The Realization of Tone in Singing in Cantonese and Mandarin

by

Murray Henry Schellenberg

B.A. (spec), The University of Alberta, 1989
M.A. The University of Victoria, 1999

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Linguistics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

February 2013

© Murray Henry Schellenberg, 2013
Abstract

There are many ways that cultures with tone languages may deal with the interaction of linguistic tone and music. Contemporary vocal music in both Mandarin and Cantonese stems from a common source, a new style of Chinese music that developed in the late nineteenth and early twentieth centuries, but the two musics realize linguistic tone differently. This thesis examines experimentally the differences in the phonetic manifestation of tone in singing in both Cantonese and Mandarin as well as examining the comprehensibility of the sung words. One set of experiments asked native speakers to sing songs containing minimal sets by tone. The second set of experiments had native speakers try to recognize the set of words extracted from the songs. Cantonese singers included a rising contour when singing words with rising tones and Cantonese listeners were attuned to this. Mandarin singers did not add in contour information and Mandarin listeners had difficulty recognizing the words out of context. The thesis also expands the discussion of singing in tone languages by examining some of the sociological and political factors which appear to have influenced the ways in which tone is expressed (or not) in these two varieties of Chinese.
Preface


The research was conducted with the approval of the Behavioural Research Ethics Board, as part of the research project entitled “Processing Complex Speech Motor Tasks”, H04-80337 (current version A011), and B04-0337, principal investigator Bryan Gick, co-investigator Janet Werker. My personal certificate for completing the interagency advisory panel on research ethics introductory tutorial for the tri-council policy statement: Ethical conduct for research involving humans (TCPS) was issued on August 22, 2007.
Table of Contents

Abstract ................................................................. ii
Preface ............................................................... iii
Table of Contents ...................................................... iv
List of Tables ........................................................ vii
List of Figures ........................................................ viii
Acknowledgements ..................................................... ix
Dedication ............................................................. x

1 Introduction .......................................................... 1
  1.1 Tone and Tone Languages ...................................... 2
    1.1.1 Describing Tones More Accurately ...................... 5
    1.1.2 Minimal Pairs and Minimal Sets ......................... 9
  1.2 Singing in tone Languages .................................. 11
  1.3 Structural vs. Phonetic Manifestations of Tone in Singing ...... 12
  1.4 Structural Manifestations and Correspondence ............. 13
  1.5 An Experimental Approach to Studying Phonetic Manifestations 16

2 Structural Manifestations of Tone ................................ 19
  2.1 Singing in Tone Languages: A Cross Cultural Comparison ... 21
  2.2 Linguistic Tone and Formal Composition ................... 26
  2.3 Phonemic Divergence: How big a deal is it? ................. 31
  2.4 Speech and Singing ........................................... 36
  2.5 Structural Manifestations of Tone in Mandarin and Cantonese ... 38
  2.6 Phonetic Manifestations of Tone in Mandarin and Cantonese ... 43
  2.7 Conclusions .................................................. 44

3 Differences Between Mandarin and Cantonese Popular Music .... 45
  3.1 Mandarin and Cantonese Popular Music .................... 45
  3.2 Linguistic Explanations for the Differences ............... 46
  3.3 Communicative Intent ........................................ 51
  3.4 Possible Non-linguistic Explanations ....................... 55
  3.5 Nationalism, Identity and Speech Tone .................... 58
  3.6 A New Music for China ...................................... 58
    3.6.1 Language Reform ...................................... 59
    3.6.2 Music Reform ......................................... 62
    3.6.3 Text Setting and Tone ................................ 65
  3.7 Cantopop ...................................................... 68
    3.7.1 Hong Kong Identity ................................... 69
    3.7.2 Development of Cantopop .............................. 71
    3.7.3 Cantopop, Language and Identity .................... 72
  3.8 Conclusions ................................................ 76
6.3.1 Participants................................................................. 128
6.3.2 Stimuli ................................................................. 128
6.3.3 Procedure .............................................................. 129
6.3.4 Results ................................................................. 131
6.3.5 High versus Low Notes ............................................. 134
6.3.6 Discussion ............................................................. 134

6.4 Conclusions .............................................................. 136

7 Conclusions ............................................................... 139
7.1 Tone and Singing in Cantonese and Mandarin ................. 139
7.2 Perception of Tone in Singing ...................................... 140
7.3 Cross-cultural Examination of Tone in Singing ............... 141
7.4 Linguistic and Cultural Influences on Tone and Singing .... 142
7.5 Strategies for Representing Tone in Singing ................... 143
7.6 Areas for future Research ............................................ 144

References ..................................................................... 145
List of Tables

Table 1.1: Various descriptive systems for Mandarin tones .................. 4  
Table 1.2: The tones of Cantonese ........................................ 7  
Table 2.1: Tone languages where the interactions between speech and  
singing have been discussed/studies .................................... 20  
Table 2.2: Percentage of correspondence between speech melody and sung  
 melody ................................................................. 22  
Table 3.1: Three continua of speech melody/song melody correspondence ... 53  
Table 4.1: The tones of Cantonese .......................................... 78  
Table 4.2: Full statistical results for slope comparisons .................... 89  
Table 4.3: The tones of the numerals one through nine in Cantonese ....... 94  
Table 5.1: Target words (second syllable only) ................................ 104  
Table 5.2: Standard deviations (Hz) for mean F0 levels of normalized pitch  
 contours ............................................................... 111  
Table 5.3: Means of raw duration scores .................................... 116  
Table 6.1: Confusion matrices for both sung and spoken conditions of the  
 Mandarin study .......................................................... 122  
Table 6.2: Post hoc cell-wise contributions for spoken tones (Mandarin) ..... 125  
Table 6.3: Frequency counts from the Academia Sinica Corpus ............ 126  
Table 6.4: Percentage occurrences of tones in Mandarin .................... 127  
Table 6.5: Confusion matrices for both sung and spoken conditions of the  
 Cantonese study .......................................................... 130  
Table 6.6: Post hoc cell-wise contributions for the sung and spoken  
conditions (Cantonese) ..................................................... 132  
Table 6.7: Confusion matrix for tone and high or low note ................... 134
List of Figures

Figure 1.1: Pitch contours for the sentences “You’re going?” and “You’re going” .......................................................... 3
Figure 1.2: The Chao (1930) system of numbers showing the four tones of Mandarin ......................................................... 6
Figure 1.3: Pitch contours of the six Cantonese tones .................. 8
Figure 2.1: Percentage correspondence between speech melody and sung melody in 9 languages (11 studies) ......................... 21
Figure 2.2: Examples 3 and 5 from Mark and Li (1966, p. 178) .......... 30
Figure 2.3: Transcription of an excerpt from “Heaven and Earth” (天與地) by Paul Wong (黃貫中) .................................................. 32
Figure 4.1: Score of the Cantonese song .................................... 83
Figure 4.2: Mean F0 values for all six tones as sung on the high note and low note with lines showing extrapolated slopes ................ 87
Figure 4.3: Mean F0 values for rising tones, level tones and falling tone on the high and low notes with lines showing extrapolated slopes . 88
Figure 4.4: Subject-by-subject graphs of rising vs. non-rising tones .................. 90
Figure 4.5: Mean F0 tracings for the numerals for rising tones vs. level tones. 96
Figure 4.6: Subject-by-subject graphs of rising vs. non-rising tones – numerals ................................................................. 97
Figure 4.7: Mean durations of tones on sung syllable [si] ..................100
Figure 5.1: Score of the Mandarin song .................................. 107
Figure 5.2: Mean tone contours with normalized duration ................ 110
Figure 5.3: Mean contours of Mandarin tones condensed into two groups . 112
Figure 5.4: Subject-by-subject graphs for 4 tones ....................... 113
Figure 5.5: Subject by subject graphs for tones conflated into two registers . 115
Figure 6.1: Identification rates for each tone (Mandarin) ................ 123
Figure 6.2: Identification rates for each tone (Cantonese) ................. 131
Figure 6.3: Identification rates for tones by note sung (Cantonese) ....... 133
Acknowledgements

I would like to acknowledge the help and support of the many people who have helped and supported me in this endeavour. My committee members, Bryan Gick, Molly Babel and Nathan Hesselink; the university examiners, Michael Tenzer and Gunnar Hansson; and the external examiner D. Robert Ladd who have all given me so much valuable advice and feedback. My cohort members, Heather Bliss, Raphael Girard, Beth Stelle and Anita Szakay who made going back to school so enjoyable; Robert Xu in Hong Kong for all his help running participants there; the many scholars who provided feedback at conference presentations of portions of this thesis and most of all to my husband, Soon Hung Loo, who was more excited about my going back to school than I was; Thank you all so much.
To Loo Soon Hung
Chapter One

Introduction

It is conservatively estimated that between fifty and sixty percent of the world’s languages are tone languages – languages where pitch contributes to the meaning of words (Yip 2002). Tone in these languages is phonemic, so that changing the pitch (or pitch contour) of a word can drastically change the meaning of the word.

Since pitch is so closely tied to meaning in these languages and pitch is also one of the main components of music, the interaction of speech melody and song melody in tone languages has fascinated scholars for years. Modern vocal music in Mandarin and Cantonese provides a particularly interesting context in which to examine tone and singing. In the late nineteenth and early twentieth century there was a very influential movement to reform music in China. This led to a new style of music which has dominated much of Chinese popular music since (Mittler 1997, Jones 2001). In the 1960s and 70s, this style of music, which had been predominately Mandarin, split off into a separate Cantonese music (Choi 1990, Wong 2001). What is particularly interesting is that Mandarin music and Cantonese music represent tone very differently. This thesis examines some
of those differences – specifically, differences in the realization of tone as produced by the singer during performance.

The remainder of this chapter provides general background and explanatory information on linguistic tone, the manifestation of tone in song and the role of laboratory experiments in this kind of research. Chapter Two takes a cross-linguistic look at singing in tone languages and examines the extent of the range of correspondence between speech melody and song melody in a variety of languages, something which has not been done before. Chapter Three discusses potential reasons for the differences. Chapters Four and Five present experimental studies that examine the extent of phonetic realization of tone by singers during singing in Cantonese (Chapter Four) and Mandarin (Chapter Five). Chapter Six presents experiments that test the ability of listeners to recognize the sung forms.

1.1 Tone and Tone Languages

All languages use pitch in various ways. In English, for example, questions can often be differentiated from statements simply by the pitch contour of the sentence (the intonation pattern) — “You’re going?” with a rising intonation pattern is a question as opposed to “You’re going” with a falling intonation

---

1 This section provides an introduction to tone in language for non-specialists.
pattern which is a statement of fact. This can be easily seen in Figure 1.1 which shows the pitch contours for these two sentences as spoken by the author.

A *tone language* is a language where the pitch (or sometimes the pitch contour) of the syllables contributes to the meaning of a word. *Tone*, as used by linguists, refers to the pitches or pitch contours used to differentiate the meanings of words in these languages. Ekwueme (1974) provides an excellent example from Igbo, a Kwa language of south-eastern Nigeria, of how just two tones can affect the meaning of a string of segments (consonants or vowels).

1. / ísí / (high; high) ‘head’
   / ísí / (high; low) ‘smell’
   / ísí / (low; low) ‘blindness’
   / ísí / (low; high) ‘you go from [somewhere] to …’
   (Ekwueme 1974, p. 338)
In Example 1, each word has the same pattern of segments – /i/ + /s/ + /i/ – but in addition there are two tones, high (represented here by an acute accent) and low (represented by a grave accent). The tone patterns are also given here in parentheses and the meaning is given in single quotes. As can be seen in the example, depending on which tone is assigned to each syllable, four possible words can be produced. To a speaker of Igbo or any other tone language, a word is not complete without its particular tone pattern. Changing the pattern of the pitches can be as disconcerting to a speaker of Igbo as, say, changing the ‘s’ sound ([s]) at the beginning of the word ‘sin’ to a ‘t’ sound, [t], would be for a speaker of English.

Tones, as previously stated, may sometimes be conveyed by varying pitch contours as well as pitches; some languages have tones that involve changing pitch over the course of the syllable. Mandarin Chinese is a good example of a language with this type of tone. Mandarin has four different tones: high level; rising; falling-rising; and falling. Each pitch contour is completed over the course of a syllable.

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>LABEL</th>
<th>CHAO NUMBERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>high level</td>
<td>Tone 1</td>
<td>55</td>
</tr>
<tr>
<td>rising</td>
<td>Tone 2</td>
<td>35</td>
</tr>
<tr>
<td>falling-rising</td>
<td>Tone 3</td>
<td>214</td>
</tr>
<tr>
<td>falling</td>
<td>Tone 4</td>
<td>51</td>
</tr>
</tbody>
</table>

Table 1.1: Various descriptive systems for Mandarin tones.
It is common for scholars who study a tone language to assign labels (usually numbers) to the tones for easy reference. For example, the Mandarin tones are usually labelled as Tone 1, Tone 2, Tone 3 and Tone 4 as shown in the second column of Table 1.1.

1.1.1 Describing Tones More Accurately – The Chao Numbers

Labels such as “falling” or “rising” are often not sufficient to describe the tones in a language. Based on his own observations, Chao (1930) proposed a system to describe Mandarin tones, shown in the third column of Table 1.1, which is still used today and which has been subsequently used to describe tone systems all over Asia. He proposed five levels delimited by roughly equidistant musical intervals as shown in Figure 1.2. The lowest tone is labelled as 1 and the highest as 5. The abstract use of the numbers one through five instead of specific notes or pitches is deliberate. Tones are always relative, they are not absolute. Each individual speaker naturally uses different pitches for tones depending on the speaking range of his/her voice; but also, over the course of an utterance, a given tone may change its actual pitch due, for example, to the natural tendency for speech to fall in pitch over the course of an utterance (a phenomenon called downdrift [Connell 2001]). Other factors such as a change in emotion or health

---

2 Subsequent instrumental analysis (e.g. Howie 1971, Xu 1998) has shown that Chao’s original observations were remarkably accurate.
on the part of the speaker can also affect the pitch of the tones (Li et al. 2011, Ciocca et al. 2004). A listener understands which tone is being used from the larger context.

The shape of a tone is expressed in the Chao system as a sequence of numbers corresponding to the initial and final pitch level (and also the middle in the cases of tones which have combinations of falls and rises such as Mandarin tone 3). These numbers have absolutely no relation to the labels in the second column of Table 1.1. According to this system then, the Mandarin tones can be expressed as a series of numbers as shown in the third column of Table 1.1.
Table 1.2: The tones of Cantonese.

<table>
<thead>
<tr>
<th>TONE</th>
<th>DESCRIPTION</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High level</td>
<td>師 si55 – ‘teacher’</td>
</tr>
<tr>
<td>2</td>
<td>High rising</td>
<td>史 si35 – ‘history’</td>
</tr>
<tr>
<td>3</td>
<td>Mid level</td>
<td>嗑 si33 – ‘to try’</td>
</tr>
<tr>
<td>4</td>
<td>Low falling</td>
<td>時 si21 – ‘time’</td>
</tr>
<tr>
<td>5</td>
<td>Low rising</td>
<td>市 si23 – ‘market’</td>
</tr>
<tr>
<td>6</td>
<td>Low level</td>
<td>鼓 si22 – ‘yes’</td>
</tr>
</tbody>
</table>

This system is very useful in languages with more complex tone systems.

Cantonese, for example, has a tone system that involves three relative levels (also known as registers) as well as three contours. However, two of the contour tones are rising tones. The tones are laid out in Table 1.2. The Chao numbering system can provide a very quick and easy comparison between the two rising tones (tones 2 and 5). Tone 2 is described as 35 and tone 5 as 23. From this it is easy to see that the register is important for rising tones as well as for level tones in Cantonese. Figure 1.3 provides actual pitch tracings of the Cantonese tones as produced by a male native speaker of Cantonese. In the pitch tracings tone 5 can be seen to start at the same level as tone 6 and end at the level of tone 3. This particular speaker actually starts tone 2 a little lower – at the same level as tones 5 and 6 – but rises to the level of tone 1.

Some scholars, following the classical Chinese system of describing tones, make a further distinction in Cantonese and subdivide the level tones into short and long variants bringing the total number of tones up to nine. The short or
entering tones are briefer level tones that occur only – and always – on syllables that end with specific consonants ([p, t, k]) so for most phonetic research they are viewed simply as predictable variants of the long tones. Because they derive from different historical sources, however, the difference can become very important in historical or comparative research. Consequently, descriptions of the Cantonese tone system can vary in the number of tones depending on the research being conducted. In situations where it is necessary to define nine tones the Chao numbers can still be used. The three entering tones, all level, are described using a single number to reflect their brevity:
For the purposes of this thesis the historical distinction between the entering and non-entering tones is not relevant and the entering tones have been subsumed into their non-entering counterparts.

A word of caution: there is not a single system for representing tones. Linguists who work on tone languages of South America, for example, use a numbering system very similar to that of Chao (1930) but with the numbers reversed. They use 1 as the highest level and 5 as the lowest. It is always best to explain the system being used. As this thesis focuses on two varieties of Chinese I will use the Chao system, which is the standard for linguists studying Chinese.

1.1.2 Minimal Pairs and Minimal Sets

In the course of studying the phonetics of tone it is often necessary to isolate the effects of tone in a language. It is desirable to limit the potential influence of non-tonal aspects of language (this is discussed in greater detail below) and, to this end, linguists spend a lot of time describing minimal pairs or minimal sets. A minimal pair consists of two words that are “identical in form except for a contrast in one phoneme, occurring in the same position” (Yule 2010, p. 44). The difference does not have to be tonal: ‘pat’ and ‘bat’ are a minimal pair in English.
that differs segmentally only in the initial consonant (in fact, only in the voicing of the initial consonant). In the Igbo example above /ìsì/ ‘head’ and /ìsì/ ‘smell’ are minimal pairs as they have only one difference: the tone of the second syllable. Note, however, that /ìsì/ ‘head’ and /ìsì/ ‘blindness’ have two differences (the tone on syllable one and the tone on syllable two) and are therefore not a minimal pair.

A minimal set is simply an expansion of a minimal pair. In Mandarin, for example, a minimal set of the Mandarin tones would be a set of words that all have the same segmental sequence and differ from one another solely on the basis of tone. The classic example is on the syllable [ma].

3. 媽 ma55 ‘mother’
    麻 ma35 ‘hemp’
    馬 ma214 ‘horse’
    罵 ma51 ‘scold’

The examples on the syllable [si] given in Table 1.2 constitute a minimal set by tone for Cantonese. The presence of minimal pairs and minimal sets like these is one of the most compelling arguments available for the presence of lexical tone in a language. Speakers differentiate these sequences and the only difference, in these cases, is the tone. Therefore tone must play a role in the lexical structure of the words.
1.2 Singing in Tone Languages

On the surface, the idea that speech melodies in a tone language dictate song melodies seems a fairly reasonable assumption: if changing the pitch can change the meaning, the song melody must, perforce, preserve the meaning and, consequently, the speech melody. This assumption has shown up frequently in the literature over the years. Schneider (1961, p.204), for example, states that “speechtone [sic] and musical tone must be definitely correlated.” Rycroft (1959, p. 28) says that “the setting of words to music in a ‘tone language’ either places limitations upon melodic freedom … or else makes word selection a more exacting matter.” Jones (1959) and Schneider (1942, 1950, 1961) devote most of their papers to trying to provide prosodic explanations for differences between speech melody and song melody in specific languages.

Conversely, scholars like Ward (1932), Bright (1957) and Mark and Li (1966) have not assumed an essential connection between speech melody and music melody. As Mark and Li (1966, p. 167) put it: “if textual intelligibility is of such overriding concern and entirely dependent upon tonal inflection, one might ask, why do people bother to sing at all?” The assumption that speech melody and tone melody must be closely allied is challenged in Chapter Two.

Whatever the assumptions, work examining the manifestation of tone in singing relies on comparisons of speech melody and tone melody. The following
sections lay out the different ways that speech tone may be manifested and the
different methodologies that have been used to examine those manifestations.

