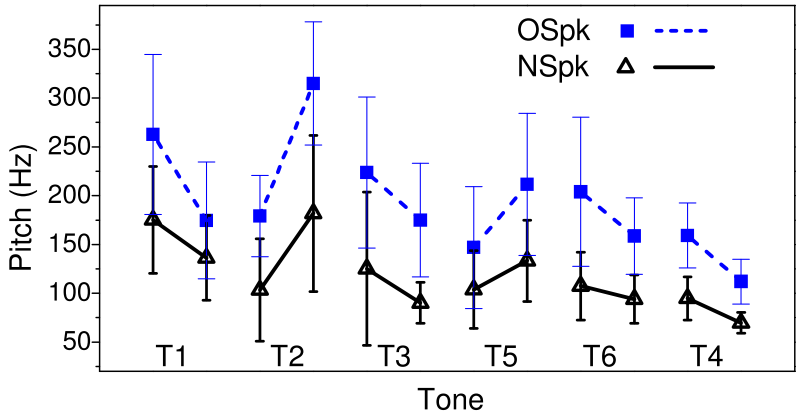

Figure 2-9 Schematic mean pitch contours in OSpk and OSng, with error-bars indicating ±1 standard deviation.

Table 2-9 Mean values of pitch and *r* in OSpk and NSpk

Tone Category	Speech Styles	Pitch (Hz)		*r*
		$f_{Beginning}$	f_{Ending}	
T1	OSpk	262.76	174.57	0.6822
	NSpk	175.27	136.28	0.7836
T2	OSpk	179.15	314.94	1.812
	NSpk	103.39	181.83	1.814
T3	OSpk	223.76	174.98	0.8052
	NSpk	125.13	90.157	0.8078
T4	OSpk	159.17	111.96	0.7181
	NSpk	94.785	69.653	0.7578
T5	OSpk	146.85	211.58	1.499
	NSpk	103.76	133.34	1.317
T6	OSpk	203.85	158.60	0.8156
	NSpk	107.39	93.970	0.8851

In Figure 2-9, the mean values of $f_{Beginning}$ and f_{Ending} in OSpk are denoted by solid square symbols and the counterparts in NSpk are represented by unfilled triangle symbols. These mean points are connected by dashed lines to approximate pitch contours. At the mean points, error-bars are drawn to show the standard deviation. The tone categories are arranged in the same order as in Figure 2-5 (Page 54), with tones arranged according to tone register from high to low, and the level/falling tones preceding the rising tones within the same tone register.

It can be observed from Figure 2-9 that the pitch register in OSpk is generally higher than in NSpk. To judge whether this difference is significant, independent-sample *t*-tests are conducted. The results for the data of *r*, $f_{Beginning}$, and f_{Ending} are listed in Table 2-10, Table 2-11, and Table 2-12, respectively.

Table 2-10 Independent-sample *t*-test results on *r* in OSpk and NSpk

Tones	*t*	*p*	*df*
T1	1.740	0.091	34
T2	0.007	0.995	8
T3	0.030	0.976	12
T4	0.735	0.469	24
T5	-1.735	0.102	16
T6	1.289	0.222	12

Table 2-11 Independent-sample *t*-test results on $f_{Beginning}$ in OSpk and NSpk

Tones	*t*	*p*	*df*
T1	-3.767	0.001	34
T2	-2.528	0.035	8
T3	-2.369	0.035	12
T4	-5.830	0.000	24
T5	-1.743	0.104	16
T6	-3.043	0.010	12

Table 2-12 Independent-sample *t*-test results on f_{Ending} in OSpk and NSpk

Tones	*t*	*p*	*df*
T1	-2.199	0.035	34
T2	-2.924	0.019	8
T3	-3.634	0.003	12
T4	-6.031	0.000	24
T5	-2.795	0.013	16
T6	-3.693	0.003	12

In terms of pitch sloping, Figure 2-9 and the mean values of *r* in Table 2-9 both reveal that the falling or rising tendency of the intrinsic tone contour is strengthened in OSpk compared with the tone contour in NSpk, i.e., the falling tones are more declining and rising tones are more ascending in OSpk. This is consistent with what we observed from the graphs of Figure 2-6 (Page 65), Figure 2-7 (Page 66), and Figure 2-8 (Page 67) in Section 2.3.1. However, the *t*-test results reveal that the difference of *r* between OSpk and NSpk is not significant: the *p*-values are all above the threshold of 0.05 in Table 2-10. This indicates that the pitch slopes of syllables in OSpk versus NSpk have some differences but the differences are not statistically significant.

In terms of pitch register difference, it can be observed from Figure 2-9 that all the mean pitch contours of OSpk are located above the mean pitch contours of NSpk. Only the -1 standard deviation of OSpk and +1 standard deviation of NSpk partly overlap. The independent-sample *t*-test results listed in Table 2-11 and Table 2-12 show that all the *p*-values except the beginning point of T5 (p = 0.104) are below 0.05. This verifies that the pitch height of most pitch points is significantly higher in OSpk than in NSpk. Although the beginning point of T5 has considerable overlap between OSpk and NSpk, the mean value of OSpk is still above NSpk, as displayed in Figure 2-9.

Therefore, the key factor to distinguish OSpk and NSpk is pitch register. Pitch register of OSpk is evidently higher than NSpk. Pitch contour sloping shows a slight difference but is not significant.

2.4 Results and discussions on NSpk and NSng

In this experiment, we have more control over the materials that we need. The sample size in this part of the experiment is also much larger than those for the first two pairs of speech styles. Thus, the difference between NSpk and NSng will be more fully studied. In addition to the analysis on the pitch patterns of syllables (in Section 2.4.1) as well as the whole utterances (in Section 2.4.2), we also studied the utterance-final intonation by comparing the differences between utterance-final and non-final syllables in these two styles of speech, which will be discussed in Section 2.4.3. The present section (Section 2.4) is related to and anticipates Experiment II of Chapter 3, as both are comparisons on the speech styles of normal speaking and normal singing. The research on the data of the informant ZSX here is a pilot study, which can help us to make preliminary conclusions. More informants were to be recruited and the experimental methodology was to improve in Experiment II to further explore the different intonation patterns in normal speaking and normal singing.

2.4.1 Calculation and analysis at the syllable levels

Similar to the method used in Section2.2.2 and Section 2.3.2, we first drew the mean pitch distributions of NSpk and NSng in Figure 2-10, with Table 2-13 listing the mean values of $f_{\text{Beginning}}$, f_{Ending} and r, and then carried out statistical tests (results listed in Table 2-14 to Table 2-16) to examine whether the differences observed from Figure 2-10 and Table 2-13 are significant.

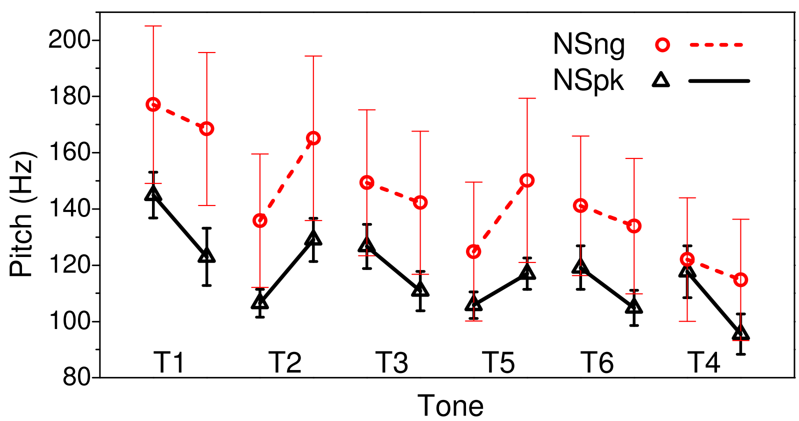

Figure 2-10 Schematic mean pitch contours in NSpk and NSng, with error-bars indicating ±1 standard deviation.

Table 2-13 Mean values of pitch and *r* in NSpk and NSng

Tone Category	Speech Styles	Pitch (Hz)		*r*
		$f_{Beginning}$	f_{Ending}	
T1	NSpk	144.89	122.91	0.8502
	NSng	177.10	168.47	0.9515
T2	NSpk	106.48	129.01	1.213
	NSng	135.84	165.09	1.218
T3	NSpk	126.71	110.82	0.8765
	NSng	149.35	142.26	0.9520
T4	NSpk	117.73	95.52	0.8158
	NSng	122.03	114.75	0.9395
T5	NSpk	105.78	116.99	1.107
	NSng	124.83	150.13	1.208
T6	NSpk	119.10	104.80	0.8815
	NSng	141.10	133.91	0.9489

Table 2-14 Independent-sample *t*-test results on *r* in NSpk and NSng

Tones	*t*	*p*	*df*
T1	-16.643	0.000	385
T2	-0.410	0.683	152
T3	-14.917	0.000	293
T4	-13.041	0.000	154
T5	-8.193	0.000	156
T6	-14.746	0.000	290

Table 2-15 Independent-sample *t*-test results on $f_{\text{Beginning}}$ in NSpk and NSng

Tones	*t*	*p*	*df*
T1	-15.349	0.000	385
T2	-10.597	0.000	152
T3	-10.036	0.000	293
T4	-1.594	0.114	154
T5	-6.892	0.000	156
T6	-10.220	0.000	290

Table 2-16 Independent-sample *t*-test results on f_{Ending} in NSpk and NSng

Tones	*t*	*p*	*df*
T1	-21.767	0.000	385
T2	-10.475	0.000	152
T3	-14.331	0.000	293
T4	-7.473	0.000	154
T5	-10.168	0.000	156
T6	-14.159	0.000	290

In Figure 2-10, the mean values of pitch points in NSpk and NSng are represented by unfilled triangles and unfilled circles, respectively. The mean values of $f_{\text{Beginning}}$, and f_{Ending} are connected to approximate the pitch contours of tones, with solid lines for NSpk and dashed lines for NSng. Error-bars are also shown to indicate the standard deviation. Like Figure 2-5 (Page 54) and Figure 2-9 (Page 69), the tone categories are arranged according to the tone registers and tone shapes, and in the order of T1 -> T2 -> T3 -> T5 ->T6 ->T4.

Compared with the data of MCS in Figure 2-5 (Page 54) and the data of YSF in Figure 2-9 (Page 69), the overall tessitura of ZSX, which is displayed in Figure 2-10, is much narrower. Even when both the lay informant ZSX and the professional Cantonese Opera actor YSF are in the same speech style of normal speaking, ZSX only occupies a tessitura from about 90 to 150 Hz, while YSF ranges from about 50 to 275 Hz.

For the differences between NSpk and NSng, firstly we compare the pitch contour sloping. From Figure 2-10, we can observe that the pitch contours of level and falling tones decline more sharply in NSpk than in NSng. For rising tones, T5 has steeper slope in NSng than in NSpk, like the performance of MCS, but T2 does not show evident difference between ZSX's NSpk and NSng.

The independent-sample *t*-tests of *r* values were applied to help us judge whether the tones show significant difference in these two types of speech, the results of which are listed in Table 2-14. It turns out that *p*-values of all tones are below 0.001 except T2 ($p = 0.683$). Thus, *r* is significantly higher in NSng than in NSpk except for T2. The pattern indicates that falling/level tones have significantly less declining pitch contours and one of the rising tones T5 has a significantly more ascending pitch contour in NSng. Although for the other rising tone T2, the difference between NSpk and NSng is not significant, it is still slightly more ascending in NSng.

Pitch register difference between NSpk and NSng are also compared here. It can be observed that the pitch contours of NSng generally lie above NSpk in Figure 2-10. The error bars of NSng are more dispersed than NSpk, which indicates that NSng has wider pitch span.

Independent-sample *t*-tests results listed in Table 2-15 and Table 2-16 show that all the *p*-values are less than 0.001 except for the beginning point of T4 ($p = 0.114$). The results indicate that the difference of pitch distribution between these two styles of speech is significant for all the tones except the beginning point of T4. For the beginning point of T4, although the difference is not statistically significant, the mean value of NSng is still slightly higher than NSpk. As the pitch slope of T4 in NSng is much less sloping than in NSpk, actually the mean register for the entire pitch contour of T4 in NSng lies far above its counterpart in NSpk, like the pattern of other tones. The overall pattern of all the tones shows that ZSX tends to use a significantly higher pitch register in NSng than in NSpk.

Therefore, the overwhelmingly major pattern shows that both pitch slope and pitch register are significantly different in NSpk and NSng. Compared with NSpk, NSng has an anti-declining effect for pitch slope, as well as a higher pitch register.

2.4.2 Calculation and analysis at the utterance level

To reveal the utterances' pitch direction in NSpk and NSng, we also adopted the method used in Section 2.2.3, i.e., we compared the pitch of adjacent same-tone syllables within a phrase, judged whether the pitch difference of first and second syllables (represented by the parameter u) is positive or negative, and identified the main pattern of u in NSpk and NSng, respectively.

The comparison results are listed in Table 2-17, which show that in NSpk, the dominant result for both f_{Max} and f_{Min} is $u > 0$. In NSng, the dominant result of f_{Max} is $u < 0$, while the comparison of f_{Min} does not show a strong preference, as the number of $u > 0$ pairs and the number of $u < 0$ pairs are quite close (41 versus 46).

Table 2-17 Statistical analysis on adjacent same-tone syllable pairs in NSpk and NSng

u	NSpk		NSng	
	f_{Max}	f_{Min}	f_{Max}	f_{Min}
> 0	54	61	30	41
= 0	0	0	1	0
< 0	33	26	56	46

The patterns observed in NSpk and NSng are completely parallel to our comparison on MCS's OSpk and OSng. Preceding syllables tend to have higher pitch than following syllables in speaking while this tendency is obviously weakened in singing. Thus, speaking utterances show a strong preference for declination, while singing utterances have more freedom not to follow this tendency, and even exhibit a kind of upward trend in the f_{Max} comparison.

2.4.3 Comparison between final and non-final syllables

Most of the lyrics presented to ZSX are declarative sentences. Through the analysis on utterance-final syllables and non-final syllables, we tried to find out whether speaking and singing exhibit sentence-final declination and, if they do, whether this sentence-final declination involves pitch register lowering or pitch slope declining.

Table 2-18 Mean values of pitch and *r* of final and non-final syllables in NSpk

Tone Category	Final / Non-final	Pitch (Hz)		*r*
		$f_{Beginning}$	f_{Ending}	
T1	Final	144.40	112.75	0.7816
	Non-final	144.96	124.22	0.8591
T2	Final	101.21	124.32	1.228
	Non-final	107.86	130.4	1.209
T3	Final	123.87	105.48	0.8541
	Non-final	127.33	112.00	0.8814
T4	Final	120.23	90.77	0.7623
	Non-final	116.92	97.06	0.8331
T5	Final	102.36	112.83	1.103
	Non-final	106.66	118.06	1.108
T6	Final	115.09	98.04	0.8537
	Non-final	119.60	105.63	0.8849

Table 2-19 Independent-sample *t*-test results on *r* of final and non-final syllables in NSpk

Tones	*t*	*p*	*df*
T1	-4.680	0.000	191
T2	1.049	0.298	75
T3	-2.227	0.028	142
T4	-3.561	0.001	76
T5	-0.412	0.682	81
T6	-2.463	0.015	144

Table 2-20 Independent-sample *t*-test results on $f_{Beginning}$ of final and non-final syllables in NSpk

Tones	*t*	*p*	*df*
T1	-0.299	0.766	191
T2	-5.651	0.000	75
T3	-2.053	0.042	142
T4	1.376	0.173	76
T5	-3.623	0.001	81
T6	-2.228	0.027	144

Table 2-21 Independent-sample *t*-test results on f_{Ending} of final and non-final syllables in NSpk

Tones	*t*	*p*	*df*
T1	-5.302	0.000	191
T2	-2.870	0.005	75
T3	-4.590	0.000	142
T4	-3.563	0.001	76
T5	-3.716	0.000	81
T6	-4.925	0.000	144

Table 2-18 lists the mean $f_{Beginning}$, f_{Ending} and r values of final and non-final syllables in NSpk. The corresponding independent-sample *t*-test results on final and non-final syllables are listed in Table 2-19 through Table 2-21.

The r values of final syllables are smaller than non-final syllables for level and falling tones in Table 2-18. The r values of rising tones do not have this tendency: $r_{Final} > r_{Non\text{-}final}$ for T2, while $r_{Final} \approx r_{Non\text{-}final}$ for T5. Also, *t*-tests show that the r values are significantly different between final and non-final syllables for level and falling tones while the difference is not obvious for rising tones.

The data of $f_{Beginning}$ and f_{Ending} in Table 2-18 show that the pitch of final syllables is lower than that of non-final syllables for both points in all tones except the beginning points of T1 and T4. These two points also show exceptions in *t*-tests and do not have significant difference between final and non-final syllables. For the other tones, which represent the predominant pattern, at both the beginning point and at the ending point, the pitch values of the final syllables are significantly lower than those of the non-final syllables in NSpk.

Therefore, for level/falling tones, $f_{\text{Beginning}}$, f_{Ending} and r values drop dramatically when syllables are at the final position, which indicates that the falling tendency is evidently strengthened at the end of utterances in normal speaking. The r values of rising tones do not have obvious difference, thus the utterance-final declination effect on rising tones are only reflected by pitch register lowering rather than pitch contour slope change.

The counterpart data of NSng are listed in Table 2-22. Independent-sample *t*-test results on final and non-final syllables are listed in Table 2-23 to Table 2-25.

Different from the NSpk data in Table 2-18, $f_{\text{Beginning}}$, f_{Ending} and r values for NSng do not show a fixed direction between final syllables and non-final syllables in Table 2-22. Moreover, *t*-test results in Table 2-23, Table 2-24 and Table 2-25 indicate that the differences between these two types of syllable are not statistically significant among most tones. Although the *p*-values of $f_{\text{Beginning}}$ and f_{Ending} of T3 are less than 0.05, they are very close to 0.05: for $f_{\text{Beginning}}$, $p = 0.047$; for f_{Ending}, $p = 0.048$. The exception of T3 does not override the dominant pattern that final syllables and non-final syllables are almost equal in singing, both in terms of pitch slope and pitch register.

Table 2-22 Mean values of pitch and *r* of final and non-final syllables in NSng

Tone Category	Final / Non-final	Pitch (Hz)		*r*
		$f_{\text{Beginning}}$	f_{Ending}	
T1	Final	180.71	172.98	0.9569
	Non-final	176.64	167.89	0.9508
T2	Final	133.78	161.50	1.205
	Non-final	136.38	166.04	1.221
T3	Final	140.37	133.50	0.9507
	Non-final	151.31	144.17	0.9523
T4	Final	115.49	108.94	0.9429
	Non-final	124.13	116.62	0.9384
T5	Final	120.46	142.87	1.193
	Non-final	126.01	152.10	1.212
T6	Final	142.30	135.91	0.9540
	Non-final	140.95	133.66	0.9483

Table 2-23 Independent-sample *t*-test results on *r* of final and non-final syllables in NSng

Tones	*t*	*p*	*df*
T1	0.751	0.453	192
T2	-0.667	0.507	75
T3	-0.329	0.742	149
T4	1.037	0.304	76
T5	-0.648	0.519	73
T6	0.842	0.401	144

Table 2-24 Independent-sample *t*-test results on $f_{Beginning}$ of final and non-final syllables in NSng

Tones	*t*	*p*	*df*
T1	0.640	0.523	192
T2	-0.387	0.700	75
T3	-2.004	0.047	149
T4	-1.501	0.137	76
T5	-0.796	0.429	73
T6	0.206	0.837	144

Table 2-25 Independent-sample *t*-test results on f_{Ending} of final and non-final syllables in NSng

Tones	*t*	*p*	*df*
T1	0.824	0.411	192
T2	-0.550	0.584	75
T3	-1.996	0.048	149
T4	-1.359	0.178	76
T5	-1.126	0.264	73
T6	0.351	0.726	144

To summarize, the comparison between final syllables and non-final syllables shows that there is obvious final declination effect in NSpk. For level/falling tones, the effect involves both lower pitch register and stronger declining pitch contour shape. For rising tones, the declination effect is only reflected by pitch register lowering. NSng is free from the requirement of utterance-final declination, and there is no obvious difference between the final syllables and non-final syllables.

2.5 A perception test

Recently linguists have paid more and more attention to the role of listeners in speech communication (Ohala, 1981). In addition to the acoustic measurements on the differences between speaking and singing, here we also conducted a small-scale perception test to test the findings of our acoustic experiment from the perspective of listeners.

2.5.1 Experimental method of the perception test

For the experimental method of our perception test, the most important information is how we prepared the stimuli. Our stimuli consisted of two parts, utterance stimuli and syllable stimuli. The utterance stimuli required participants to judge whether the stimuli were in a speaking or singing style. The utterance stimuli can be further divided into two types.

The first type was synthesized utterances transformed from MCS's recordings. The synthesizing method was based on our conclusions of the above acoustic experiments. One of MCS's OSng utterances was modified to the OSpk pattern (OSng-OSpk): we dragged down each syllable a little to make their pitch contour more descending than the original shape; we also reduced the pitch register successively through the utterance to make the register of every syllable lower than its preceding syllable. Reversely, one of MCS's OSpk utterances was transformed to OSng pattern (OSpk-OSng): we modified the pitch contour of each syllable to make "level" tones really level and rising tones more rising; we also lifted up the pitch register of every syllable such that it has similar pitch register as its preceding syllable. The synthesis of OSpk-OSng was in a conservative way and settled for a non-declination version rather than an updrift version as such. Utterance-final intonation was not involved in synthesizing the stimuli.

The second type of utterance stimuli were original utterances, including the original utterances of the first type, and several pairs of NSpk and NSng utterances by ZSX. Here listeners serve as inspectors for our hypothesis on singing and speaking. When choosing the utterance stimuli of ZSX, it is required that the NSpk and NSng versions have similar pitch register, and the pitch contours of paired syllables have a lot of overlap between the two versions. If listeners' opinion is consistent with ours, it corroborates our hypothesis that pitch slope rather than pitch register plays the key role for listeners to distinguish speaking and singing.

The syllable stimuli were all extracted from ZSX's utterance stimuli. Thus, all syllable stimuli were segmented from continuous natural speech. NSpk and

NSng versions were presented in pairs. The original order of text was shuffled and the perception test of the syllable stimuli was arranged prior to the utterance stimuli. In this way, participants' judgment on syllables was not influenced by the content of the utterance stimuli. There were 36 pairs of syllables, and the order of NSpk and NSng versions was randomized. According to Wong (1999), the accuracy of tone perception for isolated syllables is quite low, while in carrier sentence, the situation is much better as the carrier sentence provides a reference. Therefore, we inserted the question number (which was also recorded by ZSX) before each pair of syllable stimuli as a reference.

There was a 0.5-second silence between the question number and the first stimulus. A 5-second break was inserted between the first and the second syllable stimuli so that participants can have enough time for decision on the first stimulus. On the questionnaire sheet, the corresponding options were three to four Chinese characters which are only different in tones. Participants were required to identify what they heard and write down their judgment on the two stimuli respectively.

When the sound stimuli were ready, we compiled them into a long sound file, presented the sound file to the participants and asked them to fill in the corresponding questionnaire, which is attached in Appendix E. Ten native informants of Cantonese participated in this experiment.

2.5.2 Results and discussions on the perception test

For the original utterances by ZSX, the judgment of all participants was the same as our hypothesis. For MSC's utterance stimuli, all participants agreed that the original OSng utterance was in singing style and the OSng-OSpk utterance was in speaking style. 70% of the participants considered that the original OSpk utterance was in speaking style, while 80% of the participants judged the OSpk-OSng utterance as a singing utterance.

Therefore, most of the participants' judgments met our expectation. This verifies our judgment that speaking and singing is not biased towards our experimental purposes but is consistent with most Cantonese native informants' intuitions. As a clear majority of listeners judged the synthesized OSpk-OSng utterance to be in a singing style, and the OSng-OSpk utterance was universally judged by the listeners as a speaking utterance, our observation in acoustic experiments can be taken to have captured the essence of the difference between speaking and singing.

The result of the perception test on syllable stimuli is more complicated. Although we tried to adopt utterances with similar pitch register in both NSpk and NSng when selecting ZSX's utterance stimuli, some of the syllables abstracted from these utterances still have a pitch gap between the NSpk and NSng versions. When there is a pitch gap, the NSng version usually has higher pitch than the NSpk version. Consequently, the NSng version tends to be misheard as a higher tone.

Paired-samples *t* tests were conducted to compare the judgment accuracy of the speaking and singing versions of syllables. If all the 36 pairs are taken into account, the correlation $r = 0.069$ ($p = 0.691$); $t = 1.017$; $p = 0.316$; $df = 35$. The data show that there is not evident correlation between these two groups, and the judgment accuracy of the NSpk group and the NSng group is not significantly different.

If the pairs with an obvious pitch gap are excluded, there are 19 pairs of syllables left for statistical analysis. The result then becomes: $r = 0.742$ ($p = 0.000$); $t = 0.960$; $p = 0.350$; $df = 18$. Thus, with the factor of pitch gap taken out, the correlation of judgment accuracy between the NSpk and NSng versions is strong (*r* is high) and significant (the *p* value of *r* is below 0.001), and there is no significant difference between these two groups of data (the *p* value of the *t* test is $0.350 > 0.05$).

The perception results of the syllable stimuli indicate that the declination intonation of speaking and non-declination intonation of singing do not affect the judgment accuracy of tones. The two versions are judged to be the same tone except when there is an obvious pitch gap between the two versions.

Chapter 3
Experiment II: Speaking versus Singing

"Speaking" and "singing" in this chapter refer to normal speaking and normal singing. In Experiment II, we recruit more layman informants to participate in our acoustic studies. Experiment II was conducted under a more explicit and systematic framework – a matrix composed of the two dimensions of pitch configuration and pitch domain. The dimension of pitch configuration has the three factors of pitch slope, pitch register and pitch span. The dimension of pitch domain has two levels – the syllable level and the utterance level, where the utterance level can be further divided into two parts – the utterance-body portion and the utterance-final portion. The definition and calculation method of related parameters will also be clearly shown later in this chapter.

The results of this experiment show that speaking has a constant declining tendency for pitch slope and pitch register at both the levels of syllable and utterance, and there is an extra declination effect at the utterance-final portion. Compared with speaking, singing exhibits a kind of ascending pattern for pitch slope and pitch register in the domains of syllable level and utterance-body portion, and there is no obvious utterance-final intonation. For the factor of pitch span, defined here as the distance between the extreme values, our data shows that it is positively correlated with the steepness of pitch slope – greater steepness usually corresponds to wider pitch span. As the pitch slope downtrend of speaking is usually stronger than the uptrend of singing, the pitch span of speaking is wider than singing at both the levels of syllable and utterance.

3.1 Method of Experiment II

3.1.1 Framework: matrix composed of pitch configuration and pitch domain

Most of the previous studies on intonation are carried out in a descriptive manner and analyse the linguistic data case by case. To more accurately reveal the different intonation patterns in speaking and singing, as well as in different sentence types of speaking, we are not going to describe the characteristics qualitatively, but analyze the distinction between them quantitatively. The framework used in Experiment I is only a rough one. Here we propose a new framework, which can more accurately characterize intonation differences. It consists in a matrix of two dimensions – pitch configuration and pitch domain. Experiment II and Experiment III will be conducted using this framework.

Presuming the canonical form of lexical tones as the default pitch status, intonation can modify the pitch configuration by the mechanisms listed below, which are also depicted in Figure 3-1:

1) Pitch slope: uptrend (+) or downtrend (-);

2) Pitch register: raised (+) or lowered (-);

3) Pitch span: widened (+) or narrowed (-).

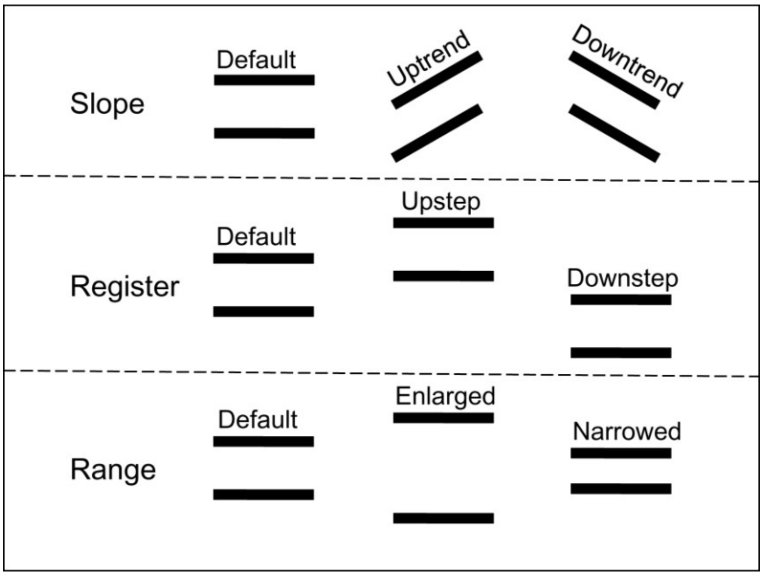

Figure 3-1 Schematic diagrams displaying pitch configuration mechanisms.

To conduct quantitative analysis, we need to define the concrete parameters for pitch slope, pitch register, and pitch span, which will be listed in Section 3.1.4.

In the dimension of pitch domain, the levels and parts for intonation to exert its influence are:

1) Syllable;

2) Utterance: A) Utterance-body portion;

B) Utterance-final portion.

At the syllable level, we mainly investigate the intonation effects on non-final syllable intervals of the utterance. Here syllable intervals refer to the pitch contours of the individual syllables (the utterance is segmented into individual syllables in our analysis).

Phonetically an utterance is a unit of speech bounded by silence. In the present thesis, an utterance coincides with an "intonational phrase". According to Ladefoged (1975/2006, p. 116),

> "The intonation of a sentence is the pattern of pitch changes that occurs. The part of a sentence over which a particular pattern extends is called an intonational phrase."

An utterance divides into two parts for intonation analysis – the utterance-body portion and the utterance-final portion. The utterance-body portion covers all the non-final syllables, treats them as a whole and is concerned with the overall trend. The utterance-final portion only refers to the final syllable of an utterance. The intonation of this part is called utterance-final intonation. The utterance-final portion is a special and important position of an utterance, as it may carry extra burden for expressing a certain kind of intonation effect. The intonation at the utterance-final portion in the present thesis is actually the boundary tone posited at the intonation phrase break (Wong, Chan, & Beckman, 2005).

In previous literature on intonation, only intonation at the utterance level has been studied. Another domain – the effect of intonation on individual syllables has been neglected in previous studies but will be rigorously dealt with in our present study.

With pitch configuration as one dimension and pitch domain as the other, a two-dimensional matrix can be built up. This matrix not only constitutes the frame for the analyses in Experiment II and Experiment III in our present study, but may also serve as a general framework for experimental studies on intonation in tone languages.

In the following sections, we will analyze pitch slope, pitch register and pitch span at different pitch domains. To be more succinct, the analyses are referred to in a short-hand fashion as "[pitch domain] [pitch configuration]", e.g., "syllable slope" means the analyzed pitch domain is the syllable level and the analyzed pitch configuration is the pitch slope factor.

3.1.2 Questionnaire and instructions

The materials of our recording questionnaire consist of two parts, as shown in Appendix F.

Part 1 contains three sets of syllables in isolated citation form, which minimally embody the six-way tonal contrast. Each syllable is to be repeated three times. Thus, there are nine items for each tone category. The informants were asked to pronounce the syllables clearly at a steady speed, with a short pause after every syllable. In this way, we can be certain that these syllables are in "isolated citation form" and are not influenced by adjacent syllables.

Part 2 is a comparison between speaking and singing. There are nineteen utterances of lyrics taken from popular Cantonese pop songs, children's songs, and opera songs. Utterances 1 to 16 were selected for investigation at the syllable level. Each of them has all the six tones within the same utterance so as to make sure that every tone category gets enough items for statistical purpose. Utterances 17 to 19 were adopted for the observation at the utterance level. They contain a series of reduplicative phrases. A reduplicative phrase is a perfect same-tone sequence for our purpose. Studying these monotone sequences can help us study the pitch configuration of the whole utterance.

In this chapter, the following abbreviations are used to mark these three speech styles: isolated citation form – **ICF**; speaking utterances – **SPK**; singing utterances – **SNG**.

To make the recording procedure more efficient, reading aloud of the lyrics was arranged before singing. Original sound tracks of the songs were provided and the informants could listen to the songs repetitively until they were familiar with the melody and were able to sing in tune by themselves. The informants were required to repeat both speaking and singing once. The better version was chosen for measurement and further analysis.

3.1.3 Informants, recording and measurement

There were twelve young native informants of Cantonese participating in this experiment, six male and six female informants aged twenty to thirty. Seven of the informants were recorded in a soundproof room at the Hong Kong Polytechnic University and the other five informants were recorded in quiet rooms at their homes. All the informants were recorded with a Marantz PMD620

Professional Handheld Digital Audio Recorder. The microphone was set at mono channel, the sampling rate was 44,100 Hz, and the resolution was 16-bit. The tessitura differences of individual speakers (especially between-gender differences) were normalized by the application of the Z-Score method (with details in Section 3.1.4).

As before, the acoustic measurements were carried out by using the Praat software (Boersma & Weenink, 2009). For every syllable, measurement was applied to the vocalic portion. We measured f_0 at every 10% of the duration, and then connected these 11 evenly-distributed points to approximate the pitch contour. By this method, time-normalized f_0 contours can be obtained.

This time-normalized method is conventional for tone studies. In experimental studies, it is a norm to display the time-normalized pitch contour based on the mean data. In the descriptive framework of tones, the factor of duration is also disregarded. For instance, the tone transcription method (five-point-scale) proposed by Chao (1930/2006) marks the pitch of the beginning, ending and transition points of a tone, but it does not provide the duration information. When describing tones, linguists are concerned about how much the pitch changes rather than how long this change takes. For example, for the two rising tones T2 and T5 in Cantonese, linguists are concerned about the relative height of their beginning and ending points rather than the duration of these two tones. In the present study, we studied the realization patterns of tones in different contexts, thus the time-normalized convention is adopted.

On some occasions when we hope to observe the duration difference, e.g. in the comparison between the non-final syllables and final syllables in DN and QN[6], the duration difference is shown by the "real-time" tone contours, such as in Figure 4-5 (Page 136) and Figure 4-7 (Page 137). The difference between "real-time" and "time-normalized" tone contours is that for "time-normalized" tone contours, all the contours are of the same length; while for "real-time" tone contours, the length of a tone contour is proportionate to the mean duration of syllables of that tone category.

[6] Generally speaking final syllables are longer than non-final syllables in both DN and QN.

3.1.4 Calculation methods

The raw data of f_0 values, which were in the unit of Hz, were converted to semitones (abbreviated as ST) by

$$f_{\mathrm{ST}} = \frac{12}{\log_{10} 2} \log_{10} \frac{f_{\mathrm{HZ}}}{100}. \qquad \text{Formula 3-1}$$

The reference frequency is set as 100Hz here. The use of semitones can reduce the cross-speaker variation compared with Hz.

To further normalize individual differences, Z-Scores (Jassem, 1971; Menn & Boyce, 1982; Rose, 1987; Zhu, 2005) were adopted here. The calculation method is

$$z_i = \frac{f_{\mathrm{ST}_i} - m}{s}, \qquad \textbf{Formula 3-2}$$

where z_i stands for the Z-Score of point i, $f_{\mathrm{ST}i}$ for semitone value of pitch point i, m for mean value of all the pitch points in the sample (including syllables in isolated citation form, speaking and singing utterances, and utterances of different sentence types in Experiment III) of an informant, and s for standard deviation.

Z-Scores actually locate the targeted pitch points according to their relative height within the tessitura of a certain informant. It is a kind of relative pitch that normalizes individual absolute tessitura differences. After such normalization, further calculations and analyses can be carried out for the data of all informants.

In the dimension of pitch domain, when analyzing data at the syllable level, the final syllables were excluded. In the analysis of final intonation, final syllables and non-final syllables were separated for calculation (using the same method as calculating data at the syllable level), and further comparison and analysis was conducted to examine whether final syllables and non-final syllables have significant differences. To reveal the patterns of utterance-body intonation, only the reduplicated phrases in Utterance 17 to 19 were investigated as they are monotone sequences and can therefore directly reflect the trend of utterance.

In the dimension of pitch configuration, certain parameters were chosen to represent pitch slope, pitch register and pitch span, which are listed below.

Pitch slope: The slope of the linear regression of Z-Score. The linearly fitted Z-Score is given by

$$z = at + b, \qquad \textbf{Formula 3-3}$$

where a is the slope and our focus, t is the normalized time percentage, and b is the intercept.

If $a > 0$, the fitted line is rising; if $a < 0$, the line is falling; if $a = 0$, the line remains horizontal. The pitch slope tends to be steeper if the absolute value of a is greater. Thus, for rising slope ($a > 0$), a greater a indicates a steeper slope; for falling slope ($a < 0$), a greater a reflects a flatter slope, and vice versa.

Syllable slope and utterance slope have different sampling methods. For syllable slope, linear regression is conducted on all eleven pitch points within the syllable; for utterance slope, linear regression is conducted on the median points of every syllable within the utterance.

Actually pitch slope can be defined in two ways. One is to take duration into consideration and the other is not. The first way may sound more accurate, but in the present study it is more practical to adopt the second way.

Firstly, as a quantitative study, the concern of the thesis is the patterns reflected by the mean values of the large amount of pitch data. The basis for calculating mean tone contours is the time-normalized method, as introduced in Section 3.1.3 (Page 88). Without normalizing time, the mean tone contours cannot be calculated. To be consistent, it is more convenient to regard pitch slope as the slope of the time-normalized pitch contour.

Secondly, when describing the pitch slope of a tone, it is the convention that the duration is disregarded. For example, the rising slope of T2 (25, high-rising) can be regarded to be steeper than T5 (23, mid-rising) in Cantonese, without regard to their actual duration. The underlying assumption is the prosodic feature of syllable isochrony. Cantonese is a syllable-timed language (Bauer & Benedict, 1997; Cheung, 1986; Mok, 2008) and syllable is regarded as an isochronic unit, i.e. the bearing unit of tone is isochronic. Thus, it is needless to refer to duration.

Thirdly, most lyrics selected for the questionnaire are from songs of moderate speech rate, and generally there are not great duration differences between speaking and singing. The duration effect can be balanced off among the aggregate data. Thus, we disregarded the duration differences when calculating pitch slope.

Pitch register: The median (represented as c) of pitch points within a syllable or an utterance.

As pitch register is used to denote the relative height of an overall pitch space, two kinds of parameters can be taken into consideration – the mean value and the median value. The median value was chosen to avoid the fluctuation of boundary pitch points. Occasionally the pitch points at the boundaries may have extremely high or low values, but the overall pitch level is actually not that high or low. For this reason, the median value is more stable than the mean value.

Pitch span: The distance *d* between extreme values of the pitch points within a syllable or an utterance:

$$d = z_{\text{Max}} - z_{\text{Min}}. \qquad \textbf{Formula 3-4}$$

From the definitions of pitch slope and pitch span given above, it can be inferred that these two parameters are interrelated: a steeper slope associates with a wider span, and vice versa. However, pitch slope cannot substitute for pitch span because pitch span can not signify the direction of pitch slope (whether the slope is falling or rising). Conversely, pitch span cannot be substituted as it is defined to outline the magnitude of the pitch space. Together with the factor of pitch register, which locates the relative height, a pitch space can be fixed.