1.3 **Structural vs. Phonetic Manifestations of Tone in Singing**

In some of their recent work on the Dinka language, Ladd and Remijsen (in
progress) have proposed a classification for the possible manifestations of lexical
tone in singing. They divide the possibilities into *structural* manifestations – those
which are prescribed by the musical melody – and *phonetic* manifestations –
related characteristics of tone that are included during performance but not
“required” by the prescribed music. The phonetic manifestations are further
subdivided into *compensation* – characteristics of lexical tone which may be
deliberately maintained or even exaggerated by the singer; and *residue* –
manifestations of secondary characteristics of tone which are transferred over in
a more reflexive or automatic way.

*Structural* in this context refers to the musical structure, the form of the
music as created by the composer/creator of the song as opposed to the linguistic
structure. This is the sense of the word that will be used in this thesis. Structural
manifestations are usually quantified in terms of correspondence between the
speech melody and the song melody. Various methods that have been used are
discussed in the next section.
1.4 Structural Manifestations and Correspondence

Various metrics have been used to try and quantify structural correspondence between speech melodies and song melodies on a structural level. This section describes four methods which have been used in the literature to quantify correspondence: native speaker observation, note shape/tone shape comparison, F0 (pitch) comparison and contour comparison.

1. **Native speaker observation** – Chao 1924, 1956, Ekwueme 1974, Mugovhani 2007

   Native speakers of a language listen to songs and provide opinions as to the degree of correspondence. It has not been clear, under this metric, what specific factors are driving the listener’s judgements.

2. **Note shape/ tone shape comparison** – Chan 1987a, Wee 2007

   This has been used but only in conjunction with other methods. The shape of the tone, for example a falling tone, is compared with the shape of the musical contour assigned to that syllable. This only works for languages that have a variety of contour tones and assumes a fairly melismatic\(^3\) melody.

   Correspondence in these situations has been interpreted as a direct match in the shapes but also, as in Chan (1987a), a melisma that moves first down and then back up has been interpreted to represent a falling-rising tone (based on the shape of the whole melisma), a falling tone (based on the first part of the melisma).

---

\(^3\) A *melisma* is a situation where multiple musical notes are assigned to a single syllable.
or a rising tone (based on the second part of the melisma). A mismatch is either a direct opposition of movement (one melody moves down and the other moves up) or a situation where one melody remains level and the other moves.

3. *F0* (pitch) comparison – Bright 1957, Yung 1983, Lau 2010 (analyses 1 and 3)

   This method looks for absolute representations of tone within the music. This metric requires the examiner to divide the range of notes in the song melody into levels that reflect the tone system of the language. The tone of each syllable is compared to the fundamental frequency of the note(s) on which it is set to see if there is a correspondence between the register of the tone and the relative placement of the musical note in the overall range of the musical melody. A match under this metric occurs when a low tone, for example, is set to a note in the lowest portion of the melodic range and a mismatch is when it falls outside of the predetermined range. Studies which have found matches using this metric (Yung 1983, Lau 2010) have observed that a limited range of notes usually corresponds to a particular tone rather than a single note and that there is usually some overlap.


---

4 Fundamental frequency or *F0* is the acoustic correlate of pitch, which is an auditory-perceptual phenomenon.
In this metric the comparison is not between syllables and notes but compares the transitions between successive syllables and successive notes. This method is based on the idea that tones are relative and that the expression of tone in speech is always in relation to what came before it and what comes after it. This method compares an actual spoken phrase with its sung correspondent. Each transition – the movement from one syllable/note to the next – is categorized as rising, falling or remaining level and a match is when the transition from one syllable to the next moves in the same direction as the transition from the note on which the first syllable is set to the note on which the second syllable is set. A mismatch is when the transitions do not go in the same direction.

There is a slightly different variation of this method which has also been used (Ward 1932, Agawu 1988, 1995). In this variation, the sung contour is compared with a simplified tone contour – a representation of tones on a set number of levels. The studies where this method has been used all involve tone systems that have three level tones. The tone contour is derived not from a spoken rendition of the words but from a theoretical representation of the tones on three unchanging levels. The main difference from the first variation is that the theoretically based contour does not include the natural downturns of speech.
Studies of structural manifestations for individual languages exist in sufficient numbers that a preliminary comparison across languages is possible. This will be done in Chapter Two. Phonetic manifestations, conversely, are relatively understudied and comprise the main focus of this thesis.

1.5 An Experimental Approach to studying Phonetic Manifestations

Phonetic manifestations have received much less study than structural manifestations. The main focus of this thesis is an experimental examination of phonetic manifestations in sung Cantonese and Mandarin. There are, of course, numerous ways in which such an analysis can be carried out, all of which have their advantages and shortcomings. The choice to adopt a designed laboratory experiment was made in this case for a number of reasons. As the focus of the investigation was the phonetic manifestation of tone, it was necessary to eliminate other potential influences on fundamental frequency (F0) and duration. Three potential phonetic confounds are the intrinsic F0 of vowels, the influences of onset consonants on vowel F0 and the intrinsic duration of vowels.

The intrinsic F0 of vowels is the natural tendency in spoken language for high vowels like [i] and [u] to have a higher F0 than low vowels such as [a], a phenomenon first observed in German by E. A. Meyer in 1896-97 (Whalen and Levitt (1995) provide a comprehensive history). Intrinsic F0 differences have been found in a wide variety of languages; Whalen and Levitt (1995) survey 58
studies involving 31 languages, all of which exhibit this phenomenon. Their analysis includes ten tone languages in which intrinsic F0 is observed as a difference within individual tones; a high tone [i] or [u] will have a higher F0 than a high tone [a].

Similarly, the consonant that appears in front of a vowel can have a noticeable effect on the F0 of the vowel. Early work on this was carried out by House and Fairbanks (1953) but the topic has continued to receive considerable attention. As an example, vowels after voiceless stops are higher in F0 than those following voiced stops. Hombert et al. (1979) analyse the pitch contours associated with different onsets showing that a consonant can strongly influence the shape of the subsequent vowel’s contour and discuss the role of this phenomenon in the development of tones.

A third potential confound is intrinsic vowel duration. Lehiste (1970) found that low vowels like [a] are intrinsically longer than high vowels like [i] or [u]. Like intrinsic F0, she found intrinsic vowel duration to hold across languages, including tone languages.

There are also potential confounding influences from the music. The basic physiology of singing affects the F0 contour of notes so performance practices such as portamenti (sliding between pitches) influence the shape of the following note (Fujisaki 1981).
In order to counter these potential confounds it is useful to look at a minimal set that occurs with the same melody. This ensures that issues such as those discussed above will be held constant and that any differences may be more confidently assumed to be due to the one factor that is different: tone.

In terms of singing and tones, it can be seen that in an analysis of the phonetic manifestation of tone the structure of the syllable and the musical context need to be controlled. If an analysis were to be carried out on random syllables from a song any phonetic differences noted may be due to tone, but they may also be due to other factors. It is possible, of course, to counter these confounds in other ways. Ladd and Remijsen (in progress), for example, use large numbers of naturally occurring samples with great success. The extremely high number of tokens allows for the potential confounds to be “averaged out,” as it were. Circumstances often dictate the choice of methodology. In the studies for this thesis there happened to be a fairly large number of participants who could all learn and perform the same song readily available. The musical styles under investigation also allowed for the creation of special songs that would pass as “natural” songs. The experiments looking at the phonetic manifestations of tone in Cantonese and Mandarin start in Chapter Four. Chapter Two first takes a broader, cross-linguistic look at how linguistic tones manifest structurally in music composition/creation.
Chapter Two

Structural Manifestations of Tone

Languages are not consistent in either the extent or the manner in which they express tone in singing. This chapter takes a cross-linguistic look at structural manifestations of tone in singing, representations of tone prescribed within the structure of the music, by collecting together the evidence from studies looking at various languages and examining the variety of ways in which composers deal with the inter-relationship between lexical tone and singing when putting a song together. Language is not the sole determinant of music in tone languages. Rather, composers/song creators seem to accommodate language when it is convenient but they are willing and able to override linguistic requirements when needed. Different cultures, however, seem to make different demands on their composers.

The next section of this chapter provides a preliminary cross-cultural comparison of the degree to which different cultures parallel speech melody and song melody structurally. Section 2.2 diverges slightly and looks at the influence of tones on the formal composition process. Section 2.3 looks briefly at the impact of deviating from the prescribed speech melody, and the final sections look more closely at manifestations of tone in Cantonese and Mandarin.
<table>
<thead>
<tr>
<th>Language</th>
<th>Country/Region</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burmese</td>
<td>Burma</td>
<td>Williamson (1981)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wong and Diehl (2002), Ho (2006),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cheung (2007), Lau (2010),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zhang (2011, to appear)</td>
</tr>
<tr>
<td>Lushai (Mizo)</td>
<td>India</td>
<td>Bright (1957)</td>
</tr>
<tr>
<td>Mandarin</td>
<td>China</td>
<td>Chao (1924, 1956), Schneider (1950),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chan (1987a), Stock (1999),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wee (2007)</td>
</tr>
<tr>
<td>Tai Phake</td>
<td>India</td>
<td>Morey (2010)</td>
</tr>
<tr>
<td>Thai</td>
<td>Thailand</td>
<td>List (1961), Mendenhall (1975),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Saurman (2006)</td>
</tr>
<tr>
<td>Wu-Ming Tai</td>
<td>China</td>
<td>Mark and Li (1966)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixtec</td>
<td>Mexico</td>
<td>Pike (1946)</td>
</tr>
<tr>
<td>Navaho</td>
<td>U.S.A.</td>
<td>Herzog (1934)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anyi</td>
<td>Côte d’Ivoire</td>
<td>Gibbon et al. (2011)</td>
</tr>
<tr>
<td>Dinka</td>
<td>South Sudan</td>
<td>Ladd and Remijsen (in progress)</td>
</tr>
<tr>
<td>Ewe</td>
<td>Ghana</td>
<td>Schneider (1942), Jones (1959),</td>
</tr>
<tr>
<td>Fanti</td>
<td>Ghana</td>
<td>Ward (1932)</td>
</tr>
<tr>
<td>Hausa</td>
<td>Nigeria</td>
<td>Richards (1972), Leben (1983)</td>
</tr>
<tr>
<td>Igbo</td>
<td>Nigeria</td>
<td>Ekueme (1974)</td>
</tr>
<tr>
<td>Shona</td>
<td>Zimbabwe</td>
<td>Schellenberg (2009)</td>
</tr>
<tr>
<td>Venda</td>
<td>South Africa</td>
<td>Blacking (1967)</td>
</tr>
<tr>
<td>Xhosa</td>
<td>South Africa</td>
<td>Starke (1930)</td>
</tr>
<tr>
<td>Zulu</td>
<td>South Africa</td>
<td>Rycroft (1959, 1979)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duna Pikono</td>
<td>Papua New Guinea</td>
<td>Solis (2010)</td>
</tr>
</tbody>
</table>

Table 2.1: Tone languages where the interaction between speech and singing have been discussed/studied. The second column indicates the general geographic region where the language is spoken and the third column indicates the sources.
2.1 Singing in Tone Languages: A Cross Cultural Comparison

As can be seen in Table 2.1, the interrelationship between speech melody and song melody has been examined for a quite varied group of languages. With one exception, however, each study examines a single language (the exception is Chan 1987a, which looks at both Mandarin and Cantonese). A cross-cultural comparison has not been done. Are there patterns which show up across all languages?

If song melodies of musical cultures with tone languages are, in fact, constrained by the speech melody, the prediction would be that all cultures with tone languages should show very high, if not absolute correspondence between the two melodies.

![Bar chart showing percentage correspondence between speech melody and sung melody in 9 languages (11 studies). Correspondence by chance is 33.3% indicated by the dashed line. Details about individual studies are given in Table 2.1.](image)

Figure 2.1: Percentage correspondence between speech melody and sung melody in 9 languages (11 studies). Correspondence by chance is 33.3% indicated by the dashed line. Details about individual studies are given in Table 2.1.
Figure 2.1 shows the percentage of correspondence for nine different languages from eleven studies for which statistical results were either given or calculable. The different colours represent general geographical regions: grey is used for languages from Africa; black for languages from eastern Asia and white for India. The range of correspondence is extremely wide: from 48% for Kalami to 92% for both Cantonese and Zulu. In all cases, this is well above the 33.3% chance (using chi-squared tests, all have p ≤ 0.01) but the range suggests that language is not driving the composition of songs in all tone languages.

The studies used for this analysis are a subset of the academic studies listed in Table 2.1; details are provided in Table 2.2. The comparison was limited

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

Table 2.2: Proportions of correspondence between speech melody and sung melody. Results for languages marked with an asterisk have been calculated post hoc from data published in the papers cited.

A chi-squared (goodness of fit) test compares observed frequencies (in this case the number of times the melodies correspond) with a theoretical model (here a mathematical calculation of chance – 33.3%). The p-value (p for ‘probability’) indicates the likelihood that the result is purely by chance. A p-value of less than 0.01 means that there is less than 1% probability that the result is random.
to this subset of studies as they were studies that either presented results
calculated using the contour comparison method described in Section 1.4 in the
previous chapter – direct comparisons of the transitions in the contour of the
spoken version of song lyrics and the contour of the sung version – or provided
representations from which such comparisons could be made. Studies that
employed other methods of comparison were excluded because, although these
are all valid ways to examine tone in singing, it was felt that comparing across
methodologies would introduce too many variables into the comparison. It must
be noted that in this smaller sample of studies used, the results come from
extremely different sample sizes and somewhat differing statistical methods
have been used to arrive at the results shown. The comparison should only be
taken as a very rough guide.

The language names marked with an asterisk in Table 2.2 are those
languages where the studies did not provide statistical results. In these instances,
the results have been calculated from graphic representations of the melodies
provided in the papers. The analysis for those data follows the methodology
explained in Schellenberg (2009) and tracks correspondence between contours
based on transitions from one frequency to the next.

This methodology is as follows: from any given syllable, there are three
possible transitions or directions in which the fundamental frequency can move:
up, down or across (remain the same). It is this sequence of transitions which creates the contour. The transitions are coded for directionality: up, down or level. If the second syllable is within 1.5 Hz of the first it is coded as being level; this number was chosen as the frequency differential limen for the range in which most songs are sung is approximately 1 – 2 Hz (Durrant and Lovrinic 1984, p. 225). Pitch discrimination threshold for an average adult is about 11.0 cents, or 1 Hz in the 146-147 Hz range (Seashore 1967, p. 54-55). The direction of the corresponding contours between each syllable in the sung and spoken versions is then compared. If both the contours are the same then the correspondence is labelled as parallel. If one goes up and the other goes down they are labelled as opposing and if one stays the same and the other does not they are labelled as non-opposing. In all cases, potential bias is towards correspondence. If there happens to be a melisma in the sung version where a single syllable is set to more than one note of music all the transitions associated with that syllable are noted and if one of them matches the spoken version it is counted as parallel.

The column in Table 2.2 labelled “parallel” gives the proportion of occurrences where the transitions are parallel between the speech melody and song melody. For example, if the speech melody goes up, so does the song melody. One possible way to provide a greater variety of structural representation of tone might be to only disallow transitions that go in opposite
direction. That is to say, a transition that goes down in speech could be represented by either a falling transition or a level transition in the music but not by a rising transition. However, when both parallel and non-opposing transitions are grouped together as shown in the column in Table 2.1 labelled “not opposing”, not all of the languages look the same. As a strategy this seems to hold across some of the languages, but certainly not all. It is noteworthy that in every language there are at least some cases where speech and song melodies move in opposing directions; no language in this sample shows correspondence of 100%.

There are also a number of papers in which correspondence has been studied but which report only general, non-statistical conclusions and for which there is insufficient data given to calculate numerical results. Of these, only one, Ekwueme’s (1974) analysis of Igbo (a Niger-Congo language spoken in Nigeria), reports a high level of correspondence. Fanti (Ward 1932), Mixtec (an Oto-Manguean language spoken in Mexico – Pike 1946, 1948), Lushai (Bright 1957) and Venda (a Bantu language spoken in South Africa – Blacking 1967) are all reported as having relatively low levels of correspondence. Herzog (1934) reports for Navaho (an Athabaskan language spoken in the south-western United States) that one song type (gambling songs) has high correspondence and another (healing songs) has extremely low correspondence.
2.2 Linguistic Tone and Formal Composition

There are some composers who have felt it imperative to follow the speech melody when writing vocal music in tone languages. Agawu (1984) gives an analysis of the work of Ghanaian composer Ephraim Amu (1899-1995) which examines the influence of linguistic factors in Amu’s compositions. Amu received formal musical training in the Western European tradition (studying music theory at the Royal College of Music in London) but had also, from the earliest part of his career, been strongly influenced by traditional Ghanaian music and performance. Agawu, after examining the (varying) influence of speech melody on song melody in traditional Ghanaian music, tracks an ever increasing influence of speech melody on sung melodies through Amu’s career noting compositional developments (increasing polyphony and less use of “vertical” harmony,\(^6\) for example) that allowed Amu to mirror speech melody in his songs. In a later published interview (Agawu and Amu 1987), Amu is adamant about the need to maintain not only speech melody in music but also speech rhythm, remarking that such maintenance “makes the song easy to learn, and people can understand you when you sing. They have no difficulty in getting the words” (Agawu and Amu 1987, p. 57). Agawu, however, points out an interesting anomaly. In the course of the interview, Amu provides examples

\(^6\) Vertical harmony refers to harmony where there are clear, simultaneous chord progressions that carry through all voices as opposed to polyphony which has multiple interwoven melodies.
from indigenous Ewe and Twi songs which he claims show strict correspondence between speech melody and sung melody. Agawu, also a speaker of these languages, comments in a note that Amu’s sung examples “did not do what he intended them to do … In many cases, the contrary relationship was evident” (Agawu and Amu 1987, p. 63, note 10). Agawu (loc. cit.) goes on to point out that this, in itself, is not a serious problem as “musical constraints often exert a more fundamental effect on musical structure than speech tones, per se”, something that had been shown quite clearly for these languages in the earlier paper (Agawu 1984).

In 1968, linguist and ethnomusicologist David Rycroft won the competition to write the national anthem for the then-emerging nation of Swaziland. Rycroft (1970) is a description (and self-analysis) by the composer of the process involved in writing the winning entry. Rycroft is well known for his work on the interaction between speech melody and sung melody so it is not surprising that he includes a detailed section on how the speech tones of the texts⁷ influenced his composition. He was very conscious of mirroring speech melodies throughout the song and included linguistic features such as on-glides

---

⁷ There were, in fact, two separate texts from which the contestants could choose; the government having been unable to decide. The same texts were supplied to all contestants; Rycroft decided to set and submit both texts, only one of which was short-listed.
triggered by depressor consonants\(^8\) marked with grace notes in the score for the benefit of non-native speakers. He also included quite extensive performance notes with the piece which are almost exclusively related to linguistic aspects of the piece.

It is interesting to note that while both Amu and Rycroft were African born, their formal musical training was very much in the Western European tradition. They also both had a strong academic awareness of the languages in which they composed and were very familiar with the idea of tone as an abstract concept in language, so this awareness may well have coloured their composition techniques. Mugovhani (2007) analyses the works of six Venda choral composers. He includes as part of his analyses whether or not the composer has followed the tone patterns of spoken Venda; none of the composers he examines appear to pay much attention to this aspect of the Venda language. He attributes this in large part to the use of a Western musical style.

\(^8\) A depressor consonant is a consonant that affects the speech tone in a syllable which follows it; usually causing a high tone to appear as a low tone. In Zulu “high-toned syllables beginning with such consonants commence with a brief rising on-glide” (Rycroft 1979, p. 306).
The traditional composition techniques employed in Beijing Opera also involve an academic awareness of tone (Liu 1974). The song form used in Beijing Opera is based on the tz’u poetic form which “features asymmetrical verses of long and short lines that are composed of various numbers and combinations of even and oblique tones following prescribed patterns” (Liu 1974, p. 80). ‘Even’ and ‘oblique’ tones refer to traditional groupings of tone patterns that were deemed similar; ‘even’ is usually associated with the level tone and ‘oblique’ is associated with all other tones. There are, consequently, two kinds of “rhyming” in the tz’u form: similar vowel sounds and similar tone patterns. The poetic structure requires different rhymes in different specific locations. The expectation is that the sung melody will always reflect the spoken melody, but Stock’s (1999) analysis shows that this expectation is frequently violated in performance.