3.2 Results of mean pitch contours of speaking and singing utterances

According to the method introduced in Section 3.1.4, we calculated the Z-Score value at every measuring point for every informant. We also calculated the mean value of all the informants at every measuring point. The mean pitch points within a syllable constitute the mean pitch contour of the syllable. Figure 3-2 to Figure 3-10 display the mean speaking and singing pitch contours of all the syllables we measured for Utterance 1 to 19. For Utterance 1 to 16, all the syllables in the utterance were measured and displayed in the figures. For Utterance 17 to 19, only the reduplicated phrases and the final syllables were measured and displayed; the other syllables of these utterances were not adopted for calculation and analysis.

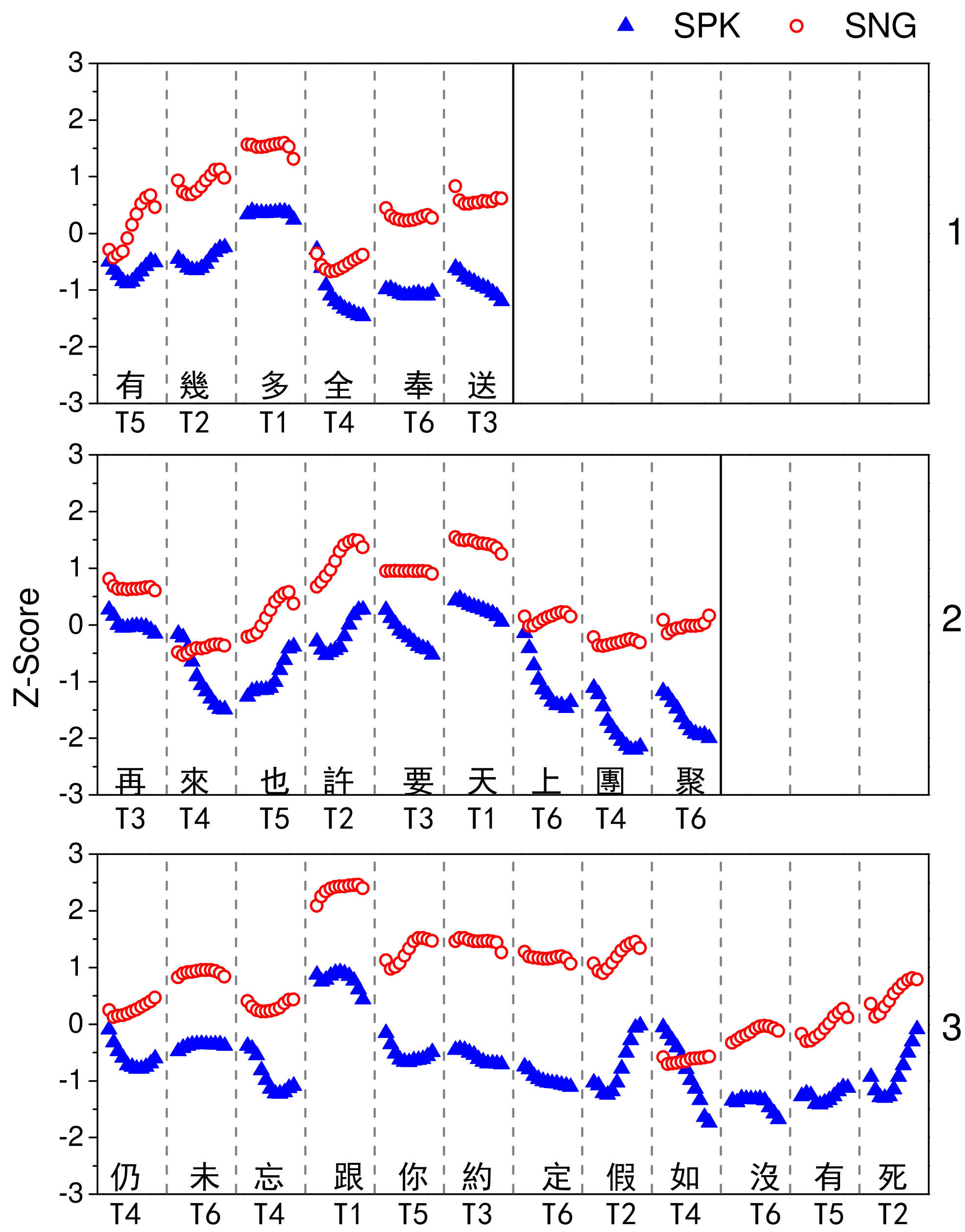

Figure 3-2 Mean SPK and SNG pitch contours of Utterance 1 to 3.

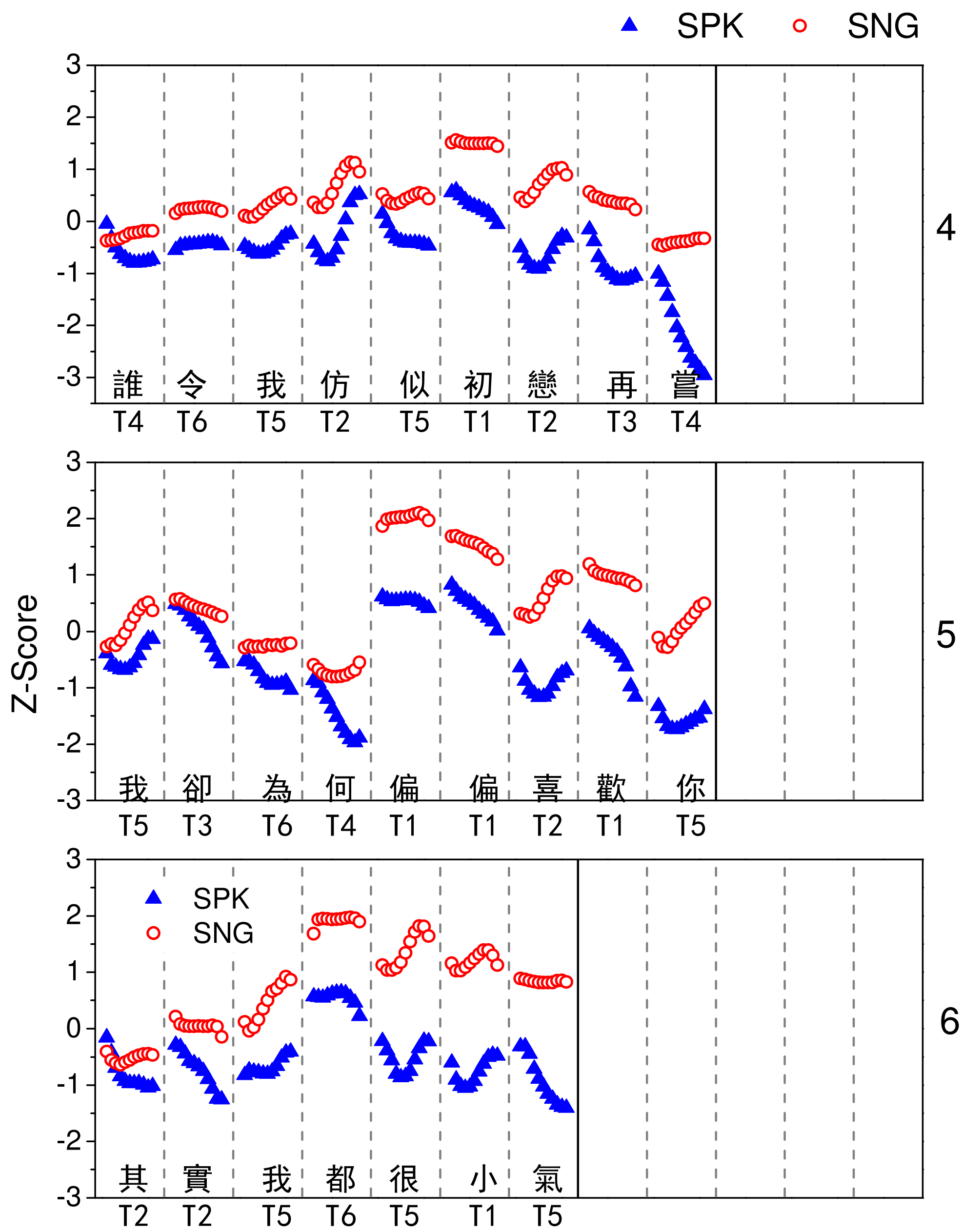

Figure 3-3 Mean SPK and SNG pitch contours of Utterance 4 to 6.

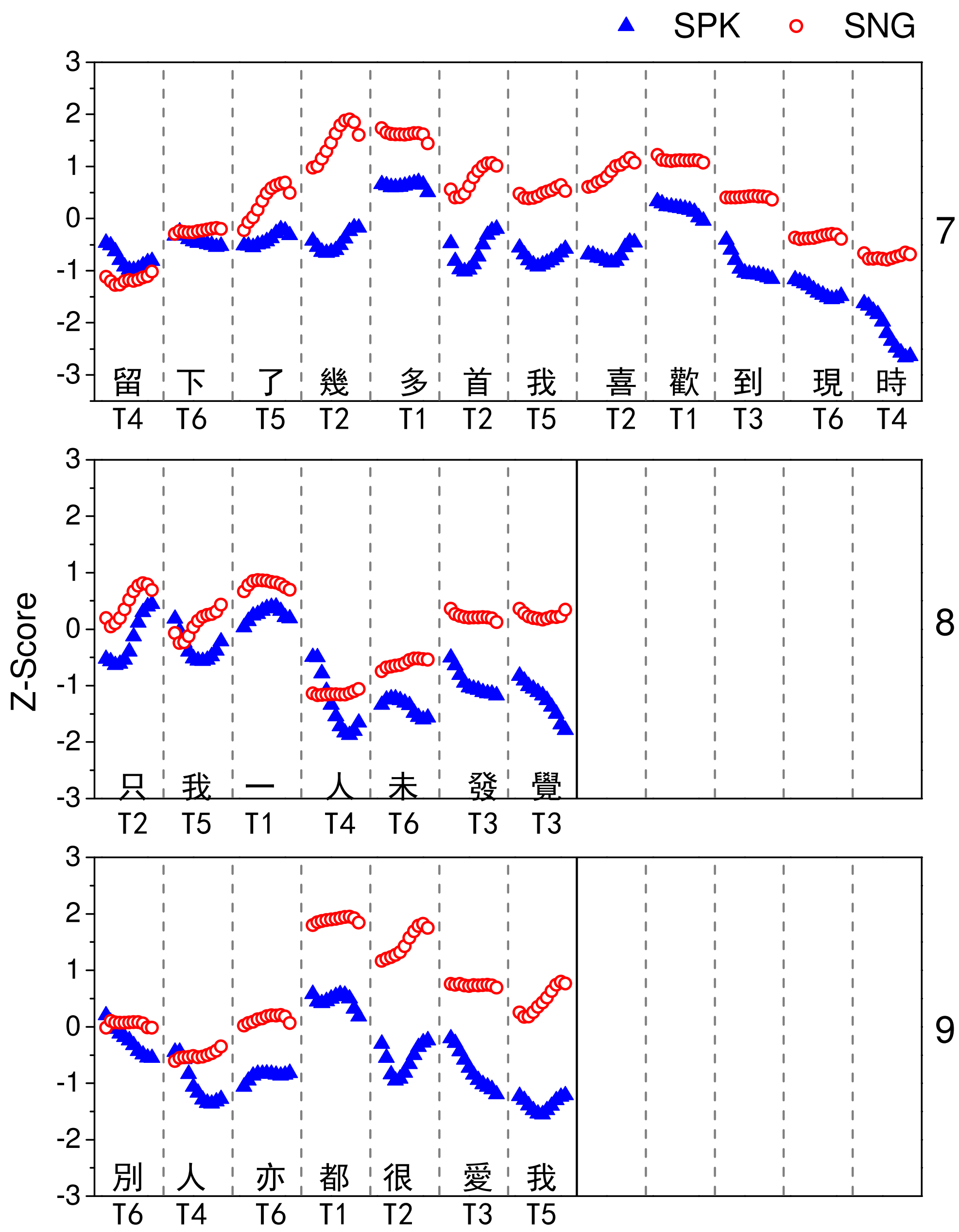

Figure 3-4 Mean SPK and SNG pitch contours of Utterance 7 to 9.

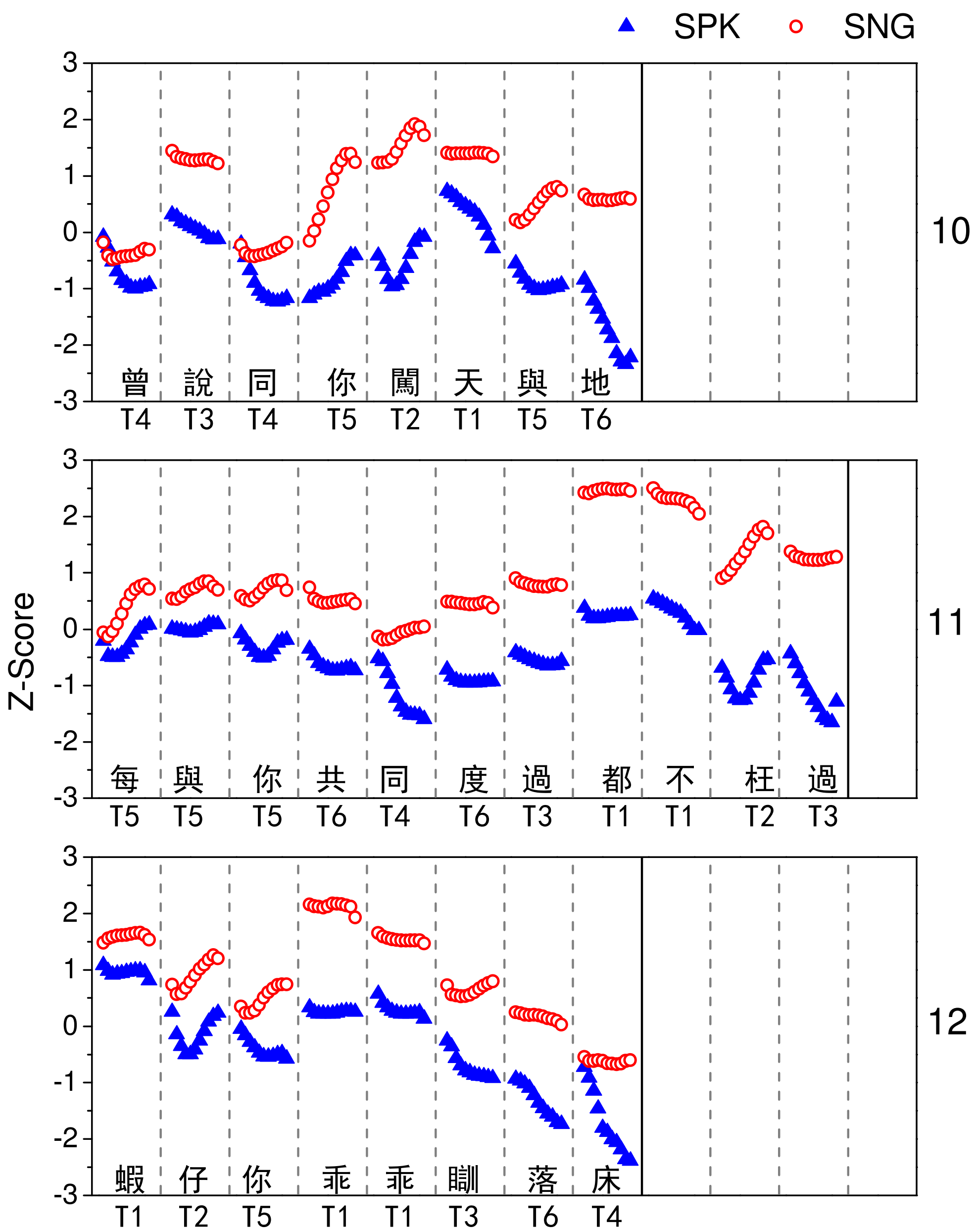

Figure 3-5 Mean SPK and SNG pitch contours of Utterance 10 to 12.

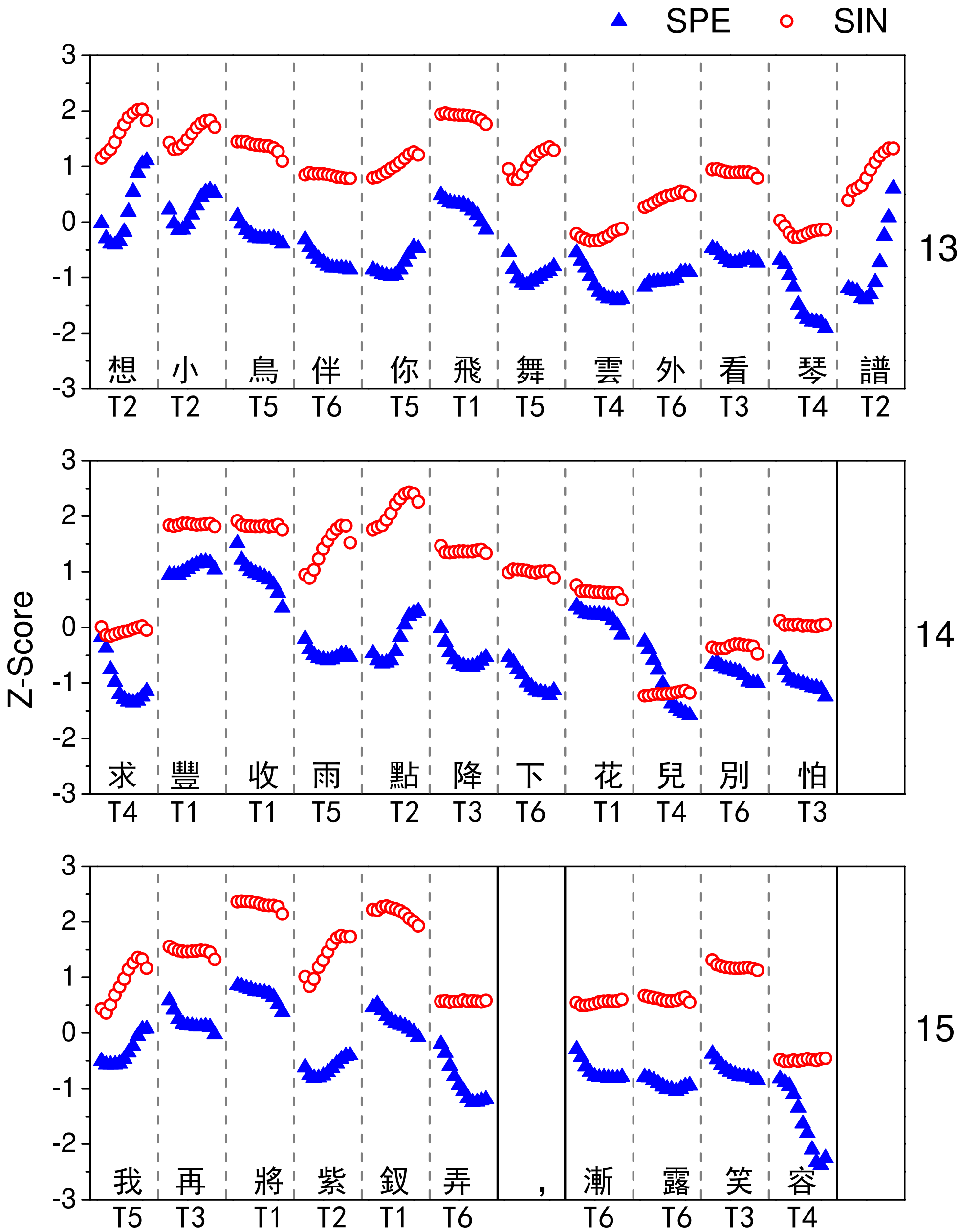

Figure 3-6 Mean SPK and SNG pitch contours of Utterance 13 to 15.

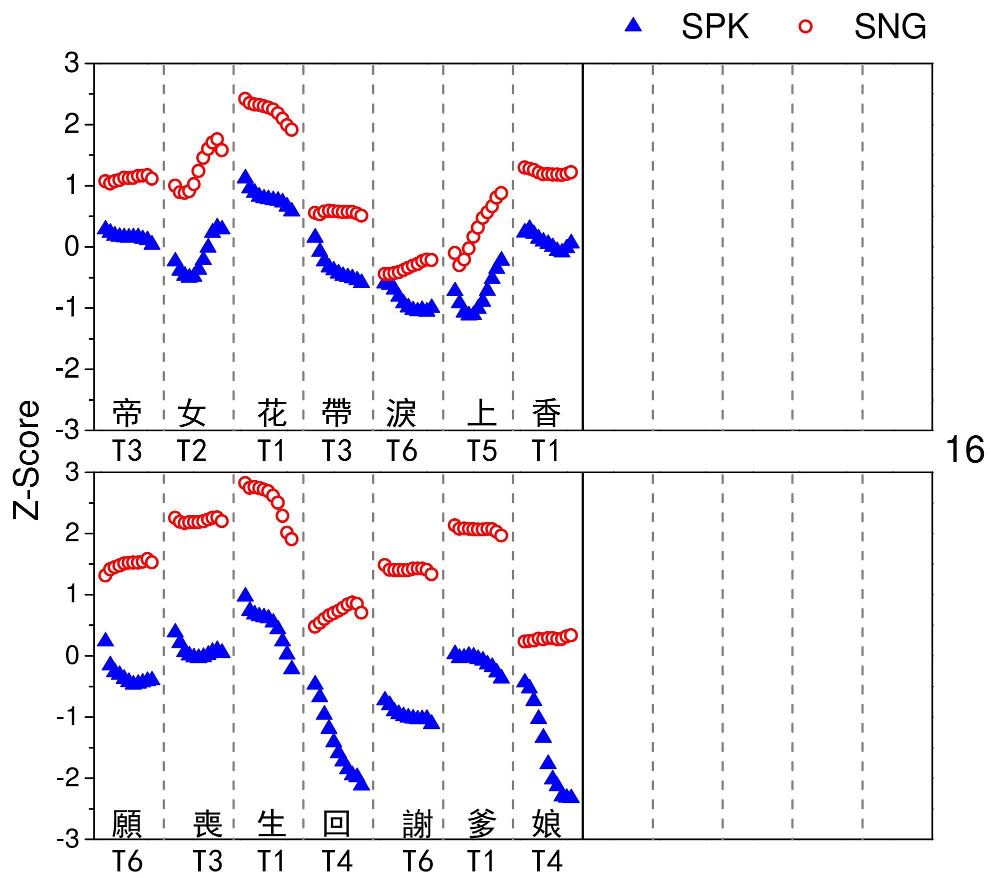

Figure 3-7 Mean SPK and SNG pitch contours of Utterance 16.

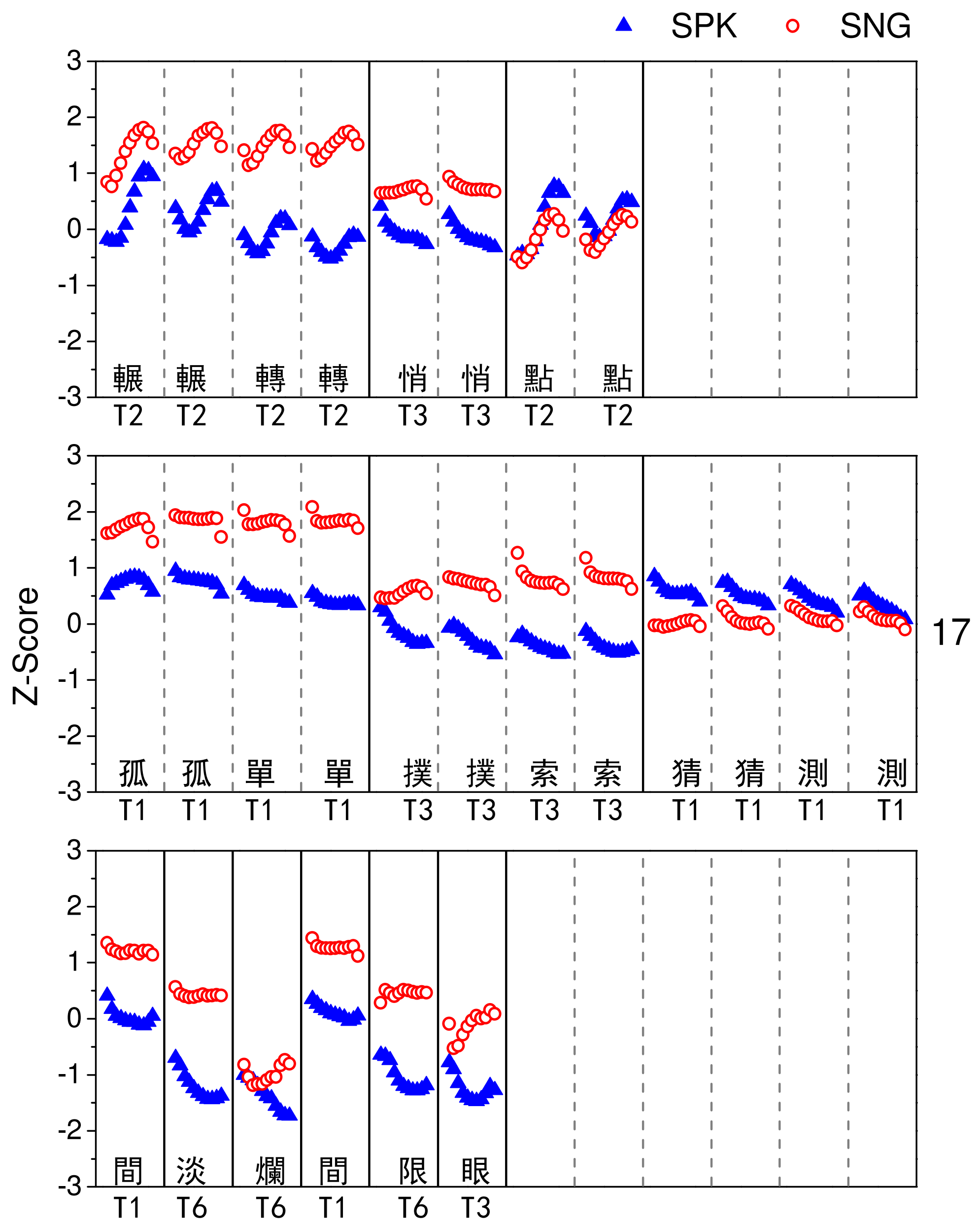

Figure 3-8 Mean SPK and SNG pitch contours of Utterance 17.

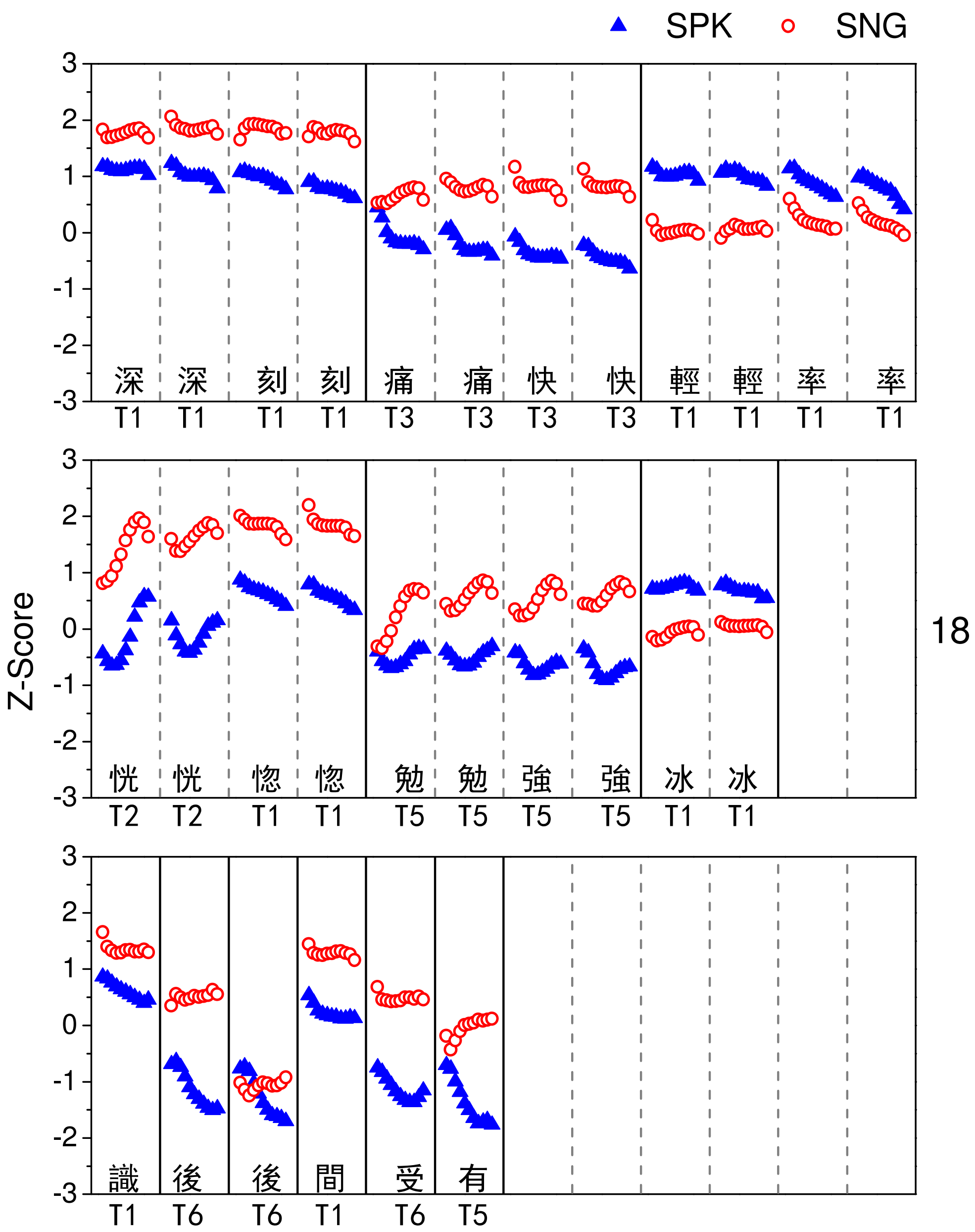

Figure 3-9 Mean SPK and SNG pitch contours of Utterance 18.

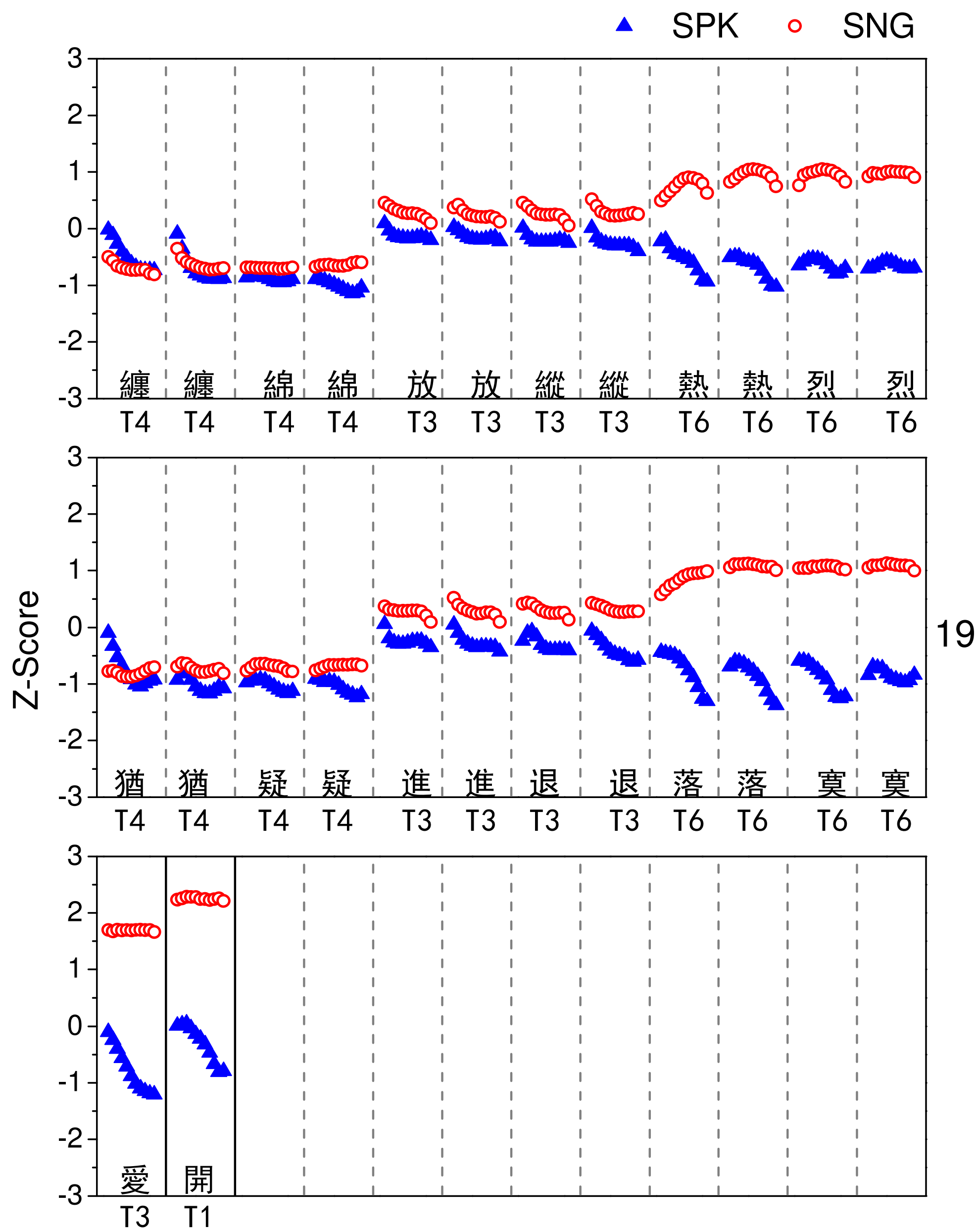

Figure 3-10 Mean SPK and SNG pitch contours of Utterance 19.

In each figure, the vertical axis represents the Z-Score value and the horizontal axis represents time. The speaking version and singing version are juxtaposed in each figure to show the difference between these two speech styles. The speaking pitch points are denoted by solid triangles while the singing pitch points are denoted by unfilled circles. The corresponding Chinese characters and tone categories are listed at the bottom of the graphs. For the sake of visual clarity, we place twelve time-normalized syllable intervals in each figure. If an utterance has less than twelve syllables, the intervals on the right will be unfilled.

If an utterance has more than twelve syllables, it will be divided into two or more parts, e.g. Utterance 16 is divided to two parts and displayed in two rows, as shown in Figure 3-7. The dashed lines are used to separate different syllables within an utterance. The solid lines indicate the end of an utterance or signify that the adjacent syllables belong to different phrases.

From Figure 3-2 to Figure 3-10 above, we can get a first impression that the speaking pitch contours usually have a more declining pitch slope than their singing counterparts. In addition, the speaking pitch contours are usually located lower in the graph. These are only rough overall impressions. Below we are not going to analyze the utterances case by case. Rather, we would like to provide a more systematic and accurate analysis on the differences between speaking and singing. Our analysis will be conducted in the framework of the matrix composed of pitch configuration and pitch domain. The results and discussions in Section 3.3 to Section 3.5 below will be arranged in the order of the pitch domains: syllable level, utterance-body portion, and utterance-final portion. Within each section, detailed analysis of pitch configuration factors, including pitch slope, pitch register and pitch span will be carried out.

3.3 Results and discussions of syllable-level measurements

Figure 3-11 displays the mean pitch contours of different tones in ICF, SPK, and SNG. It can be observed that T1, T3, T4 and T6 show declination in ICF and SPK but tend to keep level or even exhibit a slight rising tendency in SNG. Compared with ICF, the declination effect is more obvious and stronger in SPK. T2, which is the high-rising tone, has a steeper slope in ICF than in SPK and SNG. The mid-low rising tone T5 has a steeper slope in ICF and SNG than in SPK.

As T1 and T4 are tones with both level and falling variants, we are not surprised to see that their mean pitch contours show a falling tendency in ICF and SPK while remaining horizontal in SNG. One may suppose that T1 and T4 tend to use falling variants in ICF and SPK but use level variants in SNG. However, T3 and T6, which are traditionally regarded as "level tones" (Bauer & Benedict, 1997; Chao, 1947; Cheung, 1986; Jones & Woo, 1912), also show obvious declination in ICF and SPK in Figure 3-11. The comparison of the pitch slope of T1, T4, T3 and T6 shows that T1, T3 and T6 are nearly parallel, in ICF and in SPK alike, while T4 has much stronger falling pitch contour. The falling pitch contours of T3 and T6 reflect that the traditionally putative "level tones" are not so "level" in their acoustic pitch contours. As T3 and T6 do not have falling variants, the falling slope in ICF and SPK is actually caused by the declination intonation in speaking. In the three graphs of Figure 3-11, ICF and SPK are more similar to each other than they are to SNG.

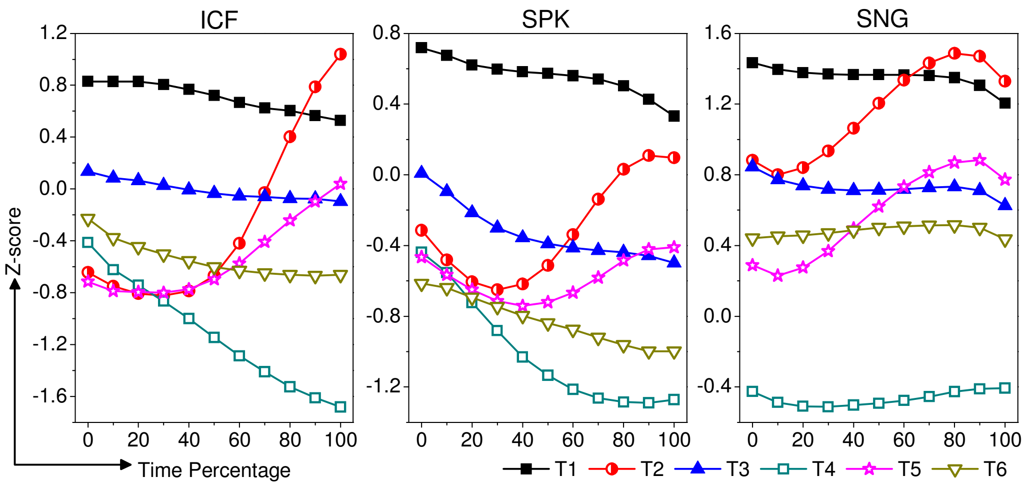

Figure 3-11 An overview of mean tone contours in ICF, SPK (non-final syllables) and SNG (non-final syllables).

The patterns of the two rising tones seem a little complicated. ICF seems to have the steepest rising slope for both T2 and T5. In contrast, the slope of rising tones in SPK looks the most reduced. There is a split in SNG: the slope of T2 in SNG is also flattened, like the one in SPK; while the slope of T5 in SNG is much steeper than SPK and nearly paralleled to the one in ICF. It should also be noted that the beginning points of T2 and T5 are differently distributed in these three graphs. The beginning points of T2 and T5 are quite close in ICF and SPK but have an evident gap in SNG. The beginning point of T2 in SNG is notably higher than the other two speech styles. This is probably the reason why T2 in SNG has a less rising slope than in ICF.

Figure 3-11 provides a direct overview of what the syllable contours look like in different speech styles, especially the slope and the relative location of the tone contours within a tone-system. More detailed and accurate analysis will be conducted below to explore the properties of pitch slope, pitch register and pitch span.