This style of poetic structure was borrowed and used as a song form by speakers of the Wu-Ming dialect of Tai (Mark and Li 1966), but there is an interesting twist in that the different songs examined by Mark and Li were all sung to variants of the same melody. They found that while the song melody was usually retained, at key points in the text where tone was important the

---

9 There is controversy over whether to refer to this form of theatre as “Peking Opera” or “Beijing Opera”. Guy (1995) addresses this issue in detail and argues that on the basis of historical precedent and statistical prevalence “Peking Opera” has become the de facto English term. Reactions from readers of earlier versions of this thesis has convinced me that this is not so and that “Peking Opera” is regarded by native English speakers as archaic. I have chosen, therefore, to use the term “Beijing Opera” which is used by scholars such as Jonathan Stock (p.c.).
melody frequently (but not always) diverged to accommodate the text. This can be seen in Figure 2.2 which presents two of the examples. These phrases are sung to the same melody but at the end of the phrase, in a position where poetic constraints require tone-based rhyming, the music changes to reflect the tonal contour of the words. Note, however, that the contour of the music does not change significantly to accommodate the tone of the second word of each phrase; a position where there are no poetic constraints.

Other scholars have observed that linguistically significant points may be accommodated in sung melodies. Wee (2007) suggests that speech melodies and sung melodies in Mandarin Chinese correspond at positions of metrical prominence; that is, that syllables on the most prominent beats of a bar will also be linguistically salient syllables and that the melodies will match at those points. Richards (1972) found a similar tendency for higher levels of correspondence in phrase-initial position but, unfortunately, did not investigate whether there was a contributing linguistic factor.
An interesting compositional technique is found in the polyphony of the Aka (a Bantu language) in Central Africa. Unlike the music of their neighbours, Aka music is structured of four basic parts that move contrapuntally. If all parts were to sing the same words the melody of some parts would be considerably divergent from the speech melody. To counteract this, only one part is sung to words and the other parts are sung to nonsense syllables (Fürniss 2006).

There is some indication that a composer’s native language may influence his or her composition style even in non-tone languages and even in orchestral music. Hall, Jr. (1953) proposed that much of Elgar’s popularity in England (and unpopularity elsewhere) was due to similarities between his melodic style of large leaps and descending lines and the large pitch range and frequent occurrence of falling intonation patterns in spoken British English. Patel, Iverson and Rosenberg (2006) followed up on this idea and found that musical themes by French and English composers from the turn of the twentieth century show strong correlations to certain prosodic features — namely rhythm and interval variability — of the composers’ native languages.

2.3 Phonemic Divergence: How big a deal is it?

A common assumption about singing in tone languages is that, because tone is phonemic in these languages, a song melody that diverges from the spoken
melody will lead to grievous misunderstanding. But how much does phonemic divergence actually affect comprehension?

There are many actual examples of mismatches in tone languages. One is the Cantonese pop song “Heaven and Earth” (天與地) by Paul Wong (黃貫中).

The lyrics for one line in the song read:

4. 彷似流水不會斷
   fong35 ci13 lau21 seoi35 bat55 wui13 tyun13
   ‘incessant like flowing water.’

A musical transcription is given in Figure 2.3. The word lau21 (流) means ‘flowing’. The expectation in Cantonese music (see Chapter Three for more detail) would be that the melody should move downwards for the third word. In this case, the melody rises in a manner that would be allowable for a level tone causing it to sound like lau22 (meaning ‘leaking’; 漏); the sentence potentially would come across as ‘incessant like leaking water’ but the context provides plenty of indications of the intended meaning and native Cantonese listeners

![Figure 2.3: Transcription of an excerpt from “Heaven and Earth” (天與地) by Paul Wong (黃貫中). The deviation from the tonal pattern is shown in bold.](image)
have no problem getting the intended meaning of ‘incessant like flowing water’.\textsuperscript{10}

It is difficult to find a truly parallel example in English, but something similar may be found in the song “All I Want for Christmas Is My Two Front Teeth” by Donald Gardner. The final line of the song is “then I could wish you Merry Christmas” but, when sung with a lisp, as intended, is actually sung as “then I could with you Merry Chrithmath.” Native speakers of English have no difficulty in understanding the lyrics, even though there have been three significant phonemic changes ([s] and [ʃ] are replaced by [θ]). These phonemic mismatches have resulted in changing the word “wish” into “with”, a word which exists independently in English. Native English speakers, however, never seem to parse the word as “with” – most of them probably do not even notice the change. There is sufficient context to make the meaning abundantly clear. The change of the word “Christmas” to “Chrithmath” causes even less problem as there is no word in English with which it can compete. Although “Chrithmath” does not exist as a word in English its presence does not leave the listener baffled as to the meaning of the sentence. Again, the context makes it obvious what the intended word is.

Psycholinguistic experiments by Schirmer et al. (2005) have shown that changing a tone in an utterance in a tone language is only as disruptive to the

\textsuperscript{10} I am indebted to Zoe Lam of the Department of Linguistics at the University of British Columbia for drawing this example to my attention.
hearer as changing a segment, such as was demonstrated in the previous section. There is even some suggestion that speakers of tone languages may be more accepting of changes/errors in tones than of comparable changes of segments (Cutler and Chen 1997). Changing a few tones in an utterance does not appear to make much difference to comprehension. Chao (1956, p. 52) claims that “thanks to the liberal amount of redundancy that is usually present in all languages … Chinese without tones, as often spoken by foreigners [is] … intelligible, provided that it is perfect in pronunciation, construction, and use of words.”

A few other scholars have also remarked on the impact of mismatches between spoken and sung melody in tone languages. Bright, commenting on the lack of correspondence between song melody and spoken melody in Lushai (a Tibeto-Burman language spoken in north-eastern India, also known as Mizo), notes that:

we may wonder whether the complete distortion of lexical tone in singing makes it difficult to understand the words of songs. The informant’s answer is ‘It is not difficult for us.’ That is, the tone contrasts in Lushai may be completely obliterated without seriously obscuring the meaning of sentences. (1957, p. 28)

Ward (1932, p. 710) in discussing Fanti (a dialect of Akan, a Niger-Congo language) singing on the African Gold Coast similarly observes that “normally, of course, incorrect tones do not render words unintelligible … an ordinarily intelligent hearer can always understand, except in a fairly small number of cases
in which difference of tone does involve difference of meaning.” Ward raises an important point here in that the number of tonal minimal pairs in most tone languages does not seem to be particularly high. The number of minimal pairs of the same part of speech is much smaller still. This is not to say that they cannot be found, but their proportion of usage in song lyrics in a situation where misunderstanding could occur is not likely to be high and, consequently, their potential role in controlling song melody is, at best, minimal. The fact that tone is phonemic does not imply that changing the tone will change the meaning of the word; it merely implies that changing the tone has the potential to change the meaning.

Apart from the above anecdotes, the only works that look specifically at the perception of tone in music are Wong and Diehl (2002) and Vondenhoff (2009). Wong and Diehl (2002) found that listeners of Cantonese did, in fact, use the song melody to determine the meaning of tonally ambiguous words in a short sung phrase. However, the study used a carrier phrase, a sentence into which all potential words can fit, which is a common practice in psycholinguistic and phonetic research. This is done to remove any potential effect from neighbouring words. The phrase used in this case was *ha yat go zi hai ____* (‘the next word is ____’). The effect of using such a carrier phrase is that context is kept constant, leaving the listeners with only the song melody to fall back on. In
this case, the listener must, out of necessity, use the only available means of
disambiguating the sentences: the sung melody. Songs, however, are rarely
isolated sentences composed of highly ambiguous words; listeners usually have
a multitude of other cues to draw on to understand the lyrics. However, it is
interesting to find that, in the absence of all other indicators, listeners are willing
to use music for linguistic purposes.

Vondenhoff (2009) in her Master’s thesis used the same procedure to
examine perception in Mandarin songs and found that Mandarin listeners did
not use the melody to attempt to determine the target words in the carrier phrase;
her participants performed no better than chance in their word recognition. She
did, however, have some logistical problems and had to use stimuli that she had
produced herself and admits that use of a non-native speaker may have
contributed to the results.

2.4 Speech and Singing

It is apparent from the levels of correspondence between speech melody and
sung melody shown in Figure 2.1 that song melodies in cultures with tone
languages reflect, in a general way, the speech melodies of their languages.
While diverging from the speech melody occasionally does not significantly
impair comprehension, a general free-for-all is not the best choice for a tone
language, either. Matching melodies will certainly enhance comprehension, so it
is in a culture’s best interests to match, all other things being equal. As Herzog (1934, p. 465) points out, “speech-melody may furnish music with raw material, or with suggestions for further elaboration”.

However, when things are not equal, music “trumps” language. The research on singing in tone languages is full of comments on melodic mismatches which simply cannot be explained by linguistic rules. For example, Ekwueme (1974, p. 350) writes “even in Igbo music there are instances in which the aesthetic demands of the music override the requirements of linguistics”. In a very similar vein, Stock (1999, p. 184), writing on Beijing Opera, says: “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”.

There is other evidence that musical requirements override language requirements. Shaw (2008), for example, found that scat singers regularly violate the phonological rules of English syllable structure for musical reasons. For example, unusual consonant clusters such as [bw], [lj] or [dl] are used by scat singers to “enhance the melodic pitch contour, the musical phrasing, the auditory interpretation, or the distinctive trademarking of individual artistic style” (Shaw 2008, p. 147). Music creators are willing to accommodate language, as long as it doesn’t interfere with the music.
2.5 Structural Manifestations of Tone in Mandarin and Cantonese

The main focus of this thesis is the manifestation of tone in singing in Cantonese and Mandarin. Earlier studies have examined how and to what extent Mandarin and Cantonese express tone structurally in music. Generally speaking, Cantonese has lexical tones well represented in vocal music whereas Mandarin does not.

Chan 1987a (which incorporates the work reported in Chan 1987b) analyses 6 pop songs each in both Cantonese and Mandarin. Her primary analysis involves comparing the tones of words in the same phrasal position across stanzas. That is, do words that appear in the same position in different verses of the same song have the same tone? Looking at exact matches in Cantonese – words in corresponding verses having the same tone – she finds there is a match 69% of the time. She finds, however, that certain tones pattern together. The high level and high rising tones group together as do the mid level and low rising tones. Grouping these together to produce a simplified system of four levels she found a match 90.7% of the time. She also found a trend that higher numbers of stanzas means higher numbers of mismatches (where a mismatch is having words from different registers in corresponding positions in the lyrics).

A similar analysis for the six Mandarin songs finds the exact matches to be 49.6%. She looks for possible groupings of tones, particularly into the historic
even/oblique distinction that pairs tones 1 and 2 together and tones 3 and 4 together but finds no evidence to support this grouping. She argues that there are “permitted pairings” of tones 1 and 2, tones 1 and 4, and tones 2 and 4 and that these pairings raise the matching to 71.9%. This suggests that tones 1, 2 and 4 constitute one group and tone 3 constitutes a separate group by itself which implies some kind of special status for tone 3.

She expands her analysis to the acoustic analysis of the recordings of the melodies of the Cantonese songs but is hampered by logistical restrictions. She is unable to provide statistical analysis for her examination of the Cantonese songs because the data is taken from commercial recordings and the sung line cannot be isolated for acoustic analysis. Based on her observations she concludes that “relative pitch levels are, by and large, preserved in the songs” (p. 140). She also notes a tendency for less matching in songs with what she terms imported melodies (foreign songs fitted with new Cantonese lyrics).

Her analysis of the Mandarin melodies is based on published scores rather than on recordings. Here she compares the contour of the tone (the tone shape) with the contour of the musical note or notes that correspond to that syllable (the note shape). Using her published results it is possible to calculate this correspondence as 35% for a simple one-to one comparison with no groupings. She suggests that complex melismas can be interpreted in various ways to
provide the proper tonal contour for competing tones in different stanzas but she concludes (p. 163) that “in Mandarin, the tendency is to allow the overall melody to dominate.”

Wong and Diehl (2002) analyse four Cantonese pop songs. They find a conflation of tones into registers similar to that found by Chan (1987a) but limited to only three registers; they found that the low level tone (tone 6) and the low falling tone (tone 4) conflate into a single level. They provide only their results so the actual methodology is not clear. They conclude that “an ordinal mapping between musical note and tone group occurs such that the direction of pitch change in two consecutive musical notes is the same as in the two consecutive tone groups attached to them” (p. 204). From this it is possible to infer that they used a contour based analysis. They note that in these songs, “tone sequences preserve only the direction of F0 change” [italics in original] as opposed to the characteristic F0 ratio found in careful Cantonese speech (p. 204). Their analysis found a correspondence of 91.81% in the four songs.

Elaine Lau (2010) examines 6 modern Cantonese children’s songs. She carries out three different analyses. In the first analysis she compares a conflated tone system with absolute pitch levels within the pitch range of the song. She uses slightly different register divisions from those found in either Chan (1987a) or Wong and Diehl (2002). In her system, the high register includes tones 1 and 2,
the mid level includes only tone 3 and the low level includes tones 4, 5 and 6. She categorizes the low rising tone with the low level tone and the low falling (creaky) tone rather than with the mid level tone. She divides the melodic range of each song into three equal divisions and compares the tone register of a given syllable with the pitch range divisions of the song. She conducts separate statistical analyses for each level and finds statistical significance for each level. High register tones are generally sung in the upper range of the melody; mid register tones are sung in the middle range and low register tones are sung in the lower range.

Her second analysis is a contour matching analysis looking at the matching of transitions between corresponding syllables in the spoken and sung melodies. Here she finds a mean percentage of correspondence across the six songs of 79.2% (sd = 10.2).

Her third analysis uses the same technique as her first analysis to look specifically at the two rising tones to try and determine if there is a specific structural representation for these in the music. Although she does not present statistical results for this analysis, she observes there may be a tendency for contour tones to have an interaction with their neighbouring tones. This appears to be related to the end point target pitch of the contour tone and seems to support the register system found in Chan (1987a) and Wong and Diehl (2002).
The tonal representation in Mandarin music has received much less formal attention with scholars occasionally remarking on its paucity (e.g. Chen 2007, Chao 1956). Chan’s (1987a) findings are discussed above. Similar results are reported in Vondenhoff (2009, p. 22) from an earlier unpublished study she conducted which found “less than 40% of melodic note sequences had the same direction as those of the lexical tones on the corresponding syllables.” A different approach was taken by Wee (2007) who examined ten Mandarin folk songs and proposed a two-register system like that used by Chan (1987a) where tones 1 and 2 join to form a high register and tones 3 and 4 join to form a low register (this corresponds to the even/oblique categorization of Classical Chinese). He found correspondence only on the primary accent of a musical measure. His measure of correspondence comes from two possible criteria: a match of melodic shape between multiple notes assigned to the syllable (a *melisma*) and the contour of the tone or a register-based transition. A corresponding register-based transition is one where the final note at the end of the syllable is either higher (in the case of a high register tone) or lower (in the case of a low register tone) than the first note of the following syllable in the same measure. Looking only at positions of metrical prominence and using his criteria he reports a surprising 97.2% “accuracy” rate. Unfortunately, in several of the examples he provides, his two criteria are in direct conflict, for instance, a downward melisma followed by a
rise to the first note of the following syllable in the same measure. As long as one of the criteria holds he maintains that a correspondence exists. Of more concern are the situations where the melisma is in direct conflict with the shape of the tone but, because there is no other syllable in the measure, he counts this as a match because the criterion for a register-based correspondence is “vacuously satisfied” (Ibid., p. 136).

2.6 Phonetic Manifestations of Tone in Mandarin and Cantonese

Very little has been done to look at phonetic manifestations of tone in Cantonese and Mandarin. Chan (1987a) observes rising contour information added in by singers during the performance of Cantonese songs. She further suggests that: “in faster-paced songs, the tendency is for the tonal contours to be levelled out. This is accomplished by sacrificing the initial pitch rise when time is short” (p. 140). She provides narrowband spectrogram tracings of two short excerpts, one of 7 syllables and the other of 4 syllables. In these, the rising tones exhibit very clear rises and appear strikingly different in shape and slope from the level and falling tones. There is only one example of the falling tone and it appears very similar to the level tones.

Chao (1956, p. 57) suggested that singers in Mandarin may use grace notes (brief extra-metrical notes) “in order to ‘smuggle in’ the tone, if not already
suggested in the main melody.” This single line is the only discussion the author has been able to find regarding the phonetic realization of tones in Mandarin.

2.7 Conclusions

It does not appear that the (undeniably important) melody of speech in a tone language imposes itself on the music of the culture, rather, as Herzog (1934, p. 466) puts it: “the melodic element which is strong in tone languages, intrudes upon the music of the peoples speaking such languages”. Cultures choose to manifest tones to varying degrees and in varying ways in music. Mandarin and Cantonese provide an excellent opportunity to examine the manifestation of tone in singing. The musics are related but they incorporate different degrees of structural realization of tone. The differences will be the focus of the next chapter.
Chapter Three
Differences between Mandarin and Cantonese Popular Music

3.1 Mandarin and Cantonese Popular Music

From the preceding discussion it becomes apparent that, at least at the structural level, modern songs in Mandarin and Cantonese treat tone very differently. Where Cantonese songs exhibit a high correspondence between speech melody and song melody as well as possibly having rising contours included by the singers, Mandarin songs appear to show little correspondence in text setting.\footnote{Text setting simply refers to the act/art of setting words to music.} This is particularly surprising in that both musics stem from a common stock. Popular music, with predominately Mandarin lyrics, arose in the early decades of the twentieth century and was hugely popular in the large urban centres, particularly in Shanghai which was the main centre of its production for the first half of the twentieth century (Mittler 1997, Jones 2001, Wong 2001). Following the rise to power of the Chinese Communist Party in 1949, the production for this type of music moved to Hong Kong (Wong 2001). In the late 1960s and early 1970s Cantonese popular music branched directly out of the existing Mandarin music (Wong 2001, McIntyre et al. 2002). Why, then, are the two musics, otherwise quite similar and stemming from a common source, so different in the ways that they treat tone? The rest of this chapter examines possible explanations...
for this difference. The next section suggests some possible linguistic explanations and the last part of the chapter examines some possible non-linguistic explanations.

### 3.2 Linguistic Explanations for the Differences

Cantonese and Mandarin although frequently called dialects of Chinese are not mutually intelligible and are usually treated as separate languages by linguists. There are several features which may potentially play a role, consciously or unconsciously, in the decisions composers make when they set these languages to music. Aspects of language such as the number of tones in the language, neighbourhood density of tones, homophone density and the general functional load of tones in the language may all influence how and to what extent the tones of the language are manifested in music. The communicative intent of the songwriter/singer may also play a role.

One of the most quickly noticeable tone-related differences between Cantonese and Mandarin is the actual number of tones: Cantonese has six tones whereas Mandarin has only four; the fact that there are more tones to distinguish means that Cantonese has more potentially confusable pairs of words. It is interesting, however to compare the tone systems of Cantonese and Zulu which, as shown in Figure 2.1, exhibit very similar, and very high, levels of structural representations of tone in singing. Where Cantonese has a complex six-tone
system, Zulu has the simplest tone system possible: two level tones, high and low.

Another linguistic issue is the proportion of monosyllabic words in the two languages. Mandarin tends to prefer bisyllabic words; Chen et al. (1993) found that bisyllabic words made up 65.60% of the Mandarin corpus they looked at while monosyllabic words accounted for only 9.52%. Qian et al. (2004) found that bisyllabic words accounted for 54.61% of the words in the Cantonese corpus they looked at but that monosyllabic words accounted for 41.94%. This would suggest that there more chances of confusing syllables in Cantonese; Mandarin listeners would have greater opportunities for using context to extract meaning for ambiguous syllables at the word level. It must be noted that the Qian et al. (2004) corpus is much smaller than that used by Chen et al. (1993) – 6400 words for the Cantonese versus 54,500 words for the Mandarin.