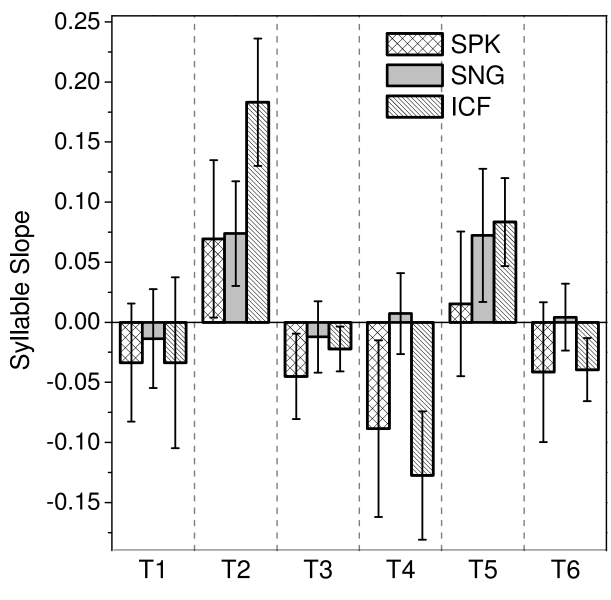

Figure 3-12 Mean pitch slope of syllables in SPK, SNG and ICF, with error-bars indicating ±1 standard deviation.

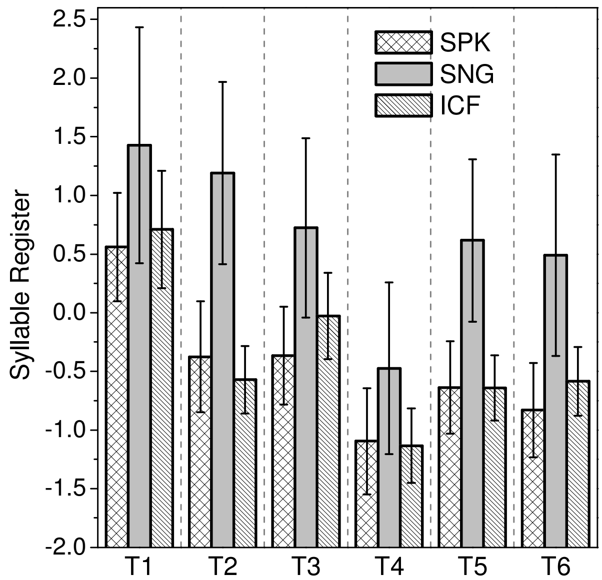

Figure 3-13 Mean pitch register of syllables in SPK, SNG and ICF, with error-bars indicating ±1 standard deviation.

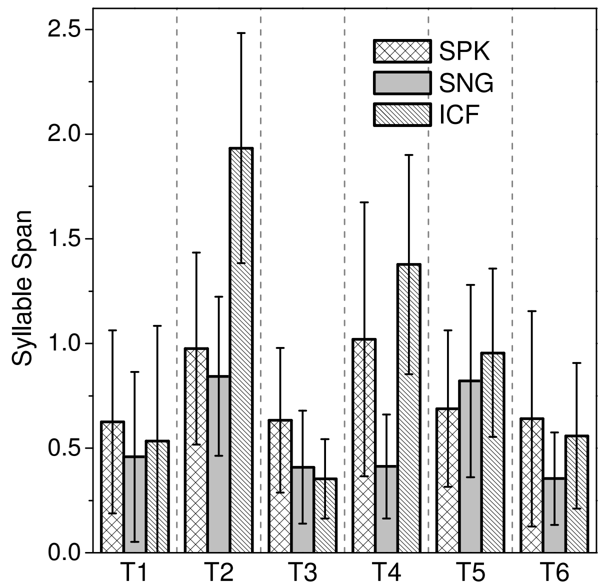

Figure 3-14 Mean pitch span of syllables in SPK, SNG and ICF, with error-bars indicating ±1 standard deviation.

Figure 3-12, Figure 3-13, and Figure 3-14 are bar-charts displaying the mean and standard deviation of syllable slope, syllable register and syllable span in SPK, SNG and ICF. In these figures, the *x* axis stands for the nominal variables of tone categories from T1 to T6. The *x* axis is posited at the value "0" in Figure 3-12. For syllable slope, whether the data are above or below zero is very important as it signifies different directions of the pitch slant: above for rising slope and below zero for falling slope. Having the *x* axis placed at the zero value can clearly signify whether the data are above or below zero. In Figure 3-13 for the data for syllable register, the *x* axis is located at the bottom of the graph. The median of some syllables may be below zero (in Z-Score), but it does not matter whether the data are plus or minus. The *x* axis placed at the bottom of the data can make all the bars uni-directional, thus more clearly displaying the relative height of different speech styles. For the data of syllable span in Figure 3-14, the *x* axis is also placed at the bottom of the graph. The calculation method dictates that the mean value must be above zero. The bar-charts on utterance-body portion and utterance-final portion will also follow this paradigm.

Table 3-1 Brown-Forsythe test on SPK, SNG and ICF

Tones	df_1	Slope		
		$df_{2\ slope}$	F_{slope}	p_{slope}
T1	2	230.063	17.711	0.000
T2	2	466.937	196.497	0.000
T3	2	886.274	162.775	0.000
T4	2	445.500	362.533	0.000
T5	2	688.239	130.587	0.000
T6	2	661.636	154.334	0.000
Tones	df_1	Register		
		$df_{2\ register}$	$F_{register}$	$p_{register}$
T1	2	761.936	245.849	0.000
T2	2	676.536	954.690	0.000
T3	2	742.896	495.405	0.000
T4	2	665.323	149.880	0.000
T5	2	598.482	655.119	0.000
T6	2	662.107	625.806	0.000
Tones	df_1	Span		
		$df_{2\ span}$	F_{span}	p_{span}
T1	2	295.490	16.620	0.000
T2	2	306.219	218.488	0.000
T3	2	826.850	100.264	0.000
T4	2	377.482	199.045	0.000
T5	2	468.470	18.966	0.000
T6	2	486.232	58.656	0.000

In addition to the bar-charts, further analysis was conducted to examine whether the means of the parameters in the groups of SPK, SNG and ICF are significantly different. First we carried out a One-Way ANOVA. The factor is speech style, and the dependent variables are syllable slope, syllable register, and syllable span of T1 to T6, respectively. The result of Levene's test of homogeneity of variances shows that $p < 0.05$ for all the parameters of all the tones, which indicates that the variances cannot be assumed to be equal. Thus, a Brown-Forsythe test was conducted instead. The result shows that the *p*-values of all the parameters of all the tones are below 0.001, as listed in Table 3-1. It can be inferred that all the data between groups for SPK, SNG and ICF are significantly different.

Post Hoc (Games-Howell) multiple comparisons were then carried out, as listed in Table 3-2.

Table 3-2 Post Hoc (Games-Howell) multiple comparisons on SPK, SNG and ICF

Tones	Slope			Register			Span		
	SPK-SNG	SPK-ICF	SNG-ICF	SPK-SNG	SPK-ICF	SNG-ICF	SPK-SNG	SPK-ICF	SNG-ICF
T1	0.000	1.000	0.015	0.000	0.013	0.000	0.000	0.240	0.367
T2	0.550	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
T3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.038
T4	0.000	0.000	0.000	0.000	0.594	0.000	0.000	0.000	0.000
T5	0.000	0.000	0.052	0.000	0.995	0.000	0.000	0.000	0.012
T6	0.000	0.848	0.000	0.000	0.000	0.000	0.000	0.140	0.000

Table 3-3 Paired samples *t*-tests on SPK and SNG at the syllable level

Tones	*df*	Slope		Register		Span	
		t_{slope}	p_{slope}	$t_{register}$	$p_{register}$	t_{span}	p_{span}
T1	563	-8.593	0.000	-18.064	0.000	7.190	0.000
T2	347	-1.259	0.209	-32.069	0.000	4.700	0.000
T3	443	-15.124	0.000	-27.186	0.000	10.906	0.000
T4	335	-21.768	0.000	-13.129	0.000	16.613	0.000
T5	311	-15.911	0.000	-28.442	0.000	-4.230	0.000
T6	395	-14.250	0.000	-29.773	0.000	10.682	0.000

As our research focus is on the comparison between SPK and SNG and as these two speech styles were recorded with the same text (lyrics), paired samples *t*-tests were also carried out on the paired speaking syllables and singing syllables. The results of the tests are shown in Table 3-3. Together with the data displayed in Figure 3-12 to Figure 3-14 and the statistical results listed in Table 3-1 to

Table 3-3, we can further analyze whether there are certain constant patterns and whether the patterns are of significance.

In Figure 3-12 (the data for syllable slope), the bars of rising tones are above zero while level/falling tones are below-zero in most circumstances. The exceptions are T4 and T6 in SNG – they have above-zero bars. This is consistent with what we observed in Figure 3-11 (Page 103): the mean pitch contours of T4 and T6 have slight rising slope in SNG. It can also be observed from this bar-chart that for level/falling tones, SNG always has the greatest value among the three groups. Moreover, SNG is significantly greater than the other two groups, as the *p*-values of T1, T3, T4 and T6 are all below 0.05 for "Slope: SPK-SNG" and "Slope: SNG-ICF" in Table 3-2.The comparison between SPK and ICF does not exhibit a certain direction: T1 ($p = 1.000$ in Table 3-2) and T6 ($p = 0.848$ in Table 3-2) are comparable; for T3, SPK < ICF ($p < 0.001$ in Table 3-2); for T4, SPK > ICF ($p < 0.001$ in Table 3-2).

For the two rising tones, the pitch slope value of SNG is also greater than SPK, but is smaller than ICF. That rising tones in ICF show the greatest rising slope may be due to the fact that syllables are more fully pronounced in ICF and the pitch contours can climb to the peak of the presupposed target. The slope of T5 is only slightly smaller in SNG than in ICF ($p = 0.052$ in Table 3-2).

There are two driving forces acting on the pitch slope of rising tones in SNG. The first force is the ascending intonation effect of SNG, which makes the pitch slope of rising tones more ascending. The second force involves the mechanism of adding an improvised lower ornamental note before the targeted note for composing a rising pitch contour (Cheung, 2012a; Cheung, 2007; Ho, 2008; Yung, 1989). For T5, as the intrinsic distance between the beginning point and ending point is not great (T5 is transcribed as 23, thus the rising contour of it only needs to climb up 1 tone-register), it does not need a much lower ornamental note to realize its rising contour. For T2, the intrinsic distance between the beginning point and ending point is much greater (T2 is transcribed as 25, thus the rising contour of it has to climb up 3 tone-registers). Because of the limited distance between the ornamental note and the targeted note, the actual beginning point of T2 is dragged up in SNG (as can be observed from Figure 3-11 (Page 103), the beginning point of T2 is apparently higher in SNG than the other two speech styles). Influenced by the second driving force, the pitch slope of T2 is not ascending as much as what we expected.

To summarize the syllable slope difference between SNG and SPK, the syllable slope of SNG is always greater than SPK, and this difference is significant for all tones ($p_{slope} < 0.001$ in Table 3-3) except T2 ($p_{slope} = 0.209$ in Table 3-3). This reveals that syllable slope is strongly influenced by the declination effect in SPK – syllables of level/falling tones have steeper fall, while rising tones have reduced rise. The declination effect is weakened or at times even reversed in SNG – level/falling tones tend to have flatter falling contour or at times even exhibit slight rising contour, while rising tones tend to have greater rising slope.

Figure 3-13 (the data of syllable register) shows a neater pattern. SNG is higher than SPK as well as ICF for all the tones. Games-Howell multiple comparisons listed in Table 3-2 further show that SNG is significantly higher than the other two speech styles as the *p*-values are all below 0.001 for "Register: SPK-SNG" and "Register: SNG-ICF" in Table 3-2. The comparison between SPK and ICF does not exhibit a fixed pattern among the six tones: T4 ($p = 0.594$ in Table 3-2) and T5 ($p = 0.995$ in Table 3-2) have comparable means; for T1, T3 and T6, SPK < ICF ($p < 0.05$ in Table 3-2); for T2, SPK > ICF ($p < 0.001$ in Table 3-2). Therefore, Cantonese informants tend to use a much higher register for syllables in singing than normal speaking, including continuous speech and isolated citation form, while the comparison between continuous speech and isolated citation form does not show a fixed pattern.

Syllable span has a strong correlation with syllable slope. A steeper syllable slope is usually accompanied by a wider syllable span. As shown in Figure 3-14, results for SPK are higher than for SNG except for T5. The data in "Span: SPK-SNG" of Table 3-2 and the paired-samples *t*-test results in Table 3-3 confirm that the differences between SPK and SNG are significant ($p < 0.001$ in Table 3-2 and $p_{span} < 0.001$ in Table 3-3) for all the tones.

SNG also has a narrower syllable span than ICF for all the tones except T3. From the mean tone contours displayed in Figure 3-11 (Page 103), it can be observed that the main part of T3 in SNG is more horizontal than in ICF but the beginning and ending parts curve more in SNG, making the syllable span of T3 in SNG wider than in ICF.

The two rising tones (T2 and T5) and T4 are more sloping in ICF than in SPK, correspondingly the syllable spans of these tones have wider spans in ICF. For the other tones (T1, T3 and T6), syllable span in ICF is narrower than in SPK.

3.4 Results and discussions of utterance-body-part measurements

When referring to Figure 3-2 (Page 93) through Figure 3-10 (Page 101), we noticed that, as the ultimate pitch contour output is a combination of tones and intonation, if the utterance consists of syllables of different tones it is hard to observe the intonation patterns. Thus, for the analysis in the utterance-body portion, only the reduplicated phrases in Utterance 17 to 19 are involved here.

To provide a direct impression of the difference between SPK and SNG in the utterance-body portion, some reduplicative phrases from Section 3.2 are selected for display in Figure 3-15, which cover examples of T1 to T6. It can be observed from Figure 3-15 that nearly all the pitch contours of SNG are above SPK except the first two syllables of T4. Pitch register is higher in SNG than in SPK, both at the syllable level and at the utterance level.

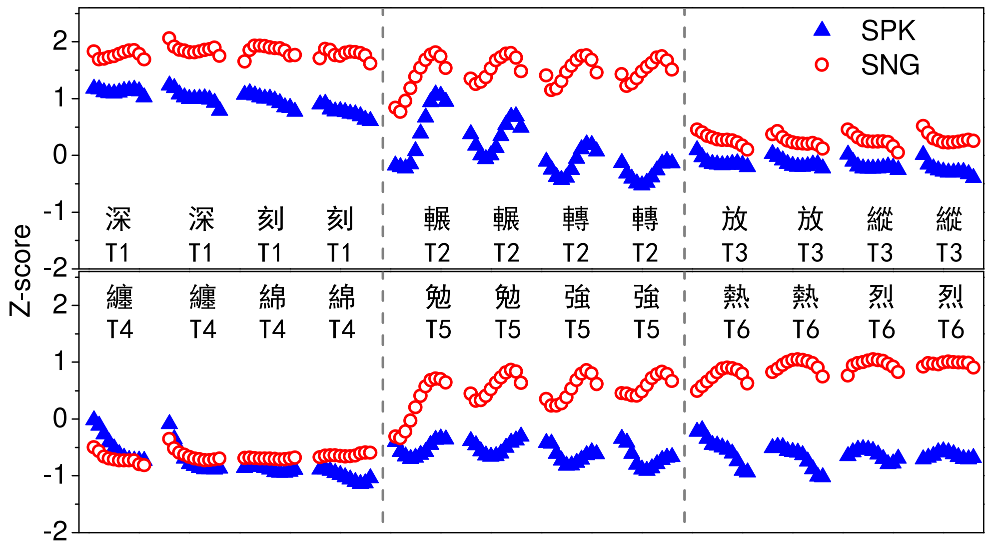

Figure 3-15 Examples of mean SPK and SNG pitch contours of reduplicative phrases from Utterance 17 to 19.

The downdrift effect in SPK is very obvious in Figure 3-15. Not only is the slope of every constitutive syllable influenced by the speaking declination, but the slope of the whole phrase also exhibits a strong downward tendency. Meanwhile, the SNG phrases tend to keep all the syllables at the same level or even show a slight uptrend. For example, this utterance uptrend is noticeable in the graphs of T5 and T6 reduplicative phrases in Figure 3-15.

As the whole utterance trend goes down for the SPK version and keeps horizontal or goes up for the SNG version, SPK and SNG appear to diverge and the distance between these two versions gets progressively wider. In most cases, the utterance slope has a steeper slope in SPK than in SNG. It can thus be inferred that the utterance span is wider in SPK than in SNG.

It should also be noted that not all the pitch points within a reduplicative phrases keep in line. For SPK, the declination at the syllable level and at the utterance level does not have the same gradient of slope, thus the phrase exhibits downdrift pattern rather than a straight downtrend line. Similarly, SNG has a terracing updrift pattern rather than a smooth uptrend line. This is the reason why the linear regression was only conducted on the median points of every constituent syllable rather than all the pitch points within the phrase, as introduced in Section 3.1.4. In this way, the pitch sloping effect at the syllable level can be suppressed, and the calculated slope from linear regression can more accurately represent the utterance slope.

Similar to the paradigm of the investigation at the syllable level, through bar-charts and related statistical analysis, a more detailed analysis of the utterance-body portion will be provided below. The three bar-charts in Figure 3-16, Figure 3-17 and Figure 3-18 display the mean and standard deviation values of utterance slope, register and span respectively, which directly show the contrast between the two groups SPK and SNG. Table 3-4 lists the paired samples *t*-tests on the corresponding parameters in SPK and SNG.

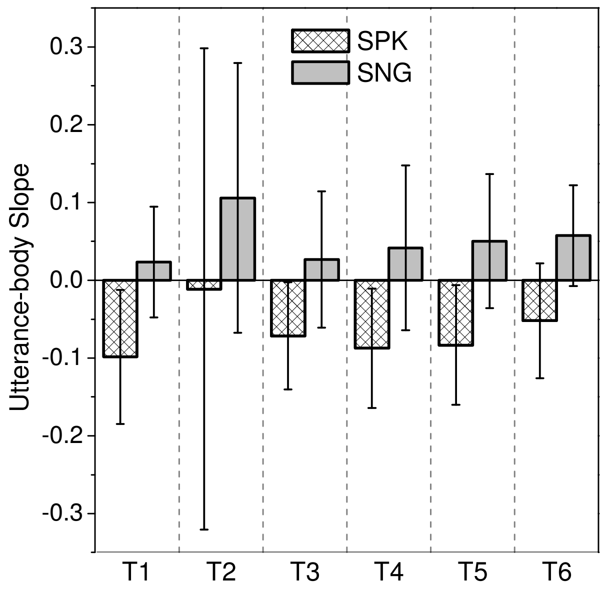

Figure 3-16 Mean pitch slope of utterance body in SPK and SNG, with error-bars indicating ±1 standard deviation.

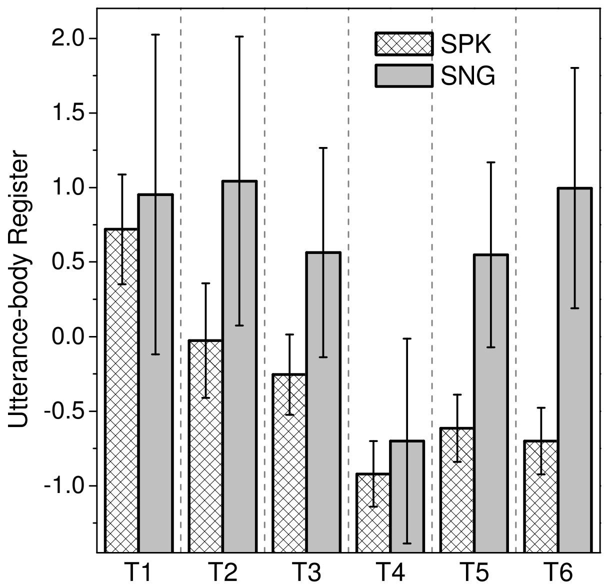

Figure 3-17 Mean pitch register of utterance body in SPK and SNG, with error-bars indicating ±1 standard deviation.

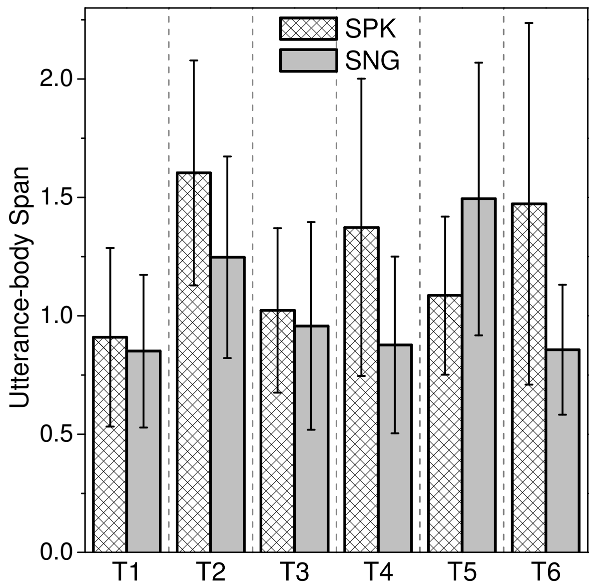

Figure 3-18 Mean pitch span of utterance body in SPK and SNG, with error-bars indicating ±1 standard deviation.

Table 3-4 Paired samples *t*-tests on SPK and SNG in the utterance-body portion

Tones	*df*	Slope		Register		Span	
		t_{slope}	p_{slope}	$t_{register}$	$p_{register}$	t_{span}	p_{span}
T1	71	-10.320	0.000	-1.718	0.090	1.098	0.276
T2	35	-2.017	0.051	-5.795	0.000	3.680	0.001
T3	59	-6.334	0.000	-8.378	0.000	1.114	0.270
T4	23	-5.920	0.000	-1.607	0.122	3.897	0.001
T5	11	-4.297	0.001	-6.558	0.000	-1.692	0.119
T6	23	-6.115	0.000	-10.043	0.000	3.550	0.002

The utterance slopes s in Figure 3-16 how clearly distinct patterns between SPK and SNG. For all six tones, the bars of SNG are all above zero while the bars of SPK are all below zero, which denotes that the utterance slope has an upward tendency in SNG but a downward trend in SPK. The shape of the lexical tones does not affect the direction of utterance slope. The style of speech is the only factor determining whether the bars are above or below zero. Although the error bars in both speech styles are long (especially T2 in SPK), which indicates that the variation is great among the data within a group, the consistent distinction of the mean-bars shows that the utterance-body of SPK has a downtrend while SNG has an uptrend.

In Table 3-4, paired samples t-tests have the results of $p_{\text{slope}} \leq 0.001$ for all tones except T2 ($p_{\text{slope}} = 0.051$). The exception of T2 may be due to the fact that the within-group variation of T2 is too great and there is much overlap in the area of ±1 standard deviation, as displayed in Figure 3-16. Except for T2, the slope data of utterance body in SNG is significantly greater than SPK for the other tones. Together with the factor of syllable slope, SNG has an updrift pattern while SPK exhibits a downdrift pattern.

One thing that should also be mentioned is that the SNG bars are usually shorter than the SPK bars within the same tone category, except for T2 (the mean bar of SPK is shorter than SNG but the error bar of SPK is much longer than SNG) and T6 (the bars of SPK and SNG have similar length). This suggests that the absolute values of the means are usually greater in SPK than in SNG. Thus, the downtrend in SPK is generally steeper than the uptrend in SNG. This agrees with what we directly observed from the examples in Figure 3-15 (Page 111).

Given the updrift pattern in SNG, some linguists and musicians may wonder whether the successive same-tone syllables still belong to the same musical note. Does SNG sound out of tune? Should the latter syllable be transcribed with a higher grade than the former syllable? To answer these questions, we also conducted a quantitative analysis on the semitone values of successive same-tone syllables in SNG – if the difference between successive same-tone syllables is less than one semitone, it can be assumed that the successive same-tone syllables belong to the same musical note and their difference only reflects variation which can hardly be noticed by listeners. To be consistent with the method of calculating the utterance slope, here we also chose the median point as the representative pitch of each syllable. The calculation formula is:

$$g = f_{\text{ST}_{i+1}} - f_{\text{ST}_i} \qquad \textbf{Formula 3-5}$$

In our statistical analysis, there are 396 pairs of successive same-tone syllables in the SNG reduplicated phrases. Among them, 238 pairs are $g > 0$, and the remaining 158 pairs are $g < 0$, which indicates that the updrift pattern is more likely in SNG. The absolute values of g were also obtained and we found that $|g|$ is less than 1 (with the mean value of 0.4905) for most pairs of syllables. The result of the one-sample *t* test also supports that our data are significantly less than 1: $t = -21.409$, $df = 395$, $p < 0.001$. Therefore, the variance of the updrift pattern is generally within a semitone, and the successive same-tone syllables can still be regarded as pairs of syllables on the same musical note.

In Figure 3-17 (the data of utterance register), all the tones have higher SNG values than SPK values. Paired samples *t*-test results in Table 3-4 show that the utterance register of SPK and SNG shows a significant difference for T2, T3, T5 and T6 ($p_{register} < 0.000$); but the difference is not significant for T1 ($p_{register} = 0.090$) and T4 ($p_{register} = 0.122$). Thus, the utterance register is always higher in SNG than in SPK, and in most circumstances this difference is significant.

Figure 3-18 (the data for utterance span) shows that SNG is lower than SPK for all the tones except T5. This reminds us of the pattern of syllable span – in Figure 3-14 (Page 105), SNG also has lower values than SPK except for T5. Thus, syllable span and utterance span exhibit the same pattern. As our calculation method of utterance span involves all the pitch points within the reduplicative phrase, syllable slope and syllable span may also exert an influence here. Paired samples *t*-test results in Table 3-4 show that the utterance spans of SPK and SNG have significant differences for T2, T4 and T6 ($p_{span} \leq 0.002$); but the difference is not significant for T1 ($p_{span} = 0.276$), T3 ($p_{span} = 0.270$) and T5 ($p_{span} = 0.119$). Therefore, SPK has a wider utterance span than SNG, but this difference is not as significant as at the syllable level.

3.5 Results and discussions of utterance-final-part measurements

In the utterance-final portion, we will test whether the utterance-final syllable exhibit certain consistent patterns compared with non-final syllables. SPK and SNG will be analyzed separately, i.e., we will examine whether utterance-final effects exist in SPK and SNG, respectively.

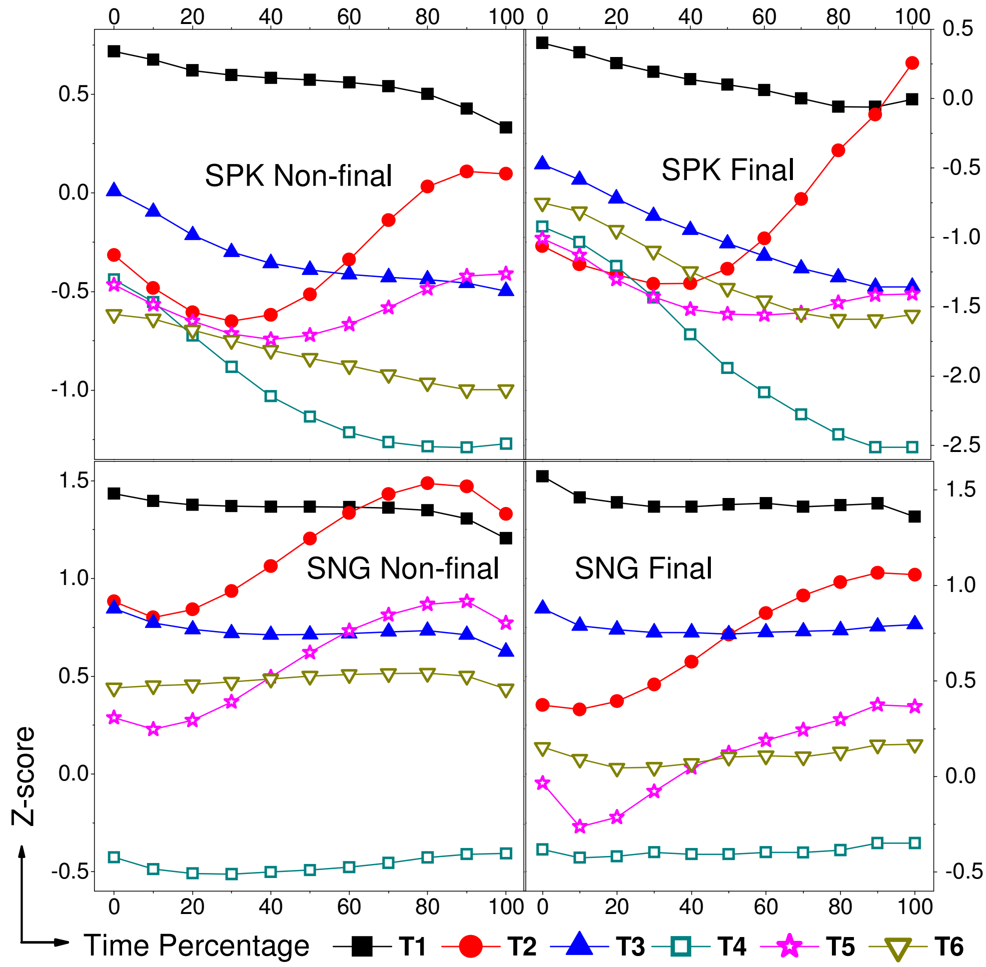

Figure 3-19 An overview of the mean tone contours at utterance-final and non-final positions in SPK and SNG.

Firstly we would also like to provide an overview of the mean final and non-final tone contours in SPK and SNG, which are displayed in Figure 3-19. Let us compare final and non-final syllables in SPK first. T1 looks similar in both final and non-final positions. T3, T6 and T4 are more obviously declining in final position. The tone shape of the mid-rising tone T5 astonishingly becomes falling for final syllables. Only the later part of T5 shows a slight rise, from which we can see the hint of its original rising tone shape. Different from all the other tones including T5 which show extra declination effects in final position, the contour shape of T2 is more rising for final syllables.

The two graphs of "SPK Non-final" and "SPK Final" in Figure 3-19 also show different space distances between tones: the bottom half space occupied by T3, T6, T5 and T4 seems more crowded in the graph of "SPK Final". Especially the distance between T6 and T3 seems to be narrowed in the tonal system of the final syllables in SPK.

The patterns between final and non-final syllables are quite different in SNG. Pitch slope does not have evident difference between final and non-final syllables. The tone contours in the two graphs "SNG Non-final" and "SNG Final" of Figure 3-19 are nearly parallel. The relative location in the non-final tonal system and final tonal system is also similar for T1, T3 and T4. For T2, T5 and T6, they seem to be located at lower levels in the graph of "SNG Final".

It should be noted that the four tone registers of Cantonese lexical tones are not evenly distributed in the three graphs for "SPK Non-final", "SPK Final" and "SNG Non-final" in Figure 3-19. The distance between T1 and T3 is usually much wider than the distance between T3 and T6. In the graph "SNG Non-final" in Figure 3-19, there is also much space between T1 and T3 and between T6 and T4, while the distance between T3 and T6 is quite close. Compared with these three graphs, the four tone registers are much more evenly distributed in the graph "SNG Final" of Figure 3-19. The distances between T1 and T3, T3 and T6 are similar, while the distance between T6 and T4 is a little narrower.

Below we will have further analysis of the pitch slope, pitch register, and pitch span between final and non-final syllables in SPK and SNG, respectively, following the paradigm in Section 3.3 and Section 3.4. The comparison between final and non-final syllables in SPK is displayed in Figure 3-20 to Figure 3-22 and the related statistical results are listed in Table 3-5; while the counterparts of SNG are shown in Figure 3-23 to Figure 3-25 and Table 3-6.

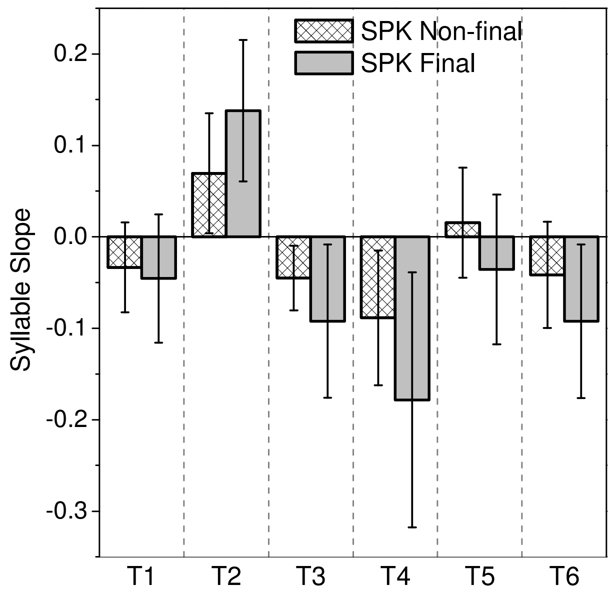

Figure 3-20 Comparison of mean pitch slope between final and non-final syllables in SPK, with error-bars indicating ±1 standard deviation.

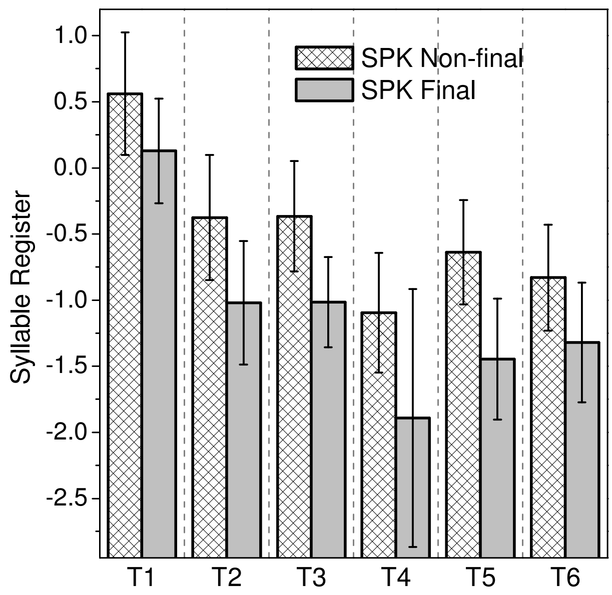

Figure 3-21 Comparison of mean pitch register between final and non-final syllables in SPK, with error-bars indicating ±1 standard deviation.

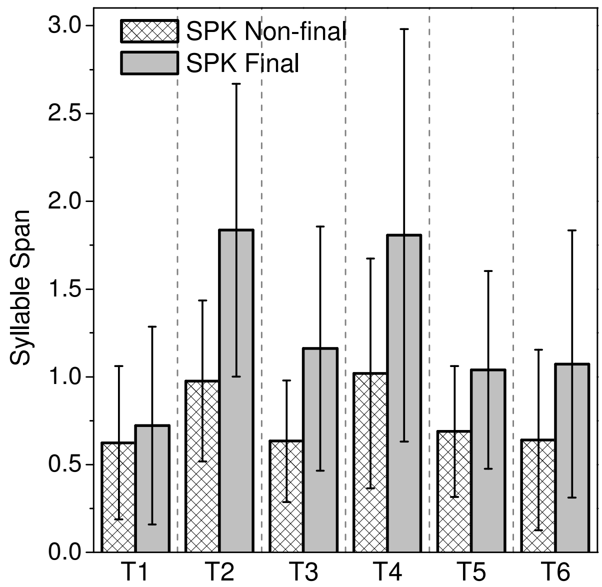

Figure 3-22 Comparison of mean pitch span between final and non-final syllables in SPK, with error-bars indicating ±1 standard deviation.

To test whether the mean values of final and non-final syllables are significantly different, One-Way ANOVA or Brown-Forsythe tests were conducted, as listed in Table 3-5 (for SPK) and Table 3-6 (for SNG). Firstly Levene's test of homogeneity of variances was carried out. If $p > 0.05$, equal variances can be assumed, so that a One-Way ANOVA, which is more powerful than the Brown-Forsythe test, can be conducted. If $p < 0.05$, the hypothesis of equal variances is rejected and the Brown-Forsythe test is conducted instead of the One-Way ANOVA. In Table 3-5 and Table 3-6, the df_2 data reflects which test was performed: if df_2 is an integer, the test is a One-Way ANOVA; if df_2 contains a decimal, the test is a Brown-Forsythe test.

For the syllable slope comparison in SPK, from Figure 3-20, it can be observed that the bars of the final syllables are located lower for utterance-final syllables for all tones except T2. This reveals that final syllables exhibit greater declination effects than non-final syllables in terms of syllable slope, which agrees with what we observed from Figure 3-19 (Page 117). It is especially interesting that T5 as a rising tone also shows a negative mean value for the slope of final syllables. The final declination overrides the original rising tone shape of T5 such that the final output pitch contour exhibits a fall. Statistical results in Table 3-5 indicate that the syllable slope differences between final and non-final syllables are significant ($p_{\text{slope}} < 0.001$) except for T1 ($p_{\text{slope}} = 0.161$). Therefore, the dominant pattern is that final syllables have significantly greater downward pitch slope than non-final syllables in SPK.

T2 is the only tone that does not show an extra declining pitch slope and, in fact, exhibits a greater climbing tendency at the final position. To explain the exception of T2, we went back to the original data. There were only two final syllables of T2 in our questionnaire, Utterance 3 and Utterance 13. Utterance 13 can be considered as a question but we did not notice this when we designed the stimuli. Thus, many informants added extra interrogative rising intonation (as discussed in Section 1.2.2) in the final syllable of Utterance 13. Consequently, the mean slopes of the final T2 syllables were much greater in this utterance than in the Utterance 3. It can be inferred that this is a great interference to the ultimate pitch slope of T2 in the final position.

Table 3-5 One-Way ANOVA / Brown-Forsythe test on final and non-final syllables in SPK

Tones	Test	df_1	Slope		
			$df_{2\ slope}$	F_{slope}	p_{slope}
T1	Brown-Forsythe	1	80.142	2.005	0.161
T2	One-Way ANOVA	1	370	23.900	0.000
T3	Brown-Forsythe	1	75.198	22.157	0.000
T4	Brown-Forsythe	1	64.983	23.637	0.000
T5	One-Way ANOVA	1	358	26.972	0.000
T6	Brown-Forsythe	1	136.303	35.066	0.000
Tones	Test	df_1	Register		
			$df_{2\ register}$	$F_{register}$	$p_{register}$
T1	One-Way ANOVA	1	634	57.293	0.000
T2	One-Way ANOVA	1	370	41.917	0.000
T3	Brown-Forsythe	1	108.671	210.106	0.000
T4	Brown-Forsythe	1	63.619	38.609	0.000
T5	One-Way ANOVA	1	358	167.169	0.000
T6	One-Way ANOVA	1	502	120.352	0.000
Tones	Test	df_1	Span		
			$df_{2\ span}$	F_{span}	p_{span}
T1	Brown-Forsythe	1	82.291	1.986	0.163
T2	Brown-Forsythe	1	23.969	24.988	0.000
T3	Brown-Forsythe	1	76. 795	39.831	0.000
T4	Brown-Forsythe	1	65.691	25.482	0.000
T5	Brown-Forsythe	1	53.527	17.433	0.000
T6	Brown-Forsythe	1	134.673	31.028	0.000

The syllable register of final and non-final syllables clearly shows a distinct contrast in Figure 3-21. Bars of non-final syllables are higher than bars of final syllables for all the tones. Statistical results listed in Table 3-5 show that the pitch register of final and non-final syllables are significantly different ($p_{register} < 0.001$). Thus, the pitch register of final syllables is always significantly lowered in SPK.

As we discussed before, syllable span is closely related to the steepness of syllable slope. We notice that the bars of final syllables are longer than non-final syllables in the data of syllable slope shown in Figure 3-20 (Page 119), thus the final syllables are more sloping than non-final syllables. Correspondingly, the bars of final syllables are higher than non-final syllables in the data of syllable span in Figure 3-22. Similar to the statistical results of syllable slope, the syllable span difference between final and non-final syllables is significant ($p_{span} < 0.001$) except T1 ($p_{span} = 0.163$) in Table 3-5. Generally speaking, final syllables have significantly wider pitch span than non-final syllables in SPK.

The data for final syllables are significantly different from those for non-final syllables in SPK, i.e., there is obvious final declination effect in SPK. To be more specific, the realization of final declination involves a more steeply declining pitch slope, a lower pitch register and a wider pitch span.

The comparison between final and non-final syllables displays a totally different picture in SNG.

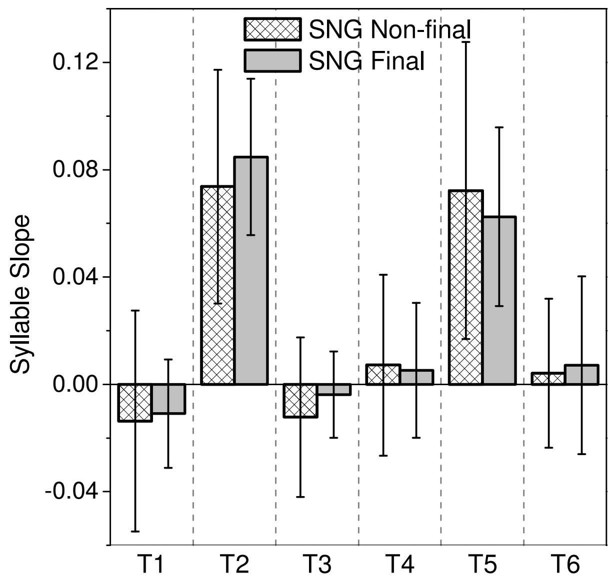

Figure 3-23 Comparison of mean pitch slope between final and non-final syllables in SNG, with error-bars indicating ±1 standard deviation.

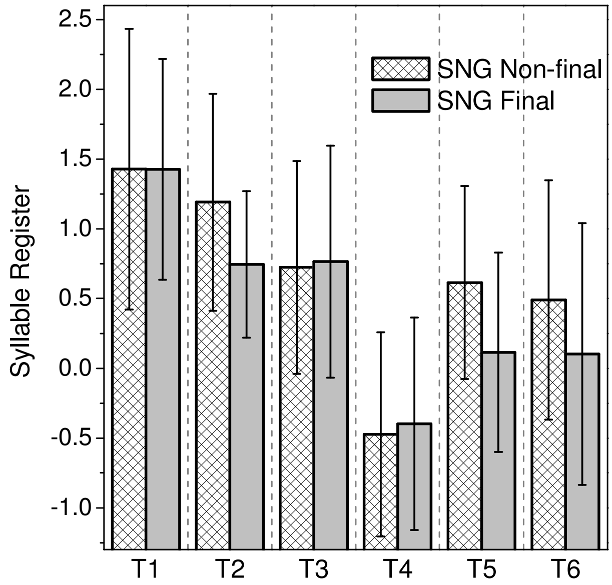

Figure 3-24 Comparison of mean pitch register between final and non-final syllables in SNG, with error-bars indicating ±1 standard deviation.

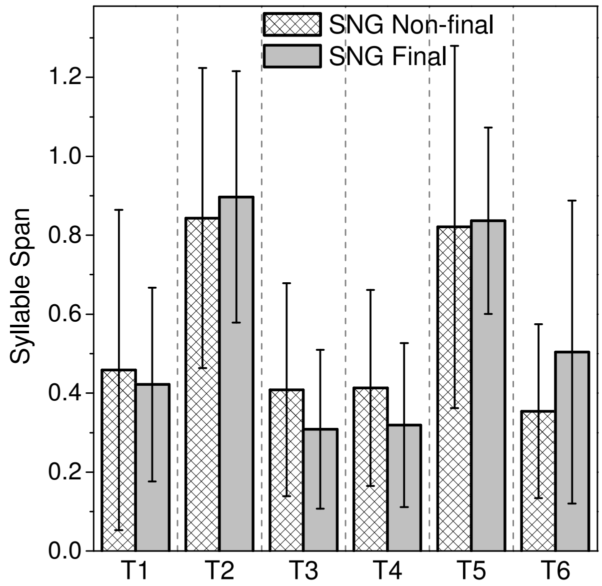

Figure 3-25 Comparison of mean pitch span between final and non-final syllables in SNG, with error-bars indicating ±1 standard deviation.

In Figure 3-23 (the data for syllable slope), the bars of the final syllables are located higher for T1, T2, T3 and T6, but lower in T4 and T5. The bars of final and non-final syllables do not seem to have much difference between each other. Statistical results in Table 3-6 confirm that most tones do not show a significant difference between the final and non-final group (p_{slope} > 0.05), the only exception is T3 (p_{slope} < 0.001). Looking back at the mean tone contours comparison of SNG in Figure 3-19 (Page 117), we find that the ending part of non-final syllables of T3 has a more obvious downward bend, which probably makes the pitch slope data of the non-final syllables of T3 seem more declining. In fact, the overall tone contours of T3 are paralleled for final and non-final syllables. As the exception of T3 can be explained, we can conclude that pitch slope is equal for final and non-final syllables in SNG.

Table 3-6 One-Way ANOVA / Brown-Forsythe test on the final and non-final syllables in SNG

Tones	Test	df_1	Slope		
			$df_{2\ slope}$	F_{slope}	p_{slope}
T1	Brown-Forsythe	1	160.170	0.896	0.345
T2	One-Way ANOVA	1	370	1.495	0.222
T3	Brown-Forsythe	1	163.374	12.643	0.000
T4	Brown-Forsythe	1	101.082	0.268	0.606
T5	Brown-Forsythe	1	92.806	2.887	0.093
T6	One-Way ANOVA	1	502	0.893	0.345
Tones	Test	df_1	Register		
			$df_{2\ register}$	$F_{register}$	$p_{register}$
T1	Brown-Forsythe	1	102.724	0.000	0.996
T2	Brown-Forsythe	1	30.392	15.022	0.001
T3	One-Way ANOVA	1	514	0.182	0.670
T4	One-Way ANOVA	1	394	0.542	0.462
T5	One-Way ANOVA	1	358	21.617	0.000
T6	One-Way ANOVA	1	502	16.607	0.000
Tones	Test	df_1	Span		
			$df_{2\ span}$	F_{span}	p_{span}
T1	One-Way ANOVA	1	634	0.560	0.455
T2	One-Way ANOVA	1	370	0.453	0.501
T3	Brown-Forsythe	1	116.794	13.929	0.000
T4	One-Way ANOVA	1	394	7.647	0.006
T5	Brown-Forsythe	1	111.843	0.137	0.712
T6	Brown-Forsythe	1	126.818	15.040	0.000

The comparison of pitch register does not show a fixed direction either: T2, T5 and T6 have lower bars for final syllables; T3 and T4 are the opposite and show the pattern of "final > non-final"; T1 has nearly equal bars for both groups in Figure 3-24. Statistical results in Table 3-6 reveal that for T1 ($p_{register} = 0.996$), T3 ($p_{register} = 0.670$) and T4 ($p_{register} = 0.462$), the difference between final and non-final syllables are not significant; while for T2, T5 and T6, the differences are significant ($p_{register} \leq 0.001$). Thus, the messy pattern suggests that pitch register is not influenced by the utterance-final effect in SNG.

The pattern of pitch span is also irregular between final and non-final syllables in SNG. The final syllables of T1, T3 and T4 have narrower pitch span than non-final syllables, while final syllables of T2, T5 and T6 have wider span in Figure 3-25. Statistical results in Table 3-6 show that for T1 ($p_{span} = 0.455$), T2 ($p_{span} = 0.501$) and T5 ($p_{span} = 0.712$), the difference between final and non-final syllables are not significant; while for T3, T4 and T6, the differences are significant ($p_{span} < 0.01$). The random pattern indicates that final syllables and non-final syllables do not exhibit clear pitch span difference in SNG.

Therefore, final syllables do not have outstanding behavior compared with non-final syllables in SNG, whether from the perspective of pitch slope, of pitch register, or of pitch span. The absence of boundary tones is also an important characteristic of SNG.

In sum, in the utterance-final portion, SPK and SNG exhibit different patterns: SPK shows significant utterance-final declination effects while SNG does not exhibit any notable utterance-final intonation.

Chapter 4
Experiment III: Different Sentence Types

Experiment III was carried out in the same framework and with the same informants as in Experiment II, but with a focus on internal differences within the normal speaking speech style. Two cross-cutting dichotomies were dealt with here: one is declarative versus question; the other is absence versus presence of an SFP. Thus, there are four sentence types: declarative without an SFP, declarative-derived question without an SFP, declarative with an SFP, and declarative-derived question with an SFP.

The stimuli were so designed as to be composed of same-tone syllables. The experimental results indicated that for all four sentence types, the utterance-body intonation is declining, and all the non-final syllables of level/falling tones show descending pitch contours. For utterances without an SFP, there is a final-lowering effect on declarative sentence, which involves a slightly more declining pitch slope and a significantly lower pitch register. The final intonation of questions without an SFP is very prominent among the four sentence types: it has a sharp final rise, while the pitch register is nearly the same as the non-final syllables. For utterances with an SFP, the SFP carries the burden of utterance-final intonation and the syllable immediately before SFP does not exhibit anything extraordinary. There are also other minor differences among the four sentence types, which can be revealed through systematic study within the framework of a matrix composed of pitch configuration and pitch domain.

4.1 Questionnaire and instructions

As mentioned in Section 1.4, the recording process of Experiment II and Experiment III was conducted in the same session and with the same informants. Both experiments also adopted similar measurement and statistical methods. On these aspects of experimental methodology, Experiment II and Experiment III overlap and one can refer back to Section 3.1.3 and Section 3.1.4 for detailed information. Here we will not repeat them, but will only introduce the design of our questionnaire and instructions for Experiment III.

In the questionnaire for Experiment III, a set of utterances derived from same-tone-sequence trunks is designed to test the influence of different sentence types. The sentence types investigated here actually makes a 2*2 matrix. One dimension is declarative sentence versus declarative-derived question; while the other dimension is with a sentence-final particle (**SFP**) versus without it. Table 4-1 lists these sentence types and their abbreviations.

Table 4-1 Matrix of the sentence types investigated in the present study

	Declarative	Declarative-derived Question
Presence of SFP	**DN**: Declarative without an SFP	**QN**: Declarative-derived question without an SFP
Absence of SFP	**DP**: Declarative with an SFP	**QP**: Declarative-derived question with an SFP

As the first step, a set of trunks were designed. The trunks are in the format of "Subject (3 syllables) + Adverbial Modifier (2 syllables) + Verb (2 syllables) + Object (2 syllables)". This format is designed to meet the following criteria: (1) Structure consistency: all the designed trunks must be in the same syntatic structure to ensure that they have the same prosodic structure. (2) Naturalness: the resultant utterance structure must be one that is frequently used in daily conversation so that the informants can produce them fluently and naturally. (3) Equal trunk length: different lengths of utterances may lead to different pitch slopes of utterances (Shih, 1997), to make the trunks of the six tones comparable, the length is controlled to be the same for all these trunks. Furthermore, we maintain a consistent number of syllables for every component in the utterance structure, in order to avoid the phrasing effect differences caused by different lengths of corresponding syntactic units. On the other hand, we would like to maximize the utterance length, thus the subject is three syllables (rather than two syllables) long and a two-syllable adverbial modifier is added rather than the simple "Subject + Predicate + Object" structure. (4) Non-stopped syllables: all the syllables of the designed utterance are non-stopped syllables, so as to avoid the intrinsic differences between stopped and non-stopped syllables.

In previous studies on the context effects of adjacent tones, such as Xu (1997), the stimuli were designed to be composed of a sequence of sonorants, so that the pitch variations at the boundary between the two adjacent syllables can be shown. In our present study, our focus is not on the pitch boundary between the two adjacent syllables. We just need to measure the pitch contour of every individual syllable. Thus, it is not necessary that all the syllables are composed entirely of sonorants.

In accordance with the above criteria, six trunks for the six tones were designed, as listed below:

(1) T1: 張新輝今朝裝修餐廳

(2) T2: 史廣海九點表演小品

(3) T3: 宋志慶過去愛好寄信

(4) T4: 何明豪時常懷疑程洋

(5) T5: 呂偉敏以往領養母馬

(6) T6: 謝士健二號冒認鄧穎

Then the targeted variations were made to the trunks by adding the required SFP or telling the informants what kind of intonation we intended to obtain. Take the utterances of T3 as an example. The trunk is "宋志慶過去愛好寄信 [sʊŋ^{33} tsiː^{33} hɪŋ^{33} kɔː^{33} hɵy^{33} ɔːi^{33} hou^{33} kei^{33} sɵn^{33}]" ("SONG Zhi-qing liked sending letters in the past"). For the DN sentence, it is in a declarative intonation, and is just the trunk. We told the informants that this sentence is to tell a fact that SONG Zhi-qing liked sending letters in the past. For the QN sentence, it is a question, and is also only the trunk. We told the informants that this sentence is to ask whether SONG Zhi-qing liked sending letters in the past. For the DP sentence, we told the informants that this is a declarative sentence again, but adds the SFP [aː^{33}] (which strengthens the declarative function in Cantonese) after the trunk. For the QP sentence, we let the informants know that this is also a question, but with the interrogative SFP [aː^{21}] at the end. For all these four types of utterances, we provided a detailed demonstration with concrete examples to the informants to make it clear what kinds of intonation were required. In the recording session, if we found that the informants had pronounced the wrong intonation or missed the SFP, we asked them to read the sentences aloud again until they were in the correct intonation and pronounced fluently.

The utterances were grouped together according to sentence type rather than tone, so that informants could feel more comfortable and keep a certain type of intonation consistent with all the tones. The informants were also required to read each sentence twice and we chose the better version for analysis.

4.2 Results of mean pitch contours of different sentence types

The mean pitch contours of the whole utterances are displayed in Figure 4-1, with the DN, QN, DP and QP versions juxtaposed within the same graph. It can be observed that the final syllables of QN sentences (sharp rising) as well as the SFPs of DP and QP (occupying much lower pitch register and wider pitch span) have a large deviation from the pitch span of the syllables comprising the utterance body. As a result, the span of the display is so wide that the utterance bodies of the DN, QN, DP and QP sentences are too close and have too much overlap in Figure 4-1. To allow for more effective comparison, we excluded the final syllables of DN and QN and the SFPs of DP and QP, whereby enlarging the display range to show the difference among the trunks of DN, QN, DP and QP in Figure 4-2. In Figure 4-1 and Figure 4-2, solid triangles represent pitch points of DN, solid circles denote those for QN, unfilled triangles denote those for DP, and unfilled circles denote those for QP.

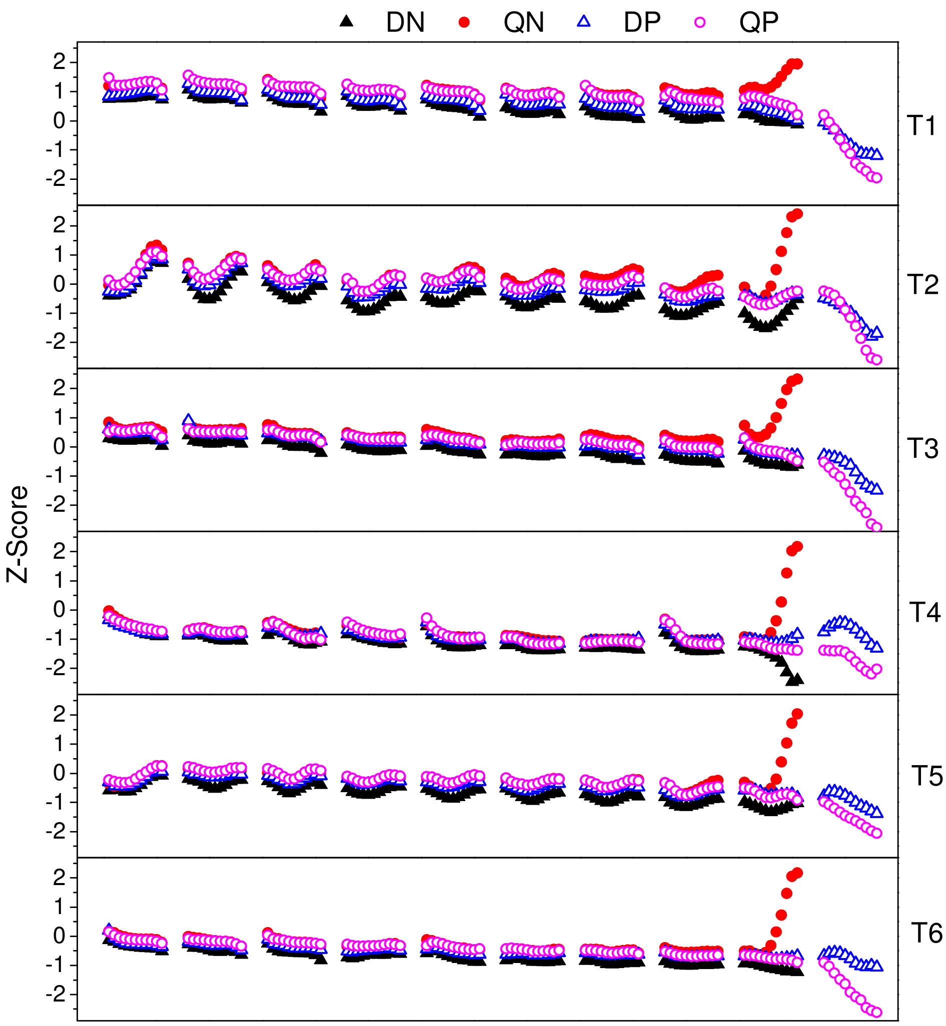

Figure 4-1 Mean pitch contours of the whole utterances (DN, QN, DP and QP).

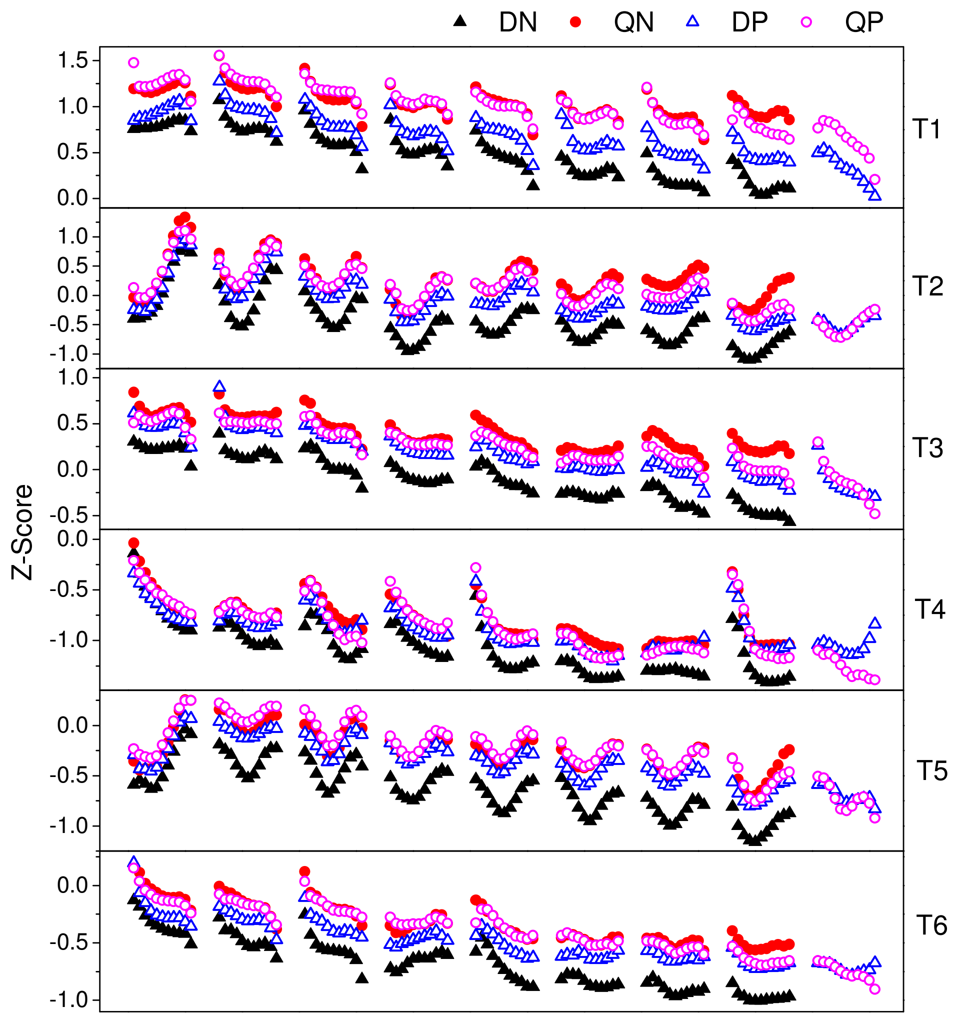

Figure 4-2 Focused scope on the non-final syllables of DN, QN, DP and QP.

In Figure 4-1, as mentioned before, the utterance body does not show any clear difference among the different sentence types because of the wide scope range. These four sentence types can be distinguished by the contour at the "tail", namely the last syllable of DN and QN and the SFP of DP and QP. QN has a sharp rising pitch contour for the ending syllable. It is ideally analyzed as having a boundary tone H. The pitch contour of the last syllable of DN is the lowest among the four pitch contours and has a more steeply falling pitch slope for T4. For the two sentences with SFPs, the two SFPs start at a similar pitch and diverge later, a pattern which indicates that they have different pitch slopes as well as different pitch registers – if they were to differ only in pitch register, they should be parallel, but parallelism applies only to T4. The SFP of DN is usually much more falling than the SFP of QN.

With the more focused scope in Figure 4-2, the differences of utterance body among different sentence types can be more clearly shown. The syllable contours of different sentence types seem to be parallel, which implies that the syllable slope does not have much variation in different sentence types. For utterance slope, in either Figure 4-1or Figure 4-2, the utterance contours of all sentence types exhibit the same downward trend. It can be observed in Figure 4-2 that the utterance-body slope of DN is slightly more declining than the other sentence types, which makes the distance between DN and other sentence types growing wider along the time axis. In terms of pitch register, in the domains of both syllable level and utterance-body portion, the pitch contours of DN are always lower than the other sentence types. For the two sentence types, declaratives and questions, declaratives usually have lower register than questions. When there is no SFP, the difference between them is even bigger: DN is lower than DP, and QN is usually slightly higher than QP while sometimes partly overlapping with it. Thus, if we list the four sentence types from low to high, the order is: DN < DP < QP ≤ QN.

Apart from the overall pitch contours of the trunks, we are especially interested in the final syllable of the trunk, the pitch contours (DN, QN, DP and QP) of which are displayed separately in Figure 4-3. For "final syllable of the trunk", we mean the last syllable of the DN and QN sentences and the penultimate syllable of the DP and QP sentences, because the last syllable of DP and QP is SFP. For example, the final syllable of the trunk is "信" in the T3 sentences.

Figure 4-3 is consistent with the other tone graphs of the present study, in which the time-normalized pitch contours are displayed.

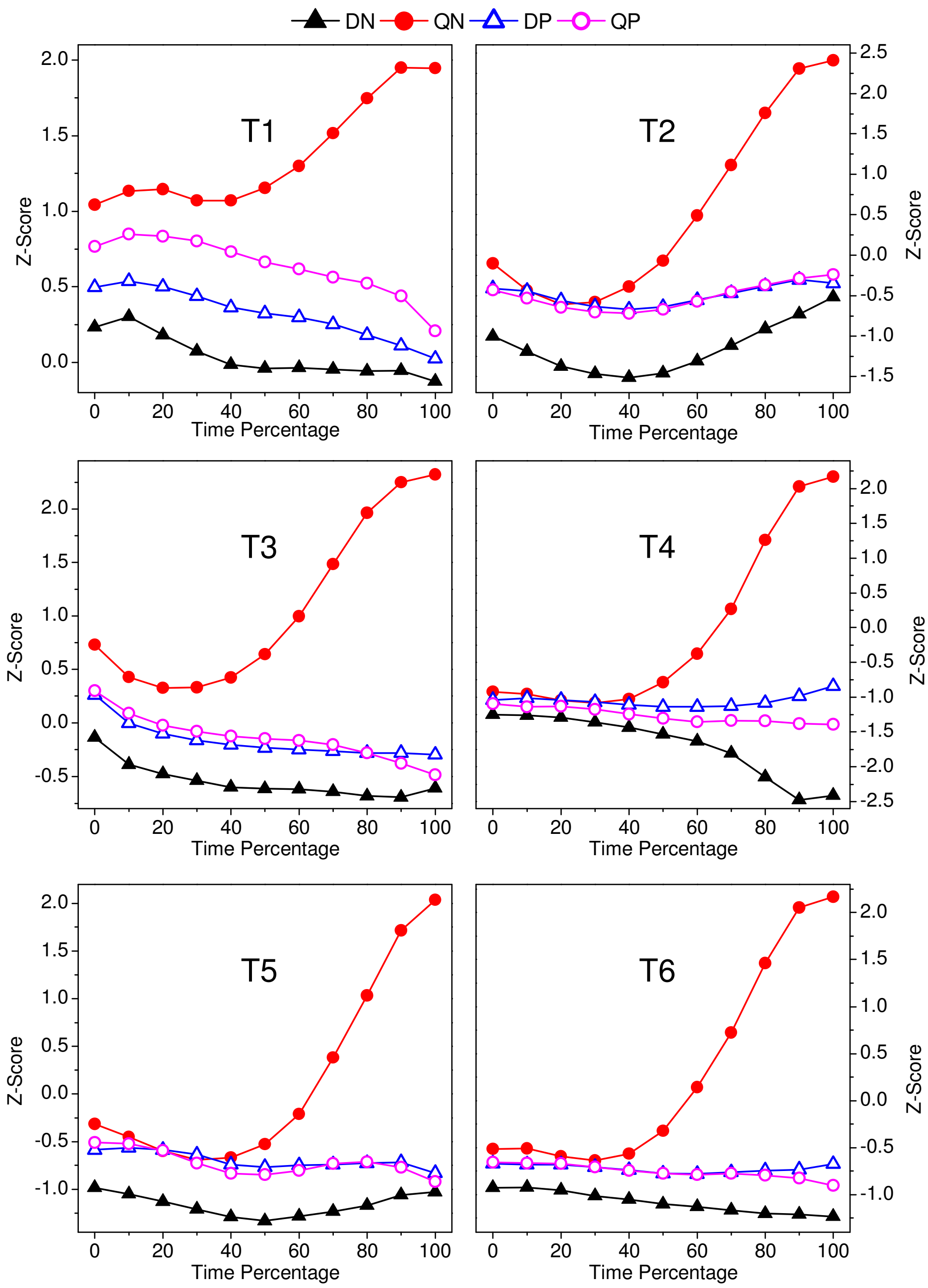

Figure 4-3 Final syllable of the trunk – pitch contours of DN, QN, DP and QP.

The pitch points in Figure 4-1 (Page 131) and Figure 4-2 (Page 132) are not connected because each graph has several syllables and the space for the pitch contour of each syllable is too narrow; in fact the pitch points themselves can already approximate the pitch contour. For the graphs which only display the pitch contour of one syllable, in which the space is wide enough, the pitch points are connected in order to better represent the pitch contour, as shown in the graphs in Figure 4-3 and the following figures Figure 4-5 to Figure 4-9.

Figure 4-3 is actually the focused scope of the penultimate syllable interval of Figure 4-1 (Page 131, for DN and QN, the last syllable interval is blank), which more clearly displays the difference of the final syllable of the trunk. It can be observed that regardless of the original tone shapes, the final pitch output of all the last syllables in QN is sharp rising pitch contour, which is markedly different from the other sentence types. The pitch contour of DN is always the lowest and at the bottom of the graphs. The pitch contours of DP and QP are very close and nearly overlapping for all tones except T1 (for T1, QP is higher than DP, and there is a certain distance between DP and QP).

As the final syllables of DN and QN show distinct characteristics in Figure 4-3, we would also like to conduct further analysis on them separately. We juxtaposed the mean tone contours of non-final syllables and final syllables in Figure 4-4 (time-normalized) and Figure 4-5 (real-time) for the analysis of DN. For QN, the comparison is displayed in Figure 4-6 (time-normalized) and Figure 4-7 (real-time).

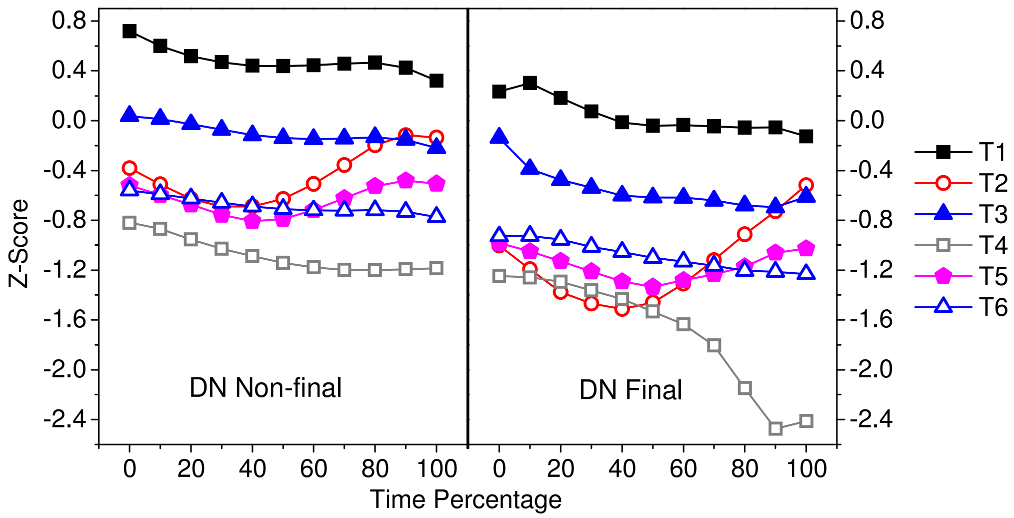

Figure 4-4 Time-normalized tone contours in non-final and final positions of DN.

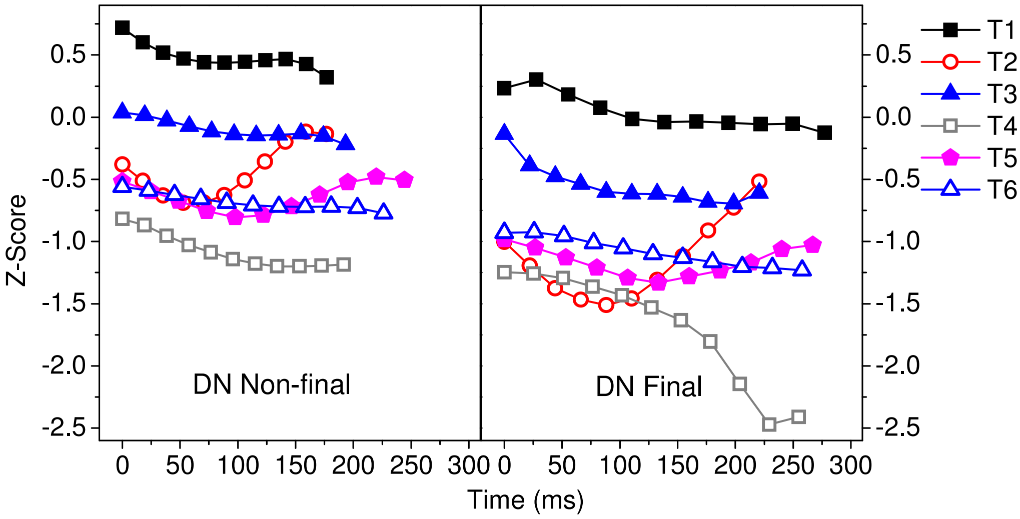

Figure 4-5 Real-time tone contours in non-final and final positions of DN.

For the "time-normalized" figures (Figure 4-4 and Figure 4-6), all the tone contours are of the same length. This type of figures is consistently presented throughout the present thesis. For the "real-time" figures (Figure 4-5 and Figure 4-7), the length of a given tone contour is proportionate to the mean duration of syllables of that tone category. The "real-time" figures can help us to observe the duration differences of tone contours. Here the "real-time" figures are also displayed in order to show the duration differences between final and non-final syllables in DN and QN.

In the graphs in Figure 4-4 and Figure 4-5, the display scale of the Z-Score axis covers the same range, which can help us observe the pitch register and pitch span differences. It can be observed that compared with the non-final mean contours, the final contours in DN significantly move downwards for every tone. The pitch slope is also slightly more declining in the final position for T1, T3 and T6. The slope declination effect is much stronger for the falling tone T4. For the rising tones, the ascending slope of T5 is slightly weakened in the utterance-final position, but the slope of T2 does not seem to show much difference between final and non-final syllables. It is shown in Figure 4-5 that the final syllables in DN are a little longer than the non-final syllables. The mean duration data show that the ratio between final and non-final syllables is 1.24 for DN. The final declination of DN is realized by "simultaneous addition" in the terminology of Chao (1933/2006), which does not require much extra duration.

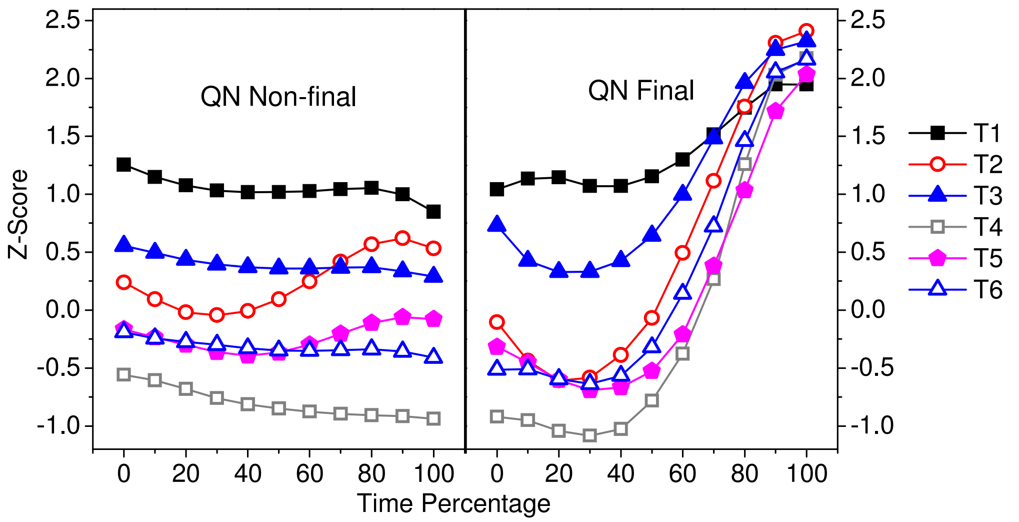

Figure 4-6 Time-normalized tone contours in non-final and final positions of QN.

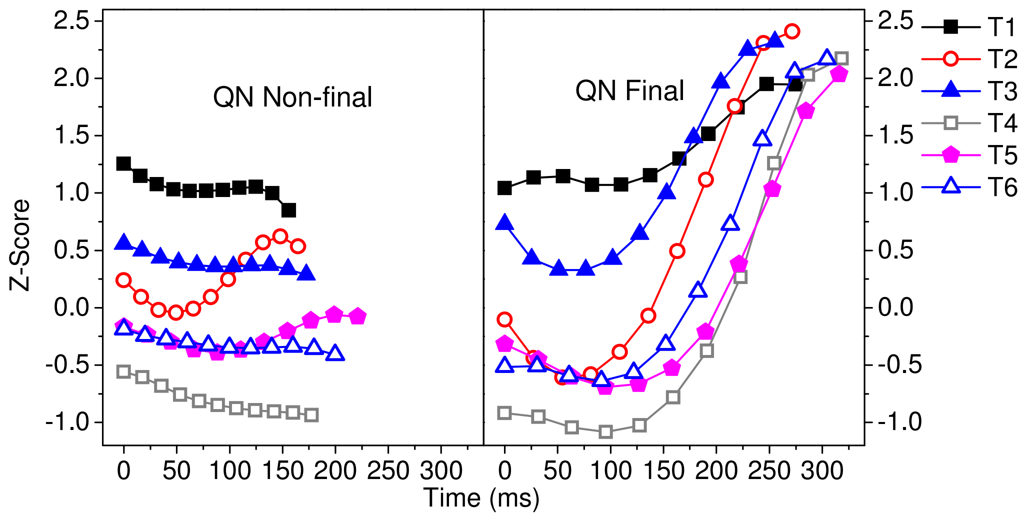

Figure 4-7 Real-time tone contours in non-final and final positions of QN.

The non-final tone contours of QN in Figure 4-6 and Figure 4-7 do not show extraordinary behavior compared with their DN counterpart in Figure 4-4 and Figure 4-5. However, in final position for QN, the mean tone contours show the outstanding characteristic of the QN utterances – a sharp rise. In the right graphs of Figure 4-6 and Figure 4-7, it can be observed that no matter where the original tone contours start, the second half (after about 40% of the time percentage, or about 100ms in the real time) of all the tone contours converge to an extremely high point. Before the transition point, the pitch register of the tone contours keeps at a similar level as the non-final syllables in QN, i.e. the pitch contours of the six tones can be divided into four levels: T1 occupies the highest level; T3 takes up the second highest level; T2, T6 and T5 cluster in the third highest level; and T4 stays at the lowest level.

This indicates that the pitch height of the first half of QN utterance-final contours remains consistent with the beginning point of the original tone contours: according to the five-point-scale transcription method, the beginning point of T1 is 5; T3 is 3; T2, T5 and T6 are 2; and T4 is 2 or 1. It can be inferred that tones locating in the final position of QN are not totally neutralized: the first half of the pitch contour conveys the original tone registers, and only the second half is neutralized towards a sharp rising pitch contour. As the sharp rise of the second half contour of the QN-tail is more prominent, the preservation of the original tone register in the first half may be neglected, as reported by previous perception tests (Ma, 2007; Ma, Ciocca, & Whitehill, 2006a; Mai, 2000).

As far as the duration difference is concerned, it can be observed from Figure 4-7 that the final syllables of QN are much longer than the non-final syllables. The mean duration data show that the ratio between final and non-final syllables is 1.59 for QN. The extra length allows more time for the realization of adding the sharp rise. The combination of tone and QN-final intonation is a kind of “successive addition” in the terminology of Chao (1933/2006).

As the pitch contours of the QN-tail for all the tones exhibit a high-rising pattern, people may wonder whether the intonation output of QN-tail is different from the intrinsic high-rising tone T2. Here in Figure 4-8 we compare the tone contours of the QN-tail (the same contours displayed in Figure 4-7) with the T2 contours in different contexts. In Figure 4-8, the tone contours of the QN-tail are represented by smaller symbols and grey lines, while T2 contours in different contexts are denoted by larger symbols and black lines. The T2 QN-tail, which is the intersection of the two comparing groups, is represented by smaller symbols but connected by a black line.