This raises questions about the possible effect of tonal neighbourhood density on the manifestation of tone in singing. Neighbourhood density refers to how many “neighbours” – words that differ from an item by only one element; added, subtracted or substituted – a particular word has (Landauer and Streeter 1973). Dense neighbourhoods (having many neighbours) have been shown to inhibit perception of a given word. If a word has many neighbours listeners take longer to process that word (e.g. Vitevitch et al. 1999, Tsai 2007). Qian et al. (2004)
provide a breakdown of tonal neighbourhood density as percentage for Cantonese. They state that nearly 30% (29.83%) of Cantonese syllables have no tonal neighbours, that is, 30% of Cantonese syllables have only one tone associated with them. A further 25% have only two tones associated with them. Only 8.32% of Cantonese syllables have 5 or 6 tones associated with them.

A similar idea is homophone density. This is the number of homophones that a given word has. The presence of homophones leads to potential ambiguity. Huff and Payne (2012) estimate Mandarin uses approximately 1255 syllables (including tone distinctions) to express roughly 6000 morphemes, where a morpheme is equated with a written character. Qian et al. (2004) estimate the number of syllables (including tonal distinctions) in Cantonese to be approximately 1700 to express about 10,000 morphemes/characters. It is apparent from these numbers that syllables have to do more than double duty in both languages. The density of homophones (true homophones that are identical not only in segmental structure but also in tone) is very high in both languages. Yip (2007) estimates that for both Cantonese and Mandarin approximately 80% of syllables (including tone) have multiple meanings associated with them. The primary method of disambiguating these meanings is through context although frequency can bias listeners’ choices (Li and Yip 1996, Yip 2007).
All of the factors discussed above relate to what has been called the functional load of tone in a language. Functional load refers to “a measure of how much use a language makes of a contrast” (Surendran and Levow 2004, p.1). Surendran and Niyogi (2003) lay out a series of measurements to compute comparative functional loads for phonological features within a language. These measures can also be used to compare across languages. Surendran and Levow (2004) use these measures to look specifically at the functional load of tone in Mandarin. They emphasize that the functional load values “should be interpreted as relative, rather than absolute, values. In other words, functional load values can only be interpreted by comparing them to other functional load values” (Surendran and Levow 2004, p. 2). Surendran’s website (Surendren no date) includes functional load calculations for a variety of phonological features in words for both Cantonese and Mandarin (as well as English, Dutch and German) computed from existing corpora. The numbers are presented as percentages and the higher the number the more important the feature is in the language. Tone in Cantonese has a value of 7 while in Mandarin it has a value of 2. This means that if tone distinctions were lost (all the words had the same tone) then the fraction of information lost would be 7% in Cantonese and 2% in Mandarin. Surendran (no date) warns, however, that the data for Cantonese comes from a much smaller corpus which is made up of child-adult speech and
that cross-linguistic comparisons using the Cantonese data should, therefore, be treated with great caution. It is possible, however, to compare within each of the two languages along the same lines used in Surendran and Levow (2004). They compare the functional load of tone with that of vowels in Mandarin and find that they are very comparable: tone has a value of 2.1 and vowels have a value of 2.2 for words. This means that if all the tones of Mandarin were realized as the same tones, the effect on comprehension would be the same as if all the vowels were realized as the same vowel. The equivalent numbers for Cantonese are given in Surendran (no date): tone has a value of 7 and vowels have a value of 6. As with Mandarin, the effect of removing tone contrasts would be the same as the effect of removing vowel contrasts although it appears that the results in Cantonese would cause greater loss of information than in Mandarin.

How do these issues relate to the manifestation of tone in singing? If a language tends to have many words with high tonal neighbourhood densities that would mean large numbers of words would have the potential to be confused with each other if the tones were not faithfully represented in the music. This might be exacerbated by the fact that, in the two languages studied here, there are exceptionally high homophone densities. Failing to represent the tones accurately would increase the number of potentially confusable syllables to an even higher level. Conversely, it could be argued that speakers of both of these
are so used to relying on context to disambiguate the high number of true homophones that already exist that they don’t need to have the tones. The functional load data suggests that not representing the tones in the songs is akin to not singing the vowels properly. It is interesting to note that it is well known that as singers increase their sung F0, the vowels, for purely acoustic reasons, become less and less intelligible until at F0 levels greater than 1000 Hz (higher than about B5 or C6, the very top of the soprano range) there are only two discernable vowels (Benolken and Swanson 1990, Hollien et al. 2000). This does not stop composers from setting the full range of vowels to these notes.

3.3 Communicative Intent

Mark and Li (1966, p. 167), while discussing the assumption that songs must reflect linguistic tone patterns when singing in a tone language, pose the following question: “if textual intelligibility is of such overriding concern and entirely dependant upon tonal inflection, one might ask, why do people bother to sing at all?” This section looks at just that question. There appear to be situations where singing (or singing-like activities) is used with differing degrees of communicative intent. What is the primary focus of the activity? How important are the words?

Herzog (1934) discusses possible reasons why ritual curing songs in Navaho have less correspondence between speech melody and song melody
than gambling songs. He comments that intelligibility of the curing songs decreases but notes that this is not necessarily a bad thing “since the mysterious in ritual always has its effect on beholder and listener … the violation of speech-tones in the body of ritualistic songs is just one difficulty among many of which he [the Navaho layman] is more conscious” (p. 463).

A few scholars have noted differences in the degree of speech melody and song melody correspondence in different musical genres within a culture. Yung (1983, p. 29) points out that: “in Chinese vocal music, the specific relationship between linguistic tones and musical tones in a song may differ from dialect to dialect, and also among musical genres within a dialectal region.” Rycroft (1979), Chao (1956) and List (1961) lay out examples of such continua of correspondence in different languages which have been charted out in Table 3.1. This table is not a direct comparison of degrees of correspondence in these three languages. The idea of “high” or “low” correspondence is relative only for each individual language; there is insufficient evidence to be able to say if song types in a given column have similar degrees of correspondence. These continua may seem to stretch the definition of ‘music’ somewhat but all three authors seem comfortable classifying the various forms included as music. As Nettl (2005, p. 17) points out “there is no interculturally valid conceptualization or definition of music.” This

---

12 The term “genre” is being used here in a very general sense.
Table 3.1: Three continua of speech melody / song melody correspondence.

<table>
<thead>
<tr>
<th></th>
<th>High Correspondence</th>
<th>Low Correspondence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nguni</strong> <em>Rycroft (1979)</em></td>
<td>War cries; praise-poetry</td>
<td>Personal solo songs</td>
</tr>
<tr>
<td></td>
<td>Traditional dance-songs</td>
<td>Modern church, school and popular music</td>
</tr>
<tr>
<td><strong>Chinese</strong> <em>Chao (1956)</em></td>
<td>Children’s songs, street vendors’ cries</td>
<td>Chanting: traditional (learned) reading style</td>
</tr>
<tr>
<td></td>
<td>Recitative in traditional Chinese drama</td>
<td>Traditional, “stereotyped” melodies</td>
</tr>
<tr>
<td><strong>Thai</strong> <em>List (1961)</em></td>
<td>Mnemonic recitation (e.g. multiplication table)</td>
<td>Traditional literary recitation</td>
</tr>
<tr>
<td></td>
<td>Classical songs</td>
<td>Popular songs</td>
</tr>
</tbody>
</table>

supports the findings of Bell Yung (1989), Feld and Fox (1994), and many other ethnomusicologists who have argued that the line between language and music is fairly fluid.

In all three cases, the genres with the highest degree of correspondence are genres which have been thought to be “not quite” singing. Rycroft (1979, p. 308) characterizes the Nguni war cries as “marginal” (although considered song by the people who use them) and praise poetry as “pseudo-melodic” (not considered song by the users). Chao (1956, p. 53) describes the children’s songs and vendors’ cries as “something intermediate [between speech and singing] ... based largely on the phonemic tones of the words, spoken in a stereotyped manner.” List (1961, p. 17) points out that recitations, both mnemonic and

---

13 A family of related Southern African Bantu languages including Zulu, Xhosa, Ndebele, and Swazi.
literary, are considered as something different from singing by the Thai people. It is particularly interesting to observe, however, that “traditional” songs in all three cases are not that high on the continua; they are second only to modern songs in their lack of correspondence in all three instances.

This is not to say that modern songs in all cultures should be assumed to have low correspondence. As discussed above, Wong and Diehl (2002) found that popular songs in Cantonese have one of the highest correspondences of any song type studied (92%). Schellenberg (2009) found in Shona (a Bantu language spoken in Zimbabwe) that the hymn Jesu Idombo had a higher correspondence (64.2%) than the traditional song Nyii Dzaibva (56.5%).

Although the evidence is very limited, it is tempting to speculate as to whether there is a connection between the communicative function of the genre and the degree of correspondence. Mnemonics and praise poetry may have predetermined pitch contours to aid memory. As well, Nguni praise poetry is traditionally recited at the same time as other singing and dancing are taking place (Rycroft and Ngcobo 1988), so there is also a need to have the poetry stand out from the ambient noise. Similarly, the vendors’ cries in Mandarin need to be heard over the street noise so the “musicality” of these two forms may well aid in projection and endurance. In all of these cases, the musical nature of these utterances has an extra-musical function. As the function becomes more focused
on the music (whatever the function of music may be for these cultures), the communicative intent of the composer/performer and, consequently, the primacy of the language seems to decrease and the focus on music takes precedence.

3.4 Possible Non-linguistic Explanations

The interaction between speech tones and sung music has been explored for a few types of traditional Chinese music. The interaction in Kunqu Opera, the oldest form of traditional Chinese opera extant, was examined by Marjorie Liu (1974). This style of opera, which gained prominence in the sixteenth century, uses a special dialect, the Chung-chou or Central dialect which reflects Classical Chinese pronunciation. Liu (1974) analyses two performances of a single aria and finds that speech tone does influence song melody but so do stress, intonation, speed and duration. She observes that musical ornaments are used to “clarify tones and rhythms” (p. 84) but, conversely, finds that the general tone contour is determined at the onset of the word and that subsequent melismatic ornamentation modifies the speech tones. In her words, “speech tones … are actually conveyed through a gestalt of relative duration, loudness, intonation and other aspects of articulation, in conjunction with pitch contour and ornamentation, rather than by pitch contour alone” (p. 84). Chao (1956) briefly discusses what he calls “the traditional form of Chinese drama” (p. 57) using the
Central dialect and supports the idea that it follows “largely the tones of the words” (loc cit.).

Jonathan Stock (1999) analyses transcriptions of historic performances of Beijing Opera, a style of Chinese opera which developed roughly at the turn of the nineteenth century from a fusion of several older styles (Stock, 1999). His analysis finds that singers of this style from the earlier part of the twentieth paid great attention to the musical structure of the arias and did not just exaggerate linguistic tone. He observes that, in terms of speech melody/song melody interaction, there is “occasional correspondence … but also frequent moments of contradiction” (p. 195) and concludes that “the subordination of musical elements to the sounds of the spoken language is neither automatic nor complete” (p. 184).

Although the origins of Cantonese opera can be traced back to the Ming Dynasty (1368-1644) drastic changes took place in the 1920s and 1930s with “the introduction of the Cantonese dialect in both sung and spoken passages, the incorporation of Western melodic instruments, the creation of new aria types, and the employment of traditional Cantonese fixed tunes and singing narratives in the vocal music” (Chau 1999, p. 1). Yung (1983, 1987) looks at the relationship between speech melody and song melody in this relative newcomer to the regional opera field. He points out that very often the performer is given a script
only a few days before the performance which gives very little time to learn a new opera. The arias are based on pre-existing musical forms so much of the actual music is improvised, influenced by a combination of the pre-existing music for that aria type and the linguistic tones of the words to be sung. Yung (1987, p. 84) indicates that:

…the linguistic tonal inflections of the text are generally retained in the singer’s musical delivery; in other words, the singer follows what shall be called the rule of contour matching between linguistic tonal inflection and melodic contour. The singer, however, may depart from this rule because he must first follow the more important rule on line-ending pitches for the retention of the identity of the aria type. The rule of contour matching is also bypassed when the duration of the note is too short, and when the melody has to move in the opposite direction because of the pitch level of the following note. The singer also uses ornamental glides to reproduce linguistic tonal inflections.

He concludes that “the role of linguistic tones in the creative process proves to be nothing more than a rough guide” (Yung 1987, p. 90).

Yung (1987, p. 91) goes on to note that “in other dialectal regions of China … the linguistic tones in vocal music do not play as large a role as in Cantonese opera – a fact which suggests that ease of linguistic communication is not necessarily the only, or even the most important, explanation for this phenomenon.” He goes on to suggest other reasons for the close relationship between speech melody and song melody in Cantonese opera. One may be purely aesthetic: given the uses and re-use of a limited set of melodies, the use of
the linguistic tones of the lyrics provides a greater variety of melody. The other reason he suggests is expediency. Heavy reliance on linguistic tone limits the need for musical notation and rehearsal – singers may be relied upon to produce good theatre without needing a lot of outside rehearsal.

3.5 Nationalism, Identity and Speech Tone

Yung’s (1987) observations about the differences between Cantonese opera and other regional varieties of Chinese opera in terms of the correspondence between speech melody and song melody seem to apply just as well to the discussion of popular music, at least when comparing the popular musics of Cantonese and Mandarin. These styles of music do not rely on improvisation, nor are they set to a limited number of melodies but the differences remain. This section will briefly argue that there are compelling socio-political factors related to nationalism and identity that may contribute to the expression or non-expression of tone in these two musics.

3.6 A New Music for China

China in the late nineteenth and early twentieth centuries was a country in turmoil. With the loss of the opium wars and the resultant forced opening of Chinese ports to Western interests, many members of the Chinese intelligentsia started to believe that the only way for China to survive was to compete with the
foreign powers on their own level. The old ways were seen as no longer effective; China needed to modernize – and modernization meant Westernization (Mittler 1997, Jones 2001).

Western approaches in areas as diverse as the military, education, literature, music and theatre were imported and touted as superior to the old Chinese ways; anything traditional was viewed as outmoded and undesirable. This spirit of reform continued through the overthrow of the monarchy in 1911 and the establishment of the Republic of China the following year (MacNair 1931, Wong 1991, Townsend 1992). It is not the purpose of this chapter to detail the rise of nationalism in China; what follows, rather, is a brief summary of some of the highlights relevant to the discussion of tone and text setting.

3.6.1 Language Reform

Language reform in Republican China was closely linked with literacy and national unification. Much of the proposed reform actually centred on the writing system. The traditional ideographs were seen as a deterrent to literacy – too much time was needed to memorize all the characters. The Chinese system was contrasted with those of the great world powers, all of which had high levels

---

14 There are many excellent resources to which the interested reader may refer for greater detail: Chen (1999) provides an excellent general overview; de Francis (1950) focuses on reform to the writing system and Ramsey (1987) provides an excellent history of the language reform process regarding pronunciation.
of literacy, all of which had phonetic components in their writing systems and many of which had forced their way into China as the result of treaties (this included Japan which, although small, had defeated China in the war of 1894-1895). Many Chinese intellectuals of the time believed that devising a single national language with a simpler writing system would be a stepping stone to China’s development back into a great power (de Francis 1950, Chen 1999).

In the traditional Chinese system the language of both literature and administration had been Classical Chinese, a scholarly, archaic version of Chinese that was exclusively written and required years of study. This was rejected in favour of the idea of a common, national, vernacular language. The problem, of course, was that there was no common, national, vernacular language. China, then as now, was a conglomeration of mutually unintelligible regional dialects. Mandarin, the language of the capital, was used as an unofficial *lingua franca* but it was not generally thought of as a national language (Ramsey, 1987).

The reformers wanted a single language that would unite the country. A Conference on Unification of Pronunciation was held in 1913, very shortly after the founding of Republican China “to establish a standard national pronunciation for the ideographs, to analyze the national pronunciation in terms of its basic sounds, and finally to adopt a set of phonetic symbols to represent
these basic sounds” (de Francis 1950, p. 55). There was serious infighting and
disagreement among the delegates, but finally a Dictionary of National
Pronunciation was proposed with pronunciations based primarily on that of
Beijing but with certain historically based modifications. Chen (1999, p. 18)
enumerates the three primary modifications that were approved:

1. Words which historically had initial voiced consonants in Classical
   Chinese ([v], [n] and [ŋ], which do not occur as initials in modern
   Mandarin, would be pronounced with the voiced initials.
2. Words which historically had dental sibilants, pronounced with palatals
   in modern Mandarin, would revert to the Classical Chinese pronunciation
   with dental sibilants ([ts], [tsʰ], [s]). Words that historically had velars, also
   pronounced with palatals in modern Mandarin, were to retain their
   modern palatal pronunciations ([tʂ], [tʂʰ], [ç]).
3. Words that historically had the fifth “entering” tone in Classical Chinese,
   missing in modern Mandarin, would be pronounced with the entering
   tone.

Because of its mixed nature this proposed “dialect” acquired the name
Blue-Green Mandarin (de Francis 1950, p. 66). In 1919 the Ministry of Education
was convinced to publish the dictionary and support the system. Chao Yuen Ren
(趙元任, 1892-1982) produced a set of recordings to accompany one of the
textbooks and taught the system in his Chinese courses at Harvard around this
time but the system was not particularly successful. Ramsey (1987, p. 9) says that
Chao was the only person who was ever able to teach the system, which sounded
“like Chinese no one had ever heard before”.15 Another scholar, Wang Pu, had

15 Because it was Chinese that had never been heard before.
tried to make a set of recordings but was unable to produce the proposed fifth tone (loc. cit.). These artificial dictionary readings remained the official standard until 1932 when the system was quietly changed to reflect the variety of Mandarin spoken in Beijing (Chen 1999).

Even though there was an officially sanctioned (albeit virtually unpronounceable) national language, adherents of many different factions continued to wrangle over whether there should be a single official language and what that standard should be (de Francis 1950, p. 66). There was little agreement on what pronunciation ought to be used. There were exacting standards for all those who sat on the original committee (Chen 1999, p. 31) and its members were some of the best trained scholars of their time. The committee, however, overlooked the fact that languages change constantly and that language legislation is difficult to enforce. Their focus, which seems to have over-ridden all other considerations, was on the need for a single language to bolster the national Chinese identity. Unfortunately, they could not agree on what that language should be.

3.6.2 Music Reform

Music reform was also seen as important in the modernization of China. Western-style music was already quite familiar to the Chinese by the turn of the twentieth century through the hymns taught by missionaries, army songs
introduced by German military instructors involved in military training and Western-style school songs imported via Japan (Wong 1982, Wang 2001a).

The music reform movement was spearheaded by scholars like Xiao Youmei (蕭友梅, 1884-1940) and Cai Yuanpei (蔡元培, 1868-1940) who both studied in Germany in the 1910s and would go on to found the Shanghai Conservatory of Music, the first conservatory in China. These musicians were all trained at a time when one of the strongest influences in European music was nationalism. The idea of nationalism in music resonated strongly with the Chinese reformers who soon desired to produce a national music for China but in the Western idiom. As Cheung (2008, p. 64) noted, “Chinese musical modernity developed with the premise that Chinese should embrace Western musical practices, while continuing native ideals and producing music and musical meanings that Chinese listeners could embrace.”

An example of the prevailing attitude can be found in Chao (1931) where the author, after carefully describing traditional Chinese music, spends the final three pages of the essay explaining how all of what he has described is an “art which has almost worked itself out” (p. 95) that may be improved and revitalized by adopting Western instruments and harmonies.

---

16 Other influential musicians studied in Japan or the United States. Cheung (2008) provides detailed histories of many of the most prominent musicians of this era, both reformers and traditionalists.
One of the major appeals of the Western musical system was the perception that Western music was both logical and technical. As Jones (2001, p. 25) points out, traditional Chinese music lacked “a tempered scale, functional harmony, counterpoint, orchestration, standardized notation, and the engineering prowess embodied by Western instruments like the piano.” This perception of Western music as technology appealed to the prevailing sense of what some scholars have termed scientism – an intellectual appreciation of science without necessarily applying its principles (Wellmuth 1944). Kwok (1965, p. 3) in his book on the prevalence of scientism in China in the first half of the twentieth century defines it as “the tendency to use the respectability of science in areas having little bearing on science itself.” The structure and rules associated with Western composition were seen as superior to the traditional Chinese methods of music creation simply because they were structures and rules and therefore must be scientific and logical. The structure found in traditional Chinese music appears to have been overlooked or regarded as too simplistic.

This combination of nationalism and the fascination with Western musical theory and harmony led to the development of a “kind of music employing a harmonic framework reminiscent of the musical language of the late nineteenth century to accompany pentatonic melodies” (Mittler 1997, p. 33) which Mittler
(loc. cit.) labels *pentatonic romanticism*. This pentatonic romanticism continued to dominate much of Chinese music throughout the twentieth century.

3.6.3 Text Setting and Tone

The confusion and indecision surrounding the burgeoning national language, the adopted Western use of melody as a musical indicator of nationalism and the fascination for Western harmony as a system appear to have had repercussions in the realm of text setting. Melody and harmony acquired primacy (Mittler 1997, p. 283 states that “pentatonic melody is considered the most essential part of Chinese musical heritage”) while linguistic tone was generally disregarded in musical contexts (Chao 1956). Other factors also existed that would have influenced the composers of the time in their approaches to text setting and the general trend for the dominance of musical melody over speech melody.

The use of folk music as an expression of nationalism was prevalent in Europe at this time (Bohlman 2004) and was adopted by Chinese composers (Wang 2001a). Many of the most prominent Chinese musicians and scholars of this era were keen collectors of Chinese folk music (a recurring theme in the various biographies in Cheung 2008) and folk melodies were used in compositions (Wang 2001a, 2001b provides several examples). However, traditional folk songs were not in a language that would have been understood by the nation as a whole but in one of the numerous regional dialects or
indigenous languages. As such, they were in direct opposition to the overall goal of national unity. Composers made use of the folk songs in various ways that used only the melodies. For example, Li Shutong (李叔同, 1880-1942) wrote new words for existing folk song melodies (Wang 2001a); Huang Zi (黄自, 1904-1938) and Xian Xinghai (冼星海, 1905-1945) incorporated folk tunes and structures into larger choral works with lyrics drawn from other sources (Wang 2001b).

The functions of the songs could strongly influence the texts that were being set. Some composers like Li Jinhui (黎锦晖, 1891-1967) wrote music that was intended to help children learn the new national language (in Li’s case, using the phonetic system devised by his brother, the phonetician Li Jinxi (黎錦熙, 1890-1978)). Li would later go on to write the first pop songs in Mandarin, songs that had a very different goal – pure entertainment (Jones 2001). Leftist composers like Nie Er (聶耳, 1912-1935) wrote songs to celebrate the worker and inspire the masses (Howard 2012). The “conservatory” composers like Xiao Youmei and Chao Yuen Ren (whom Xiao once called “China’s Schubert” (Cheung 2008, p. 332)) were proponents of the philosophy of “art for art’s sake.” Although melodic elements from folk songs could be (and sometimes were) worked into these songs, the non-national language of the originals prevented use of the text.
Setting different texts to pre-existing melodies was not a new idea. There was some precedent for Chinese texts set to pre-existing (usually foreign) melodies in the case of the earlier Western-style songs such as hymns, army songs and school songs. Mitler (1997, pp. 23-24) says of school songs that “Western melodies were matched with Chinese texts. ‘Composers’ put words to melodies and only very seldom wrote a tune themselves.” Chao (1931, p. 94) tells of his children learning “imported tunes with Chinese words” at school. Wang (2001a, p. 4-7) lists several examples of school songs with Chinese words set to foreign melodies. Wong (1982) discusses Hong Xiuquan’s (洪秀全, 1814-1868) use of Western hymn tunes set with Chinese texts for mass singing during the Taiping Rebellion (1851-1864). There appears to have been little attempt to make the tone patterns of the new words align with the contour of the melody. Chao (1924, p. 10) writing at the height of the reform movement says that “both the Chinese and the missionaries make up songs that fit neither the old [classical Chinese] nor the new [modern Chinese] system of tones.”

This is not to say that all composers ignored tone. Chao Yuen Ren paid strict attention to tone in the composition of many of his songs, most often following the traditional classical Chinese system of tones rather than the modern system of Mandarin (Chao 1956, p. 58). He states outright that “most contemporary Chinese song writers, however, pay no attention to tone” (loc. cit.).
It is worth noting that Chao was a phonetician who had a fascination with tone, a fact which most likely influenced his compositional choices.\textsuperscript{17}

The combination of a strong focus by composers on melody as an expression of nationalism and the debate over what a national language might look like makes it unsurprising that the vast majority of the founding composers of the new style of music ignored lexical tone when setting Chinese words to music. Given that there was a precedent to overlook tone not only in Western-style music but also in other existing forms like Beijing Opera (Stock 1999) this tendency is even less surprising. The pattern for the (non)representation of tone in modern Mandarin music seems to have been set from its very inception. Melody took – and continues to take – precedence over tone.

\textbf{3.7 Cantopop}

With the coming to power of the Communist Party in China in 1949 the entertainment industry, along with many other industries, moved to Hong Kong where the producers would be able to continue to make money. Once there they continued to produce the same Mandarin movies and music they had produced

\textsuperscript{17} The similarity to David Rycroft (1924-1997) who was discussed in Chapter Two is striking. Both men were linguists who worked on the phonetics/phonology of tone languages but were also gifted amateur composers who composed songs in tone languages (Rycroft is best known in the music world as the composer of the national anthem of Swaziland). Both Rycroft and Chao made linguistic tone patterns an integral part of their songs.
on the mainland, but gradually Cantonese started to exert some influence (Wong 2001).

3.7.1 Hong Kong Identity

Hong Kong was ceded to the British by the Chinese during the nineteenth century as a result of the Opium Wars. Throughout its early colonial history it was predominately a temporary immigrant community in which the transient inhabitants, rulers or colonized, maintained strong ties to the homeland – wherever that was. Some people elected to stay but most didn’t, always being replaced by new workers with ties to a homeland. With the 1949 Communist takeover in China came a surge of immigration from the Chinese mainland. China closed its borders making it harder for the Chinese outside China to maintain ties with their ancestral homes (although many continued to support their families in China). Those who stayed in Hong Kong settled, married and had children. As the post-war generation of Hong Kong-born children grew up and entered the work force they identified far less with their ancestral homes and far more with their birth home. This is usually considered the start of the Hong Kong identity. (For a more detailed history see e.g. Baker 1983).

Much has been written about the Hong Kong identity; most of it centered on how the 1997 handover of Hong Kong to China would affect it (e.g., Feng 1998, Tong et al. 1999). One of the striking commonalities of the research centring
on the handover is the basic assumption that there is a distinct Hong Kong identity. Scholars generally seem confident of its existence (although see Feng (1998) for a discussion of opposing views) but less has been written about the initial development of this identity.

Wong (2001 p. 39) comments that “during the 60s, as a new generation grew up in Hong Kong and as new immigrants from the mainland gradually adjusted themselves to the Hong Kong style of living, Hong Kong people gradually developed a new sense of identity and self-confidence.” Choi (1990) explores in depth how the Hong Kong identity emerged and expressed itself in a wide variety of media. He states that:

Since its social and cultural severance from China in the 1950s, Hong Kong has gradually evolved its separate culture and identity which, today [1990], is already very different from that of the Mainland. This culture is manifested in the distinctive outlooks, aspirations, lifestyles, and, of course, language patterns of the local residents. It is also articulated through various cultural products: popular songs, films, TV programmes, popular books and even comics (p. 538).

Siu (1996, p. 184) states that “the indigenization of popular music and film have reinforced and made visible the outlooks, aspirations and expressions of an increasingly distinctive culture of a generation made in Hong Kong.” The Hong Kong identity, therefore, is of recent origin starting just after the Second World War and not really hitting its stride until the 1970s.
3.7.2 Development of Cantopop

The Chinese entertainment industry that had been based in Shanghai moved to Hong Kong when the Communist Party came to power in China in 1949. The industry continued to produce Mandarin music and movies in the same style as they had done in Shanghai (Wong 2001).

Although this music remained very popular with older audiences, by the 1960s the younger generation were generally not attracted to what McIntyre et al. (2002, p. 229) describe as the “gushy, lullaby lyrics and Lydian\textsuperscript{18} musical arrangements” of this older style of music. They were more interested in the rock ‘n’ roll music being imported from the United States and England. McIntyre et al. (2002 p.228) describe how “Hong Kong fans of Western rock artists such as Elvis Presley, the Beatles and the Rolling Stones formed their own groups to imitate these Western styles. Shortly thereafter, they found a new challenge in writing their own music using Cantonese lyrics while retaining Western instrumentation and rhythmic patterns.” The existing (Mandarin) music industry was quick to pick up on this and soon the Cantonese popular music scene was up and running. By the late 1970s the earlier rock-inspired music had given way

\textsuperscript{18} Lydian refers to a scale or mode that has a different arrangement of whole tones and semitones than the major or minor scales Western listeners are used to. To hear the Lydian scale on a piano start on F and play all the white notes up to the next F. An excellent example of a popular tune written in the Lydian mode is the theme song from The Simpsons.
to a more romantic, pop style of music, strongly influenced by the theme songs of popular Cantonese television programs (McIntyre et al. 2002).

3.7.3 Cantopop, Language and Identity

Those identifying as Hongkongese wanted to distinguish themselves from both the colonial British and the Mainland Chinese. Music and language were the primary means of expressing that distinction. McIntyre et al. (2002, p. 239) describe Cantopop as something that “defined a cultural space that distinguished itself from the culture of the rest of Greater China.” Witzleben (1999, p. 252) points out that “when Cantopop arose in the 1970s, it was unquestionably oppositional in terms of language and ethnicity. On the Chinese mainland, Cantopop is also linguistically oppositional in that it is sung in a dialect rather than in the national language.”

Language was, and continues to be, a major contributing factor in establishing the Hong Kong identity. Gu and Ho (2012, p. 4) point out that “Hong Kong remains the only place in China where Cantonese, the mother tongue of the majority of its residents, is the official medium of instruction for primary schools and even some secondary schools.” Mandarin is widely spoken but is viewed with some suspicion and even dislike (Gu and Ho 2012). As Bolton (2011, p. 68) points out “modern Hong Kong was essentially founded by refugees fleeing from the control of a Communist regime and the horrors of the Cultural
Revolution, and the identification of Hong Kong people with the national language of the PRC is tempered by the experience of the last six decades.”

There is also a complex relationship with English, the language of the former colonialist controllers. Tsang and Wong (2004) summarize this nicely:

English, as one of the official languages alongside Chinese, is both the language of power and the language of educational and socio-economic advancement. Chinese people in Hong Kong mostly began to learn English in childhood. Despite this, English is not used as a second language in which people operate easily and spontaneously for general social and daily communication. It is, to a great extent, restricted to specialized academic and professional uses. Whereas Cantonese has its practical use in everyday communication, the influence of English is, to a certain extent, evident in code-mixing. (pp. 768-769)

English is found in Cantopop but almost always in the form of code-mixing or code-switching (switching back and forth between two or more languages or dialects) that very clearly reflects this secondary role that English plays in the Hong Kong identity (Chan 2009).

Given the importance of both music and language in the expression of the Hong Kong identity it is not surprising that songwriters place a strong emphasis on text setting that complements the words. There is little that distinguishes Cantonese music without words from music of mainland China (Mittler 1997). The Cantonese language was and continues to be one of the primary means by which Hong Kong residents can distinguish themselves from both the colonial
British and the mainland Chinese (Tong et al. 1999). Cantonese language played (and continues to play) a pivotal role in maintaining the identity of both the performers and the consumers of the music and, as such, demands much greater attention in the composition process than comparable music produced elsewhere. It is, therefore, arguably important that the language be easily recognized and understood when sung if the music is to be readily identified as Cantonese music.

Cantopop developed in the shadow of Cantonese Opera as well, which also has a history of using language and music as indicators of identity. Ferguson (1993, p. 70) describes how in the early years of the twentieth century Cantonese gradually replaced the previously used archaic stage dialect, something which did not happen in other regional opera styles. This use of the vernacular in a traditional art form has been a significant factor in Cantonese identity in the twentieth century (Ferguson 1993). Cantonese Opera has very clear procedures for the manifestation of tone in the music. Yung (1983) found, for example, that for one particular aria type (the Seven Syllable Melody) not only is there relative matching of contour but there is a strong tendency to what he calls absolute matching – the use of particular pitches to represent particular groups of tones. Wong and Diehl (2002) and Chan (1987a, 1987b) found that the technique of grouping Cantonese tones into registers based on the final target level and
reflecting these registers in the music used in Cantonese Opera is the same
technique used in Cantopop.

In the first half of the twentieth century Cantonese Opera had been
popular not only in Hong Kong but in areas such as Shanghai as well, and had
influenced Chinese popular music from its early days as well as being influenced
by it.

Even in the 30s, the Cantonese opera, under the leadership of Sit
Kok-sin (Xue Jue-xian) who had spent time in Shanghai, had
introduced western instruments into the opera accompaniment.
Composers such as Lui Man-sing (Lu Wen-cheng) had
Westernized Cantonese tunes and introduced elements of
Cantonese music into the dance music played in Shanghai dance
halls ... Cantonese pop songs were born under the influence of
these new musical styles, as well as inspirations drawn from Shi
Dai Qu [Mandarin popular songs] (Wong 2001, p.30)

The Hong Kong identity being confirmed by Cantopop was most readily
expressed by language use. Popular music was one of the primary means of
broadcasting and reinforcing this identity. Given that there already existed a
system for representing Cantonese tone in music it is not surprising that
composers and performers adopted this system and used it to give their
expressions of identity voice leaving Cantopop with a higher correlation between
speech melody and song melody.
3.8 Conclusions

Although Mandarin and Cantonese popular music both stem from the same musical roots, composers of these two styles deal with lexical tone differently. The complex socio-political environments in which both of these musics developed can be argued to have had a lasting effect on their approaches to the realisation of speech tone.

The strong influence of Western musical nationalism on the composers of the incipient music of a reforming China combined with controversy over whether there needed to be a national language and, if so, what that language would sound like, meant that melody became the driving force of Mandarin music.

The importance of the Cantonese language as a differentiator from both the British and the mainland Chinese meant that language became one of the most important expressions of identity in mid-century Hong Kong. Consequently, lexical tone becomes much more important in the music that expresses that identity, not only in composition but also in performance. The choice on the part of composers to express lexical tone in songs is likely to be the result of a complex combination of influences.
The structural manifestation of tone is the purview of the composer/creator but the singer can also play a role. What role does the singer play in Cantonese and Mandarin?
Chapter Four

Phonetic Realization of Tones in Sung Cantonese

4.1 Introduction

For this chapter, an experiment was carried out to examine whether the phonetic manifestation of tone during singing in Cantonese as observed by Chan (1987a) and Yung (1983) occurs systematically. Separate analyses were carried out to look at the slope of the contour and to determine if the durational correlates of tone were maintained while singing.

4.1.1 Cantonese Tones

As previously stated in Chapter One, Cantonese has six tones. Table 4.1 repeats Table 1.1 and outlines the six tones of Cantonese. The structural realization of the tones for composing popular songs in Cantonese modifies the tone system slightly. Both Chan (1987a) and Wong and Diehl (2002) find a conflation of the six Cantonese tones into registers according to the final F0 level of the tone.

<table>
<thead>
<tr>
<th>TONE</th>
<th>DESCRIPTION</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High; level</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>High; rising</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>Mid; level</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>Low; falling</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>Low; rising</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>Low; level</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 4.1: The six Cantonese lexical tones.
Wong and Diehl (2002) find a three register system for music composition:

- high — tones ending on level 5 (tones 1 and 2)
- mid — tones ending on level 3 (tones 3 and 5)
- low — tones ending on level 1 or 2 (tones 4 and 6)

This reflects the tones of Cantonese speech 92% of the time in the four songs they examine. Chan (1987a) does not find the conflation of the low level tone with the low falling tone and posits a four register system similar to that found by Yung (1983) for Cantonese Opera. Both Chan (1987a) and Wong and Diehl (2002) found that contour information for tones is not included in the composition of pop songs.

Considering sung Cantonese, Chan (1987a) observes through the use of pitch tracings that rising contours are apparent during the singing of syllables with rising tones. This suggests a phonetic component to the realization of tones in sung Cantonese in addition to the structural realization discussed above. Specifically, Chan’s (1987a) observations suggest that rising tones are distinguished from level tones phonetically by the inclusion of an upward (positive) sloping trajectory of the F0 contour, especially in the early part of the note. However, Chan’s (1987a) tracings do not suggest a distinction between the level and falling tones in terms of slope. Tone 4 is often associated with creaky voice (Lam and Yu 2010, Vance 1977) and has been found to pattern with the
level tones. Vance (1977) found that synthesized stimuli with low F0 and a falling contour were rarely identified as tone four by Cantonese listeners. Lam and Yu (2010) found that tone-four words without creak were more likely to be perceived as tone six than tone four. Chan (1987a) finds that the low level tone and the low falling tone are differentiated structurally by a fourth register (à la Yung 1983) while Wong and Diehl (2002) find a conflation of the low level tone with the low falling tone.

For this chapter, an experiment was carried out to test whether such phonetic manifestation occurs systematically during singing in Cantonese, with specific focus on the prediction that rising tones will be phonetically distinct from non-rising tones in terms of the F0 slope. In order to test for the phonetic representation of tone information in sung Cantonese, the first part of this experiment looks at a six-way minimal set (by tone) on [si] embedded in a song. These words were contained within a carrier phrase to minimize influence from external factors. The disadvantage of a carrier phrase is that it unifies context, leading to a possible confound of over-compensation such that the subject may exaggerate the tonal information to enhance comprehension in the absence of context. A second analysis attempts to counterbalance this confound by looking at a different set of words (numerals) in a more context-full environment. Tones in Cantonese have a secondary characteristic of duration associated with them.
(Kong 1987); the third analysis looks at whether these durational correlates of tone are maintained in singing.

4.2 Analysis One – Cantonese Minimal Set

4.2.1 Participants

Participants were 12 native speakers of Cantonese (6M, 6F, mean age = 45.17, sd = 16.17). All were residents of Vancouver, BC (Canada) and all were also fluent in English. Ten of the participants were choral singers, two of whom had received formal training in singing. The two non-choral singers reported regular singing. The participants reported an average of 3.3 hours of singing per week (sd = 1.2).

4.2.2 Stimuli

The target stimuli for this study were the minimal set of [si] on all six tones given as examples in Table 4.1. These six words were embedded in a specially written Cantonese song – a carrier phrase set to music and repeated 9 times. The music for the song consisted of 2 variations of a melody written by Patrick Wong (Wong and Diehl 2002) to match the spoken contour of the phrase (excluding the target words and the numerals). The two variations were put together in 4 pairs along with a separate concluding phrase to form a short song with the form AABBAABBBC. The target word is sung on two different notes: in phrase A the
note is E and in phrase B the note is C#. The score is given in Figure 4.1. The lyrics translate as:

The first word is (—), the second word is (—),

The third word is (—), the fourth word is (—),

The fifth word is (—), the sixth word is (—),

The seventh word is (—), the eighth word is (—),

The ninth word is (—). We are all words.

As both the numerals and the target words were to be randomized during the experiment the song needed to be adapted in such a way that structural manifestations of the tones which are expected in Cantonese music could be maintained at least part of the time. Depending on the target tone, the musical melody potentially deviates from the spoken melody in the two locations in question – the second word of each line (the numerals) and the last word of each line (the target and filler words). For each of those two situations there are two
possible musical contours: one with a falling interval and one with a rising interval. This means that for each target/filler word and for each numeral there is one musical environment that corresponds to its spoken melody and one that contradicts it. Sometimes, during the course of randomization, the target word/numeral will be in a position which complements its tone and sometimes it will be in a position that will clash. While clashes are not overly frequent in
Cantonese music, they do occur (around 8% of the time according to Wong and Diehl 2002; Lau 2010 found they clashed almost 20% of the time in children’s songs). There were 18 words used: the six target words and 12 distracter words. A single repetition of the song used 9 words so two repetitions of the song constituted one block.

4.2.3 Procedure

Each participant was provided with a printed score of the song which also contained a fictitious history of the song in Chinese to provide a plausible explanation for randomization. Participants were told that in this song the words marked by the dash are traditionally chosen by the singer and that a further challenge may be added by changing the order of the numerals – counting backwards or choosing first the even numbers and then the odd numbers. They were also told that in the study both the word choice and the order of the numerals would be random.