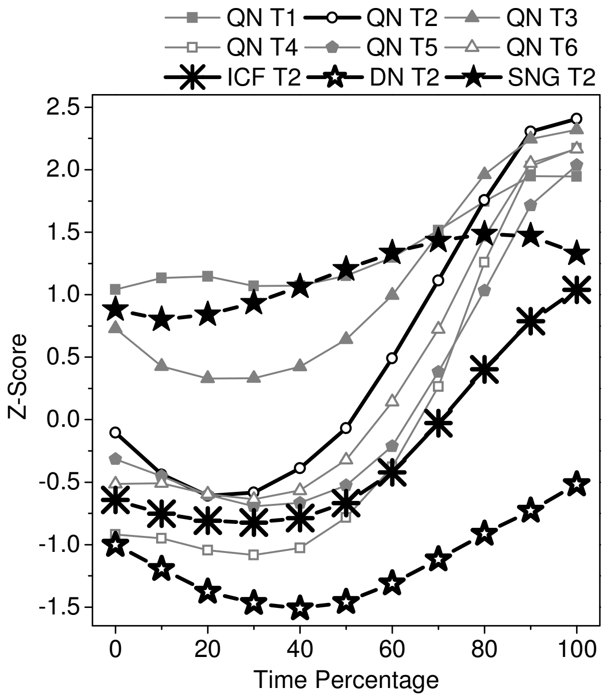

Figure 4-8 Different tones in the final position of QN versus T2 in ICF, DN and SNG.

From Figure 4-8 it can be observed that the ending points of the QN-tail for the different tones are significantly higher than the ending points of T2 in other contexts. The T2 contours in different contexts have similar starting points in the different speaking speech styles, including the ICF, DN and QN versions, while the starting point is much higher in the singing speech style. It can be inferred that the intonation effects in speaking mainly exert their effects on the rising slope and the height of the end points of T2 (QN > ICF > DN), while the starting points are less influenced and their heights are similar.

The ICF and SNG versions of T2 have similar ending points and are much higher than the DN version, but they are still notably lower than the ending points of tones at QN-tail. Thus, the intonational high-rise is always far beyond the tonal high-rise. This is probably the reason why people would not be confused as to whether a high-rising contour belongs to the intonational layer or the tonal layer. To be more accurate, the intonational QN-tail should be described as extra-high-rising.

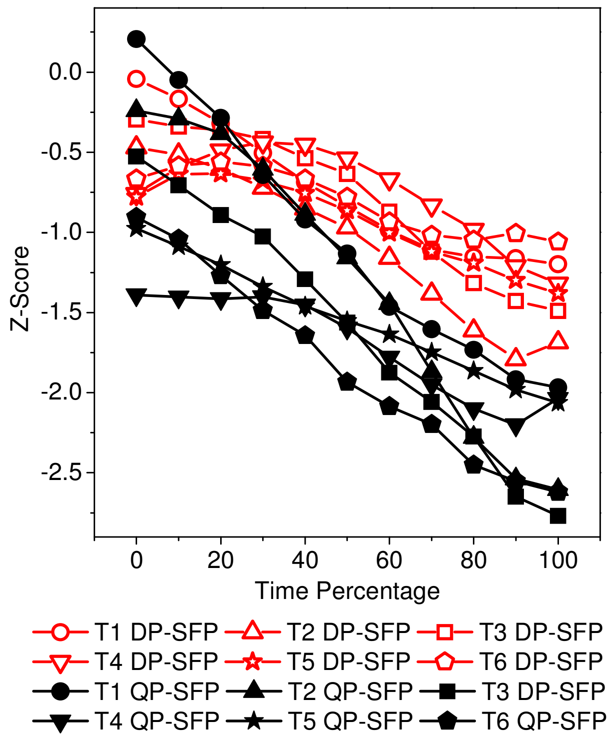

Figure 4-9 Mean SFP contours in DP and QP

Finally, we would also like to make a close examination of the SFPs in DP and QP. Their mean pitch contours are displayed in Figure 4-9. In the legend of Figure 4-9, T1 to T6 indicate the tone of the trunk that a particular SFP attaches to. The unfilled symbols represent the SFP pitch contours in DP (abbreviated as DP-SFP in Figure 4-9), while solid symbols denote the SFP pitch contours in QP (abbreviated as QP-SFP in Figure 4-9).

From the preliminary observations in Figure 4-1 (Page 131), we have already noticed that DP-SFP and QP-SFP have similar starting points but that the pitch contours of QP-SFP are usually lower and more declining than the pitch contours of DP-SFP. Figure 4-9 more clearly shows this difference. All DP-SFP and QP-SFP contours start at a similar position except the QP-SFP in the T4 utterance, which has a notably lower starting point. Excluding the DP-SFP and QP-SFP contours in T4 utterances (which have similar pitch slopes but QP-SFP has a much lower register than DP-SFP). Other contours of QP-SFP drop dramatically while their DP-SFP counterparts move downwards more mildly. Due to this slope difference, the DP-SFP group (with open symbols) and QP-SFP group (with solid symbols) have a crossing at about 30% of the time axis. Before that they have much overlap. After that all the contours of DP-SFP group are higher than the QP-SFP group.

We are also interested in whether SFP contours are affected by the progressive assimilation effect from the previous syllable. According to the ending point transcription using the five-point-scale method, the six tones in order from high to low should be: T1 = T2 > T3 = T5 > T6 >T4. The starting points (as well as about 20% of the beginning part of the pitch contours) of the QP-SFP group are almost in the same order from high to low. After the initial 20% of the contours, this order is no longer maintained. The DP-SFP group does not show the same sequence. Throughout the pitch contours, we could not find a clear pattern. It can be inferred that DP-SFP is not affected by the previous syllable, while QP-SFP is influenced by a progressive assimilation effect at the beginning part of the syllable.

4.3 Results and discussions of syllable-level measurements

Similar to the analysis in Section 3.3 to Section 3.5 for the differences between speaking and singing, here we will also analyze the differences of different sentence types in the pitch configuration - pitch domain framework.

At the syllable level, the bar-charts in Figure 4-10, Figure 4-11 and Figure 4-12 display the mean data of syllable slope, syllable register and syllable span, respectively. These bar-charts can directly show the contrast among the different sentence types.

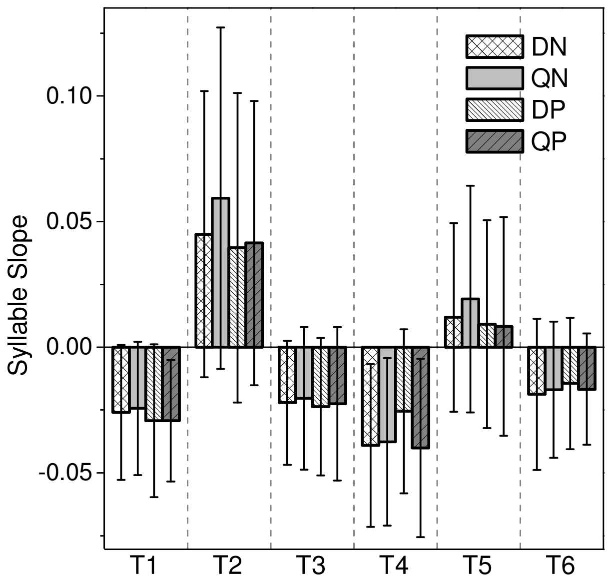

Figure 4-10 Mean pitch slope of syllable in DN, QN, DP and QP, with error-bars indicating ±1 standard deviation.

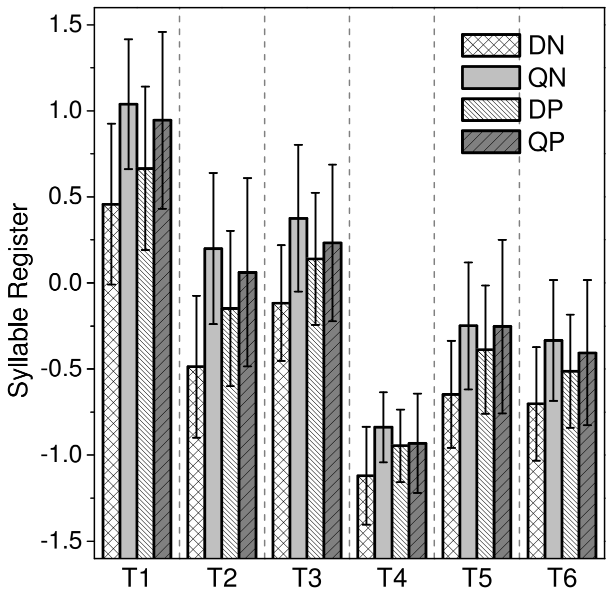

Figure 4-11 Mean pitch register of syllable in DN, QN, DP and QP, with error-bars indicating ±1 standard deviation.

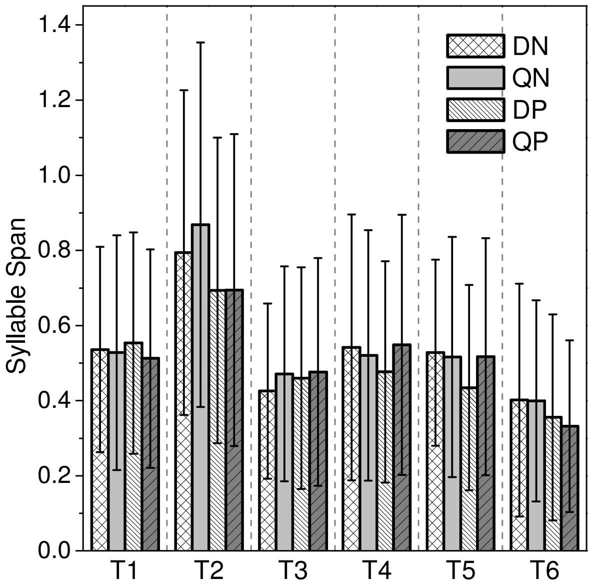

Figure 4-12 Mean pitch span of syllable in DN, QN, DP and QP, with error-bars indicating ±1 standard deviation.

To test whether the differences among different sentence types are significant, further analysis was carried out. Firstly we conducted Levene's test of homogeneity of variances. If the result was above 0.05, a One-Way ANOVA was then carried out; otherwise the Brown-Forsythe test was conducted instead. This is because the One-Way ANOVA is a more powerful statistical test than the Brown-Forsythe test, but it is only applicable when equal variances are assumed. The Brown-Forsythe test is less powerful, but it is applicable even when equal variances are not assumed. The results of the One-Way ANOVA / Brown-Forsythe tests are listed in Table 4-2.

If the *p*-value of the One-Way ANOVA / Brown-Forsythe test is greater than 0.05, the between-group difference (the data groups refer to the four sentence types) is not significant, and they can be regarded as equal. If the *p*-value of the One-Way ANOVA / Brown-Forsythe test is less than 0.05, it can be inferred that the between-group difference is significantly more than the within-group difference. In this case, at least one pair of the groups is significantly different. We then have to figure out which pair(s) of groups is (are) significantly different. For this purpose, Post Hoc multiple comparisons were carried out. There are also two kinds of Post Hoc multiple comparisons: the Tukey HSD test is more powerful but equal variance is required; while the Games-Howell test is applicable in a situation when equal variance is not assumed. Thus, if the *p*-value of Levene's test is above 0.05, a Tukey HSD test was conducted; if below 0.05, the Games-Howell test was carried out instead. The results of the Post Hoc multiple comparisons are listed in Table 4-3, except for the cases where the *p*-value of One-Way ANOVA / Brown-Forsythe test is greater than 0.05.

Table 4-2 One-Way ANOVA / Brown-Forsythe test on syllables in DN, QN, DP and QP

Tones	Test	df_1	Slope		
			$df_{2\ slope}$	F_{slope}	p_{slope}
T1	One-Way ANOVA	3	404	0.683	0.563
T2	One-Way ANOVA	3	404	2.146	0.094
T3	One-Way ANOVA	3	404	0.244	0.865
T4	One-Way ANOVA	3	404	4.400	0.005
T5	One-Way ANOVA	3	404	1.388	0.246
T6	One-Way ANOVA	3	404	0.475	0.700
Tones	Test	df_1	Register		
			$df_{2\ register}$	$F_{register}$	$p_{register}$
T1	Brown-Forsythe	3	391.324	34.820	0.000
T2	One-Way ANOVA	3	404	39.608	0.000
T3	One-Way ANOVA	3	404	25.560	0.000
T4	Brown-Forsythe	3	366.093	21.292	0.000
T5	Brown-Forsythe	3	366.231	21.999	0.000
T6	One-Way ANOVA	3	404	19.178	0.000
Tones	Test	df_1	Span		
			$df_{2\ span}$	F_{span}	p_{span}
T1	One-Way ANOVA	3	404	0.240	0.868
T2	One-Way ANOVA	3	404	3.852	0.010
T3	One-Way ANOVA	3	404	0.645	0.587
T4	One-Way ANOVA	3	404	1.016	0.385
T5	One-Way ANOVA	3	404	2.300	0.077
T6	One-Way ANOVA	3	404	1.616	0.185

Table 4-3 Post Hoc (Tukey HSD / Games-Howell) multiple comparisons on syllables of DN, QN, DP and QP

Tones	Test	Slope					
		DN-QN	DN-DP	DN-QP	QN-DP	QN-QP	DP-QP
T4	Tukey HSD	0.990	0.020	0.997	0.048	0.957	0.008
Tones	**Test**	**Register**					
		DN-QN	DN-DP	DN-QP	QN-DP	QN-QP	DP-QP
T1	Games-Howell	0.000	0.010	0.000	0.000	0.894	0.000
T2	Tukey HSD	0.000	0.000	0.000	0.000	0.156	0.006
T3	Tukey HSD	0.000	0.000	0.000	0.000	0.053	0.339
T4	Games-Howell	0.000	0.000	0.000	0.001	0.039	0.973
T5	Games-Howell	0.000	0.000	0.000	0.043	1.000	0.122
T6	Tukey HSD	0.000	0.001	0.000	0.003	0.491	0.130
Tones	**Test**	**Span**					
		DN-QN	DN-DP	DN-QP	QN-DP	QN-QP	DP-QP
T2	Tukey HSD	0.637	0.353	0.358	0.637	0.023	1.000

With the bar-charts in Figure 4-10 to Figure 4-12 and the statistical results in Table 4-2 and Table 4-3 in place, detailed and accurate analysis at the syllable level can now be provided.

For syllable slope, from Figure 4-10 we can observe that the data of different sentence types do not show clear differences. The only observable consistent pattern is that the values of QN are always slightly greater than those of DN. The One-Way ANOVA / Brown-Forsythe test results listed in Table 4-2 show that for most tones, p-values are above 0.05, which indicate that they do not have a significant difference and can be regarded as equal. The odd one out is T4, for which $p = 0.005$. Figure 4-10 and the results of Post Hoc multiple comparisons listed in Table 4-3 show that for syllable slope of T4, DP is significantly greater than DN, QN and QP, while DN, QN and QP do not have significant difference between each other.

The pattern of syllable register is more regular. As displayed in Figure 4-11, the bars of different tones show the following pattern: DN < DP < QP < QN, which also agrees with what we observed in Figure 4-2 (Page 132). The One-Way ANOVA / Brown-Forsythe test results listed in Table 4-2 are all less than

0.001, which means at least one pair of sentence types are significantly different. The results of the Post Hoc multiple comparisons listed in Table 4-3 indicate that except for QP-QN and QP-DP, all the pairs of sentence types have significantly different syllable register. It should also be noted that in the sequence of DN < DP < QP < QN (as can be observed from Figure 4-11), QP is between DP and QN, and thus possibly the data of QP partly overlap with those of DP and QN. For QP and QN, the syllable register of all tones except T4 is not significantly different. For QP and DP, only T1 and T2 show significant difference between these two sentence types, and for other tones the data of QP and DP can be regarded as equal.

The comparison of syllable span does not show a clear sequence of the four sentence types in Figure 4-12. The bars of DN, QN, DP and QP have similar heights. The One-Way ANOVA / Brown-Forsythe test results listed in Table 4-2 further support our observation as all the *p*-values except T2 are greater than 0.05, i.e., for most tones there is no significant difference among different sentence types. For T2, the results of Post Hoc multiple comparisons listed in Table 4-3 show that only QN and QP have significant difference among the sentence types. This exception is probably caused by an accidental (as opposed to regular) effect rather than a certain factor.

To sum up the overall pitch configuration at the syllable level, syllable slope and syllable span do not have fixed order and significant differences among DN, QN, DP and QP. There is a regular pattern for syllable register, which shows the sequence of DN < DP < QP < QN. Apart from the two pairs of sentence types QP-DP and QP-QN, the difference between the other pairs is statistically significant.

4.4 Results and discussions of utterance-body-part measurements

For the analysis in the utterance-body portion, we will again first display the bar-charts of utterance slope, utterance register, and utterance span in Figure 4-13, Figure 4-14, and Figure 4-15, respectively. The bar-charts are followed by further statistical analysis, which test whether the differences are significant or not.

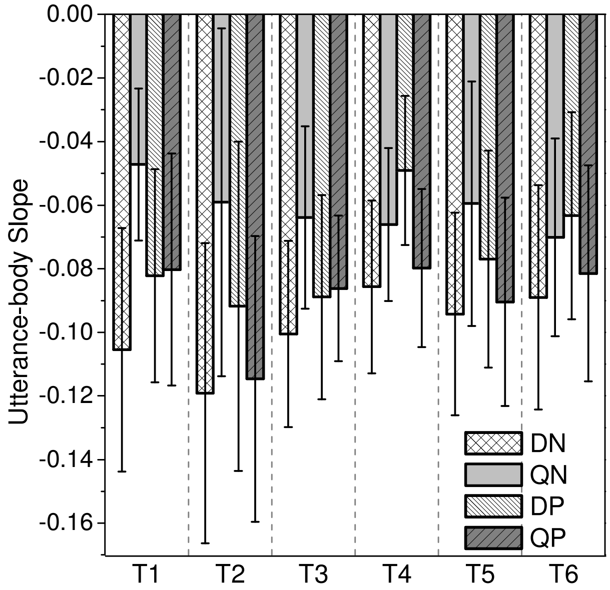

Figure 4-13 Mean pitch slope of utterance body in DN, QN, DP and QP, with error-bars indicating ±1 standard deviation.

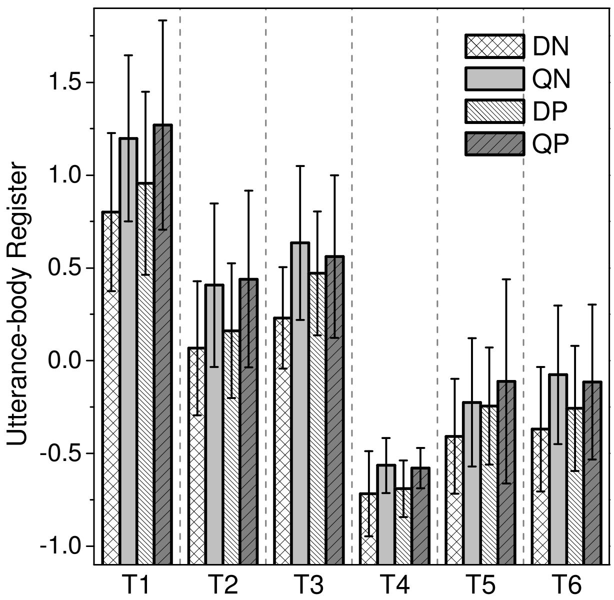

Figure 4-14 Mean pitch register of utterance body in DN, QN, DP and QP, with error-bars indicating ±1 standard deviation.

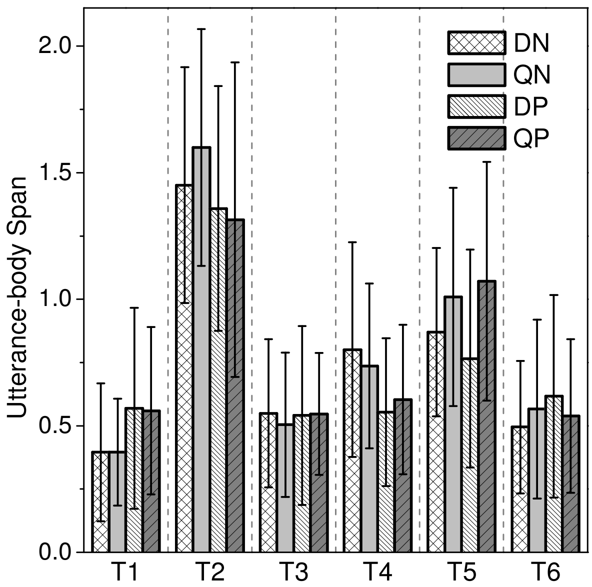

Figure 4-15 Mean pitch span of utterance body in DN, QN, DP and QP, with error-bars indicating ±1 standard deviation.

Table 4-4 One-Way ANOVA on utterance body in DN, QN, DP and QP

Tones	df_1	Slope		
		$df_{2\ slope}$	F_{slope}	p_{slope}
T1	3	44	6.132	0.001
T2	3	44	3.653	0.020
T3	3	44	3.481	0.024
T4	3	44	5.125	0.004
T5	3	44	2.505	0.071
T6	3	44	1.428	0.247
Tones	df_1	Register		
		$df_{2\ register}$	$F_{register}$	$p_{register}$
T1	3	44	2.388	0.082
T2	3	44	2.348	0.086
T3	3	44	2.686	0.058
T4	3	44	2.610	0.063
T5	3	44	1.158	0.336
T6	3	44	1.613	0.200
Tones	df_1	Span		
		$df_{2\ span}$	F_{span}	p_{span}
T1	3	44	1.181	0.328
T2	3	44	0.724	0.543
T3	3	44	0.060	0.981
T4	3	44	1.377	0.262
T5	3	44	1.290	0.290
T6	3	44	0.280	0.839

Table 4-5 Post Hoc (Tukey HSD) multiple comparisons on utterance-body slope of DN, QN, DP and QP

Tones	DN-QN	DN-DP	DN-QP	QN-DP	QN-QP	DP-QP
T1	0.001	0.336	0.268	0.065	0.089	0.999
T2	0.025	0.540	0.996	0.384	0.043	0.675
T3	0.015	0.750	0.610	0.152	0.233	0.996
T4	0.231	0.004	0.937	0.351	0.538	0.021

In the statistical tests on utterance body, for all the tested items, the Levene statistical tests of homogeneity of variances have results greater than 0.05, which indicates equal variances can be assumed. Thus, One-Way ANOVA can be conducted (the results are listed in Table 4-4), and for the Post Hoc test, we used the Tukey HSD test (the results are shown in Table 4-5).

For utterance-body slope, as shown in Figure 4-13, all the bars are below zero, which indicates that all the sentence types have a downward utterance slope. As all the sentence types belong to the speaking speech style, we are not surprised to see this result. Among the four sentence types, DN is always the smallest. The greatest bars associate with two sentence types: QN is the greatest for T1, T2, T3 and T5, and DP is the greatest for T4 and T6. DP and QP are usually between DN and QN. For T1 and T3, DP is smaller than QP; for other tones, DP is greater than QP. The One-Way ANOVA results in Table 4-4 show that for T1 to T4, p-values are less than 0.05, i.e., the between-group difference is significant for these four tones. For T5 ($p = 0.071$) and T6 ($p = 0.247$), the four sentence types do not have significant utterance-body slope difference. In Table 4-5, Tukey HSD tests were conducted on T1 to T4 for Post Hoc multiple comparisons. The p-values are less than 0.05 for "DN-QN" of T1, T2 and T3. For these three tones, the data of QN are significantly greater than DN. For T2, QN is also significantly greater than QP. T4 is different from T1, T2 and T3. For T4, DP is the greatest among the four sentence types, and the Post Hoc multiple comparisons show that DP is significantly greater than DN and QP.

In terms of utterance-body register, the bars in Figure 4-14 show that the question types of QN and QP are always higher than the declarative sentences DN and DP. For the two question types, QN is higher than QP for T3, T4 and T6, but lower than QP for T1, T2 and T5. For the two declarative sentence types, DN is always lower than DP. Thus, DN is the lowest among the four sentence types. The p-values of the One-Way ANOVA for "Register" shown in Table 4-4 are all greater than 0.05, which indicates that the utterance-body register is not significantly different among the four sentence types. Therefore, in terms of utterance-body register, the declarative sentences (especially DN, which is the lowest) are slightly lower than the question sentences.

Finally let us observe the data of utterance-body span. The data comparison in Figure 4-15 does not show a consistent pattern for the four sentence types, and the bars of different sentence types seem to have similar lengths. In Table 4-4, the *p*-values of the One-Way ANOVA for "Span" are also all greater than 0.05, i.e., the between-group difference is not significant for all the tones. Statistically speaking, the utterance-body span of different sentence types can be regarded as equal.

4.5 Results and discussions of utterance-final-part measurements

In this section, as we are discussing the utterance-final intonation, the sentences types with SFP, i.e. DP and QP, are excluded. SFP is itself a functional word that fulfills the function of expressing sentential connotations. When discussing utterance-final intonation, we are discussing the intonation effects on syllables of content words with regular tones. Thus, we only focus on the sentences without SFP, i.e. DN and QN, and observe how the final syllables of these two types of utterances carry the function of expressing intonation.

We will analyze the difference of pitch slope, pitch register and pitch span between final and non-final syllables. The analysis on DN and QN will be conducted separately. As in previous sections (Section 4.3 and Section 4.4), bar-charts of the mean data will be displayed first, followed by statistical tests. For the statistical tests, if the *p*-value of Levene's test was greater than 0.05, a One-Way ANOVA was conducted; otherwise the Brown-Forsythe test was carried out instead.

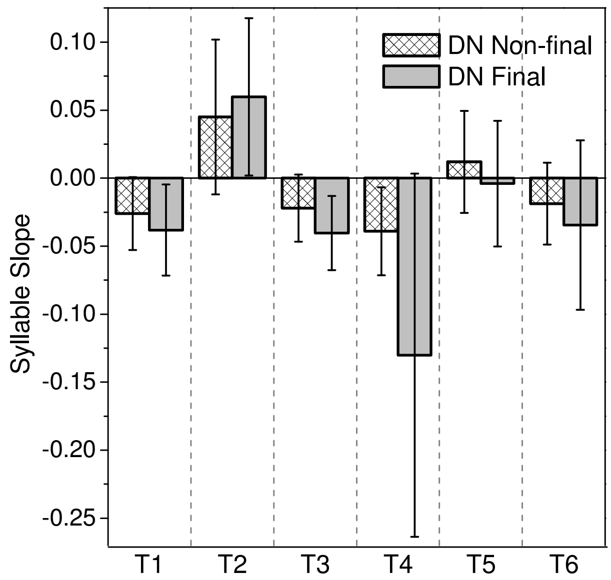

Figure 4-16 Comparison of mean pitch slope between final and non-final syllables in DN, with error-bars indicating ±1 standard deviation.

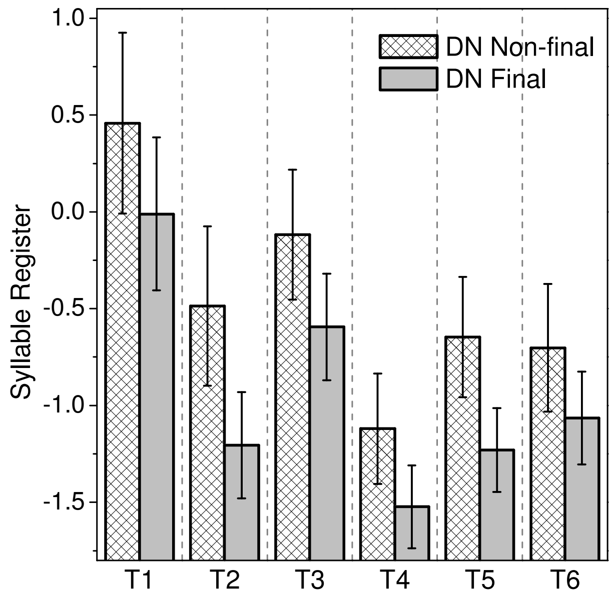

Figure 4-17 Comparison of mean pitch register between final and non-final syllables in DN, with error-bars indicating ±1 standard deviation.

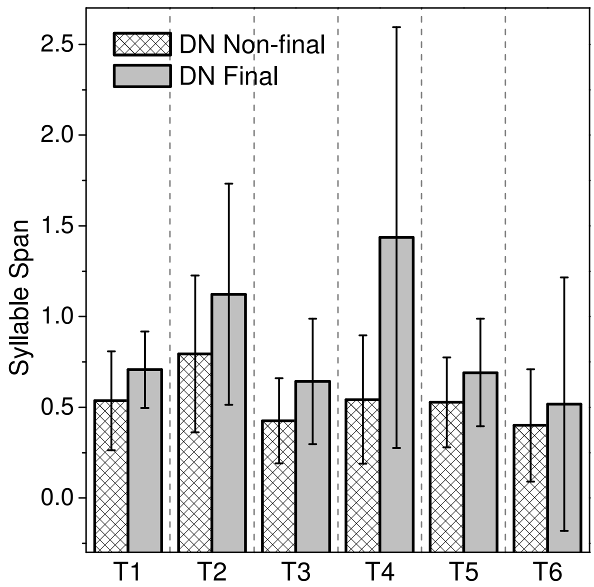

Figure 4-18 Comparison of mean pitch span between final and non-final syllables in DN, with error-bars indicating ±1 standard deviation.

Table 4-6 One-Way ANOVA / Brown-Forsythe test on final and non-final syllables for DN

Tones	Test	df_1	Slope		
			$df_{2\ slope}$	F_{slope}	p_{slope}
T1	One-Way ANOVA	1	106	2.075	0.153
T2	One-Way ANOVA	1	106	0.722	0.397
T3	One-Way ANOVA	1	106	5.687	0.019
T4	Brown-Forsythe	1	11.162	5.538	0.038
T5	One-Way ANOVA	1	106	1.820	0.180
T6	One-Way ANOVA	1	106	2.180	0.143
Tones	Test	df_1	Register		
			$df_{2\ register}$	$F_{register}$	$p_{register}$
T1	One-Way ANOVA	1	106	11.122	0.001
T2	One-Way ANOVA	1	106	34.500	0.000
T3	One-Way ANOVA	1	106	22.297	0.000
T4	One-Way ANOVA	1	106	22.396	0.000
T5	One-Way ANOVA	1	106	39.519	0.000
T6	One-Way ANOVA	1	106	13.584	0.000
Tones	Test	df_1	Span		
			$df_{2\ span}$	F_{span}	p_{span}
T1	One-Way ANOVA	1	106	4.365	0.039
T2	One-Way ANOVA	1	106	5.606	0.020
T3	One-Way ANOVA	1	106	8.195	0.005
T4	Brown-Forsythe	1	11.257	7.043	0.022
T5	One-Way ANOVA	1	106	4.423	0.038
T6	One-Way ANOVA	1	106	1.063	0.305

The bar-charts of final and non-final syllables for DN are displayed in Figure 4-16, Figure 4-17 and Figure 4-18, and the related statistical results are listed in Table 4-6.

In terms of syllable slope, as displayed in Figure 4-16, the data for final syllables are generally smaller than for non-final syllables except for T2, i.e., for most tones the pitch slope declines more steeply in the final position. Of particular interest is T5, the mid-rising tone, which has positive values for pitch slope in non-final positions but negative values for final syllables. This indicates that T5 preserves its rising slope in non-final positions but it is suppressed by the utterance-final declination, so much so that it even exhibits a declining pitch contour as the ultimate acoustic output when it is the final syllable. As the only exception, T2 has its rising slope strengthened (rather than weakened by the slope-declining effect) at the final position.

The One-Way ANOVA / Brown-Forsythe test results in the rows for "Slope" in Table 4-6 show that all tones except T3 (p = 0.019) and T4 (p = 0.038) have p-values greater than 0.05. Thus, although the utterance-final intonation of DN has a slope-declining effect on final syllables, this effect is often insignificant. This is not consistent with what we found about SPK (in Section 3.5, where the SPK sentences are also declarative sentences without SFP).

SPK and DN sentences show a similar relationship between final and non-final syllables: final syllables are more declining than non-final syllables; but the pattern shows different levels of significance for SPK and DN sentences. For SPK, the difference between final and non-final syllables is generally significant; while for more than half of the tones in DN, final and non-final syllables are not significantly different. One important difference between the SPK sentences in Chapter 3 and the DN sentences in the present chapter is that the DN sentences are composed of syllables of the same tone; while the SPK sentences are composed of syllables of different tones (every utterance covers six tones). Probably there is a mono-tone effect on the DN sentences which leads to this difference between SPK and DN. However, to make an utterance mono-tonal is the most convenient and direct way to observe utterance-body intonation. At the moment all we can do is keep in mind that there may be a kind of mono-tone effect in operation.

The pattern of syllable register is much neater in the comparison between final and non-final syllables in DN. For all the tones, the bars of non-final syllables are much longer than the final syllables in Figure 4-17. One-Way ANOVA / Brown-Forsythe tests were also conducted. The p-values are all equal to or less than 0.001 as shown in the "Register" rows of Table 4-6. Thus, the final syllables have a significantly lower register than non-final syllables. The utterance-final declination effect of DN is more obvious on pitch register than on pitch slope.

Finally we have the analysis on syllable span. Figure 4-18 shows that the syllable span of final syllables is much larger than the syllable span of non-final syllables. Possibly this is due to the fact that final syllables usually have a more tilted pitch slope. The One-Way ANOVA / Brown-Forsythe test results listed in the "Span" rows of Table 4-6 show that all *p*-values are less than 0.05 except T6. Thus, generally speaking, final syllables have a significantly larger pitch span than non-final syllables in DN.

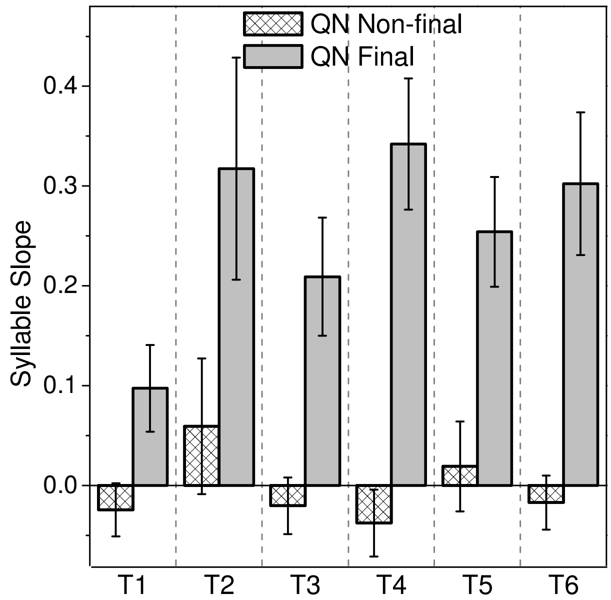

Figure 4-19 Comparison of mean pitch slope between final and non-final syllables in QN, with error-bars indicating ±1 standard deviation.

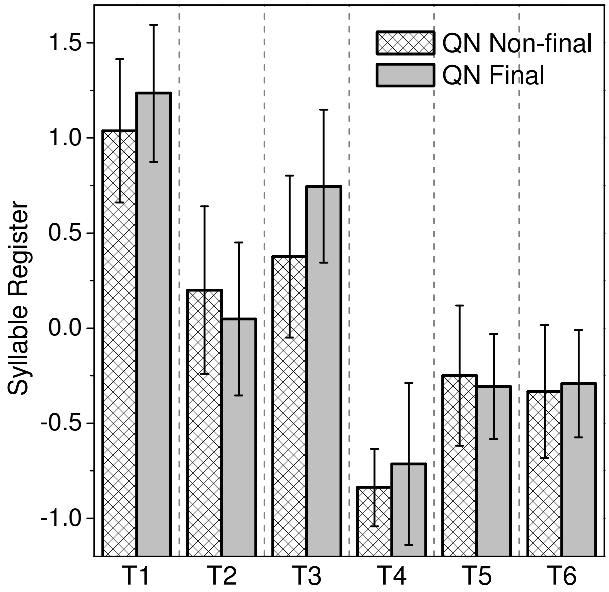

Figure 4-20 Comparison of mean pitch register between final and non-final syllables in QN, with error-bars indicating ±1 standard deviation.

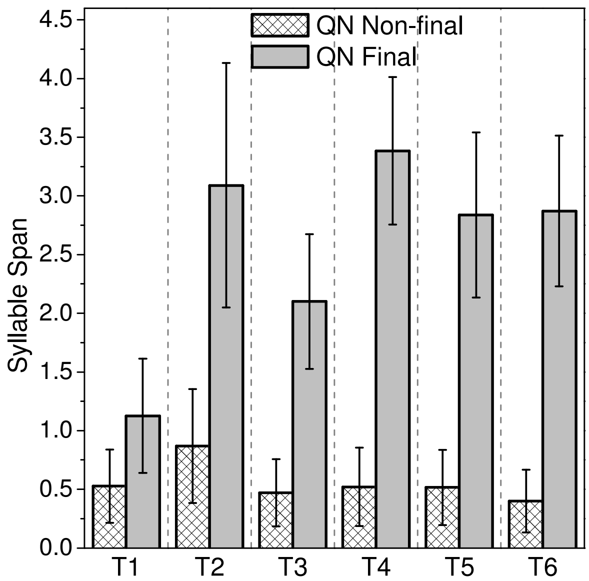

Figure 4-21 Comparison of mean pitch span between final and non-final syllables in QN, with error-bars indicating ±1 standard deviation.

Table 4-7 One-Way ANOVA / Brown-Forsythe test on final and non-final syllables for QN

Tones	Test	df_1	Slope		
			$df_{2\ slope}$	F_{slope}	p_{slope}
T1	Brown-Forsythe	1	12.049	89.768	0.000
T2	Brown-Forsythe	1	12.050	61.870	0.000
T3	Brown-Forsythe	1	11.640	175.395	0.000
T4	Brown-Forsythe	1	11.718	387.282	0.000
T5	One-Way ANOVA	1	106	275.560	0.000
T6	Brown-Forsythe	1	11.400	235.920	0.000
Tones	Test	df_1	Register		
			$df_{2\ register}$	$F_{register}$	$p_{register}$
T1	One-Way ANOVA	1	106	2.954	0.089
T2	One-Way ANOVA	1	106	1.284	0.260
T3	One-Way ANOVA	1	106	8.143	0.005
T4	Brown-Forsythe	1	11.637	0.986	0.341
T5	One-Way ANOVA	1	106	0.278	0.599
T6	One-Way ANOVA	1	106	0.166	0.685
Tones	Test	df_1	Span		
			$df_{2\ span}$	F_{span}	p_{span}
T1	One-Way ANOVA	1	106	34.041	0.000
T2	Brown-Forsythe	1	11.603	53.162	0.000
T3	Brown-Forsythe	1	11.691	93.581	0.000
T4	Brown-Forsythe	1	11.784	240.274	0.000
T5	Brown-Forsythe	1	11.575	127.737	0.000
T6	Brown-Forsythe	1	11.481	174.157	0.000

For the comparison between final and non-final syllables for QN, the bar-charts are shown in Figure 4-19, Figure 4-20 and Figure 4-21, and the results of the further statistical analysis are listed in Table 4-7.

Figure 4-19 displays the comparison of syllable slope. The non-final syllables in QN preserve the intrinsic tone-shapes, i.e., for T1, T3, T4 and T6, the data of pitch slope is below zero; while for T2 and T5, the bars of syllable slope is above zero. The pitch slope data for non-final syllables in QN look similar to their counterparts in DN (which are displayed in Figure 4-16, Page 152).

In the preliminary observation in Section 4.2, we noticed that the most outstanding characteristic of QN is that it has a sharply rising pitch contour in utterance-final position. Thus, it is not surprising to see that in Figure 4-19, the bars for final syllables are all well above zero. Although the intrinsic tone shape of T1, T3, T4 and T6 should be level or falling, meaning their syllable slope values should be zero or minus, here the intrinsic tone shapes are overridden by the sharp rise dictated by the QN-final intonation. One-Way ANOVA / Brown-Forsythe test results listed in Table 4-7 further indicate that the pitch slope difference between final and non-final syllables is significant as all the *p*-values of "Slope" are less than 0.001.