To ensure familiarity with the written characters, participants were also given a list of 36 words in Chinese with English translations which contained the six target words, the 12 distracter words and 18 other words. They were told that the words in the song would be taken from the list. They listened to a recording of the melody as often as they wished until they felt they were comfortable enough with the song to be able to sing it. They were allowed to listen to the
recording at any time during the experiment if they felt they needed to be reminded of the melody. Verbal instructions and explanations were given in English and the instructions were repeated in written Chinese on the computer during the running of the experiment. The stimuli were presented using E-prime (Schneider et al. 2007). Each screen presented a single line of the song, showing the musical score, the words to sing and an arrow to indicate if the first interval was rising or falling (a circle indicated the final phrase). The experiment was timed to present the screens to correspond with a performance tempo of 66 beats per minute. A metronome was used to help the participants maintain the tempo. Participants were instructed to push the space bar (which initiated a single repetition of the song) in time with the metronome and count 4 beats before starting to sing. The first screen appeared on the 4th beat. Screens were timed to appear on successive 4th beats until the end of the song at which time a pause screen was presented and the next repetition could begin at the subject’s initiation. Just before beginning each song repetition the subject heard the first two notes of the song to give them the starting notes.

4.2.4 Results

The recordings were segmented and the target words exported as .wav files from the main recording using PRAAT (Boersma and Weenink 2010). The entire word was segmented out from the onset of frication to the end of periodicity. The pitch
contours were examined, the voiced portions were segmented out and several octave errors were corrected by hand. For visualization purposes, the pitch contours of the voiced sections were adjusted for target F0 – nine of the participants sang in the same F0 range but three of the female participants sang an octave higher. The results for these three participants were divided by 2 to place them in the same octave range as the others. The results were also normalized for duration: F0 values were extracted at eleven equally spaced intervals across the duration of the vowel. The data (using the original F0 measurements) were then exported into R (R Development Core Team, 2008) for statistical analysis. Raw duration values were also recorded.

A total of 381 tokens were analyzed out of a possible 432 tokens (6 tokens x 6 tones x 12 participants). Most missing tokens were lost due to melody memory errors; singers would forget the melody and either stop completely or relapse to humming to “find” the melody again. Some tokens were also lost to synchronicity errors: participants would get out of sync with the timing of the computer, lose the general thread of the song and stop singing. One singer (S13) did not sing the song quite as written but modified melody B so that the final two notes formed a descending contour. As a result, all target words for this singer are on a single note (although the numerals were sung on two different notes as written in the score).
Figure 4.2: Mean F0 values for all six tones as sung on the high note (a) and low note (b) with lines showing extrapolated slopes. Error bars are excluded for clarity. Rising tones have solid lines; level tones have dashed lines and the falling tone has a dot/dash line.

Figure 4.2 presents the mean F0 values across all singers for the beginning, middle and end of each tone. The first and last time points (0 and 11) were removed from all tokens to limit effects from both the musical and phonetic contexts. Figure 4.2a presents the values for tones sung on the higher note and 4.2b presents those for tones sung on the lower note. Chan’s (1987a) observations suggest that rising tones are distinguished from level tones phonetically by the inclusion of an upward (positive) sloping trajectory of the F0 contour, especially during the early part of the note.
Based on our prediction that rising tones should be distinguished phonetically from non-rising tones, the two rising tones were combined into one group and the three level tones into another group. Figure 4.3 shows a representation of the extrapolated slopes based on the mean F0 values for the tones grouped into three categories: rising, level and falling. The portion of the note which appears to be exhibiting the primary difference is the first half, from points one to five, so the results of the individual slope measurements for that portion were fitted to a mixed-effects model with subject as a random-effect factor and tone shape (level, falling or rising) as a fixed-effect factor; “level” was set as the intercept. When the high notes are taken separately, the slopes of the first half of the rising tones are significantly steeper than those of the level tones.
(β = 0.7750, t = 2.006). The falling tone was not significantly different from the level tones. For the tones sung on the lower note, there were no significant differences in any of the slopes. For the pooled results (both high and low notes together) the slope of the rising tones in the first half of the note was significantly different from that of the level tones (β = 0.87984, t = 2.188). Falling contours were again not significantly different from the level tones. There were no statistically significant differences for any of the slopes associated with the second half of the notes. Full statistical results are given in Table 4.2.

<table>
<thead>
<tr>
<th>a) HIGH NOTES</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLOPE OF FIRST HALF (1-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (level)</td>
<td>-0.2686</td>
<td>0.4852</td>
<td>-0.554</td>
</tr>
<tr>
<td>shape-fall</td>
<td>-0.3057</td>
<td>0.4601</td>
<td>-0.664</td>
</tr>
<tr>
<td>shape-rise</td>
<td>0.7750</td>
<td>0.3863</td>
<td>2.006</td>
</tr>
<tr>
<td>SLOPE OF SECOND HALF (5-9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (level)</td>
<td>-0.3766</td>
<td>0.3441</td>
<td>-1.094</td>
</tr>
<tr>
<td>shape-fall</td>
<td>0.2993</td>
<td>0.3360</td>
<td>0.891</td>
</tr>
<tr>
<td>shape-rise</td>
<td>-0.1119</td>
<td>0.3058</td>
<td>-0.366</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b) LOW NOTES</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLOPE OF FIRST HALF (1-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (level)</td>
<td>-0.19208</td>
<td>0.42481</td>
<td>-0.452</td>
</tr>
<tr>
<td>shape-fall</td>
<td>0.08299</td>
<td>0.29262</td>
<td>0.284</td>
</tr>
<tr>
<td>shape-rise</td>
<td>0.37774</td>
<td>0.28156</td>
<td>1.342</td>
</tr>
<tr>
<td>SLOPE OF SECOND HALF (5-9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (level)</td>
<td>0.37725</td>
<td>0.32905</td>
<td>1.147</td>
</tr>
<tr>
<td>shape-fall</td>
<td>-0.03990</td>
<td>0.19338</td>
<td>-0.206</td>
</tr>
<tr>
<td>shape-rise</td>
<td>0.04324</td>
<td>0.18038</td>
<td>0.240</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c) POOLED NOTES</th>
<th>Estimate</th>
<th>Std Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLOPE OF FIRST HALF (1-5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (level)</td>
<td>-0.25700</td>
<td>0.26672</td>
<td>-0.964</td>
</tr>
<tr>
<td>shape-fall</td>
<td>-0.09318</td>
<td>0.25375</td>
<td>-0.367</td>
</tr>
<tr>
<td>shape-rise</td>
<td>0.87984</td>
<td>0.40222</td>
<td>2.188</td>
</tr>
<tr>
<td>SLOPE OF SECOND HALF (5-9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (level)</td>
<td>0.09666</td>
<td>0.31047</td>
<td>0.311</td>
</tr>
<tr>
<td>shape-fall</td>
<td>0.01848</td>
<td>0.16655</td>
<td>0.111</td>
</tr>
<tr>
<td>shape-rise</td>
<td>-0.22750</td>
<td>0.17522</td>
<td>-1.298</td>
</tr>
</tbody>
</table>

Table 4.2: Full statistical results for slope comparisons.
Figure 4.4: Subject-by-subject graphs of rising vs. non-rising tones – target words on [si].

4.2.5 Individual Variation

Figure 4.4 presents graphs for the individual singers. These graphs present the singers’ mean F0 values for the syllables sung on the high note. As the question is
whether rising tones are distinguished from the other tones, the results for the falling tone have been combined with those for the level tones giving one curve for the rising tones and one for the non-rising tones. Examining the individual results suggests that the use of a rising contour to mark the rising tone is not a technique used by all Cantonese singers. It can be seen that the rising contour is being articulated by only some of the singers, particularly S16 but also, to a lesser degree, by S10, S12 and S21, although for S12 the curve for the rising tones shows a similar pattern to the curve for the non-rising tones. Over half of the singers appear to have a slightly more positive slope for the rising tones.

4.2.6 Discussion

The results suggest that some singers modify the slope of the F0 contour to make a distinction between rising tone contours (tones 2 and 5) and non-rising tone contours (tones 1, 3, 4 and 6) when singing in Cantonese. In other words, they include an extra rising contour when singing words that have a rising tone. The individual variation found, however, suggests that this is not a universal feature of singing in Cantonese.

The contours do not appear when the note is in the lower part of the voice range. This is probably not related to the contour in the structure of the music. Chan (1987a) shows in one of her tracings a syllable with a rising tone at the bottom of a falling musical contour where there appears to be a very clear rising
contour added in by the singer. This absence of the contour on the lower note may be similar to a phenomenon found in the intrinsic F0 of vowels. Intrinsic F0 differences show a strong tendency to disappear in the lower range of the voice (Whalen and Levitt 1995, Connell 2002). Subtle pitch contrasts seem to be minimized in the lower tessitura. The singers in this study were all required to sing the song in the same key. Although it is not possible to quantify, when listening to the recordings, the impression is that the range of the song was low for many of the singers. This suggests an interesting direction for future study of this phenomenon.

When they are used, the added contours appear to be very robust – strong enough, in fact, to show through the results of the non-contour singers in the statistical results. The mean extent of the rise at the frequency of the upper note in this song is equivalent to approximately 41 cents or nearly half a semitone. Forty-one cents in the octaves where this song is being sung is well above the pitch discrimination threshold of the average adult which, at this frequency, is about 1 Hz or 11 cents (Seashore 1967, pp. 54-55; Levitin 2006, p. 28).

Furthermore, this rising contour appears to be produced during the first part of the syllable. This appears to be a different strategy from that employed in spoken Cantonese. Wong (2006) found that tones in spoken Cantonese align to the end of the syllable. Similar to spoken Cantonese, however, the Cantonese
singers in this study do end near the target pitch, first marking the contour, then they singing the target note (i.e. starting flat and rising to the target, rather than starting at the target and going sharp).

Unfortunately, examinations for creaky voice as an indicator of tone 4 proved not feasible in this study because, as previously mentioned, for many of the singers the lower sections of the song were in the lower range of many of the singers’ voices. Many of the singers exhibited a substantial amount of creak throughout the recordings so it is not possible to tell if the creak is there for linguistic or physiological reasons.

4.3 Analysis Two – Cantonese Numerals

If, as discussed in Chapter Two, tonal mismatches do not disturb listeners, then the question arises as to whether the phenomenon observed in the preceding analysis may be due to a lack of context. Carrier phrases, although useful in many ways, eliminate the possibility of understanding from context. Is it possible that singers recognize this lack of context and add part of the tonal information, the rising contour, as a means of counter-balancing the lack of context? Would they do the same thing if context were available?
4.3.1 Stimuli

To examine this question a second analysis was done using the same recordings as the first study. In this analysis, however, the numerals were the focus of analysis. Within the song, the numerals one through nine appear in the lyrics within the phrase:

5. 第 [一] 個 字
dai [yat] go zi
‘number [one] word’

It is assumed that this phrase, in conjunction with the repetitive nature of the song, provides sufficient context for the second word to be unmistakeably interpreted as a number whether the tone is correctly realized or not.

As shown in Table 4.3, the numerals from one to nine in Cantonese provide examples of five of the six tones but focusing on the numerals meant it was not possible to control for vowel quality or syllable structure. For the purposes of this experiment, only the numerals which have open syllables were analysed:

<table>
<thead>
<tr>
<th>NUMERAL</th>
<th>TONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>一 yat</td>
<td>1</td>
</tr>
<tr>
<td>二 yi</td>
<td>6</td>
</tr>
<tr>
<td>三 saam</td>
<td>1/3</td>
</tr>
<tr>
<td>四 sei</td>
<td>3</td>
</tr>
<tr>
<td>五 ng / m</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NUMERAL</th>
<th>TONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>六 luk</td>
<td>6</td>
</tr>
<tr>
<td>七 cat</td>
<td>1</td>
</tr>
<tr>
<td>八 baat</td>
<td>3</td>
</tr>
<tr>
<td>九 gau</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.3: The tones of the numerals one through nine in Cantonese.
yi ‘two’ (tone 6), sei ‘four’ (tone 3) and gau ‘nine’ (tone 2). The tokens for the
numeral 5 ‘m’ (tone 5) were also included – this is a syllabic nasal but was
included as the four words provide examples of two level tones (tones 6 and 3,
low [2-2] and mid [3-3] respectively) and two rising contour tones (tones 5 and 2,
low rising [2-3] and high rising [3-5], respectively). The numerals with closed
syllables were excluded as they have the shortened variants of the level tones. As
all numerals appeared in a single repetition of the song, one repetition
constituted one block.

4.3.2 Results

Twelve repetitions of 4 tones by 12 participants resulted in 576 tokens; 13% of the
data was removed due to participant error. As in the previous study, participant
error resulted from memory failure or synchronicity errors.

Figure 4.5 shows the mean extrapolated pitch contours for all four tones
for all participants. For statistical analysis, the two rising tones were grouped
together and the two level tones were grouped together. The slope was taken
from an interpolated straight line from time-point 2 to time-point 5 and from
time-point 5 to time-point 8. More of the outer edges of the contours was omitted
than in the previous study in order to minimize any effect from the musical
transition curves or from possible effects of neighbouring consonants as these
factors could not be controlled. Statistical analysis of this was carried out in R
For the tones sung on the higher note, there was a main effect of level vs. contour for the slope between points 2 and 5 ($\beta = -2.518; t = -2.271$). For tones sung on the lower note there was again no significant difference between the slopes for the rising tones and those of the level tones for the first half of the note but they did show significance in the second half of the note ($\beta = -0.5523; t = -2.296$). For the pooled notes there was a marginally significant difference between the slopes of the rising tones and the level tones for the first half of the note ($\beta = -1.5799; t = -1.986$) but no significance for the second for the second.
Figure 4.6: Subject-by-subject graphs of rising vs. non-rising tones – numerals.

4.3.3 Individual Variation

Figure 4.6 shows the individual graphs for all singers. The results for singer S16 include only one curve as, due to the randomization process, all of her level tone numerals occurred only on the lower note. As with the target word results, a range of patterns appears: subjects S10 and S11 both exhibit strong differences in
slope between the rising and level tones. Several other subjects (S13, S14, S18, S21) exhibit relatively flat contours for the level tones that are not necessarily reflected by the extrapolation of the slope from point 1 to point 5 (e.g., S14, S15, S18, S21). It is unclear if the frequently lower value for point 1 is due to the tone contour or to the fact that the different syllables had different onsets. Although it was not possible to control for such factors the results are consistent with the findings of the previous analysis. The fact that there is context to disambiguate the numbers does not seem to prevent some singers from including the rising contour when they sing.

4.3.4 Discussion

The results of this analysis appear very similar to those of the first analysis. Some singers in Cantonese include the rising contour marker for rising tone even when context could be expected to disambiguate any potential misunderstanding. The patterns found match those of the previous experiment and support the observations of Chan (1987a) and Yung (1983) that Cantonese singers produce the rising contours of rising tones in the first part of the syllable. They also mirror the findings of the first analysis where the effect is not universal: it appears to be a strategy employed by some singers but not by all. It is also found only on the upper note. The effect does not seem to occur in the lower part of the voice range.
4.4 Analysis Three – Duration

Although duration in music is usually dictated by the song writer/composer, not unlike pitch, variations of duration are one of the primary markers of “expressiveness” (Kendall and Carterette 1990), and are thus somewhat susceptible to influence by extramusical factors. Ladd and Remijsen (in progress) propose that the retention or even exaggeration of secondary features of tone in singing may be used to represent tone in singing. One possibility for this in Cantonese is duration. The durational correlates of tone in Cantonese have been well documented (Kong 1987). The six tones cannot be ranked in a definite durational hierarchy but Kong (1987) found patterns of duration across speakers: tone 2 (high rising, 35) is the longest tone and tones 1 (high level, 55) and 4 (low falling, 21) are the shortest. Of the three level tones, tone 3 (mid-level, 33) is the longest. This section looks at the duration measures of the sung minimal tonal set to see if the durational correlates of tone are maintained in singing.

4.4.1 Methodology and Results

The same recordings were used as in the previous two analyses. The target words were the six-way tonal minimal set on [si] given earlier in Table 4.1 and used in the first analysis. The recordings were segmented and the target words extracted from the original recording using PRAAT (Boersma and Weenink 2010).
Duration measures were for the duration of the voiced portion of the word from the offset of frication to the end of periodicity. The duration measures of each occurrence of the target words were extracted by PRAAT script and exported into R (R Development Core Team 2008). A total of 361 tokens were analyzed out of a possibility of 396 tokens (6 tokens x 6 tones x 11 participants). Figure 4.7 plots the mean duration measures of each tone across participants. The mean duration of all the tones was 908.31 ms (sd=103.15). Kong (1987) found the mean duration of the voiced portion of the syllable [si] when spoken in Cantonese to be 252.39 msec (sd=22.55). A repeated measures ANOVA was conducted to compare the effect of tone on duration. There was no significant effect. Similarly, when comparing only the three level tones, no significant difference was found.

Figure 4.7: Mean durations of tones on the voiced portion of the sung syllables [si] with standard deviations.
4.4.2 Discussion

It appears that singers in Cantonese do not carry the durational differences of spoken tones over to singing. A visual examination of Figure 4.7 indicates that the patterns found by Kong (1987) are not reflected in the singing data. The singing data does not even trend towards the patterns Kong (1987) found for spoken data. Tone 2, which Kong (1987) found to be the longest tone, is almost the same length as tone 1, which he found to be the shortest. Similarly, tone 3, which Kong (1987) found to be longest of the three level tones is here the shortest of the three. The duration differences associated with tone in Cantonese appear to be completely neutralized in singing.

4.5 Conclusions

The results suggest that rising contour appears to be a component of tone that is transferred over to singing in Cantonese, but only by some singers. As register is usually represented musically in Cantonese (Wong and Diehl 2002), the addition of a contour by the singer may help to distinguish tones further. It does not appear that contour information is used to correct mismatches between the sung and spoken melody – the rising contours are present even when the song melody matches the spoken melody. If the singers were only correcting mismatched speech and song melodies, there should be a marked difference between tones 2
and 5: there should be a correction only on the higher note for tone 5 and on the lower note for tone 2. No distinction appears between the two rising tones in singing. It is possible that voice quality could be used as a marker of tone 4 (the low falling tone) but the pitch range of the song used for this study precludes examination of that question. The low range of the song also appears to have affected the realization of the rising contours in the lower part of the singers’ voices. The realization of contour appears to be similar to the intrinsic F0 of vowels in that it is a phenomenon that does not surface in the lower range of the voice.

Cantonese music reflects tone both structurally and phonetically. The following chapter explores the proposal that Mandarin music handles tone somewhat differently.
Chapter Five

Phonetic Realization of Tones in Sung Mandarin

5.1 Introduction

Modern Mandarin music reflects very little of the speech melody structurally (Chan 1987a, Chen 2007, Vondenhoff 2009) although Wee (2007) suggests a tendency for it to do so in positions of musical prominence. Little is known of the phonetic manifestation of tone in Mandarin music. Chao (1956, p. 57) suggests that Mandarin singers may “smuggle in” tone. It is assumed that he refers to some kind of phonetic manifestation of tone similar to that found for Cantonese in the previous chapter. This chapter provides the results of an experiment extending the analysis used in the Cantonese singing studies of the previous chapter to Mandarin to examine if Mandarin singers include tonal information while singing. As with the previous chapter, the analysis in this chapter will also include an examination of whether Mandarin singers maintain the durational correlates of tone when singing.
5.2 Methodology

5.2.1 Participants

Participants were 10 native speakers of Mandarin (6F, 4M; mean age = 21.5, sd = 3.37) who were all residents of Vancouver. They were all fluent in English, as well. Five participants were choral singers; one was not a choral singer but had taken singing lessons for 5 years; three participants had neither lessons nor choral experience but reported singing on their own for at least one hour per week and one subject did not provide information about his singing experience.

5.2.2 Stimuli

As previously stated in Chapter One, Mandarin has four tones. The target stimuli for this study were a minimal set of shi [ʂi] on all four tones, given in Table 5.1; the target syllable is the second member of a compound word where the first syllable carries tone 2 except for the tone two word (昔時 xi⁵⁵ši³⁵ ‘time’) which has a tone one/tone two sequence. These syllables were included in the 4 stanzas

<table>
<thead>
<tr>
<th>Tone</th>
<th>Character</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>wúshí</td>
<td>1 high level 55 吾師</td>
<td>‘teacher’</td>
</tr>
<tr>
<td>xíshí</td>
<td>2 rising 35 昔時</td>
<td>‘time’</td>
</tr>
<tr>
<td>qíngshí</td>
<td>3 fall-rise 214 情史</td>
<td>‘love history’</td>
</tr>
<tr>
<td>chéngshí</td>
<td>4 falling 51 城市</td>
<td>‘city’</td>
</tr>
</tbody>
</table>

Table 5.1: Target words (second syllable only).
of a specially written poem which was then set to music. The poem was written by Chenhao Chiu and Yuan Lu, both trained linguists and both native speakers of Mandarin. The Chinese lyrics can be found in Figure 5.1; the English translation of the lyrics is as follows:

1. 當我沿著路走
   Dāng⁵⁵ wo²¹⁴ yan³⁵ zhe lu⁵¹ zou²¹⁴
   As I was walking down the road
   我看不見四周
   Wo²¹⁴ kan³¹ bu³⁵ jian⁵¹ si⁵¹ zhou⁵⁵
   I couldn’t see what was around me
   疲憊極致
   Pi³⁵ bei⁵¹ ji³⁵ zhi⁵¹
   I was so tired
   我張眼卻看到城市
   Wo²¹⁴ zhang⁵⁵ yan²¹⁴ que⁵¹ kan⁵¹ dao⁵¹ cheng³⁵ shi⁵¹
   I opened my eyes and saw the city

2. 當我沿著街走
   Dāng⁵⁵ wo²¹⁴ yan³⁵ zhe lu⁵¹ zou²¹⁴
   As I was walking through the city
   我看不見四周
   Wo²¹⁴ kan³¹ bu³⁵ jian⁵¹ si⁵¹ zhou⁵⁵
   I couldn’t see what was around me
   孤單極致
   Gu⁵⁵ dan³⁵ ji³⁵ zhi⁵¹
   I was so alone
   我張眼只看到吾師
   Wo²¹⁴ zhang⁵⁵ yan²¹⁴ zhi²¹⁴ kan⁵¹ dao⁵¹ wu³⁵ shi⁵¹
   I opened my eyes and saw the teacher

3. 當我陪伴他走
   Dāng⁵⁵ wo²¹⁴ pei³⁸ ban⁵¹ ta⁵⁵ zou²¹⁴
   As I was walking with him
   我看不見四周
   Wo²¹⁴ kan⁵¹ bu³⁵ jian⁵¹ si⁵¹ zhou⁵⁵
   I couldn’t see what was around me
想念極致
Xiang\textsuperscript{214} nian\textsuperscript{51} ji\textsuperscript{35}zhi\textsuperscript{51} 
I miss her so much

我張眼即看到情史
Wo\textsuperscript{214} zhang\textsuperscript{55} yan\textsuperscript{214} ji\textsuperscript{35} kan\textsuperscript{51} dao\textsuperscript{51} qing\textsuperscript{35} shi\textsuperscript{214} 
I opened my eyes and saw my love history

4. 當我向過往走
Dang\textsuperscript{55} wo\textsuperscript{214} xiang\textsuperscript{51} guo\textsuperscript{51}wang\textsuperscript{214} zou\textsuperscript{214} 
As I was walking through the past

我看見了四周
Wo\textsuperscript{214} kan\textsuperscript{51}jian\textsuperscript{51} le\textsuperscript{51} si\textsuperscript{51}zhou\textsuperscript{55} 
I started to see what was around me

恐懼極致
Kong\textsuperscript{214} ju\textsuperscript{51}ji\textsuperscript{35}zhi\textsuperscript{51} 
I am so afraid

我閉眼卻揮不去昔時
Wo\textsuperscript{214} bi\textsuperscript{51} yan\textsuperscript{214} qu\textsuperscript{51} hui\textsuperscript{55} bu\textsuperscript{35}qu\textsuperscript{51}xi\textsuperscript{55}shi\textsuperscript{35} 
I closed my eyes but only saw my time.

The target syllables appear as the second half of the final word of each stanza: city, teacher, love history, time. Modern Mandarin words are usually bisyllabic so it was not possible to create a completely identical carrier phrase that would also work in the context of a song, but the word choice was controlled as closely as possible. The preceding syllable has tone two with the exception of the last word (昔時 ‘time’) which is a tone one/tone two combination and the sense of the final line of each stanza, the carrier sentence, was kept fairly constant.

The music was composed by the author, a linguist with training in composition.

The song was written in the style of shidai qu (時代曲), also known as guoyu laoge (國語老歌) – Mandarin popular songs produced in Shanghai in the 1930s.
Figure 5.1: Score of the song used for the Mandarin study.
and ’40s – a style of music which is still well known and quite popular.\textsuperscript{19} It is strongly formulaic in structure; Chapter 5 of Chen (2007) provides a clear articulation of the formula and this was used as the template and guide for the composition of the music. The completed song was evaluated by four native speakers of Mandarin who were played the song and judged it to be comparable to the original style. The score is given in Figure 5.1.

5.2.3 Procedure

A mock-karaoke system was set up using E-prime (Schneider et al. 2007). Two lines of the song were displayed at a time on the computer screen while a recorded musical track played the melody with an accompaniment over computer speakers placed just behind the monitor. As each character on the screen was to be sung, it changed colour indicating to the subject when to sing that word. The next two lines of the song appeared just prior to their occurrence in the musical track. Instructions were printed in Chinese on the computer screen prior to the recording session and a pause screen was shown at the end of each repetition of the song. Each repetition was started by the subject at his/her discretion by pressing the space bar on the computer keyboard.

\textsuperscript{19} The popularity and familiarity of this music can be compared with that of big band music in North America, a genre with which it shares many common inspirations and characteristics.
Participants were first presented with a printed version of the song (identical to Figure 5.1) and listened to the musical track while watching the words on the computer screen. They were allowed to repeat this process as often as they wished until they felt comfortable enough with the song to be able to sing it. They were then recorded both speaking and singing the song 6 times each with a break between each repetition.

Participants were recorded at a sampling rate of 44,100 Hz using an AKG C520 head-mounted directional microphone and a Sound Devices USBPre pre-amp. The speakers playing the musical track were placed behind the microphone which either strongly limited the recording of the musical track or failed to pick it up altogether. Recordings were made in Audacity (Audacity Development Team 2006) on a Macintosh Classic notebook and saved as .wav files.

5.2.4 Analysis

The recordings were segmented and the target words exported as .wav files from the main recording using PRAAT (Boersma and Weenink 2010). The entire word was segmented out from the onset of frication to the end of periodicity. The pitch contours were examined, the voiced portions were segmented out and several pitch errors were corrected by hand. The results were also normalized for duration: F0 values were extracted at eleven equally spaced intervals. The data
were then exported into R (R Development Core Team 2008) for statistical analysis. Raw duration values were also recorded.

5.3 Results

Mean contours normalized for duration are shown in Figure 5.2; the first and last points were not included in the analysis to limit influences from the onset consonant and phrase final position. Error bars have been omitted for clarity but standard deviations are given in Table 5.2. The extraordinarily high F0 for point 9 of tone 1 appears to be due to two of the singers (S04 and S05). Both of these singers have a very wide vibrato which appears to have affected some of the
For each individual token, two slope measurements were computed: early slope (from time points 1 to 5) and mid slope (from time points 3 to 7). To minimize the effect of vibrato, the slopes were extrapolated; that is, they were all treated as if they were straight lines.

The distribution for these results was reasonably normal so a repeated measures ANOVA was conducted to compare the effect of tone on slope. There was no significant effect ($F(2,5)=0.094$, $p=0.912$).

In the interest of consistency, these results were also fitted to a mixed-effects model similar to that used in the previous study with subject as a random-effect factor and tone as a fixed-effect factor; to achieve comparisons across all

<table>
<thead>
<tr>
<th>POINT</th>
<th>TONE 1</th>
<th>TONE 2</th>
<th>TONE 3</th>
<th>TONE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.33</td>
<td>1.77</td>
<td>2.83</td>
<td>1.13</td>
</tr>
<tr>
<td>2</td>
<td>3.54</td>
<td>1.98</td>
<td>1.77</td>
<td>0.49</td>
</tr>
<tr>
<td>3</td>
<td>0.07</td>
<td>3.46</td>
<td>0.99</td>
<td>0.42</td>
</tr>
<tr>
<td>4</td>
<td>0.28</td>
<td>1.98</td>
<td>0.85</td>
<td>0.92</td>
</tr>
<tr>
<td>5</td>
<td>1.63</td>
<td>2.40</td>
<td>1.06</td>
<td>0.07</td>
</tr>
<tr>
<td>6</td>
<td>2.62</td>
<td>1.20</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>7</td>
<td>1.34</td>
<td>1.48</td>
<td>2.69</td>
<td>1.20</td>
</tr>
<tr>
<td>8</td>
<td>2.19</td>
<td>0.42</td>
<td>2.90</td>
<td>0.99</td>
</tr>
<tr>
<td>9</td>
<td>2.19</td>
<td>2.55</td>
<td>0.57</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Table 5.2: Standard deviations (Hz) for mean F0 levels of normalized pitch contours.
combinations the intercept was set variously as tone 1, 2 and 4. No significant interactions were found.

Following the Cantonese study in the previous chapter where the tones conflated down into three groups depending on the contour of the tones, it is possible to group the Mandarin tones into two groups: tones 1 and 2; and tones 3 and 4. This grouping is a traditional grouping referring back to the Classical Chinese even/oblique categorization (Chan 1987a, Mark and Li 1966). Chan (1987a) hypothesized that it may play a role in Mandarin music but found no evidence of it playing a role in the songs she analysed and Wee (2007) used it to

Figure 5.3: Mean contours of Mandarin tones condensed into two groups. Error bars are omitted for clarity but standard deviations are printed on the graph. The numbers above the graph are the standard deviations for the upper register and those below the graph are for the lower register.
make a register-style division. Figure 5.3 gives the mean F0 values for the Mandarin tones condensed into these two registers. The results were fitted to a mixed effects model with subject as a random-effect factor and tone as a fixed-effect factor. No significant interactions were found.

Figure 5.4: Subject by subject graphs for four tones (Mandarin).
5.3.1 Individual Variation

Figure 5.4 provides individual plots for each subject showing the contours for all four tones. There appears to be little patterning although singers S08 and S12 appear to have different levels for each tone. To examine this, a repeated-measures ANOVA was run (for all subjects) with tone as the independent variable and the midpoint F0 measure (time point 6) as the dependent variable to see if there was a significant difference between the fundamental frequencies of each tone. This was not statistically significant.

Figure 5.5 shows the individual graphs for the tones conflated into two groups. Here, again, there does not appear to be individual strategies; the contours are quite similar. Singer S04, whose samples cover a very large range, has an extremely high final F0 for the last time point of tone 1. She was one of the two singers in the study who was a choral singer and was also one of only two singers in the study who had received formal voice training (the other was singer S09). She had a very noticeable vibrato which gradually increased over the course of the notes she sang. Singer S05, also a choral singer but with no formal voice training, also exhibited a similar vibrato although the pitch modulation was not as extreme as singer S04.
5.3.2 Duration

The durational correlates of tone in Mandarin are known to be quite robust. Studies of the durational differences associated with tone in Mandarin have shown tone 4 to be the shortest and tone 3 to be the longest with tones 1 and 2 falling in between (Xu 1998, Chang and Yao 2007). Li and Guo (2012) found that the durational correlates of tone were not only maintained in whispered speech.
Table 5.3: Means of raw duration scores.

<table>
<thead>
<tr>
<th>TONE</th>
<th>MEAN (MSEC)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1418.474</td>
<td>263.8451</td>
</tr>
<tr>
<td>2</td>
<td>1080.955</td>
<td>202.5875</td>
</tr>
<tr>
<td>3</td>
<td>1349.826</td>
<td>218.0666</td>
</tr>
<tr>
<td>4</td>
<td>1400.964</td>
<td>262.5157</td>
</tr>
<tr>
<td>all</td>
<td>1313.485</td>
<td>272.961</td>
</tr>
</tbody>
</table>

in Mandarin but exaggerated. This section examines whether they are maintained in singing as well. The mean raw duration scores are given in Table 5.3. A repeated measures ANOVA was conducted to compare the effect of tone on duration. A significant effect was found ($F(3,36) = 5.669, p = 0.00275$). A post hoc Tukey’s HSD test determined that there were significant differences between the durations of tones 1 and 2 ($p = 0.0057$); tones 2 and 4 ($p = 0.0072$); and tones 2 and 3 ($p = 0.0263$). All other pairings were not significant.

5.4 Discussion

The results from the slope analysis suggest that singers in Mandarin do not include tonal contour information while they are singing. While the casual listener may observe contour changes over the course of a sung syllable in Mandarin it is most likely that these are present for musical rather than linguistic reasons.

The duration results show that tone two is shorter than the other three tones but the setting of this syllable in the music is on a note of shorter duration.
(bar 32, second note), so this difference is most likely a result of the music and not of language. The musical duration of the note on which the tone 2 syllable is set (a dotted quarter note) should be 75% of the duration of the note on which the other three tones are set (a half note). The mean duration for tone 2 given in Table 5.3 is 77.8% of the mean of the other three tones (1389.75). To test this, the durations were normalized by multiplying all the non-tone-2 measures by 0.75 and re-running the statistics. With the normalized durations there were no significant differences found confirming that the durational distinctions of spoken tone are neutralized in sung Mandarin.

Although all four of the target words were sung on the same musical note, a visual examination of Figure 5.4 might suggest that, for some singers, they may have different fundamental frequencies. However, this proved not to be a statistically significant difference. Furthermore, the pattern does not match the F0 levels found in spoken Mandarin tones. Nor does it match the levels of the Zhongzhou or “Central” dialect used in Beijing Opera where tone 1 = 44; tone 2 = 11; tone 3 = 53 or 55; and tone 4 = 24 or 424 (Chan 1987a).

5.5 Conclusions

The results of this examination of Mandarin singing suggest that singers in Mandarin do not mark tone phonetically, either by contour or by duration. This is in contrast with Cantonese singers who include missing tonal contours. It
appears that there is very little of the tone information transferred over to music in Mandarin singing. These studies have looked only at two aspects of tone: slope and duration. The next study looks at whether the words produced by the singers in these studies are actually recognizable to listeners.
Chapter Six

Perception of Tone in Singing

6.1 Introduction

The previous two chapters found that Mandarin singers do not sing tonal elements of words but that some Cantonese singers include a rising contour when singing syllables that contain a rising tone. Of course, it is possible that there are cues other than those measured in these studies that convey tone information to listeners, raising the question as to whether listeners in these languages can recognize the individual words produced. Wong and Diehl (2002) showed that Cantonese listeners use the musical contour to interpret ambiguous sung sentences. Vondenhoff (2009) found that Mandarin listeners did not. These studies used carrier phrases sung (by a single singer) to differing melodies. Listeners would hear the different phrases and choose which word they thought they heard. This is an excellent way to examine the question of structural manifestations of tone incorporated by composers but leaves open the question of phonetic manifestations produced by singers. The songs used in the previous two chapters provide material which can be used to test whether or not listeners are able to interpret the tones in individual words sung in these two languages. Two experiments were carried out to determine whether listeners could
recognize the words produced by the singers, one for Mandarin and one for Cantonese.

6.2 Mandarin

6.2.1 Participants

Thirteen listeners (7F, 6M; mean age 24.6, sd=2.47) participated. All were native speakers of Mandarin living either in Vancouver or in Hong Kong.

6.2.2 Stimuli

Stimuli consisted of audio recordings of a four-way minimal set on the syllable [ɕi] – spoken and sung words – taken from the previous study. There were 10 different speakers/singers with six repetitions of each word from each of the two conditions: sung and spoken. There were six tokens missing in the original, so six usable tokens were doubled to make up the missing complement. Two of these were sung tokens (one each for tone two and tone three) and four were spoken tokens (one for each of the four tones); these were replaced with tokens of the same tone by the same speaker. This provided a total of 480 tokens (4 tones x 6 repetitions x 2 conditions x 10 speakers). A coding error in the script used meant that one sung token for tone 2 for all singers was inadvertently omitted and replaced by a random token from the same singer during the running of the experiment. This resulted in an uneven distribution of tones for the overall sung
tokens (tone 1 = 809, tone 2 = 684, tone 3 = 814, tone 4 = 813). This did not create a
large enough disparity to affect the statistical results.

6.2.3 Procedure

The experiment was run in the Interdisciplinary Speech Research Laboratory at
University of British Columbia and in the phonetics laboratory of the Chinese
University of Hong Kong using E-Prime (Schneider et al. 2007). Participants were
given the choice of using either Traditional or Simplified Chinese, written
instructions being provided on the computer. The two forms of the experiment
were otherwise identical. The experiment involved a four-way forced choice:
four characters appeared in random placement on the screen; the subject heard
the stimulus once over headphones and had to make a choice by clicking with
the mouse on the appropriate character. Once the subject made a choice the
process was repeated until all the tokens for a given speaker had been played.
Participants always heard one speaker as a single block, always with the spoken
tokens first (to acclimatize the subject to a speaker’s pronunciation) followed by a
short break and then the sung tokens were presented. Order of presentation of
speakers was randomized across participants, as was the presentation of tokens
within each block. The results were exported to R (R Development Core Team
2008) for analysis.
Table 6.1: Confusion matrices for both sung (a) and spoken (b) conditions of the Mandarin study. ‘Tone’ is the tone of the word produced and ‘Resp’ is the response.

6.2.4 Results

Listeners made correct choices as to which tone they heard 9% of the time in the sung condition and 41% of the time in the spoken condition (chance is 25%).

Confusion matrices for the stimuli and responses for both the sung condition and the spoken condition are given in Table 6.1. The numbers across the top of each table are the tones of the target words produced by the singers/speakers in the experiment in Chapter Four. The numbers down the left side are the tones chosen by listeners in this study. The number in each cell is the number of times a listener chose the tone on the left when hearing the tone produced (across the top). Figure 6.1 gives a graphical representation of the identification rates for each tone. Each set of four bars in Figure 6.1 represents a tone produced by the
Pearson chi-squared tests of independence were run for both matrices. The Pearson chi-squared test of independence compares two vectors – in this case the intended tone as produced by the singers, and the perceived tone as
chosen by the listeners – and determines if there is a relationship between the
two. The null hypothesis of this test is that the two vectors are independent. The
findings for the spoken condition were significant ($\chi^2(9, n = 3120) = 1568.5$, $p < 0.0001$) while those for the sung condition were not. The Pearson chi-squared test of independence only indicates if the two vectors are independent or not. If they are found to be dependent it is necessary to run a post-hoc test to examine where the interactions occur.

Table 6.2 shows the results of a standardized Pearson residual cell-wise post hoc analysis for the spoken condition. This is carried out for tests of independence which are found to be significantly related (i.e., dependent) to determine the contribution of each pairing to the overall result. The post hoc test is a cell-wise assessment of whether the absolute value of the standardized Pearson residual (the difference between the expected frequency and the observed frequency) is significantly higher than the standard residual minimum for a given probability, $p < 0.001$ in this case (Arppe 2012). If the value for the individual cell is greater than the minimum value then the value of that cell is significantly higher and is marked with a ‘+’ sign; that cell contributes more than its “fair share” and its contribution is significant. Similarly, those cells marked with ‘–’ are significantly lower and their contribution is significantly less (Arppe 2012).
Table 6.2: Post hoc cell-wise contributions for the spoken Mandarin condition. Cells marked + indicate a standard residual value for that cell which is significantly higher than the overall minimum standard residual value and those marked – are significantly lower; p < 0.001.

6.2.5 Discussion

The chi-squared test results support the null hypothesis that in the sung condition for Mandarin the tones produced by the singers and those perceived by the participants are independent of each other; the tones perceived are not dependent on the tones produced. For the speaking condition, conversely, the test of independence supports the hypothesis that they are not independent of each other but that they are related or dependent. The post hoc test in Table 6.2 shows that the cells whose contributions are significantly higher are those corresponding to the tones being correctly identified – although there is some confusion with tone 1 being perceived as either tone 2 or tone 3. Generally speaking, as expected, identification of spoken syllables is more accurate than that of sung syllables.