It should also be noted that the QN-final syllable slope is not simply the algebraic sum of the intrinsic tone slope (or QN-non-final slope) and a constant factor. The mean bars of the final syllables are in the order of T1 < T3 < T5 <T6 < T2 < T4. The two rising tones T2 and T5, which are supposed to have the greatest intrinsic tone slope values, do not show the highest bars for the pitch slope of final syllables. The pitch slope of QN-final syllables is more related to the register of the beginning point of the intrinsic tone shapes. The two are in fact negatively correlated: if the beginning point of the intrinsic tone shape is low, the value of the QN-final pitch slope is high, and vice versa. It can be found that the data displayed in Figure 4-19 are quite consistent with what we observed from Figure 4-7 (Page 137) which shows the mean QN-final pitch contours of T1 to T6. In Figure 4-7, the beginning part of the pitch contours are in the order of T1 > T3 > T5 / T6 / T2 > T4 (T5, T6 and T2 are very close and cluster together), which is nearly the opposite order to the pitch-slope bars of the final syllables in Figure 4-19. The key to interpreting this phenomenon is that all the QN-final syllables share the similar extremely high ending point regardless of their original tone category and that the original tone category is somewhat preserved in the first half of the final syllable. Thus, the lower the beginning point, the more tilted the pitch slope.

The comparison of pitch register for QN is shown in Figure 4-20. The comparison between final and non-final syllables shows divergence: for T1, T3, T4 and T6, the mean bars of the final syllables are longer than those of the non-final syllables; for T2 and T5, it is the opposite. In the One-Way ANOVA / Brown-Forsythe tests, the rows of "Register" in Table 4-7 show that the *p*-values of all tones are greater than 0.05 except T3 (p = 0.005). This indicates that in terms of pitch register, final syllables and non-final syllables in QN do not show significant differences in most cases.

Finally is the comparison of pitch span. As the QN-tail showed the outstanding characteristic of sharp rising in the preliminary observation in Figure 4-1 (Page 131), we already expected that final syllables have much wider pitch span than non-final syllables. The bar-chart of Figure 4-21 and the One-Way ANOVA / Brown-Forsythe tests on “Span” in Table 4-7 further support our observation. In Figure 4-21, the mean bars of the final syllables of all the tones are much longer than the non-final syllables. The p-values of “Span” in Table 4-7 are all less than 0.001, which shows that the pitch span difference between final syllables and non-final syllables in QN is significant.

To sum up the utterance-final intonation of DN and QN, it is interesting that for DN, the utterance-final intonation mainly affects pitch register rather than pitch slope; while for QN, the utterance-final intonation mainly influences pitch slope rather than pitch register. For both types of sentences, the pitch span of the final syllables is significantly wider than that of non-final syllables.

Chapter 5
Overall Discussion and Conclusion

In Chapter 2 though Chapter 4, we reported the methodologies and results of the three experiments we conducted. In this chapter, we will aggregate all these experimental results for an overall discussion. We will firstly have a detailed discussion on the separate experiments, and then move on to the overall discussion. For the overall discussion, we will firstly look at the interrelationship of the three experiments, and then conduct general discussions along the two dimensions of the framework proposed in the present study – pitch configuration and pitch domain. Further discussions that are of special interest to us will come next, including further remarks on declination versus non-declination, further remarks on intonation at the syllable level, and further remarks on utterance-final intonation and SFP. Finally comes the overall conclusion of the present study and suggestions for future research.

5.1 Discussion of individual experiments

5.1.1 *Discussion of Experiment I*

It is remarkable and interesting that Cantonese lexical tones can be identified in both speaking and singing, in both normal and operatic settings. Through systematically studying the three pairs of speech styles – opera speaking (OSpk) versus opera singing (OSng), opera speaking (OSpk) versus normal speaking (NSpk), and normal speaking (NSpk) versus normal singing (NSng), we endeavored to find out which are the essential factors that differentiate speaking and singing, and whether these factors are solely for the distinction of speaking and singing, or whether they can also be used to indicate the characteristics of other speech styles (i.e., whether a certain factor is applicable to the distinction between speaking and singing, as well as the distinction between opera speech and normal speech).

Before discussing the pitch configuration and pitch domain of different speech styles, we would like to make it clear where the tone-realizing part lies within the pitch contour of a syllable. In most cases, when a syllable is not extremely long, the tone is carried by the whole syllable pitch contour. However, when a syllable is extremely long, say over 500 milliseconds (seldom observed in normal speaking; generally only found in opera performance and singing, especially at the utterance-final position), the pitch contour of the syllable is usually curving and only part of it serves to represent the tone of the syllable.

The tone-realizing portion for the lengthened curving syllables is different in speaking and singing. In singing, lexical tones can be realized in one of the following ways: (1) lexical tones are realized at the beginning part and the remaining pitch contour is free to change its movement trajectory; (2) the whole syllable pitch contour is utilized, and the lexical tone is approximated by two successive level pitch contours. The first way is adopted much more frequently than the second one. In opera speaking, different from the lengthened curving syllables of singing, the tone-realizing part is the latter portion of the syllable pitch contour: there is a preparatory part that moves in the opposite direction at the very beginning and the lexical tones are realized after this preparation. In our measurements, calculations and analyses, only the tone-realizing part is involved.

For the analysis of pitch configuration and pitch domain, in order to show the comparison among the three pairs of speech styles more clearly and directly, we list our findings in Table 5-1.

Table 5-1 The comparison among the three pairs of speech styles

Performer	Style of speech		Syllable comparison: whether significant differences exist		Utterance trend	Utterance-final intonation
	Speaking	Singing	Pitch slope	Pitch register		
MCS	OSpk	OSng	√	×	Declination versus Non-declination (even uptrend)	NA
YSF	OSpk	NA	×	√	Both declination	Pitch slope of final syllables: OSpk more tilted than NSpk (as displayed in Figure 2-6 to Figure 2-8, Page 65 to 67)
	NSpk					
ZSX	NSpk	NSng	√	√	Declination versus Non-declination (even uptrend)	NSpk: Final declination NSng: No final-intonation

As shown in Table 5-1, the styles of speech can influence pitch configuration in the domain of the individual syllables, the body of the whole utterance, as well as in the special position of the utterance-final portion of the utterance.

At the syllable level, it can be seen that the data of both MCS and ZSX exhibit significant differences of pitch slope between speaking and singing – level and falling tones have steeper slopes in speaking than in singing. The pitch slope of rising tones is also influenced by the speaking declination effect, which reduces their rising slope in speaking. This effect is significant for both T2 and T5 in MCS's data while only significant for T5 in ZSX's data.

In terms of pitch register, as a professional actor, MCS maintains the pitch register of both speaking and singing at nearly the same level. ZSX is an ordinary person and tends to make singing a more formal and important style of speech by raising the whole pitch register up.

YSF also uses a higher pitch register for OSpk than for NSpk, but the pitch slope does not show significant difference between these two subcategories of speaking. It can be inferred that higher pitch register may be used to emphasize that a speech style is extraordinary, such as singing (compared with speaking) and opera performance (compared with normal speech). As for the distinguishing factor between speaking and singing, it appears to be the differing gradient of syllable slope, as revealed in the data of both MCS and ZXS.

At the utterance level, because of the limited data in this pilot experiment, utterance-body intonation and utterance-final intonation were not systematically studied in all. For the data of MCS, only the utterance-body intonation has been studied. For the data of YSF, only preliminary observations have been made of pitch contours of utterances. Only the data of ZSX are relatively complete and analysis on both utterance-body intonation and utterance-final intonation has been carried out.

From all these analyses at the utterance level, it can be seen that both utterance-body intonation and utterance-final intonation are different in speaking and singing: speaking shows declination while singing exhibits non-declination. For utterance-body intonation, the data of both MCS and ZXS show that speaking utterances conform more strictly to the downdrift pattern, while singing utterances have more freedom not to follow this trend, and sometimes even show a kind of uptrend. In our preliminary observation of YSF's pitch contours of the utterances, both OSpk and NSpk show strong declination.

Thus, for speaking speech styles, there is a kind of consistent declination intonation at both the syllable and utterance levels. Whether for syllables or utterances, the later part has lowered pitch in speaking. This lowering effect is further strengthened at the utterance-final position.

For utterance-final intonation, the pitch slope of final syllables in OSpk is usually more tilted than in NSpk, the mechanism of which may be the stressing effect of OSpk. The comparison between NSpk and NSng shows that their mechanisms are different. Through related calculations and statistical analyses on final and non-final syllables of ZXS's recordings, we found that there is an evident final declination effect in NSpk while the final and non-final syllables are regarded as equal in NSng.

Our investigation of the final declination effect of NSpk involves both the aspects of pitch register and pitch slope. We found that for falling/level tones, the final syllables have a significantly lower pitch register and a more steeply declined pitch slope; while for rising tones, the utterance-final declination effect is only reflected in pitch register lowering rather than pitch slope modification.

The studies in Experiment I help us to get a preliminary impression of how speaking styles are different from singing styles, but the data were limited, especially for the analysis at the utterance level. To make our conclusions more convincing, we should strengthen the framework of our studies and engage more informants in further experiments. These improvements were made in Experiment II.

5.1.2 *Discussion of Experiment II*

Experiment II continues the research topic of the intonational differences between speaking and singing speech styles. Here we only focus on normal speech styles of speaking (SPK) and singing (SNG): twelve layman native informants were enrolled to participate in our experiment and asked to read aloud and sing the same lyrics from a deliberately designed questionnaire.

Compared with Experiment I, the research framework of Experiment II is more clearly spelled out. The framework makes use of the two dimensions of pitch configuration and pitch domain. For the dimension of pitch configuration, intonation effects can influence the pitch contours in the following three ways: (1) pitch slope ascending or descending; (2) pitch register displaced upwards or downwards; (3) pitch span being widened or narrowed. As for pitch domain, the above pitch configuration factors can affect the domains of: (1) syllable level; (2) utterance level, which includes the utterance-body portion and the utterance-final portions.

To facilitate the measurement and calculation of the three pitch configuration factors, we have defined the corresponding parameters. Pitch slope is defined as the slope of the linear regression of the fitted pitch contour. For syllables, linear regression was conducted on all eleven pitch points within the syllable; for utterances, linear regression was conducted on the median points of every syllable within the utterance. Pitch register is represented by the median of pitch points within a syllable or an utterance. Pitch span is defined as the distance between the two extreme values of the pitch points within a syllable or an utterance.

Pitch slope, pitch register and pitch span were calculated at different pitch levels and for different portions. At the syllable level, only the non-final syllables were included for calculation. In the utterance-body portion, only the data of the reduplicative phrases were investigated as they are monotone sequence and can directly reflect the trend of utterance. In the utterance-final portion, to test whether there exists a certain type of utterance-final intonation, we compared the final syllables with the non-final syllables to observe whether there is consistent and significant difference between them. The calculation method is the same as that used at the syllable level.

If most tones (despite occasional exceptions to the overall pattern) exhibit a certain pattern in a certain kind of comparison, we believe the pattern is regular and is caused by the intonation effect of the speech style. If the comparison on different tones shows randomness without any dominant pattern, this situation is considered to be caused by effects other than the intonation effects of the speech style. The results of Experiment II (data analysis conducted in Section 3.3, Section 3.4, and Section 3.5) can be simplified and used to fill in the matrix of pitch configuration and pitch domain, as shown in Table 5-2 (Page 168).

In the matrix in Table 5-2, the symbols "+", "-" and "0" demonstrate the relative characteristics of two objects in comparison, and the presence of a "*" indicates that a "+" or a "-" characteristic is statistically significant ($p < 0.05$). "+" stands for greater pitch slope values, higher pitch register, or wider pitch span; "-" denotes smaller pitch slope values, lower pitch register, or narrower pitch span; "0" represents a pattern that is not fixed and a difference that is not significant.

As for the "Objects in comparison", at the "Syllable" level and "Utterance-body" part, the comparison is between the two speech styles of SPK and SNG.

“SPK: SNG” means that SNG is the reference object and, compared with SNG, the data of SPK are higher (+), lower (-) or similar (0). As SPK and SNG are paired in comparison, and “+”, “-” and “0” are relative values, if we want to get the values of SNG, the results are just the opposite: if the value of “SPK: SNG” is a “+”, the value of “SNG: SPK” is a “-”, and vice versa; if the value of “SPK: SNG” is a “0”, the value of “SNG: SPK” is also a “0”. In the “Utterance-final” part, the two objects in comparison are final and non-final syllables in a certain speech style. The symbols of “+”, “-” and “0” represent the relative values of final syllables compared with non-final syllables, i.e., the effects of utterance-final intonation. The utterance-final intonation effects of SPK and SNG are separately listed in the last two rows of Table 5-2.

To enhance the understanding of pitch slope variation, here is a reminder on how to interpret the pitch slope values. Pitch slope actually includes two aspects, first the direction (whether it is falling or rising), and second the steepness. For the direction of pitch slope, if the pitch slope value is below zero, the pitch contour is falling; if above zero, rising. For the steepness of pitch slope, it is indicated by the absolute value: the larger, the steeper. Thus, for rising pitch contour, the pitch slope value is above zero, and if the value is greater, the slope is steeper. For falling pitch contour, the pitch slope value is below zero, and if the value is greater, the absolute value is smaller, and the slope is less steep.

Here we should also differentiate the tone layer and the intonation layer effects on pitch slope. The “+” “-” or “0” in the “Slope” column only refer to the intonation effect. If there is a “-” in the “Slope” column, it indicates that the intonation effect makes the pitch slope value smaller. For falling tones, the pitch slope value itself is supposed to be below zero, and the smaller pitch slope value suggests the pitch slope is more descending. For rising tones, the pitch slope value itself is above zero, and the smaller pitch slope value implies the pitch slope is less ascending. A smaller pitch slope value indicates a kind of pitch slope declination effect – falling tones are more descending and rising tones are less ascending. On the contrary, if there is a “+” in the “Slope” column, there is a kind of pitch slope ascending effect, which makes the falling pitch contour less falling, and rising pitch contour more rising.

It should be noted that at the syllable level, all the level/falling tones have negative pitch slope values in SPK, but T4 and T6 have positive pitch slope values in SNG (as shown in Figure 3-12, Page 104). For rising tones, they have positive pitch slope values in both SPK and SNG, but the values in SPK are smaller (as shown in Figure 3-12, Page 104). In the utterance-body portion, SPK has negative pitch slope values and SNG has positive pitch slope values for all the tones (as shown in Figure 3-16). In the utterance-final portion, SPK shows a strong extra pitch slope declination effect in the final syllables – even the rising tone T5 acquires a declining pitch contour (as displayed in Figure 3-19), thus the pitch slope values for all the tones are below zero except for T2 (as shown in Figure 3-20).

Table 5-2 Matrix for the intonation effects of speaking and singing

Objects in comparison	Shape / Domain		Slope	Register	Span
SPK:SNG	Syllable		-*	-*	+*
	Utterance-body		-*	-*	+*
Final syllable: other syllables	Utterance-final	SPK	-*	-*	+*
		SNG	0	0	0

The matrix in Table 5-2 shows a neat distinction between speaking and singing. Speaking shows a constant declination effect for pitch slope and pitch register (i.e. downtrend of pitch slope and lowering of pitch register) at both the syllable and utterance levels (including both utterance-body and utterance-final portions). Although not directly shown in the matrix in Table 5-2, it can be inferred from the relative values that singing exhibits the opposite patterns: singing exerts consistent ascending effects on pitch slope and pitch register (i.e. uptrend of pitch slope and upstep of pitch register) in both the syllable and utterance body domains. The overall intonation pattern of speaking is downdrift while singing tends to be updrift. Different from the utterance-final declination in speaking, singing does not have an effect on the utterance-final syllables. Occasionally the intonation effect of speech style may override the original contour shape of lexical tones. For instance, at the syllable level, as shown in Figure 3-11 (Page 103), the falling/level tones T4 and T6 have slightly rising pitch contours in singing. Another example is that the mid-rising tone T5 has a falling pitch contour in final position in speaking (as shown in Figure 3-19, Page 117).

As we have defined the parameter for calculating pitch span as the distance between the extreme values, here pitch span shows a positive correlation with the steepness of pitch slope. The greater the steepness is, the wider the pitch span is. The pitch slope downtrend of speaking is usually stronger than the uptrend of singing. As a result, the pitch span of speaking is wider than that of singing in both the syllable and utterance body domains with the exception of T5 (as shown in Figure 3-14 (Page 105) and Figure 3-18 (Page 114), respectively). In the utterance-final portion, final syllables in speaking also have a wider pitch span than non-final syllables (as shown in Figure 3-25, Page 124), which earns a "+*" in the matrix.

Apart from the direct comparison of the two continuous speech styles of speaking and singing, we also observed the pitch contour patterns of tones in isolated citation form (ICF). In the comparison at the syllable level, the pitch contours of ICF and SPK are more similar to each other than they are to SNG, as displayed in the three graphs in Figure 3-11 (Page 103). We found that the level/falling tones have obvious falling pitch contours in ICF and SPK but tend to remain level or even exhibit a slight rise in SNG.

ICF can be regarded as the shortest utterance in speaking and inevitably carries the declination intonation of speaking. Traditionally people might regard ICF as exhibiting "pure tones": previous experimental studies on lexical tones were usually carried out by asking people to read aloud lexical items on a questionnaire separately, and the graphs of the acoustical data were used to represent the canonical forms of tones. Our studies here remind us that ICF is a special subcategory of speaking, and one cannot assume that the "pure tone" is without the influence of intonation. Compared with ICF, the declination effect is more obvious and stronger in SPK, which indicates that the declination effect of speaking is strengthened in continuous speech. The only exception is T4, which has a greater falling pitch slope in ICF than in SPK. Probably this is influenced by the longer duration in ICF (although not presented in this thesis, we actually measured the duration data and found that the syllables in ICF are usually much longer than the syllables in SPK). As a consequence, the falling tone T4 in ICF has more time to go downwards and fully reach its final target pitch point. T1 has a similar pitch slope to T3 and T6 in graphs of both ICF and SPK, which indicates that T1 in present-day Cantonese is a level tone (like T3 and T6) rather than a falling tone (like T4). The declining pitch slopes of T1, T3, and T6 in ICF and SPK are caused by the speaking intonation effect rather than the canonical tone shape.

For rising tones, the ascending pitch slopes in ICF have the steepest slopes among the three contexts of ICF, SPK and SNG. Especially, for T2 the pitch contour in ICF rises much more dramatically than in the continuous speech styles of SPK and SNG. This may be caused by the fact that syllables in ICF are much longer and more fully pronounced, thus the pitch contours of rising tones can climb to the peak of the presupposed target. In continuous speech, the syllables are much shorter than in ICF, and the pitch of the rising tones may have gone only part-way towards the supposed peak before having to stop and turn to carry the tone of the next syllable, thus making the pitch slope of rising tones more reduced in SPK and SNG. As the ending peak of T2 is supposed to be the highest in the tone space, it may need more time to fully reach the target. Thus, compared with T5, the pitch slope difference of T2 is even greater between ICF and the continuous speech styles.

In sum, the results of Experiment II systematically show the intonation differences between speaking and singing, which are clearly displayed in the matrix listed in Table 5-2. Speaking shows a consistent declination effect by way of pitch slope descent and pitch register displacement downwards in all domains. Singing breaks the declination requirement of speaking: the syllable slope tends to be level or even a little ascending, and the utterance-body slope is ascending. Singing is a more conspicuous speech style compared with speaking – and thus also has a higher pitch register at the syllable level and in the utterance-body portion. In the utterance-final portion, singing does not show any extra intonation effect: final syllables are no different from non-final syllables. In Experiment II, we also compared ICF with the continuous speech styles of SPK and SNG. The results told us that ICF is more similar to SPK and also shows the declination effect of speaking. Another interesting phenomenon of ICF is that the falling and rising tones can be more fully realized in ICF than in the continuous speech styles.

5.1.3 *Discussion of Experiment III*

Experiment III focused on the internal differences of normal speaking. Four sentence types in normal speaking were studied, including declarative without sentence-final particles (**DN**), declarative-derived question without sentence-final particles (**QN**), declarative with a sentence-final particle (**DP**), declarative-derived question with a sentence-final particle (**QP**). Through the comparison of these four sentence types, we hoped to find out: (1) the different ways to convey certain intonational meanings in sentences with sentence-final particles (**SFP**) and without SFP; (2) the different intonation patterns of declaratives and questions.

In the preliminary observation of the mean pitch contours of the whole utterances in Figure 4-1 (Page 131), we had already noticed that for all syllables except those located in the last two syllable intervals in the figure, the syllable pitch contours across the utterance body have great overlap for the three sentence types of DP, QP, and QN; while DN sentences are slightly lower than the above three types. A common characteristic of all the sentence types is the pitch slope declination effect at the syllable level and in the utterance-body portion, which is also the representative characteristic of the speaking speech style.

The main difference among the four sentence types appears in the utterance-final portion and the SFP, while there are only some concomitant minor differences at the syllable level and in the utterance-body portion.

For DN, the utterance-final intonation effect is obvious. The final syllables have an obviously lower pitch register and a slightly more declining pitch slope compared with non-final syllables, as displayed in Figure 4-4 (Page 135) and Figure 4-5 (Page 136). In terms of duration, the final syllables are only a little longer than non-final syllables. It can be inferred that the utterance-final intonation effect of DN is superimposed on the whole final syllable. It is a kind of "simultaneous addition" in the terminology of Chao (1933/2006).

For QN, the most outstanding characteristic is the sharp rise in the utterance-final position. Figure 4-6 (Page 137) and Figure 4-7 (Page 137) show that the intonational sharp rise actually starts in the later part of the final syllable. The first half of QN utterance-final contours still preserves the pitch register feature of the beginning point of the original tone shape, but all tones converge to reach the same extremely high pitch at the end. In terms of duration, the final syllables are much longer than the non-final syllables. Thus, the superimposing method of the QN-tail is of a different nature compared with the DN-tail: it is “successive addition” rather than “simultaneous addition” in the terminology of Chao (1933/2006). Figure 4-8 further suggests that the intonational sharp rise of the QN-tail has a much higher ending point than the tonal high-rising of T2 in different contexts, which prevents them from becoming confused with each other.

Figure 4-3 shows the penultimate syllable interval of Figure 4-1 (Page 131), which is the location of the final syllable of the trunk (and also the final syllable of DN and QN). The syllable contours of DP and QP do not show much difference between them and they are both located between DN and QN (as displayed in Figure 4-3, Page 136), which implies that adding an SFP after the trunk can discharge the intonation burden on the trunk-final syllable.

For the SFPs of DP and QP, from Figure 4-9 (Page 140) we can observe that the QP-SFP has a slightly lower beginning point and a much more steeply declining pitch slope than the DP-SFP. In terms of assimilation effect, QP-SFP is influenced by progressive assimilation at the beginning part while DP-SFP is not affected.

Further detailed statistical analysis was conducted in the framework of a matrix composed of pitch configuration and pitch domain. For the four sentence types, we need to compare them pair by pair. The number of pairs is: $C(4, 2) = 4! / ((4-2)!*2!) = 6$. Thus, there are six tables summarizing our comparison results on the sentence type pairs: DN and QN (Table 5-3), DN and DP (Table 5-4), DN and QP (Table 5-5), QN and DP (Table 5-6), QN and QP (Table 5-7), DP and QP (Table 5-8).

The symbols and paradigm here are the same as Table 5-2 in Section5.1.2. In the “Utterance-final” part, the comparison of final syllables and non-final syllables is not applicable to the sentence types of DP and QP, and these cases are marked as “NA” in Table 5-4 to Table 5-8.

Table 5-3 Matrix of pitch configuration and domain for DN and QN

Objects in comparison	Shape / Domain		Slope	Register	Span
DN:QN	Syllable		-	-*	0
	Utterance-body		-*	-	0
Final syllable: other syllables	Utterance-final	DN	-	-*	+*
		QN	+*	0	+*

Table 5-4 Matrix of pitch configuration and domain for DN and DP

Objects in comparison	Shape / Domain		Slope	Register	Span
DN:DP	Syllable		0	-*	0
	Utterance-body		-	-	0
Final syllable: other syllables	Utterance-final	DN	-	-*	+*
		DP	NA		

Table 5-5 Matrix of pitch configuration and domain for DN and QP

Objects in comparison	Shape / Domain		Slope	Register	Span
DN:QP	Syllable		0	-*	0
	Utterance-body		-	-	0
Final syllable: other syllables	Utterance-final	DN	-	-*	+*
		QP	NA		

Table 5-6 Matrix of pitch configuration and domain for QN and DP

Objects in comparison	Shape / Domain		Slope	Register	Span
QN:DP	Syllable		0	+*	+
	Utterance-body		0	+	0
Final syllable: other syllables	Utterance-final	QN	+*	0	+*
		DP	NA		

Table 5-7 Matrix of pitch configuration and domain for QN and QP

Objects in comparison	Shape / Domain		Slope	Register	Span
QN:QP	Syllable		0	+	0
	Utterance-body		+	0	0
Final syllable: other syllables	Utterance-final	QN	+*	0	+*
		QP	NA		

Table 5-8 Matrix for revealing the intonation effects of DP and QP

Objects in comparison	Shape / Domain		Slope	Register	Span
QP:DP	Syllable		0	+	0
	Utterance-body		0	+	0
Final syllable: other syllables	Utterance-final	QP	NA		
		DP	NA		

At the syllable level and within the utterance-body portion, as shown in the matrices listed in Table 5-3 to Table 5-8, there are only minor differences among the four sentence types. In terms of pitch slope, only the utterance-body slope of DN is significantly smaller than QN. For the slope comparison of other pairs, there are just small differences or no differences (marked as "0"). The pitch span differences are even smaller. Among the six pairs compared, all show "0" except for the pair of QN-DP: at the syllable level, QN has a slightly wider span than

DP. The pitch register differences are a little more obvious compared with pitch slope and pitch span. Generally speaking, as displayed in Figure 4-2 (Page 132), declaratives have a lower pitch register than questions, and this difference is more prominent if the utterance is without an SFP, i.e., DN has the lowest pitch register while QN has the highest pitch register.

In the utterance-final portion, DN and QN show very distinct utterance-final intonation patterns. The final syllables of DN have a more declining pitch slope and a significantly lower pitch register than non-final syllables (as shown in Figure 4-5, Page 136). The utterance-final intonation of QN involves a sharp rising pitch contour regardless of the original tone shapes while keeping the pitch register at a level similar to the non-final syllables (as shown in Figure 4-7, Page 137). As the steepness of the pitch slope is much more tilted for final syllables in both DN and QN, both sentence types get "+*" for "Span". For DP and QP, they express the intonational meanings by use of SFPs, and so the utterance-final intonation is not applicable to them. We will have further discussions on the relationship between utterance-final intonation and SFP in Section 5.3.3.

5.2 Overall discussion of the results of the three experiments

5.2.1 *The interrelationship of the three experiments*

Although the experimental studies in the present thesis were divided into three experiments, these experiments were systematically organized and closely interrelated.

The research scope of the three experiments conducted in the present study progressed from broad to narrow: in Experiment I, we compared three pairs of speech styles – opera speaking versus opera singing, opera speaking versus normal speaking, and normal speaking versus normal singing; in Experiment II, we only focused on one pair of speech styles – normal speaking and normal singing; in Experiment III, we further narrowed our research scope to only one speech style – normal speaking, and concerned ourselves with the internal differences within this speech style. With this organization and order, we can systematically unveil the effects of intonation on Cantonese lexical tones in speaking and singing.

Among the three experiments, Experiment I and Experiment II shared the same research question: what is the key factor differentiating speaking and singing? Experiment II and Experiment III shared the same research framework of the matrix consisting of the two dimensions of pitch configuration and pitch domain.

In fact, Experiment I already contained an embryonic form of the matrix we formally proposed in Experiment II and Experiment III. As Experiment I was a pilot study, we did not clearly spell out the framework at the very beginning. From the observation and analysis on the data of different speech styles, we noticed that different speech styles may affect the pitch contours of lexical tones by variations of pitch register (inferred from the distribution of beginning and ending pitch points) and pitch slope; and we also conducted analysis at both the levels of syllable and utterance.

Experiment III did not digress from the central research question we discussed in Experiment I and Experiment II. In Experiment III, although we only concentrated on the intonation effects of four sentence types in speaking, this investigation also served to examine the research problem of the differences between speaking and singing: it told us what the internal differences within speaking can be, and if we compare them with the external differences between speaking and singing, we can be clearer about the coverage of speaking and the boundary between speaking and singing.

Therefore, the common research question and research framework link the three experiments together and make them closely interrelated.

5.2.2 *Discussion of the dimension of pitch configuration*

In the dimension of pitch configuration, there is a close interrelationship among pitch register, pitch span, and pitch slope. Pitch register and pitch span are interrelated by their inherent nature. They are both factors of pitch space: pitch register locates the relative height while pitch span outlines the magnitude. Pitch slope is a factor of pitch configuration and is different from pitch span in terms of nature, but they are interrelated by their calculation methods. As defined in Section 3.1.4, pitch slope is the slope of linear regression of the pitch points concerned, while pitch span is the distance between extreme values. Regardless of the direction of pitch slope, the steepness of pitch slope positively correlates with pitch span: steeper slope means wider span, and vice versa.

The three factors of pitch configuration make different contributions to differentiation of the intonation patterns of speaking and singing. As the properties of pitch span and the steepness of pitch slope are correlated, our discussion only focuses on pitch register and pitch slope.

There is evidence showing that pitch register is not an indispensable factor to distinguish speaking and singing. In Experiment I, through the comparison between OSpk and NSpk, we found that for the same speaker YSF, the pitch register of opera speaking is significantly higher than that of normal speaking. Zhu (2012, pp. 3-4) also found that in many provincial operas in China (including Cantonese opera), "high fundamental frequency far above one's normal pitch range can be observed". However, people regard opera speaking as a special kind of speaking rather than a kind of singing. This means that higher pitch register is not solely used in singing, but also in certain types of speech. An important function of higher pitch register is to emphasize that a speech style is conspicuous or more emotional. Compared with normal speaking, singing is a special kind of speech style that can express more emotional feelings. For this reason its pitch register is higher than that of speaking. Therefore, higher register is a concomitant characteristic of singing, but not the defining factor for making a distinction between speaking and singing.

It is assumed that pitch slope is the distinctive factor that differentiates speaking and singing. This assumption can be verified by the perception test conducted in Experiment I. In this test, we could successfully switch between the two speech styles (i.e. OSpk and OSng) by merely adjusting the pitch slope, which supported our assumption.

Along with the cross-speech-style comparison in Chapter 2 and Chapter 3, the internal studies of the normal speaking speech style in Chapter 4 further support the pitch slope declination characteristics of speaking. There are some variations among different sentence types, but the common property of all the speaking utterances is the consistent declination of pitch slope at the syllable level and in the utterance-body portion. The non-declination of pitch slope in singing does not figure in speaking sentence types.

Therefore, pitch slope is a more important factor to differentiate speaking from singing; pitch register is a concomitant characteristic conveying the importance and emotion of a speech style.

5.2.3 *Discussion of the dimension of pitch domain*

The framework proposed in this study is a two-dimensional matrix. Apart from the dimension of pitch configuration, the dimension of pitch domain is also very important. For a certain kind of pitch configuration, if it is acting on different pitch domains, there will also be variation to the matrix, which can be used to indicate different kinds of intonation effects. A typical example to illustrate the importance of pitch domain is the comparison of pitch slope between the singing speech style and the QN sentence type of speaking speech style.

From the studies in Chapter 2 and Chapter 3, we noticed that there is a kind of positive intonation force of pitch slope in singing speech styles acting on the domains of syllable and utterance-body portion. At the syllable level, this positive force enhances the level/falling tones to have a level or even slightly rising pitch contour, and enhances the rising tones to have more ascending pitch contours compared with their counterparts in speaking. In the utterance-body portion, if we line up the same-tone syllables in singing utterances, it can be observed that the contour of the utterance body also shows a slightly upward trend. In singing speech styles, the positive intonation force of pitch slope does not exert its effect on the utterance-final portion – the utterance-final syllable has pitch slope data similar to those of non-final syllables.

For the QN sentence type in the normal speaking speech style, there is also a kind of positive intonation force on pitch slope. In terms of domain, this force acts at the utterance-final portion. Although compared with the DN sentence type, the pitch slope is relatively greater for non-final syllables and the utterance body of QN, the absolute force of the intonational pitch slope in QN is negative, i.e., there is still a declination effect for the pitch slope at both the domains of syllable level and utterance-body portion.

Therefore, although there is a positive intonation effect on pitch slope for the singing speech styles and the QN sentence type, the different domains of this positive force make these two kinds of intonation different. For the singing speech styles, the positive intonational force of pitch slope is applicable at the syllable level and utterance-body portion. For the QN sentence type, the positive intonational force of pitch slope exerts its effect at the utterance-tail part. The dimension of pitch domain is indispensible for indicating a certain kind of intonation.

5.3 Further remarks

5.3.1 *Further remarks on declination versus non-declination*

Only a small amount of previous literature has referred to the non-declination phenomenon in singing speech. Fujisaki (1983, p. 48) put forth that singing is expected to sustain the mean fundamental frequency at a constant value over the time interval of a note, and there is no f0 declination within a note. The singing speech discussed in Fujisaki (1983, p. 48) did not require tone-preservation, and his research focus was not the relationship between tone and intonation. Chao (1956/2006, pp. 599-601) mentioned a style of speech called "singsong". Singsong is neither ordinary speech nor singing to a musical melody, but something intermediate, which is based largely on the phonemic tones of the words, spoken in a stereotyped manner. In singsong, each tone receives its usual full value practically without intonational variation. The popular impression of singsong is that it is "intoned". Now we can say that it is speech minus the element of intonation. The singsong style of speech is by no means limited to tonal languages.

Fujisaki (1983, p. 48) and Chao (1956/2006, pp. 599-601) had an acute sense on the special speech styles of singing and "singsong", and they noticed that the universal declination intonation is removed in these speech styles. Our present experimental study supports their observations and further demonstrates by what methods and at what levels this non-declination tendency exerts its effect. In fact, our matrix in Table 5-2 shows that "non-declination" in singing is a conservative description; Cantonese singing even shows a slight updrift pattern. We are especially interested in the "singsong" speech style described in Chao (1956/2006, pp. 599-601) because it is not only limited to tonal languages. Further investigation should be carried out on this speech style in more languages to see whether it is non-intonation (non-declination) or with slight ascending intonation, like Cantonese singing.

Our comparison between declination in speaking and non-declination in singing can also provide hints to the controversial problem of the physiological causes of declination. It has been debated whether declination is the result of automatically reducing subglottal pressure, or it is purposefully controlled by laryngeal muscles. Our present study agrees with the argument of Ohala (1978, p. 32) that declination has a linguistic purpose and is caused by active laryngeal changes in vocal cord tension. Apart from signaling clause and sentence boundaries, declination also serves as a kind of linguistic code to distinguish speaking from singing. Declination in speaking is a default intonation while non-declination in singing is a kind of marked intonation. If we ascribe the cause of declination totally to the automatic reduction of subglottal pressure, declination should be inevitable, even in singing, which is not true for Cantonese singing. The non-declination, or the slightly ascending intonation of Cantonese singing here can act as a reference object for the declination of daily speaking, thus strengthening the statement that declination is linguistically purposeful.

The introduction to singing in tonal languages (in Section 1.3.1) told us that in some tonal languages, lexical tones are preserved in singing, and the singing melody has a downward trend. This downward trend in singing is possibly due to the purpose of mimicking the prosodic characteristics of the most natural speech style – speaking. Ekwueme (1974b, p. 347) also made the following point:

> "It is only natural for any good artist to strive to imitate nature, and in Igbo aesthetic judgment the success of the artist is measured by the proximity in the semblance of his creation to that of nature."

As a result of the similarities between speaking and singing, they are also easily interchangeable:

> "because of the close relationship between melodic lines and speech tones, it is quite easy to break from singing into spoken words, or vice versa. Some songs have certain syllables, words, or even whole phrases spoken instead of sung. (Ekwueme, 1974b, p. 343)"

Among the literature which mentioned the downtrend melodic characteristics of singing, the downtrend characteristics were based on the authors' impressions of which they only provided a rough description. None of them carried out acoustic studies and displayed the pitch contours on paper. As the lexical tones are preserved in both speaking and singing in those languages, while native informants can clearly tell apart the two speech styles, it is reasonable to infer that there are still some structured intonational differences between the two speech styles. According to the descriptions in previous studies, a typical characteristic of the downward-melodic-singing is that for the same-tone syllables, latter syllables have lower pitch than former syllables. For instance, Starke (1930, p. 46) described the tone realization patterns in Xȍsa:

> "in both song and speech, a tendency for the intonation level to be lowered, so that a final high tone (for example) is always lower than an initial high tone."

It can be inferred that the utterance-body intonation in the singing of these languages is also declining. Probably the structured differences between singing and speaking in these languages are reflected at the syllable level – there may also be a kind of non-declination effect on the syllables in singing, similar to what we found in Cantonese. Further empirical studies should be conducted to test this assumption.

5.3.2 Further remarks on intonation at the syllable level

For the two pitch domains of intonation – the syllable level and the utterance level, the utterance level may be more noticeable and has been more extensively studied. Intonation at the syllable level can be easily neglected, as the bearing unit of it overlaps completely with that of lexical tones. Our discussion below will be more focused on the declination intonation at the syllable level, which has been overlooked in previous literature. As pitch contours are the algebraic sum of tone and intonation (Chao, 1933/2006, 1956/2006), even syllables in isolated citation form are still not "pure tones" because they are also affected by the speaking intonation.

In most people's instinctive imagination, "level tones" should have a "level" pitch contour output. They may also consider the pitch status of syllables in isolated citation form as the one that most closely approximates canonical forms of tones. In fact, asking native speakers to pronounce the syllables in isolated citation form is the most common method to study the tone system of a certain language or a dialect. If we regard the acoustic pitch output of them as the canonical form of the tones, we actually ignore the declination intonation at the syllable level.

In most schematic diagrams of Cantonese lexical tones, such as Fox, Luke, and Nancarrow (2008), Li, Lee, and Qian (2002), Lee et al. (2002), and Peng and Wang (2005), “level tones” have a constant horizontal pitch contour. Among the three graphs in Figure 3-11 (Page 103), SNG is the most similar to the schematic diagrams of previous studies, i.e. singing tones are actually the closest approximation to the canonical form of tones in real life. This may not accord with most people’s imaginations, but actually happens in tone acquisition for second-language learners. Ho and Ho (2007) observed that foreign learners very often follow a static framework prescribed in textbooks or conceived of by themselves in terms of musical scales too rigidly when learning tone languages in general and Cantonese in particular, which always make their speech too melodious (Ho & Ho, 2007, p. 65). They further pointed out that a musical scale is static and a lexical tone scale is dynamic (Ho & Ho, 2007, p. 66). From our present study, we can provide a clearer explanation: the transcription of lexical tones by the five-point-scale method or directly by musical notation can only represent the canonical form of lexical tones; but in real life speaking, the declination intonation is superimposed on the pitch configuration of every syllable. If we rigidly stick to the musical scales for the pronunciation of lexical tones, we will produce a non-declining version of them, which more closely approximates the realization of lexical tones in singing.

If we look at the tone-system graphs based on acoustic data of some previous experimental studies (Francis et al., 2008; Khouw & Ciocca, 2007; Li, Lee, & Qian, 2002; Peng & Wang, 2005; Rose, 2000), we can observe that T3 and T6 also have slight falling contours. However, this characteristic did not catch the attention of the researchers because they had other research focuses. In our present study, this characteristic is very important. The syllables in isolated citation form can be regarded as the shortest utterances in speaking and they also carry the speaking declination intonation.

Acoustically T3 and T6 fall slightly but perceptually they are heard to be level, even to many professional linguists’ ears. Very little traditional linguistic literature has mentioned that T3 and T6 are level tones with a falling tendency with the exception of Hashimoto (1972). The contradiction between the acoustic form (slight falling) and listeners’ perception (level) of pitch contours is probably caused by the fact that listeners automatically filter out the default declination intonation and just recognize the canonical forms of tones. This mechanism is similar to that found in previous perception studies on declination intonation (Gussenhoven & Rietveld, 1988; Pierrehumbert, 1979).

There was a famous metaphor that compares the relationship between tones and intonation to ripples riding on the top of waves (Chao, 1968, p. 39). There is no doubt that intonation at the utterance level can be compared to "waves". From our present study, "ripples" should not only be likened to lexical tones, but also to intonation at the syllable level.

5.3.3 *Further remarks on utterance-final intonation and SFP*

In previous experimental studies on utterance-final intonation of Cantonese, e.g. Fox, Luke, and Nancarrow (2008); Ma (2007); Ma, Ciocca, and Whitehill (2004, 2006a, 2006b); Mai (2000); Vance (1976); Xu and Mok (2011), investigators usually tended to use content words as the final syllable and avoided using SFPs in their designed utterances. No previous work has been done to systematically compare utterances with and without an SFP. The systematic acoustic experiment conducted in Chapter 4 has filled this gap – from this experiment we can get a more profound understanding of the relationship between utterance-final intonation and SFPs.

For intonation patterns at the syllable level and in the utterance-body portion, the stimuli with an SFP and those without do not show prominently distinct differences. They all obey the constraint of declination intonation for the speaking speech style. At the syllable level and in the utterance-body portion, among the four sentence types of DN, QN, DP and QP, there are minor differences in pitch register and even smaller differences in pitch slope: the register-lowering and slope-declining effect are strongest for DN and weakest for QN, while medial for DP and QP. To some extent, the presence of an SFP also neutralizes the polar differences of declaratives and questions.

Among the four sentence types, the most remarkable distinction happens at the end of the utterances. For utterances without an SFP, the final syllable of DN has a significantly lower pitch register and a slightly more declining pitch contour compared with non-final syllables. The final syllable of QN keeps the pitch register unchanged but has a sharply rising pitch slope, regardless of the pitch contour shape of the original lexical tone.

There was another research gap in previous studies on utterance-final intonation in Cantonese: researchers usually just provided acoustic analysis on the final syllable of the utterance, but seldom clarify whether the utterance-final intonation is acting on the final syllable or the final content word. In Experiment III, the pitch contour of every syllable within an utterance was displayed, which also helps us to observe the boundary of the utterance-body portion and utterance-final portion. It can be observed quite clearly that only the final syllable carries the utterance-final intonation. As the final content word in our designed stimuli is a disyllabic word, it is clear that the domain of the utterance-final intonation is at the final syllable rather than the final content word.

For utterances with the presence of an SFP, the last syllable of the content words in the utterance does not show any extraordinary characteristic – it is just the same as the other non-final syllables. Thus, it can be concluded that if the final syllable of an utterance belongs to a content word, it is affected by the utterance-final intonation; if the final syllable of an utterance is an SFP, it carries the function of the utterance-final intonation but has a different nature: it is composed of an independent segment or a sequence of segments, not just a floating tone at the final syllable (i.e. boundary tone).

According to traditional transcription, the lexical tones T3 and T4 are assigned to the SFPs of DP and QP, respectively. This is because DP-SFP and QP-SFP are lexicalized and given lexical tones that closely approximate their pitch contours. Nevertheless, the essence of the pitch contour of an SFP is intonational rather than tonal – as stated by Cheung (1986, p. 256), “the intrinsic pitch configuration of S[F]P is not the same thing as tone: it is intonational.” This argument can be supported by the fact that the pitch contour of an SFP is not the result of the algebraic sum of the utterance-final intonation and the “tone”. A typical example is the question sentence. It is supposed that the final syllable should have sharp-rising pitch shape, regardless of the original tone shape. However, the QP-SFP has a low-falling rather than a sharp-rising pitch contour. It can be inferred that the intrinsic nature of the pitch contour of QP-SFP is already intonational, so that it can directly carry the function of the utterance-final intonation, and does not need to go through the intonational change of becoming sharp-rising.

Cheung (1986, p. 256) also pointed out that "S[F]P may be segmentless." This can help us to bridge SFP and utterance-final intonation. It can be assumed that the default form of an utterance is with an SFP to indicate a certain sentence type, and that this SFP may be either segmental or segmentless. The segmental form of SFP is easy to understand, i.e., the SFP is an independent segment, or sometimes it consists of more segments. If there is not an independent segment for expressing the SFP, it can be segmentless but having a floating contour (or "tonal pragmatic morpheme" in the terminology of Wong, Chan, and Beckman (2005)) acting on the final syllable of the utterance, i.e., the utterance-final intonation.

For declarative sentences, this floating contour is realized in the superimposing manner of "simultaneous addition", which involves the pitch register lowering and pitch slope declining over the whole utterance-final syllable. For question sentences, this floating contour is realized in the superimposing manner of "successive addition", which keeps the original tone register for the first half of the utterance-final syllable, but adds an extra-high-rising pitch contour to the second half. The different manners of superimposing the DN-final floating contour and the QN-final floating contour can be reflected by the differing lengths of the final syllables in DN and QN. The DN-tail is only a little longer than non-final syllables in most cases, and the superimposing manner of "simultaneous addition" does not require extra syllable length. The situation of the QN-tail is quite different. The QN-tail is usually much longer, and if the informants want to emphasize the original tone of the final syllable, they can even pronounce the whole original tone before adding an extra length of sharp rising pitch contour at the end. Thus, there is more time for the QN-tail to realize the "successive addition".

Although the superimposing methods in DN and QN are different, both the utterance-final intonations of DN and QN can be regarded as a kind of segmentless SFP, which is a floating contour acting on the final syllable. In this way, the concepts of utterance-final intonation and SFP can be connected.

5.4 Conclusions

Three experiments were conducted in the present study to investigate the effects of intonation on Cantonese lexical tones in speaking and singing.

In Experiment I, speaking and singing in opera speech and normal speech were compared, i.e., opera speaking versus opera singing, opera speaking versus normal speaking, and normal speaking versus normal singing. The experimental results showed that, at the syllable level, the consistent difference between speaking and singing is slope difference, for both opera speech and normal speech. The pitch slope of syllables is more declining in speaking speech styles. At the utterance level, speaking speech styles show a declining trend while singing speech styles show a non-declining tendency. For normal speaking versus normal singing, there are also pitch register differences. Normal singing is higher than normal speaking. However, the register differences are not unique to speaking and singing – opera speaking also has higher pitch register than normal speaking. The function of higher pitch register is to emphasize that the speech style is more conspicuous or more emotional.

In Experiment II, we only focused on the comparison between normal speaking and normal singing, but refined our questionnaire and recruited more native informants. Experiment II was carried out in the framework of a two-dimensional matrix consisting of pitch configuration and pitch domain. The results show that the intonation patterns at the syllable level and in the utterance-body portion are quite consistent within speaking and within singing. Speaking has a more declining pitch slope, a lower pitch register, and a wider pitch span; while singing has a more ascending pitch slope, a higher pitch register, and a narrower pitch span. If intonation at the syllable level is compared to "ripples" (traditionally "ripples" refer to lexical tones, but here we only discuss the intonation effects), and intonation in the utterance-body portion is compared to "waves", we can conclude that "ripples" and "waves" have the same direction. The "ripples" and "waves" constitute a downdrift pattern in speaking and an updrift pattern in singing. In the utterance-final portion, speaking shows a strong final declination effect, while singing does not exhibit an obvious final intonation effect, i.e. final syllables are not significantly different from non-final syllables in singing.

In Experiment III, the scope was confined to the internal differences of normal speaking, and we compared the intonation patterns of four sentence types. Two cross-cutting dichotomies were dealt with here: declarative versus declarative-derived question, and the absence versus presence of an SFP. The overall pitch contours show that the four sentence types firstly belong to the speaking speech style – the declination effect is obvious at the syllable level and in the utterance-body portion. In these two pitch domains, the four sentence types only show minor differences in pitch register and even smaller differences in pitch slope. The most remarkable distinction among the four sentence types happens at the end of the utterances. The final syllable of DN (declarative without an SFP) is affected by the final declination, which has a significantly lower pitch register and a slightly more declining pitch contour compared with non-final syllables. The final syllable of QN (question without an SFP) shows the outstanding characteristic of a sharply rising pitch shape, which even overrides the force of the original lexical tone, i.e., the tonal distinction may be neutralized. For utterances with an SFP, the SFP itself carries the function of indicating whether an utterance is a declarative or a question. Therefore, if the final syllable of an utterance belongs to a content word, it is affected by the utterance-final intonation; if the final syllable of an utterance is an SFP, there is no hint of the utterance-final intonation in the final content word. It can be assumed that all the utterances have an SFP, either in the segmental or segmentless form. The segmental form is the traditionally observed form, which is tied to at least one independent segment. The utterance-final intonation can be regarded as a kind of segmentless SFP, which is a floating contour superimposed on the final syllable of the utterance. In this way, the SFP and the utterance-final intonation can be bridged.

In sum, through the systematic investigation of the intonation patterns of speaking and singing in Cantonese, we discovered that Cantonese singing is a kind of speech that has non-declination intonation (or even a slightly ascending intonation), thus serving as a reference object for the universal declination intonation in speaking. Our present study enriches the knowledge of intonation typology and provides further insight into the relationship between lexical tones and intonation. The matrix consisting of pitch configuration and domain proposed here also establishes a new model for experimental studies on intonation in tonal languages. The intonation effects at the syllable level, which have been neglected by previous studies on intonation, are especially emphasized in our model. As well, some byproducts of our studies also fill in research gaps in previous studies: e.g. the relationship between SFPs and utterance-final intonation, the declination effect on syllables in isolated citation form, etc.

5.5 Further investigations

Further research in several areas can be conducted to extend the findings of the present study, e.g., the linguistic investigation on singing in tonal languages, further development of the framework involving pitch domain and pitch configuration, the superimposing ways of utterance-final intonation, etc.

For linguistic investigation of singing in tonal languages, there are many interesting research questions worth further study.

Firstly, in our present study, we have compared the intonation effects of speaking and singing speech styles in Cantonese. The same research methodology can be applied to other tonal languages which preserve lexical tones in singing. As introduced in Section 1.3.1, some other tonal languages also have lexical tones fully or partly preserved in certain kinds of singing speech styles, but no acoustic analysis has been reported for the different realization patterns of the lexical tones in speaking and singing. By applying our research methodology to these languages and conducting similar studies, we can also test whether the patterns summarized in our present study are linguistically universal. As mentioned in Section 5.3.1, some languages were reported to preserve tones in singing but their singing melody has a downward trend. The singing intonation of those languages is presumably declining through the utterance-body portion, which is different from what we found for Cantonese. However, we suspect that the singing intonation effect is horizontal or even slightly ascending at the syllable level, which would differentiate singing speech from speaking speech in those languages. It is suspected that the non-declining intonation at the syllable level is a universal phenomenon of singing in tonal languages. To test this assumption, we need empirical studies on more languages.

Secondly, few previous studies have included acoustic analysis on Cantonese opera performance (including opera speaking and opera singing). In Experiment I of the present thesis, we conducted preliminary studies and found some interesting results. However, the data we used were somewhat limited. More Cantonese opera data should be collected for further studies, which will, in all likelihood, provide us with more interesting patterns to explore. In future studies, we should not only employ recordings of opera actors, but also opera actresses. More types of opera speech styles and song categories should also be covered. If possible, we can also invite professional opera performers to participate in our controlled experiments, which would help us more directly compare different speech styles, including opera speech versus normal speech, singing versus speaking.

Thirdly, we have only discussed pitch patterns in our present study, while the rhythmic characteristics of singing and speaking are left for future studies. We have noted that duration is also an important factor for the pitch-realization patterns of lexical tones. We observed that if a syllable receives enough length to manipulate its pitch movement, it can be more fully pronounced and reach its intended target. More investigations should be conducted to discuss the relationship between the duration pattern and the pitch patterns.

In our present study, we propose a framework for studying intonation effects on tones: a two-dimensional matrix composed of pitch domain and pitch configuration. This framework is a basic model and can be further developed to a more sophisticated structure.

In the dimension of pitch domain, there are two levels in the present framework – the syllable level and the utterance level. In fact, we could also add more levels between them – the word level and the phrase level. In our present study, we only investigated the basic levels of syllable and utterance, being the most essential elements. In future studies, to reveal more details of the intonation patterns, the pitch domains of word and phrase should also be involved.

In the dimension of pitch configuration, our basic model adopted a single-line structure to approximate the utterance-body intonation. This basic model can also be developed to a more complex double-line structure, which is composed of a top line and a base line. The double-line structure was proposed in previous studies for better representing the declination intonation (Cruttenden, 1986; Sorensen & Cooper, 1980). According to the description by Cruttenden (1986, p. 167):

> "The declination which occurs in such neutral sentences is represented by a slightly declining baseline and a more steeply declining top line, thus producing a narrowing of pitch range as the intonation-group progresses."

Sorensen and Cooper (1980, p. 407) pointed out that:

> "Unlike the baseline concept which, with a set of attributes, should be able to predict the location of all F_0 values, the topline and bottomline only predict where the peaks and valleys of F_0 should occur."

In our present study, as our main focus is on declination versus non-declination, the single-line structure could more directly reveal this difference. If future research on utterance-body intonation requires a more detailed description of the intonation patterns, the double-line structure could be adopted.

From the comparison between DN-tail and QN-tail in Experiment III, we noticed that the manners of superimposing the floating intonational contour on the final syllables can be different: the DN-tail is "simultaneous addition" while the QN-tail is "successive addition". In our discussions in Section 5.3.3, we also mentioned that "successive addition" may require extra syllable length. Thus, if the informants were not to emphasize the original tones of the QN-final syllable and to pronounce it quickly, we are curious whether the manner of superimposing would become "simultaneous addition". It is worthwhile to conduct further research on the relationship between the syllable length and the manners of superimposing intonation effects.

The above are several suggestions for further investigations. In fact, there are still more research topics arising from the present study. The present study is exploratory. Further investigations based on our current methodology and findings are expected to lead to exciting research outputs.

Appendices

Appendix A
Resources of materials used in Experiment I

1. Mun Cheen-shuih (文千歲)'s four CDs of Cantonese opera performance

CD1. 牡丹亭驚夢•白蛇傳 [Mudanting jingmeng, Baishe zhuan]. 李寶瑩 (Lee Po-ying) & 文千歲(Mun Cheen-shuih), Fung Hang Record Ltd, 1987.

CD2. 山伯臨終•朱弁回朝 [Shanbo linzhong, Zhubian huichao]. 文千歲 (Mun Cheen-shuih), Fung Hang Record Ltd, 1991.

CD3. 西廂記 [Xixiangji]. 文千歲(Mun Cheen-shuih), Fung Hang Record Ltd, 1997.

CD4. 蝴蝶杯 [Hudiebei]. 文千歲(Mun Cheen-shuih), Fung Hang Record Ltd, 1999.

2. Yuen Siu-fai (阮兆輝)'s data resource

[Website resource] http://resources.edb.gov.hk/~chiopera/sitemap.htm, visited on 12 May, 2010.

3. Cantonese popular songs adopted in the questionnaire for the informant ZSX

Children's song: (1)月光光 (Moonlight) ; (2)齊齊望過去 ("Oh Susannah").

Pop song: (1)男兒當自強 (A man of determination); (2)紅日 (The sun); (3)真的愛你 (Really loving you); (4)花火 (Fireworks).

Classical opera song: 帝女花•香夭 (Princess Chang Ping)

Appendix B

Transcription of recordings by Mun Cheen-shuih

1. Transcription of Mun Cheen-shuih's speaking utterances

Code	Type of melody	Chinese text and IPA transcription	Resource
BH1	白(Bai)	再過去果便行下至得[tsɔːi^{33} kwɔː^{33} høy^{33} kwɔː^{25} piːn^{22} haːŋ^{21} haː^{23} tsiː^{33} tɐk^{5}]	蝴蝶杯 (Hudiebei)
BH2	唸白(Nianbai)	緣定三生誠所望 [jyːn^{21} tɪŋ^{22} saːm^{53} sɐŋ^{53} sɪŋ^{21} sɔː^{25} mɔːŋ^{22}]	
BX1	白(Bai)	淚人，淚人，你唔慌唔累人啦 [løy^{22} jɐn^{21}, løy^{22} jɐn^{21}, nei^{23} m^{21} fɔːŋ^{53} m^{21} løy^{22} jɐn^{21} laː]	西廂記 (Xixiangji)
BX2	白(Bai)	紅娘姐啊，紅娘姐，我、我而家成個都生矖啦我 [hʊŋ^{21} nœːŋ^{21} tsɛː^{25} aː, hʊŋ^{21} nœːŋ^{21} tsɛː^{25}, ŋɔː^{23} jiː^{21} kaː^{53} sɪŋ^{21} kɔː^{33} tou^{55} saːŋ^{53} saːi^{33} laː ŋɔː^{23}]	
BX3	白(Bai)	小姐有禮 [siːu^{25} tsɛː^{25} jɐu^{23} lɐi^{23}]	
BX4	白(Bai)	小姐啊，小姐你好心答應我啦 [siːu^{25} tsɛː^{25} aː, siːu^{25} tsɛː^{25} nei^{23} hou^{25} sɐm^{53} taːp^{3} jɪŋ^{33} ŋɔː^{23} laː]	
BX5	白(Bai)	系邊度有賊啊 [hɐi^{22} piːn^{53} tou^{22} jɐu^{23} ts^{h}aːk^{25} aː]	
BX6	白(Bai)	系我啊 [hɐi^{22} ŋɔː^{23} aː]	

BX7	白(Bai)	紅娘姐，我有詩為證噶。喇，我、我、我、我讀俾你聽啊。[hʊŋ^{21} nœːŋ^{21} tsɛː^{25}, ŋɔː^{23} jɐu^{23} siː^{55} wɐi^{21} tsɪŋ^{33} kaː. laː, ŋɔː^{23}, ŋɔː^{23}, ŋɔː^{23}, ŋɔː^{23} tʊk^{2} pei^{25} nei^{23} tʰɛːŋ^{53} aː]	
BX8	白(Bai)	詩中之意，的的確確是你家小姐約我來噶啵 [siː^{55} tsʊŋ^{55} tsiː^{53} jiː^{33}, tɪk^{5} tɪk^{5} kʰɔːk^{3} kʰɔːk^{3} siː^{22} nei^{23} kaː^{53} siːu^{25} tsɛː^{25} jœːk^{3} ŋɔː^{23} lɐi^{21} kaː pɔː]	
BS1	打引白 (Dayinbai)	人世無緣同到老，樓臺一別 [jɐn^{11} sɐi^{22} mou^{11} jyːn^{11} tʰʊŋ^{11} tou^{33} lou^{23}, lɐu^{11} tʰɔːi^{11} jɐt^{5} pit^{2}]	山伯臨終 (Shanbo Linzhong)
BS2	詩白(Shibai)	生不結夫妻，死當同墓穴 [sɐŋ^{53} pɐt^{5} kiːt^{3} fuː^{53} tsʰɐi^{53}, sei^{25} tɔːŋ^{53} tʰʊŋ^{11} mou^{22} jyːt^{2}]	
BS3	詩白(Shibai)	寄語後世人，珍重看蝴蝶 [kei^{33} jyː^{23} hɐu^{22} sɐi^{33} jɐn^{11}, tsɐn^{53} tsʊŋ^{22} hɔːn^{53} wuː^{11} tiːp^{2}]	
BR1	詩白(Shibai)	好一首少女懷春之詩呀 [hou^{25} jɐt^{5} sɐu^{25} siːu^{33} nøy^{23} waːi^{11} tsʰøn^{53} tsiː^{53} siː^{53} jaː]	人鬼恋 (Rengui Lian)
BR2	白(Bai)	我唔信 [ŋɔː^{23} m^{11} søn^{33}]	
BR3	白(Bai)	系咁就好囉 [hɐi^{22} kɐm^{25} tsɐu^{22}	

		hou^{25} lɔː]	
BR4	白(Bai)	只要你 [tsiː25 jiːu33 nei23]	
BR5	驚白(Jingbai)	你系邊個啊？到、到底你駕雨而來，抑、抑或乘風而入？點、點解我一啲都唔知嘅？[nei^{23} hɐi^{22} piːn^{55} kɔː^{33} aː , tou^{33}, tou^{33} tɐi^{25} nei^{23} kaː^{33} jyː^{23} jiː^{11} lɔːi^{11}, jɪk^{5}, jɪk^{5} waːk^{22} sɪŋ^{11} fʊŋ^{53} jiː^{11} jɐp^{2}, tiːm^{25}, tiːm^{25} kaːi^{25} ŋɔː^{23} jɐt^{5} tiː^{55} tou^{55} m^{11} tsiː^{53} kɛː^{25}]	
BR6	白(Bai)	愛女杜麗娘之墓。你、你就是杜麗娘啊？[ɔːi^{33} nøy^{25} tou^{22} lɐi^{22} nœːŋ^{11} tsiː^{53} mou^{22}, nei^{23}, nei^{23}, tsɐu^{22} siː^{22} tou^{22} lɐi^{22} nœːŋ^{11} aː]	
BR7	泣白(Qibai)	你既然能夠夢中生情，為我而死 [nei^{23} kei^{33} jiːn^{11} nɐŋ^{11} kɐu^{33} mʊŋ^{22} tsʊŋ^{53} sɐŋ^{53} ts^{h}ɪŋ^{11}, wɐi^{22} ŋɔː^{23} jiː^{11} sei^{25}]	
BR8	泣白(Qibai)	我又豈能讓你孤魂無主，飄泊陰間呢？[ŋɔː^{23} jɐu^{22} hei^{25} nɐŋ^{11} jœːŋ^{22} nei^{23} kuː^{53} wɐn^{11} mou^{11} tsyː^{25}, p^{h}iːu^{53} p^{h}ɔːk^{3} jɐm^{55} kaːn^{55} nɛː]	
BD1	白(Bai)	咁就真系安樂曬囉。娘子，娘子 [kɐm^{25} tsɐu^{22} tsɐn^{53} hɐi^{22} ɔːn^{53} lɔːk^{2} saːi^{33} lɔː, nœːŋ^{11} tsiː^{25}, nœːŋ^{11} tsiː^{25}]	断桥产子 (Duanqiao Chanzi)

BD2	白(Bai)	小青姐，娘子點樣啊 [siːu^{25} ts^{h}ɪŋ^{53} tsɛː^{25}, nœːŋ^{11} tsiː^{25} tiːm^{25} jœːŋ^{25} aː]	
BK1	詩白(Shibai)	汴梁宮闕都已成灰，北望云天盡可哀 [piːn^{22} lœːŋ^{11} kʊŋ^{53} k^{h}yːt^{3} tou^{55} jiː^{23} sɪŋ^{11} fuːi^{53}, pɐk^{5} mɔːŋ^{22} wɐn^{11} t^{h}iːn^{53} tsɵn^{22} hɔː^{25} ɔːi^{53}]	朱弁回朝之哭主 (Zhubian Huichao zhi Kuzhu)
BK2	詩白(Shibai)	蒙塵圣主身何在？但見狂風卷雪啊來！ [mʊŋ^{11} ts^{h}ɐn^{11} sɪŋ^{33} tsyː^{25} sɐn^{53} hɔː^{11} tsɔːi^{22}, taːn^{22} kiːn^{33} k^{w}ɔːŋ^{11} fʊŋ^{53} kyːn^{25} syːt^{3} aː lɔːi^{21}]	

2. Transcription of Mun Cheen-shuih's singing utterances

Code	Type of melody	Chinese text and IPA transcription	Resource
CH1	花上句 (Huashangju)	龜山景色確迷人，印象難忘 [kwɐi^{55} saːn^{55} kɪŋ^{25} sɪk5 k^{h}ɔːk^{3} mɐi^{11} jɐn^{11}, jɐn^{33} tsœːŋ^{22} naːn^{11} mɔːŋ^{22}]	蝴蝶杯 (Hudiebei)
CH2	大花下句 (Dahuaxiaju)	倉惶唯有走江邊，暫避一時 [ts^{h}ɔːŋ^{55} wɔːŋ^{11} wɐi^{11} jɐu^{23} tsɐu^{25} kɔːŋ^{55} piːn^{55}, tsaːm^{22} pei^{22} jɐt^{5} siː^{11}]	
CH3	花下句 (Huaxiaju)	她就是被人欺負小漁娘，印象依稀 [t^{h}aː^{55} tsɐu^{22} siː^{22} pei^{22} jɐn^{11} hei^{55} fuː^{22} siːu^{25} jyː^{11} nœːŋ^{11}, jɐn^{33} tsœːŋ^{22} jiː^{55} hei^{55}]	
CH4	花 (Hua)	就借寶杯為聘結紅絲，一片深情 [tsɐu^{22}	

		tsɛː^{33} pou^{25} puːi^{55} wɐi^{11} p^{h}ɪŋ^{33} kiːt^{3} hʊŋ^{11} siː^{55}, jɐt^{5} p^{h}iːn^{33} sɐm^{55} ts^{h}ɪŋ^{11}]	
CH5	合尺花下句 (Hechihuaxiaju)	想起她仇深，難怪她悲憤，今晚夜風雨 [sœːŋ^{25} hei^{25} t^{h}aː^{55} sɐu^{11} sɐm^{55}, naːn^{11} kwaːi^{33} t^{h}aː^{55} pei^{55} fɐn^{23}, kɐm^{55} maːn^{23} jɛː^{22} fʊŋ^{55} jyː^{23}]	
CH6	花下句 (Huaxiaju)	古道揮拳懲惡霸，誰知一路惹 [kuː^{25} tou^{22} fɐi^{55} k^{h}yːn^{11} ts^{h}ɪŋ^{11} ɔːk^{3} paː^{33}, sɵy^{11} tsiː^{55} jɐt^{5} lou^{22} jɛː^{23}]	
CH7	花下句 (Huaxiaju)	為求不吃眼前虧 [wɐi^{22} k^{h}ɐu^{11} pɐt^{5} hɛːk^{3} ŋaːn^{23} ts^{h}iːn11 k^{h}wɐi^{55}]	
CX1	唱序(Changxu)	最誒慘求死又不能 [tsɵy^{33} ɛː ts^{h}aːm^{25} k^{h}ɐu^{11} sei^{25} jɐu^{22} pɐt^{5} nɐŋ^{11}]	西廂記 (Xixiangji)
CX2	花 (Hua)	待月西廂下，迎風戶半開，拂牆花影動，疑是玉人來啊 [tɔːi^{22} jyːt^{22} sɐi^{55} sœːŋ^{55} haː^{22}, jɪŋ^{11} fʊŋ^{55} wuː^{22} puːn^{33} hɔːi^{55}, fɐt^{2} ts^{h}œːŋ^{11} faː^{55} jɪŋ^{25} tʊŋ^{22}, jiː^{11} siː^{22} jʊk^{2} jɐn^{11} lɔːi^{11} aː]	
CX3	沉花下句 (Chenhuaxiaju)	多情自古空餘恨，由來好夢 [tɔː^{55} ts^{h}ɪŋ^{11} tsiː^{22} kuː^{25} hʊŋ^{53} jyː^{11} hɐn^{22}, jɐu^{11} lɔːi^{11} hou^{25} mʊŋ^{22}]	
CX4	正線花上句 (Zhengxian Huashangju)	小生雖無潘安貌，也無子建才，卻有一顆真心 [siːu^{25} sɐŋ^{55} sɵy^{55} mou^{11} p^{h}uːn^{55} ɔːn^{55} maːu^{22}, jaː^{23} mou^{11} tsiː^{25} kiːn^{22}	

		ts^{h}ɔːi^{11}, k^{h}œːk^{3} jɐu^{23} jɐt^{5} fɔː^{25} tsɐn^{55} sɐm^{55}]	
CX5	花上句 (Huashangju)	若非今宵端的雨雲來，縱有妙藥靈丹，也難療病 [jœːk^{22} fei^{55} kɐm^{55} siːu^{55} tyːn^{55} tɪk^{5} jyː^{23} wɐn^{11} lɔːi^{11}, tsʊŋ^{33} jɐu^{23} miːu^{22} jœːk^{22} lɪŋ^{11} taːn^{55}, jaː^{23} naːn^{11} liːu^{11} pɛːŋ^{22}]	
CX6	花下句 (Huaxiaju)	為慚衾影 [wɐi^{22} ts^{h}aːm^{11} k^{h}ɐm^{55} jɪŋ^{25}]	
CX7	花下句 (Huaxiaju)	此後只憑明月 [ts^{h}iː^{25} hɐu^{22} tsiː^{25} p^{h}ɐŋ^{11} mɪŋ^{11} jyːt^{22}]	
CS1	打引白 (Dayinbai)	兩吞聲啊 [lœːŋ^{23} t^{h}ɐn^{55} sɪːŋ^{55} aː]	山伯臨終 (Shanbo Linzhong)
CR1	花 (Hua)	近睹分明似儼然，遠觀自在若飛仙，他年得傍蟾宮客，不在梅邊在柳邊哪 [kɐn^{22} tou^{25} fɐn^{53} mɪŋ^{11} ts^{h}iː^{23} jiːm^{11} jiːn^{11}, jyːn^{23} kwuːn^{55} tsiː^{22} tsɔːi^{22} jœːk^{22} fei^{55} siːn^{55}, t^{h}aː^{55} niːn^{11} tɐk^{5} pɔːŋ^{22} siːm^{11} kʊŋ^{55} haːk^{3}, pɐt^{5} tsɔːi^{22} muːi^{11} piːn^{55} tsɔːi^{22} lɐu^{23} piːn^{55} naː]	人鬼恋 (Renguilian)
CR2	木魚(Muyu)	你若是鬼魅妖魔，又怎會如花俏麗呀 [nei^{23} jœːk^{2} siː^{22} kwɐi^{25} mei^{33} jiːu^{25} mɔː^{53}, jɐu^{22} tsɐm^{25} wuːi^{23} jyː^{11} faː^{55} ts^{h}iːu^{33} lɐi^{22} jaː]	

CR3	花下句 (Huaxiaju)	趁此夜靜更闌人熟睡，待我掘開芳冢 [tsʰɐn^{33} tsʰiː^{25} jɛː^{22} tsɪŋ^{22} kɐŋ^{53} laːn^{11} jɐn^{11} sʊk^{2} sɵy^{22}, tɔːi^{22} ŋɔː^{23} kʷɐt^{2} hɔːi^{55} fɔːŋ^{53} tsʰʊŋ^{25}]	
CR4	反線中板 (Fanxian Zhongban)	不管你是新呢 [pɐt^{5} kuːn^{25} nei^{23} siː^{22} sɐn^{53} nɛː]	
CR5	反線中板 (Fanxian Zhongban)	不管你是 [pɐt^{5} kuːn^{25} nei^{23} siː^{22}]	
CR6	花 (Hua)	世間難得有癡心，幸莫見嫌 [sɐi^{22} kaːn^{55} naːn^{11} tɐk^{5} jɐu^{23} tsʰiː^{55} sɐm^{55}, hɐŋ^{22} mɔːk^{22} kiːn^{22} jiːm^{11}]	
CR7	木魚(Muyu)	不管你是鬼是人，我也全無懼畏 [pɐt^{5} kuːn^{25} nei^{23} siː^{22} kʷɐi^{25} siː^{22} jɐn^{11}, ŋɔː^{23} jaː^{23} tsʰyːn^{11} mou^{11} kɵy^{22} wɐi^{33}]	
CR8	木魚(Muyu)	真誠相愛，我亦永不相遺 [tsɐn^{53} sɪŋ^{11} sœːŋ^{55} ŋɔːi^{33}, ŋɔː^{23} jɪk^{22} wɪŋ^{23} pɐt^{5} sœːŋ^{55} wɐi^{11}]	
CD1	乙反木魚(Yifan Muyu)	若非姐你為媒，怎能偕奠雁哪 [jœːk^{2} fei^{53} tsɛː^{25} nei^{23} wɐi^{11} muːi^{11}, tsɐm^{25} nɐŋ^{11} haːi^{11} tiːn^{22} ŋaːn^{22} naː]	
CD2	乙反木魚(Yifan Muyu)	恩義難忘撮合山 [jɐn^{55} jiː^{22} naːn^{11} mɔːŋ^{22} tsʰyːt^{3} hɐp^{2} saːn^{53}]	断桥产子 (Duanqiao Chanzi)
CD3	花下句	恩怨分清愁盡去，前嫌冰釋 [jɐn^{55} jyːn^{33}	

	(Huaxiaju)	fɐn^{55} tsʰɪŋ^{53} sɐu^{11} tsøn^{22} høy^{33}, tsʰiːn^{11} jiːm^{11} pɪŋ^{55} sɪk^{5}]	
CD4	花下句 (Huaxiaju)	欣然攜手返家門，共慶團圓 [jɐn^{55} jiːn^{21} kʷɐi^{11} sɐu^{25} faːn^{25} kaː^{53} muːn^{11}, kʊŋ^{22} hɪŋ^{33} tʰyːn^{11} jyːn^{11}]	
CY1	滚花(Gunhua)	人到深情 [jɐn^{11} tou^{33} sɐm^{53} tsʰɪŋ^{11}]	游园惊梦 (Youyuan Jingmeng)
CK1	反線二王轉乙反 (Fanxian Erwang to Yifan)	狼虎當朝流毒啊害，縱橫 [lɔːŋ^{11} fuː^{25} tɔːŋ^{53} tsʰiːu^{11} lɐu^{11} tʊk^{3} aː hɔːi^{22}, tsʊŋ^{53} waːŋ^{11}]	朱弁回朝之哭主 (Zhubian Huichao zhi Kuzhu)
CZ1	清歌(Qingge)	煙籠寒水月籠沙啊 [jiːn^{53} lʊŋ^{11} hɔːn^{11} søy^{25} jyːt^{2} lʊŋ^{11} saː^{55} aː]	朱弁回朝之招魂 (Zhubian Huichao zhi Zhaohun)
CZ2	清歌(Qingge)	長江夜望盡煙霞啊 [tsʰœːŋ^{11} kɔːŋ^{55} jɛː^{22} mɔːŋ^{22} tsøn^{22} jiːn^{55} haː^{11} aː]	
CZ3	清歌(Qingge)	江水濤濤流自巫山下啊 [kɔːŋ^{53} søy^{25} tʰou^{55} tʰou^{53} lɐu^{11} tsiː^{22} mou^{11} saːn^{55} haː^{22} aː]	

Appendix C
Transcription of recordings by Yuen Siu-fai

1. OSpk: 大人[taːi^{22} jɐn^{11}]

NSpk: 大人[taːi^{22} jɐn^{11}]

2. OSpk: 豈不是好啊？ [hei^{25} pat^{5} siː^{22} hou^{25} aː^{33}]

NSpk: 豈不是好？ [hei^{25} pat^{5} siː^{22} hou^{25}]

3. OSpk: 老相爺，晚生在宮中招親之時，也曾說過，家中並無一人，哪裡來的元配？你休要亂猜！老相爺，閒事閒非，你又何必多管？ [lou^{23} sœːŋ^{33} jɛː^{21}, maːn^{23} sɐŋ^{53} tsɔːi^{22} kʊŋ^{53} tsʊŋ^{53} tsiːu^{53} tsɐn^{53} tsiː^{53} siː^{21}, jaː^{23} ts^{h}ɐŋ^{21} syːt^{3} kwɔː^{33}, kaː^{53} tsʊŋ^{53} pɪŋ^{22} mou^{21} jɐt^{5} jɐn^{21}, naː^{23} lɵy^{23} lɔːi^{21} tɪk^{5} jyːn^{21} p^{h}uːi^{33}? nei^{23} jɐu^{53} jiːu^{33} lyːn^{22} ts^{h}aːi^{53}! lou^{23} sœːŋ^{33} jɛː^{21}, haːn^{21} siː^{22} haːn^{21} fei^{53}, nei^{23} jɐu^{22} hɔː^{21} piːt^{5} tɔː^{53} kwuːn^{25}?]

NSpk: 老相爺，晚生在宮中招親之時，也曾說過，家中並無一人，哪裡來的元配？你休要亂猜！老相爺，閒事閒非，你又何必多管？ [lou^{23} sœːŋ^{33} jɛː^{21}, maːn^{23} sɐŋ^{55} tsɔːi^{22} kʊŋ^{53} tsʊŋ^{53} tsiːu^{53} tsɐn^{55} tsiː^{53} siː^{21}, jaː^{23} ts^{h}ɐŋ^{21} syːt^{3} kwɔː^{33}, kaː^{53} tsʊŋ^{53} pɪŋ^{22} mou^{21} jɐt^{5} jɐn^{21}, naː^{23} lɵy^{23} lɔːi^{21} tɪk^{5} jyːn^{21} p^{h}uːi^{33}? nei^{23} jɐu^{53} jiːu^{33} lyːn^{22} ts^{h}aːi^{53}! lou^{23} sœːŋ^{33} jɛː^{21}, haːn^{21} siː^{22} haːn^{21} fei^{53}, nei^{23} jɐu^{22} hɔː^{21} piːt^{5} tɔː^{53} kwuːn^{25}?]

4. OSpk: 有何不可啊？ [jɐu23 hɔː21 pɐt5 hɔː25 aː22?]

NSpk: 有何不可？ [jɐu23 hɔː21 pɐt5 hɔː25?]

5. OSpk: 冤枉啊！ [jyːn53 wɔːŋ25 aː33!]

NSpk: 冤枉啊！ [jyːn53 wɔːŋ25 aː33!]

Appendix D
Lyrics of songs read and sung by the informant ZSX

1. "月光光 (Moonlight)"

月光光，照地堂，蝦仔你乖乖訓落床。聽朝阿媽要捕魚蝦咯，阿嫲織網要織到天光。蝦仔你快高長大咯，劃艇撒網就更在行。

2. "齊齊望過去 ('Oh Susannah')"

齊齊望過去，清溪里，有隻青蛙想跳水。齊齊望過去，小屋里，有隻豬仔真風趣。有隻了哥，吱吱喳喳想駁嘴。齊齊望過去，山窿里，有隻獅子竟飲醉。

3. "男兒當自強 (A man of determination)"

傲氣傲笑萬重浪，熱血熱勝紅日光，膽似鐵打，骨似精鋼，胸襟百千丈，眼光萬里長，誓奮發自強，做好漢。做個好漢子，每天要自強，熱血男子，熱勝紅日光。願海天為我聚能量，去開天闢地，為我理想去闖，看碧波高漲又看碧空廣闊浩氣揚，既是男兒當自強。昂步挺胸大家作棟樑，做好漢。用我百點熱，耀出千分光。做個好漢子，熱血熱腸熱，熱勝紅日光。

4. "紅日 (The sun)"

命運就算顛沛流離，命運就算曲折離奇，命運就算恐嚇著你做人沒趣味，別流淚心酸，更不應捨棄，我願能，一生永遠陪伴你。一生之中兜兜轉轉哪會看清楚彷徨時我也試過獨坐一角像是沒協作。在某年那幼小的我，跌倒過幾多幾多落淚在雨夜滂沱。一生之中彎彎曲曲我也要走過，從何時有你有你伴我給我熱烈地拍和。像紅日之火，燃點真的我，結伴行，千山也定能踏過。讓晚風，輕輕吹過，伴送著清幽花香，像是在祝福你我。讓晚星，輕輕閃過，閃出你每個希冀如浪花，快要沾濕我。

5. "真的愛你 (Really loving you)"

無法可修飾的一對手，帶出溫暖永遠在背後，縱使囉嗦始終關注不懂珍惜太內疚。仍記起溫馨的一對手，始終給我照顧未變樣，理想今天終於

等到分享光輝盼做到。春風化雨暖透我的心，一生眷顧無言地送贈。是你多麼溫馨的目光，教我堅毅望著前路，叮囑我，跌倒不應放棄。沒法解釋怎可報盡親恩，愛意寬大是無限，請準我，說聲真的愛你。

6. "花火 (Fireworks)"

來吧，伴我飛，不休不睡去飛。來盡力忘記，兩腳降落何地。若記憶憑據，到某一天告吹，回望即使太蠢，都相信我做得對。原來風雪可以使我堅壯使我堅強。假使敢夢與想，假使天真地唱，我也會笑容漂亮。原來歲月太長，可以豐富，可以荒涼，能忘掉結果，未能忘記遇上。長路若太短，花火生命更短，雙手可觸及你，有眼淚仍是暖。

7. "帝女花•香夭 (Princess Chang Ping)"

落花滿天閉月光，借一杯附薦鳳台上，帝女花帶淚上香，願喪生回謝爹娘。

Appendix E
Questionnaire of the perception test

聽辨調查問卷

一、這一小節包含 36 組音節，每組的前後兩個音節可能為同一個字，也可能不同。請判斷你聽到的前後兩個音節分別對應於下面 A、B、C 哪個選項。若選項中沒有您聽到的音節，可填寫您聽到的音節的聲調更接近於下面哪個音節：

T1 詩　T2 史　T3 試　T4 時　T5 市　T6 事

1 (　)(　)　A 熬　B 奧　C 傲
2(　)(　)　A 床　B 創　C 敞
3(　)(　)　A 熬　B 奧　C 傲
4 (　)(　)　A 歹　B 大　C 帶
5(　)(　)　A 弟　B 帝　C 低
6(　)(　)　A 鳳　B 諷　C 風
7(　)(　)　A 拐　B 怪　C 乖
8(　)(　)　A 杠　B 廣　C 光
9(　)(　)　A 紅　B 凶　C 哄（起～）
10(　)(　)　A 拐　B 怪　C 乖
11(　)(　)　A 化　B 花
12(　)(　)　A 狼　B 浪　C 朗
13(　)(　)　A 雷　B 淚　C 鋁
14(　)(　)　落（前後同音選 A，不同音選 B）
15(　)(　)　A 離　B 你　C 梨（士多啤～）
16(　)(　)　A 女（～仔）　B 女（牛郎織～）　C 類
17(　)(　)　A 喜　B 氣　C 稀
18(　)(　)　A 除　B 趣　C 催
19(　)(　)　熱（前後同音選 A，不同音選 B）
20(　)(　)　A 日　B 一

21(　　)(　　)　　A 上（～車）　B 想　C 尚
22(　　)(　　)　　A 盛　B 勝　C 星
23(　　)(　　)　　熱（前後同音選 A，不同音選 B）
24(　　)(　　)　　A 蠻　B 萬　C 晚
25(　　)(　　)　　A 仔　B 濟　C 劑
26(　　)(　　)　　A 夏　B 下（鄉～）　C 蝦
27(　　)(　　)　　A 向　B 香　C 響
28(　　)(　　)　　A 兆　B 笑　C 消
29(　　)(　　)　　血（前後同音選 A，不同音選 B）
30(　　)(　　)　　A 份　B 訓　C 分
31(　　)(　　)　　A 幼　B 有　C 柚
32(　　)(　　)　　A 陣　B 振　C 真
33(　　)(　　)　　A 只　B 職
34(　　)(　　)　　A 從　B 寵　C 沖
35(　　)(　　)　　A 住　B 注　C 豬
36(　　)(　　)　　A 仔　B 濟　C 劑

二、請判斷您聽到的句子是說的還是唱的，如果是說的請選 A，如果是唱的請選 B。

1(　　)　2(　　)　3(　　)　4(　　)　5(　　)
6(　　)　7(　　)　8(　　)　9(　　)　10(　　)
11(　　)　12(　　)

謝謝您參與本次聽辨實驗！

Appendix F

Materials of the recording questionnaire in Experiment II

1. Isolated citation form (In the order of T1 ~ T6):

(1) [siː]: 詩史試時市是

(2) [jiː]: 醫椅意疑以義

(3) [jyːn]: 淵苑怨圓遠願

2. Comparison between speaking and singing

(1) 有幾多全奉送 (“愛是永恆”) T5 T2 T1 T4 T6 T3

(2) 再來也許要天上團聚 (“小城大事”) T3 T4 T5 T2 T3 T1 T6 T4 T6

(3) 仍未忘跟你約定假如沒有死 (“約定”) T4 T6 T4 T1 T5 T3 T6 T2 T4 T6 T5 T2

(4) 誰令我仿似初戀再嘗 (“印象”) T4 T6 T5 T2 T5 T1 T2 T6 T4

(5) 我卻為何偏偏喜歡你 (“偏偏喜歡你”) T5 T3 T6 T4 T1 T1 T2 T1 T5

(6) 其實我都很小氣 (“朋友仔”) T4 T6 T5 T1 T2 T2 T3

(7) 留下了幾多首我喜歡到現時 (“飲歌”) T4 T6 T5 T2 T1 T2 T5 T2 T1 T3 T6 T4

(8) 只我一人未發覺 (“再見二丁目”) T2 T5 T1 T4 T6 T3 T3

(9) 別人亦都很愛我 (“如果我們不再見”) T6 T4 T6 T1 T2 T3 T5

(10) 曾說同你闖天與地 (“誰願放手”) T4 T3 T4 T5 T2 T1 T5 T6

(11) 每與你共同度過都不枉過 (“追”) T5 T5 T5 T6 T4 T6 T3 T1 T1 T2 T3

(12) 蝦仔你乖乖訓落床 (“月光光”) T1 T2 T5 T1 T1 T3 T6 T4

(13) 想小鳥伴你飛舞雲外看琴譜 (“叮嚀”) T2 T2 T5 T6 T5 T1 T5 T4 T6 T3 T4 T2

(14) 求豐收雨點降下花兒別怕 (“風車”) T4 T1 T1 T5 T2 T3 T6 T1 T4 T6 T3

(15) 我再將紫釵弄, 漸露笑容 (“紫釵記·劍合釵圓”) T5 T3 T1 T2 T1 T6, T6 T6 T3 T4

(16) 帝女花帶淚上香, 願喪生回謝爹娘 (“帝女花·香夭”) T3 T2 T1 T3 T6 T5 T1, T6 T3 T1 T4 T6 T1 T4

(17) 輾輾轉轉(T2 T2 T2 T2)之間(T1), 悄悄(T3 T3)愛已冷淡(T6), 點點(T2 T2)火花再不燦爛(T6), 孤孤單單(T1 T1 T1 T1)心間(T1), 撲撲索索(T3 T3 T3 T3)界限(T6), 猜猜測測(T1 T1 T1 T1)在你雙眼(T5) ("當愛變成習慣")

(18) 深深刻刻(T1 T1 T1 T1)相識(T1), 痛痛快快(T3 T3 T3 T3)過後(T6), 輕輕率率(T1 T1 T1 T1)置諸腦後(T6), 恍恍(T2 T2)惚惚(T1 T1)之間(T1), 勉勉強強(T5 T5 T5 T5)接受(T6), 冰冰(T1 T1)的心是我所有(T5) ("當愛變成習慣")

(19) 纏纏綿綿(T4 T4 T4 T4)放放縱縱(T3 T3 T3 T3)熱熱烈烈(T6 T6 T6 T6)的相愛(T3), 猶猶疑疑(T4 T4 T4 T4)進進退退(T3 T3 T3 T3)落落寞寞(T6 T6 T6 T6)的分開(T1) ("這一次意外")

Bibliography

Anderson, Steven R. (1978). Tone features. In Victoria. A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 133-175). New York: Academic Press.

Arvaniti, Amalia, & Baltazani, Mary. (2005). Intonation analysis and prosodic annotation of Greek spoken corpora. In Sun-Ah Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 84-117). Oxford: Oxford University Press.

Baart, Joan L.G. (2004). Tone and song in Kalam Kohistani (Pakistan). In Hugo Quené & Vincent van Heuven (Eds.), *On Speech and Language: Studies for Sieb G. Nooteboom* (pp. 5-15). Utrecht: Netherlands Graduate School in Linguistics.

Bao, Zhiming. (1999). *The Structure of Tone*. New York: Oxford University Press.

Bauer, Robert. S., & Benedict, Paul K. (1997). *Modern Cantonese phonology*. New York: Mouton de Gruyter.

Bauer, Robert. S., Cheung, Kwan-Hin, & Cheung, P. M. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change, 15*, 211-225.

Beckman, Mary E. (1986). *Stress and Non-stress accent*. Dordrecht: Foris.

Beckman, Mary E., & Pierrehumbert, Janet B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255-309.

Boersma, Paul, & Weenink, David. (2009). Praat : Doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved from Retrieved May 1, 2010, from http://www.praat.org/

Bolinger, Dwight. (1978). Intonation across languages. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (Eds.), *Universals of Human*

Language (Vol. 2. Phonology, pp. 471-524). Stanford, California: Stanford University Press.

Bruce, Gösta. (1977). *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Chan, Chi-Ching 陳志清. (2010). *粵曲韵轍及文字譜探微 [Studies on Rhymes and Texts in Cantonese Opera Songs]*. Hong Kong: Cang Yong Lou.

Chan, Marjorie Kit-Man. (1987a). Tone and Melody in Cantonese. *Proceedings of the Thirteenth Annual Meeting, Berkeley Linguistics Society*, 29-37.

Chan, Marjorie Kit-Man. (1987b). Tone and melody interaction in Cantonese and Mandarin. *UCLA Working Papers in Phonetics, 68*, 132-169.

Chang, Flynn Choi-Yeung. (2003). *Intonation in Cantonese*. Munich: Lincom Europa.

Chao, Yuen Ren. (1930/2006). A system of "Tone-letters". In Zongji Wu & Xinna Chao (Eds.), *Linguistic Essays by Yuenren Chao* (Vol. 30, pp. 98-103). Beijing: Commercial Press, 2006. (Reprinted from: English version included in Z. J. Wu, & X. N. Chao (Eds.), 2006, Linguistic Essays by Yuenren Chao (pp. 98-103). Beijing: Commercial Press).

Chao, Yuen Ren. (1932/2006). A preliminary study of English intonation (with American variants) and its Chinese equivalents. In Zongji Wu & Xinna Chao (Eds.), *Linguistic Essays by Yuenren Chao* (pp. 110-197). Beijing: Commercial Press, 2006.

Chao, Yuen Ren. (1933/2006). Tone and intonation in Chinese. In Zongji Wu & Xinna Chao (Eds.), *Linguistic Essays by Yuenren Chao* (pp. 198-220). Beijing: Commercial Press, 2006.

Chao, Yuen Ren. (1947). *Cantonese Primer*. New York: Greenwood Press.

Chao, Yuen Ren. (1956/2006). Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In Zongji Wu &

Xinna Chao (Eds.), *Linguistic essays by Yuenren Chao* (pp. 596-611). Beijing: Commercial Press, 2006.

Chao, Yuen Ren. (1968). *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen-Hafteck, Lily. (1999). Singing Cantonese children's songs: significance of the pitch relationship between text and melody. *Music Education Research, 1*(1), 93-108.

Chen, Matthew Y. (2000). *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.

Cheung, Kwan-Hin. (1986). *The Phonology of Present-day Cantonese.* Doctoral dissertation, University College London, London.

Cheung, Kwan-Hin. (2012a). *The mechanism of tone-melody match for tone preservation in Cantonese songs.* Paper presented at the Workshop on Innovations in Cantonese Linguistics, Ohio State University, Columbus.

Cheung, Kwan-Hin 張群顯. (2007). 粵語字調與旋律的配合初探 [Cantonese tone-melody match: an exploration]. *粵語研究 [Studies on Cantonese], 2*, 8-16.

Cheung, Kwan-Hin 張群顯. (2012b). *粵劇與粵音[Cantonese opera and Cantonese sounds]*. Paper presented at the Moving Forward - International Symposium on Chinese Linguistics and Philology, the 50th Anniversary of the Department of Chinese Language and Literature, The Chinese University of Hong Kong, Hong Kong.

Cheung, Kwan-Hin 張群顯. (2012c). 趙元任五度制標調法：理論與實際 [Chao Tone Letters: Theory and Practice]. *學術研究 [Academic Research], 2012*(5), 152-156.

Cheung, Kwan-Hin 張群顯, & Wong, Sing-Kwan 王勝焜. (2008). 粵曲梆黃說唱字調樂調配合初探 [Tone-melody match for aria types and narrative songs in Cantonese opera: a preliminary study]. *粵語研究 [Studies on Cantonese], 3*, 12-25.

Cheung, Samuel Hung-Nin. (1996). Songs and rhymes: Cantonese phonology as reconstructed from popular songs. *Journal of Chinese Linguistics, 24*(1), 1-53.

Cheung, Samuel Hung-Nin 張洪年. (2002). 21 世紀的香港粵語：一個新語音系統的形成 [Cantonese of the 21st century: formation of a new phonological system in the Hong Kong language]. *暨南學報(哲學社會科學) [Journal of Jinan University (Philosophy & Social Science Edition)], 24*(2), 25-40.

Cheung, Yat-Sing 張日昇. (1969). 香港粵語中的陰平調及變調問題 [Yin-ping tone and changed tones in Hong Kong Cantonese]. *中國文化研究所學報 [Journal of Institute of Chinese Culture Studies], 2*(1), 81-104.

Chow, Ivan. (2006). *Interactions between syllabic-level prosody and prosodic-group boundary markers in Cantonese.* Paper presented at the the 2006 annual conference of the Canadian Linguistic Association.

Chow, Man-ying. (2012). *Singing the Right Tones of the Words: The Principles and Poetics of Tone-Melody Mapping in Cantopop.* M.Phil. thesis, The University of Hong Kong.

Collier, René (1990). On the perceptual analysis of intonation. *Speech Communication, 9*, 443-451.

Connell, Bruce. (2002). Downdrift, downstep, and declination. In U. Gut & D. Gibbon (Eds.), *Typology of African Prosodic Systems* (pp. 3-12). Germany: Universität Bielefeld.

Connell, Bruce, & Ladd, D. Robert. (1990). Aspects of pitch realization in Yoruba. *Phonology, 7*(1), 1-29.

Cruttenden, Alan. (1986). *Intonation*. Cambridge: Cambridge University Press.

Crystal, David. (2008). *A Dictionary of Linguistics and Phonetics* (6th ed.). Malden, MA ; Oxford: Blackwell Publishing.

Duanmu, San. (2000). Tone: An overview. In Lisa Lai-Shen Cheng & Rint Sybesma (Eds.), *The First Glot International State-of-the-Article Book: The Latest in Linguistics (Studies in Generative Grammar 48)* (pp. 251-286). Berlin: Mouton de Gruyter.

Ekwueme, Laz E. N. (1974a). Concepts of African musical theory. *Journal of Black Studies, 5*(1), 35-64.

Ekwueme, Laz E. N. (1974b). Linguistic determinants of some Igbo musical properties. *Journal of African Studies, 1*(3), 335-353.

Esser, Jürgen, & Polomski, Andrzej. (1988). *Comparing Reading and Speaking Intonation*. Amsterdam: Rodopi.

Fok Chan, Y.-Y. (1974). *A perceptual study of tones in Cantonese.* University of Hong Kong, Hong Kong.

Fox, Anthony, Luke, Kang Kwong, & Nancarrow, Owen. (2008). Aspects of intonation in Cantonese. *Journal of Chinese Linguistics, 36*, 321-367.

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

Fujisaki, Hiroya. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (Ed.), *The Production of Speech* (pp. 39–55). New York: Springer-Verlag.

Fung, Roxana S. Y., & Wong, Cathy S. P. (2011). Mandarinization of Hong Kong Cantonese tones. In J. Shi, Z. Geng & J. Yang (Eds.), *Ontological studies of the Chinese language* (pp. 289-299). Beijing: Chinese Social Science Publisher.

Gandour, Jackson. (1981). Perceptual dimensions of tone: evidence from Cantonese. *Journal of Chinese Linguistics, 9*, 20-36.

Grice, Martine, Baumann, Stefan, & Benzmüller, Ralf. (2005). German intonation in autosegmental-metrical phonology. In Sun-Ah Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 55-83). Oxford: Oxford University Press.

Guo, Jinfu 郭錦桴. (1993). *漢語聲調語調闡要與探索 [An introduction and exploration to tone and intonation in Chinese]*. Beijing: Beijing Language College Press.

Gussenhoven, Carlos. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. (2007). Intonation. In Paul De Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 253-280). Cambridge: Cambridge University Press.

Gussenhoven, Carlos, & Rietveld, A. C. M. . (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics, 16*, 355–369.

Hashimoto, Oi-Kan Yue. (1972). *Phonology of Cantonese*. Cambridge: Cambridge University Press.

He, Yang 賀陽, & Jing, Song 勁松. (1992). 北京話語調的實驗探索 [Experimental exploration on intonation of Beijing Mandarin]. *語言教學與研究 [Language Teaching and Linguistic Studies], 1992*(2), 71-96.

Herzog, George. (1934). Speech-melody and primitive music. *Musical Quarterly, 20*, 452-466.

Hirst, Daniel, & Cristo, Albert Di. (1998). A survey of intonation systems. In Daniel Hirst & Albert Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

Ho, Rerrario Shui-Ching, & Ho, Sui-Fong. (2007). Musical scales and Cantonese level tones. In Emery Schubert, Kym Buckley, Rosemary Eliott, Brooke Koboroff, Johnson Chen & Catherine Stevens (Eds.), *Proceedings of the*

Inaugural Conference on Music Communication Science (pp. 64-67). New South Wales: HCESNet, University of Western Sydney.

Ho, Vincie Wing-See. (2006). *The tone-melody interface of popular songs written in tone languages.* Paper presented at the the 9th International Conference on Music Perception and Cognition.

Ho, Vincie Wing-See. (2008). *Fine-tuning tone-melody constraints through the investigation of mismatches in Cantonese pop music.* Paper presented at the the 13th International Conference on Cantonese and Yue Dialects.

Hornbostel, E. M. von. (1928). African Negro Music. *Africa, 1*(1), 30-62.

Huang, Xianjun 黃賢軍, Yang, Yufang 楊玉芳, & Lü, Shinan 呂士楠. (2007). 漢語語調降階的實驗研究 [Experimental studies on downstep in Chinese intonation]. *聲學學報 [Acta Acustica], 32*(1), 56-61.

Hulst, Harry van der, & Snider, Keith L. (1993). *The Phonology of Tone : the Representation of Tonal Register [Linguistic Models 17]*. Berlin: Mouton de Gruyter.

Ian, Maddieson. (1978). Universals of Tone. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (Eds.), *Universals of Human Language* (Vol. 2. Phonology, pp. 335-365). Stanford, California: Stanford University Press.

Jassem, Wiktor. (1971). Pitch and compass of speaking voice. *Journal of International Phonetic Association, 12*, 59-68.

Jin, Jian 金健. (2009). *声调感知的研究——汉语方言典型个案的实验 [A Study of Tone Perception: Typical Cases of Chinese Dialects Experiments]*. Doctoral dissertation, Sun Yat-Sen University, Guangzhou.

Jin, Jian 金健. (2010). 廣州方言和東海方言平調的感知研究 [A perception research of level tone patterns of Cantonese and Donghai dialect]. *方言 [Dialect], 2*, 145-155.

Jones, A. M. (1959). *Studies in African Music*. London: Oxford University Press.

Jones, Daniel, & Woo, Kwing-Tong. (1912). *A Cantonese Phonetic Reader*. London: The Univerversity of London Press.

Kam, Tak Him. (1977). Derivation by tone change in Cantonese: a preliminary survey. *Journal of Chinese Linguistics, 5*(2), 186-210.

Kao, Diana L. (1971). *Structure of the syllable in Cantonese*. The Netherlands: Mouton & Co., Printers, The Hague.

Kei, Joseph, Smyth, Veronica, So, Lydia K. H., Lau, C. C., & Capell, Ken. (2002). Assessing the accuracy of production of Cantonese lexical tones: A comparison between perceptual judgement and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing, 7*, 25-38.

Khouw, Edward, & Ciocca, Valter. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*, 104-117.

Killingley, Siew-Yue. (1993). *Cantonese*. Lincom Europa: München-Newcastle.

Klapuri, Anssi, & Davy, Manuel. (2006). *Signal processing methods for music transcription*: Springer.

Ladd, D. Robert. (1983). Peak features and overall slope. In Anne Cutler & D. Robert Ladd (Eds.), *Prosody: Models and Measurements* (pp. 39-52). Berlin, Heidelberg, New York, Tokyo: Springer-Verlag.

Ladd, D. Robert. (1984). Declination: a review and some hypotheses. *Phonology Yearbook, 1*, 53-74.

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

Ladefoged, Peter. (1975/2006). *A Course in Phonetics* (5th ed.). Boston: Thomson Wadsworth.

Lam, Kin-Ping 林建平. (2002). *香港粵語句調研究 [A Study of Intonation in Hong Kong Cantonese]*. Doctoral dissertation, The Chinese University of Hong Kong, Hong Kong.

Lau, Elaine. (2010). Tone-melody relationship in Cantonese. *Working Papers in Linguistics: University of Hawai'i at Manoa, 41*(3), 1-12.

Leben, William R. (1983). On the correspondence between linguistic tone and musical melody. *Proceedings of the Ninth Annual Meeting of the Berkley Lingusitics Society*, 148-157.

Lee, Tan, Kochanski, Greg, Shih, Chi-Lin, & Li, Yujia. (2002). Modeling tones in continuous Cantonese speech. *Proceedings of the Seventh International Conference on Spoken Language Processing*, 2401-2404.

Li, Fang-kuei. (1980). Laryngeal features and tone development *Bulletin of Institute of History and Philology, Academia Sinica, 51*(1), 1-13.

Li, Shuxian 李書嫻. (2008). 關於廣州話陰去調和陽去調的聽辯實驗 [Auditory Discrimination of Yinqu VS. Yangqu Tones of Cantonese]. *方言 [Dialect], 1*, 34-39.

Li, Yu-Jia, Lee, Tan, & Qian, Yao. (2002). *Acoustical F0 analysis of continuous Cantonese speech.* Paper presented at the International Symposium on Chinese Spoken Language Processing.

Liberman, Mark. (1975). *The Intonational System of English.* Doctoral dissertation, Massachusetts Institute of Technology.

Liberman, Mark, & Pierrehumbert, Janet (1984). Intonational invariance under changes in pitch range and length. In Mark Aronoff & Richard T. Oehrle (Eds.), *Language sound and structure* (pp. 157-223). Cambridge: MIT Press.

Liberman, Philip. (1967). *Intonation, Perception, and Language*. Cambridge: MIT Press.

Lindau, Mona. (1986). Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America, 80*(3), 757-764.

List, George. (1961). Speech melody and song melody in central Thailand. *Ethnomusicology, 5*(1), 16-32.

Liu, Marjory. (1974). The Influence of Tonal Speech on K'unch'ü Opera Style. *Selected Reports in Ethnomusicology, 2*(1), 63-86.

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

Ma, Joan K. Y. (2007). *The interaction between intonation and tone in Cantonese.* Ph.D. thesis, The University of Hong Kong, Hong Kong.

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

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

Ma, Joan K. Y., Ciocca, Valter, & Whitehill, Tara L. . (2006b). Quantitative analysis of intonation patterns in statements and questions in Cantonese. *Proceedings of the Third International Conference on Speech Prosody*, 277-280.

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

Mai, Yun 麥耘. (1998). 廣州話語調說略 [An introduction to Cantonese intonation] *廣州話研究與教學 (第三輯) [Research and pedagogics on Cantonese (Third Edition)]* Guangzhou: Publisher of Sun Yat-sen University.

Mai, Yun 麥耘. (2000). 實驗報告: 廣州話句末升語調對字調的影響 [An experimental report: effects on Cantonese tones by sentence final rising intonation]. In Chow-Yiu Sin & Kang Kwong Luke (Eds.), *第七屆國際粵方言研討會論文集 [Selected papers in the 7th International Conference on Cantonese and Other Yue dialects]*. Beijing: Commercial Press.

Mark, Lindy Li, & Li, Fang Kuei (1966). Speech tone and melody in Wu-Ming folk songs. In Jean Boisselier, Ba Shin & A. B. Griswold (Eds.), *Essays Offered to G. H. Luce by His Colleagues and Friends in Honour of His Seventy-Fifth Birthday. Artibus Asiae, Supplement 23: Papers on Asian History, Religion, Languages, Literature, Music Folklore, and Anthropology* (Vol. 1, pp. 167-186). Ascona, Switzerland: Artibus Asiae.

Matthews, Stephen, & Yip, Virginia. (1994). *Cantonese: A Comprehensive Grammar*. London and New York: Routledge.

Menn, Lise, & Boyce, Suzanne. (1982). Fundamental frequency and discourse structure. *Language and Speech, 25*, 341-383.

Meyer, Julien. (2007). *Acoustic features and perceptive cues of songs and dialogues in whistled speech: Convergences with sung speech.* Paper presented at the International Symposium on Musical Acoustics 2007, Barcelona.

Mok, Peggy Pik-Ki. (2008). On the syllable-timing of Cantonese and Beijing Mandarin. *Proceedings of the 8th Phonetics Conference of China (PCC 2008) and the International Symposium on Phonetic Frontiers (ISPF 2008).*

Mok, Peggy Pik-Ki, & Wong, Peggy Wai-Yi. (2010a). *Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables.* Paper presented at the Speech Prosody 2010, Chicago.

Mok, Peggy Pik-Ki, & Wong, Peggy Wai-Yi. (2010b). *Production and perception of the rising tones in Hong Kong Cantonese.* Paper presented at the the 9th Phonetics Conference of China, Tianjin.

Mok, Peggy Pik-Ki, & Wong, Peggy Wai-Yi. (2010c). *Production of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables.* Paper presented at the Speech Prosody 2010, Chicago.

Norman, Jerry. (1988). *Chinese (Cambridge Language Surveys)*. Cambridge: Press Syndicate of the University of Cambridge.

Ohala, John J. (1978). Production of tone. In Victoria. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5-39). New York: Academic Press.

Ohala, John J. (1981). The listener as a source of sound change *Papers from the Parasession on Language and Behavior* (pp. 178-203). Chicago: Chicago Linguistic Society.

Patel, Aniruddh D. , Iversen, John R., & Rosenberg, Jason C. . (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of Acoustical Society of America, 119*(5), 3034-3047.

Peng, Gang, & Wang, William Shi-Yuan. (2005). Tone recognition of continuous Cantonese speech based on support vector machines. *Speech Communication, 45*, 49-62.

Peng, Gang, Zhang, Caicai, Zheng, Hong-Ying, Minett, James W., & Wang, William Shi-Yuan. (2012). The effect of intertalker variations on acoustic -- Perceptual mapping in Cantonese and Mandarin tone systems. *Journal of Speech, Language and Hearing Research, 55*, 579-595.

Peng, Shu-hui, Chan, Marjorie Kit-Man, Tseng, Chiu-yu, Huang, Tsan, Lee, Ok Joo, & Beckman, Mary E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In Sun-Ah Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 230-270). Oxford: Oxford University Press.

Pian-Chao, Rulan. (2000). Tone and tone: applying musical elements to Chinese words. *Journal of Chinese Linguistics, 28*(2), 181-199.

Pierrehumbert, Janet B. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America, 66*(2), 363-369.

Pierrehumbert, Janet B. (1980). *The phonology and phonetics of English intonation.* Doctoral dissertation, Massachusetts Institute of Technology.

Pierrehumbert, Janet B. (1981). Synthesizing intonation. *Journal of the Acoustical Society of America, 70*(4), 985-995.

Pike, Kenneth L. (1976). *Tone languages : a technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion*. Ann Arbor, Michigan: University of Michigan Press.

Pike, Kenneth Lee. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Prieto, Pilar, Shih, Chi-Lin, & Nibert, Holly. (1996). Pitch downtrend in Spanish. *Journal of Phonetics, 24*, 445-473.

Pulleyblank, Edwin G. . (1978). The nature of the Middle Chinese tones and their development to early Mandarin. *Journal of Chinese Linguistics, 6*(2), 173-203.

Richards, Paul. (1972). A quantitative analysis of the relationship between language tone and melody in a Hausa song. *African Language Studies, 13*, 137-161.

Rose, Phil. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication, 6*, 343-351.

Rose, Phil. (2000). Hong Kong Cantonese citation tone acoustics: a linguistic tonetic study. In Michael Barlow (Ed.), *Proceedings of the 8th Australian International Conference on Speech Science and Technology* (pp. 198-

203). Canberra: Australian Speech Science and Technology Association (Inc).

Rycroft, David K. (1979). The relationships between speech-tone and melody in Southern African music. In J. P. Malan (Ed.), *South African music encyclopedia* (Vol. 2, pp. 301-314). Capetown: Oxford University Press.

Schellenberg, Murray. (2009). Singing in a tone language: Shona. In Akinloye Ojo & Lioba Moshi (Eds.), *Selected Proceedings of the 39th Annual Conference on African Linguistics* (pp. 137-144). Somerville, MA: Cascadilla Proceedings Project.

Schellenberg, Murray. (2011). Tone Contour Realization in Sung Cantonese. *Proceedings of the Seventeenth International Congress of the Phonetic Sciences*, 1754-1757.

Schellenberg, Murray. (2012a). Does language determine music in tone languages? *Ethnomusicology, 52*(2), 266-278.

Schellenberg, Murray. (2012b). *Tone realization in sung Mandarin.* Paper presented at the the 3rd International Symposium on Tonal Aspects of Language, Nanjing.

Schellenberg, Murray. (2013). *The Realization of Tone in Singing in Cantonese and Mandarin.* Doctoral Dissertation, The University of British Columbia (Vancouver).

Shao, Yi 邵宜, & Xian, Guowei 冼國偉. (2004). 吉隆坡現代粵語陽上變陰去現象解因 [A study of the phenomenon of Yangshang tone for Yinqu tone in Kuala Lumpur Cantonese]. *Guangzhou Daxue Xuebao (Shehui Kexue Ban) [Journal of Guangzhou University (Social Science Edition)], 3*(10), 15-18.

Shen, Xiao-nan. (1989). Interplay of the four citation tones and intonation in Mandarin Chinese. *Journal of Chinese Linguistics, 17*(1), 61-74.

Shi, Feng, Shi, Lin, & Liao, Rongrong. (1987). An experimental analysis of the five level tones of Gaoba Dong language. *Journal of Chinese Linguistics, 15*(2), 335-361.

Shih, Chi-Lin. (1997). Declination in Mandarin. In Antonis Botinis, Georgios Kouroupetroglou & George Carayiannis (Eds.), *Intonation: Theory, Models and Applications, Proceedings of an ESCA Workshop. European Speech Communication Association* (pp. 72 -106). Cambridge: Cambridge University Press.

Silverman, Kim, Beckman, Mary E., Pitrelli, John, Ostendorf, Mari, Wightman, Colin, Price, Patti, . . . Hirschberg, Julia. (1992). TOBI: A standard for labeling English Prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing*, 867-870.

So, Lydia K. H. (1996). Tonal changes in Hong Kong Cantonese. *Current Issues in Language & Society, 3*, 186-189.

Sorensen, John M., & Cooper, William E. (1980). Syntactic coding of fundamental frequency in speech production. In Ronald A. Cole (Ed.), *Perception and Production of Fluent Speech* (pp. 399-440). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Starke, Amy. (1930). *The Relation between the Intonation of Song and of Speech among the Amaxosa.* M.A. Thesis, University of Cape Town.

Stock, Jonathan P. J. (1999). A reassessment of the relationship between text, melody and aria structure in Beijing opera. *Journal of Musicological Research, 18*, 183-206.

Strik, Helmer, & Boves, Louis. (1995). Downtrend in F0 and Psb. *Journal of Phonetics, 23*, 203-220.

Sundberg, Johan. (1977). The acoustics of the singing voice. *Scientific American, March,* 104-116.

Sundberg, Johan. (2001). Level and center frequency of the singer's formant. *Journal of Voice, 15*(2), 176-186.

Tench, Paul. (1996). *The Intonation Systems of English*. New York: Cassell.

Ting, Ai Chen. (1971/1984). *Mandarin Tones in Selected Sentence Environments: An Acoustic Study.* Doctoral dissertation, University of Wisconsin, University Microfilms, A XEROX Company, Ann Arbor, Michigan.

Tseng, Chiu-yu. (1990). *An Acoustic Phonetic Study on Tones in Mandarin Chinese* (Vol. Special Publications No. 94). Nankang, Taipei, Taiwan: Institute of History & Philology, Academia Sinica.

Vance, Timothy J. (1976). An Experimental investigation of tone and intonation in Cantonese. *Phonetica, 33*, 368-392.

Vance, Timothy J. (1977). Tonal distinctions in Cantonese. *Phonetica, 34*, 93-107.

Varley, Rosemary , & So, Lydia K.H. . (1995). Age effects in tonal comprehension in Cantonese. *Journal of Chinese Linguistics, 23*(2), 76-98.

Wakefield, John C. (2010). *The English Equivalents of Cantonese Sentence-final Particles: A Contrastive Analysis.* Doctoral dissertation, The Hong Kong Polytechnic University, Hong Kong.

Wang, Anhong 王安紅, Chen, Ming 陳明, & Lu, Shinan 呂士楠. (2004). 基於言語數據庫的漢語音高下傾現象研究. *聲學學報 [Acta Acustica], 29*(4), 353-358.

Wang, William Shi-Yuan. (1967). Phonological features of tone. *International Journal of American Linguistics, 33*(2), 93-105.

Wee, Lian-Hee. (2007). Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics, 35*(1), 128-143.

Wells, John, C. (2006). *English Intonation: An Introduction*. New York: Cambridge University Press.

Whitaker, K. P. K. (1955). A study on the modified tones in spoken Cantonese. *Asia Major, Volume 5, Part 1*, 9-36.

Whitaker, K. P. K. (1956). A study on the modified tones in spoken Cantonese II. *Asia Major, Volume 5, Part 2*, 184-207.

Wong, Jum-Sum 黃湛森. (1981). *粵劇問題探討 [A Study on the Problems of Cantonese Opera]*. M.Phil. dissertation, The University of Hong Kong, Hong Kong.

Wong, Jum-Sum 黃湛森. (2003). *粵語流行曲的發展與興衰：香港流行音樂研究 (1949-1997) [The rise and decline of cantopop : a study of Hong Kong popular music (1949-1997)]*. Doctoral dissertation, The University of Hong Kong, Hong Kong.

Wong, Kin-Hong 黃健康, Chan, Chi-On 陳志安, & Lau, Kwok-Wai 劉國威. (2001). 信號分析技術在粵曲研究的應用 [Application of signal analysis techniques to the studies of Cantonese opera]. In Sau Yan Chan (Ed.), *粵劇音樂的討論 [Discussions on the Music of Cantonese Opera]* (pp. 105-117). Hong Kong: Research program on Cantonese opera, the Chinese University of Hong Kong.

Wong, Maurice Kuen-Shing. (1982). *Tone Change in Cantonese.* Doctoral dissertation, University of Illinois at Urbana-Champaign.

Wong, Patrick Chun Man. (1999). The effect of downdrift in the production and perception of Cantonese level tone. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, D. Granville & Ashlee C. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (Vol. 3, pp. 2395-2398). San Francisco.

Wong, Patrick Chun Man, & Diehl, Randy L. (2002). How can the lyrics of a song in a tone language be understood. *Psychology of Music, 30*, 202-209.

Wong, Patrick Chun Man, & Diehl, Randy L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language and Hearing Research, 46*(2), 413-421.

Wong, Tak-Sum. (2008). The beginning of merging of the tonal categories B2 and C1 in Hong Kong Cantonese. *Journal of Chinese Linguistics, 36*(1), 155-174.

Wong, Wai Yi Peggy, Chan, Marjorie Kit-Man, & Beckman, Mary E. (2005). An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In Sun-Ah Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 271-300). Oxford: Oxford University Press.

Woo, Nancy Helen. (1969). *Prosody and Phonology.* Doctoral dissertation, M.I.T., Bloomington: Reproduced by the Indiana University Linguistics Club

Xu, Bo Robert, & Mok, Peggy Pik-Ki. (2011). Final rising and global raising in Cantonese intonation. *Proceedings of the Seventeenth International Congress of the Phonetic Sciences*, 2173-2176.

Xu, Yi. (1994). Production and perception of coarticulated tones. *Journal of Acoustical Society of America, 95*(4), 2240-2253.

Xu, Yi. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61-83.

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

Yip, Moira. (1990). *The Tonal Phonology of Chinese*. New York: Garland Publishing INC.

Yip, Moira. (2002). *Tone*. Cambridge: Cambridge University Press.

Yu, Alan C. L. (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology, 24*, 187-214.

Yuan, Jiahong, Shih, Chilin, & Kochanki, Grep P. . (2002). Comparison of declarative and interrogative intonation in Chinese. *Proceedings of the Speech Prosody 2002 Conference*, 711-714.

Yung, Bell. (1989). *Cantonese Opera: Performance as Creative Process*. New York: Cambridge University Press.

Zhang, Caicai, Peng, Gang , & Wang, William Shi-Yuan. (2011). Inter-talker variation as a source of confusion in Cantonese tone perception. *Proceedings of the Seventeenth International Congress of the Phonetic Sciences*, 2276-2279.

Zhang, Ling. (2010). *Downdrift and non-downdrift: Cantonese lexical tones in speaking and singing.* Paper presented at the IACL-18 & NACCL-22, Harvard University.

Zhang, Ling, & Cheung, Kwan-Hin. (2011a). Cantonese lexical tones in speaking and singing. *Proceedings of the Psycholingusitic Representation of Tone Conference*, 32-35.

Zhang, Ling, & Cheung, Kwan-Hin. (2011b). *Experimental studies on Cantonese lexical tones in speaking and singing.* Paper presented at the the 16th International Conference on Cantonese and Yue Dialects, Hong Kong.

Zhang, Ling 張凌. (2006). *廣州話入聲字時長特性分析 [The duration characteristics of entering tones in Cantonese].* Paper presented at the the 7th Chinese Phonetics Seminar and International Forum for Latest Development of Phonetics, Beijing.

Zhang, Ling 張凌. (2008). 廣州話入聲音節在語流中的時長特性 [The duration characteristics of Cantonese entering tones in continuous speech]. *中國語文研究 [Chinese Languages Studies], 25*, 53-62.

Zhu, Xiaonong. (2012). Multiregisters and four levels: A new tonal model. *Journal of Chinese Linguistics, 40*(1-17).

Zhu, Xiaonong 朱曉農. (2005). *上海聲調實驗錄 [An Experimental Study in Shanghai Tones]*. Shanghai: Shanghai Educational Publishing House.

Zong, Fubang 宗福邦. (1964). 關於廣州話陰平調的分化問題 [About the separation of Yin-ping Tone in Cantonese]. *中國語文 [Chinese Language], 133*, 376-389.