The results for the singing condition as shown in Figure 6.1a are significantly different from those for the speaking condition. Tone three was the
favoured choice far more frequently, regardless of what tone was produced. The
listeners were very consistent in choosing tone 3 – but why tone 3? Listeners do
not appear to be choosing the most frequently occurring word. In a search in the
Academia Sinica Balanced Corpus of Modern Chinese (Chen et al. 1996) – a five
million word annotated corpus of Chinese available on the internet
(http://db1x.sinica.edu.tw/kiwi/mkiwi/) – the character 史 (tone 3) has the lowest
frequency count (86) compared with the other three characters. The frequency
counts are given in Table 6.3.

Nor is tone three the most frequently occurring tone in Mandarin. Wan
and Jaeger (1998) found that tone four was the most frequently occurring tone in
Mandarin with tone three occurring with almost equal frequency with tones one
and two as shown in Table 6.4.

Tone three is, however, the longest tone (Xu 1998). What may be
happening is that listeners, unsure of what tone is intended, simply register the
fact that the duration is exceptionally long (the mean duration for the sung
words was 1313.5 msec [sd = 273.0]) and choose the tone based on that.
<table>
<thead>
<tr>
<th>Tone</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>tone 1</td>
<td>20.5%</td>
</tr>
<tr>
<td>tone 2</td>
<td>19.5%</td>
</tr>
<tr>
<td>tone 3</td>
<td>19%</td>
</tr>
<tr>
<td>tone 4</td>
<td>32%</td>
</tr>
<tr>
<td>neutral tone</td>
<td>9%</td>
</tr>
</tbody>
</table>


The identification rates for speaking shown in Figure 6.1(b) show that listeners identify words with tones three and four fairly accurately but they have greater difficulty identifying tones one and two. Mandarin listeners are known to have difficulty differentiating tones two and three (Shen and Lin 1991, Hume and Johnson, 2003, Huang and Johnson, 2010). It has been argued that the partial neutralization of these two tones due to tone sandhi makes their differentiation more difficult for Mandarin speakers (Hume and Johnson 2003).

The confusion between tones one and three is more difficult to understand. The phrase-final position of the tokens may have allowed for a drop that the listeners perceived as tone. Cao and Sarmah (2007) found that if the turning point (the lowest extreme of the pitch trajectory) for a simulated tone three was extended beyond 67.5% of the total duration of the tone Mandarin speakers perceived it as tone one, so a change in pitch due to phrasal intonation over the course of the token may have been perceived as tone information when heard in isolation.

---

20 A small class of words in Mandarin are inherently toneless and acquire the tone of the word immediately preceding it. These are known as words with neutral tone.
6.3 Cantonese

6.3.1 Participants

Ten participants (8M, 2F, mean age 21.3, sd = 4.16) participated. All were native speakers of Cantonese living either in Vancouver or in Hong Kong.

6.3.2 Stimuli

Stimuli consisted of audio recordings of a six-way minimal set on the syllable [si]. There were 6 different speakers/singers with five repetitions of each word. The Cantonese study presented in Chapter Three only recorded sung tokens and, due to the various memory issues detailed in Section 4.2.4, not all subjects produced five tokens of each tone. Five participants had almost complete sets of sung tokens and their stimuli were chosen for the perception study. Two of these participants recorded corresponding spoken stimuli consisting of the target words embedded in the same carrier phrase used in the sung study at a later date and these were used as spoken stimuli. An additional female native speaker of Cantonese was also recorded both singing and speaking the target words in the carrier phrase using the same melody as in the previous experiment to provide an equal number of male and female speakers (3M/3F). The target words were extracted from the carrier phrase using PRAAT (Boersma and Weenink 2010) from the onset of frication to the end of periodicity. There were eight tokens missing in
the original so eight usable tokens were repeated to make up the missing complement. These were all sung tokens. There were six missing tone three tokens from two singers and two missing tone four tokens from one singer. The missing tokens were replaced with tokens of the same tone by the same speaker. There were five repetitions of six tones for all six speakers (180 tokens) for singing and five repetitions of six tones for three speakers for the spoken tokens (90 tokens) for a total of 270 tokens. The song used in the Cantonese study had two possible notes on which the target word was sung, one high and one low. There were 82 target words sung on the higher note and 98 sung on the lower.

6.3.3 Procedure

The experiment was run in the Interdisciplinary Speech Research Laboratory at The University of British Columbia and in the phonetics laboratory of the Chinese University of Hong Kong using E-Prime (Schneider et al. 2007). The experiment was a six-way forced choice: the six characters appeared in random placement on the screen; the subject heard the stimulus once over headphones and had to make a choice by clicking with the mouse on the appropriate character. Once the subject made a choice the process was repeated until all the tokens for a given speaker had been played. Participants always heard one speaker as a single block; the speakers with both spoken and sung stimuli were heard first followed by the speakers with only sung stimuli. For the three
speakers who had both spoken and sung stimuli, the participants were always presented with the spoken tokens first (to acclimatize the subject to a speaker’s pronunciation) followed by a short break, and then the sung tokens were presented. Presentation of speakers was randomized across listeners; presentation of tokens within each block was also randomized. The results were exported to R (R Development Core Team 2008) for analysis.

### a. Sung

<table>
<thead>
<tr>
<th>TONE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>113</td>
<td>109</td>
<td>84</td>
<td>108</td>
<td>91</td>
<td>143</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>32</td>
<td>17</td>
<td>15</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>71</td>
<td>54</td>
<td>67</td>
<td>64</td>
<td>55</td>
<td>53</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>20</td>
<td>26</td>
<td>21</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>27</td>
<td>19</td>
<td>16</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>81</td>
<td>58</td>
<td>87</td>
<td>76</td>
<td>81</td>
<td>65</td>
</tr>
</tbody>
</table>

### b. Spoken

<table>
<thead>
<tr>
<th>TONE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>130</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>110</td>
<td>2</td>
<td>3</td>
<td>39</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>6</td>
<td>72</td>
<td>3</td>
<td>11</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>4</td>
<td>12</td>
<td>137</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>21</td>
<td>7</td>
<td>1</td>
<td>94</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>9</td>
<td>47</td>
<td>6</td>
<td>4</td>
<td>72</td>
</tr>
</tbody>
</table>

Table 6.5: Confusion matrices for both sung (a) and spoken (b) conditions of the Cantonese study. ‘Tone’ is the tone of the word produced and ‘Resp’ is the response.
6.3.4 Results

Listeners made correct choices as to which tone they heard 18% of the time in the sung condition and 68% of the time in the spoken condition (chance is 16.7%).

Confusion matrices for the stimuli and responses for both the sung condition and
the spoken condition are given in Table 6.5. Figure 6.2 gives a graphical representation of the identification rates for each tone showing the breakdown of how many responses for each tone were given for each target tone. Pearson chi-squared tests of independence were run for both matrices. The findings for the sung condition were significant ($\chi^2(25, n = 1800) = 76.78, p < 0.001$) as were those for the spoken condition ($\chi^2(25, n = 900) = 2005.9, p < 0.001$). In neither of these conditions are the produced tones and the perceived tones independent. The results of the post hoc Pearson residual cell-wise analyses for both of these are shown in Table 6.6. The post hoc cell-wise assessment for the spoken conditions

**a. singing**

<table>
<thead>
<tr>
<th>Tone</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

**b. speaking**

<table>
<thead>
<tr>
<th>Tone</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>+</td>
<td></td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 6.6: Post hoc cell-wise contributions for the sung (a) and spoken (b) Cantonese conditions. Cells marked + indicate a chi square value for that cell which exceeds the overall minimum Pearson residual value for $p < 0.001$. 

132
shows correctly recognized tones are significant contributors, tones three and six are recognized as each other and tone five is recognized as tone two significantly often. The post hoc cell-wise assessment for the sung condition reveals higher than minimal residuals for tone six perceived as tone one, tone two perceived as tone two, tone three perceived as tone four, tone five perceived as tone five, and tone five perceived as tone two (6>1, 2>2, 3>4, 5>5, 5>2). There are significantly lower values for five perceived as one, three perceived as one, two perceived as six, one perceived as five, and six perceived as five (5>1, 3>1, 2>6, 1>5, 6>5).

Figure 6.3: Identification rates for tones by note sung. The set of bars on the left (H) represent all the syllables sung on the higher note and those on the right (L) those sung on the lower note. The individual bars within each set show the number of times words on that note were identified as a particular tone.
6.3.5 High versus Low Notes

In the Cantonese song used as the stimulus, the target words were sung on two different pitches. Results for responses according to the note sung are given in Table 6.7 and Figure 6.3. A Pearson chi-squared test of independence was found to be significant ($\chi^2(5, n = 1170) = 587.9, p < 0.001$). Post hoc cell-wise assessment found that the cells corresponding to tone 1 on a high note, tone 3 on a low note, and tone 6 on a low note were significant positive contributors ($p < 0.001$).

6.3.6 Discussion

The chi-squared test results suggest that in both conditions for Cantonese there is a relationship between what tone the singer sings and what tone the listener perceives. The post hoc tests in Table 6.5 suggest that distinctions in the spoken condition are fairly clear – the cells that contribute significantly in the positive direction are those that correspond to the tones being correctly identified. In fact,

<table>
<thead>
<tr>
<th>Tone</th>
<th>H</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>389</td>
<td>41</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>43</td>
<td>179</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>59</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
<td>252</td>
</tr>
</tbody>
</table>

Table 6.7: Confusion matrix for tone and high or low note. The column labelled H enumerates the tokens sung on the higher note and L those sung on the lower note. The rows show the counts for each of the tones.
those are the cells that contribute significantly with only three others contributing positively. There appears to be some confusion between tones 3 and 6 – the low level tone and the mid level tone, which are very similar in terms of F0. There is also some confusion between the two rising tones, tones 2 and 5.

The results of the Pearson’s residuals cell-wise analysis for the sung condition suggest that there is more confusion in that condition. What is interesting to note is that the pairs that contribute positively in terms of Pearson’s residuals for singing pair off according to rising versus non-rising tone groupings. Those pairs that contribute significantly in the positive direction and that are either the rising tones (2>2, 5>5, 5>2) or non-rising tones (6>1, 3>4). Those pairs whose contributions are significantly lower and which can be assumed to be less frequently confused are with only one exception (3>1), pairs that consist of one rising tone and one non-rising tone (5>1, 2>6, 1>5, 6>5). This suggests that Cantonese listeners are confusing the two rising tones with each other and also confusing the non-rising tones with each other. They do not seem to be confusing tones between the rising and non-rising groups. These are the same groupings that were distinguished by the singers in the experiment in Chapter Three.

The results from the high versus low note analysis suggest that the pitch of the note is used to identify the level tones. The high note is significantly more often identified as tone one, the high level tone. Conversely, the lower note is
significantly more often identified as either tone three (mid level) or tone six (low level). The results of the post hoc cell-wise analysis suggest that rising tones are not strongly separated according to the note sung (although visually on the graph it appears that tone 5, the low rising tone, is associated more frequently with the lower note).

6.4 Conclusions

Listeners in both languages differentiate the tones of spoken words well above chance but have more difficulty with sung words. Mandarin singers do not mark tone phonetically on their sung words by modifying either the slope or the duration and the Mandarin listeners do not appear to be able to identify those words. Cantonese listeners appear to make use of the rising contours that Cantonese singers add to mark rising tone and make use of the absolute $F_0$ of a sung word to identify level tones. The results for listeners in both languages complement the results for singers found in the earlier studies.

The results of the experiment in Chapter Three found that singers in Cantonese include rising contours when singing rising tones but that they did not distinguish between the two rising tones in their production, presumably relying instead on the register differences included in the composed music for that distinction (Wong and Diehl 2002). The results from the Cantonese perception experiment discussed in this chapter suggest that Cantonese listeners
make use of these cues in a similar and complementary way. The results of the Pearson’s residual post hoc test suggest that Cantonese listeners group the two rising tones together in one group and the three level tones and one falling tone together in another group. The analysis of the words sung on high and low notes support Wong and Diehl’s (2002) findings that Cantonese listeners rely on relative pitch to distinguish the three level tones in singing as well as in speech.

The results of the Mandarin singing experiment found that singers do not include lexical tone information in their F0 contours when singing. The results of the corresponding perception experiment suggest that Mandarin listeners were not able to distinguish sung words that differ by tone. They appear to rely on general length distinctions of tone but there do not appear to be other cues in the sung word that assist in tone recognition.

The results from these two perception studies emphasize the role that context, both linguistic and musical, plays in the comprehension of song lyrics. The register system that Cantonese composers use for structural representations relies entirely on musical context – specifically the direction of movement from the immediately preceding note. As shown in Wong and Diehl (2002), Cantonese listeners use the shape of the musical phrase to help them interpret which tone is being represented. When this musical context is removed, as in the Cantonese experiment discussed in this chapter, word recognition is low (18% correct in the
sung condition). The relatively low recognition rate even in the spoken condition (68%) suggests the overall importance of the contour, even in speech. An interesting direction for future research would be to examine how the levels of recognition might change if both the structural and phonetic manifestations of tone were included in a Cantonese perception experiment.

Mandarin popular music composers and singers appear to abandon representations of tone in songs. This leaves listeners with only the semantic context to determine the meaning of the words but, as pointed out in Chapter Three, Mandarin listeners have fewer potential ambiguities to navigate than Cantonese listeners. Here, too, it would be interesting to see to what extent Mandarin listeners are able to extract the meaning of words within song lyrics given the absence of lexical tone representation.
Chapter Seven

Conclusions

7.1 Tone in Singing in Cantonese and Mandarin

Tone can be manifested in singing in different ways. Structural manifestations of tone can be encoded into the prescribed structure of the music. Phonetic manifestations of tone can provide reflections of tones not prescribed by the structure of the music during performance. Although the musics are closely related, popular singing in Cantonese and popular singing in Mandarin manifest tone differently.

Cantonese popular music encodes a somewhat simplified version of the tone system of the language in the structure of the composed music (Chan 1987a, Wong and Diehl 2002). Furthermore, as shown in the experiments in this thesis, some performers further enhance the tonal representation of the song in the musical performance by including a rising contour for syllables with a rising tone although the phonetic enhancement does appear to disappear in the lower range of the voice. This combination of techniques allows for a strong representation of tone in music.

Mandarin popular music, conversely, has, if anything, only a very limited representation of tone in the composed music (Wee 2007) and, according to the
studies carried out in this thesis, Mandarin singers do not appear to employ any form of phonetic representation of tone.

The results from the experiments in Chapters Four, Five and Six find that Cantonese and Mandarin singers employ different strategies in regard to the manifestation of tone in singing but these strategies match those used by composers in the two languages. The examination of phonetic manifestations of tone in this thesis was carried out using laboratory experiments. Experimental approaches to the question of tone realization in singing are not common and the experiments in this thesis indicate that such an approach can be a useful and worthwhile one. It must be stated, however, that the requirements for experimental research such as large numbers of participants and musical cultures that allow novel song creation are not universally present. An experimental methodology is simply one approach among many that are available to the scholarly community.

7.2 Perception of Tone in Singing

The experiments in Chapter 6 indicate that listeners in Mandarin are not able to identify individual sung words out of context but that Cantonese listeners, although they have problems identifying individual words, use pitch and contour to help identify them. From this we may deduce that context plays a very important role in the understanding of song lyrics in both languages.
Musical context plays a very important role in the structural manifestation of tone in Cantonese.

7.3 Cross Cultural Examination of Tone in Singing

Included in this thesis was a preliminary cross-cultural overview of tone in singing. The cross-cultural comparison of ways in which tone is manifested in the singing of different cultures carried out in chapter two, although limited in scope, is something that, surprisingly, has not been done before. Previous work has focussed on individual cultures and languages which provided an excellent basis on which to make some initial comparisons. Different cultures, of course, make different choices and the classification of possible manifestations of tone proposed by Ladd and Remeissen (in progress) lays out various possible ways in which tone may be represented in singing. Most of the previous work has focussed on structural manifestations of tone, often using different metrics to calculate correspondence between the speech melody and the song melody. Comparing across studies which used contour-based comparisons indicates that structural manifestations are one option that is used but that it is not something that is used uniformly across cultures.
7.4 Linguistic and Cultural Influences on Tone and Singing

Chapter Three suggests reasons why these different strategies may have been employed. These chapters provide a start into expanding the research on singing and tone to include linguistic and non-linguistic factors that may influence how and to what extent a culture may represent linguistic tone in its singing.

In the case of Cantonese and Mandarin, the functional load of tone in the languages may have an influence on the manifestation of tone in singing. The functional load of tone in Cantonese appears to be higher than that of Mandarin although within both languages tone carries about the same functional load as vowels. There may be more competition within Cantonese which has a greater proportion of monosyllabic words than Mandarin and, possibly, a greater need for the tones to be represented in music in order to avoid ambiguity.

There are also extra-linguistic factors that may play a role in the representation of tone in singing in both Cantonese and Mandarin. Disagreement over what a national language should sound like coupled with a newly imported compositional practice that placed a premium on melody may have made it easier for composers and singers of Mandarin songs simply to ignore linguistic tone in songs. Conversely, the development of a separate, identifiable Cantonese music as part of the rise of a Hong Kong identity meant that language, the primary marker of identity, became much more important in music. Different
cultures make different choices – even two cultures and musics as closely related as Cantonese and Mandarin make different choices based on a range of reasons about the representation of tone in music.

7.5 Strategies for Representing Tone in Singing

The results of the experimental chapters provide support for the Ladd and Remijsen (in progress) classification of possible manifestations of tone in singing with a distinction between structural and phonetic manifestations of tone in songs. The results for Cantonese emphasize the idea that the various strategies described by Ladd and Remijsen (in progress) are not exclusive of each other. Modern day Cantonese composers, for example, employ structural manifestations in the form of register levels of the tones being reflected in the composed score (Chan 1987a, Wong and Diehl 2002) while Cantonese singers singing those songs may employ phonetic manifestations not present in the score in the form of rising contours added in during the performance.

Mandarin composers, on the other hand, seem to disregard structural manifestations of tone (Chao 1956, Chen 2007). Mandarin singers also appear not to take the opportunity to express elements of lexical tone while singing.
7.6 Areas for Future Research

Although it has received some academic attention, the interaction between speech and singing with regard to lexical tone is still relatively understudied. There are still only a very limited number of descriptive studies that chart correspondence between speech melody and song melody for tone languages. With the clearer divisions laid out by Ladd and Reijsen’s (in progress) classification system, it would be interesting to see how phonetic and structural representations interact in different languages.

With regard to the interaction of tone and music in Cantonese and Mandarin, a few suggestions have already been made as to possible areas for future research. One is to explore the apparent connection between the phonetic manifestation of rising contours and voice range in Cantonese. This observation that the phenomenon does not occur in the lower range of the voice warrants further investigation. It would also be interesting to run a perception experiment in Cantonese with sung syllables that have gradual changes in slope to see how much slope is necessary for listeners to perceive the slope a manifestation of tone. Another possible avenue for research is to expand the perception experiment to include longer phrases rather than individual words to see how the addition of context, both musical and semantic, aids in the recognition of words in song lyrics.
References


Chen, S. W. 2007. The music industry and popular song in 1930s and 1940s Shanghai: A historical and stylistic analysis, Ph. D. dissertation, Department of Film and Media Studies, Stirling University.


Ciocca, Valter, Whitehill, Tara L. and Ma, Ka Yin Joan. 2004. The impact of
cerebral palsy on the intelligibility of pitch-based linguistic contrasts.

Typology of African Systems Workshop.* Bielefeld University, Germany.
http://www.spectrum.uni-bielefeld.de/TAPS/Connell.pdf (accessed 6 January
2012).

Connell, Bruce. 2002. Tone languages and the universality of intrinsic F0:

Cutler, Anne, and Chen, Hsuan-Chih. 1997. Lexical tone in Cantonese spoken-


Williams and Wilkins.

Ekwueme, Lazarus N. 1974. Linguistic determinants of some Igbo musical

Feld, Steven, and Aaron A. Fox. 1994. Music and Language. *Annual Review of
Anthropology.* 23. 25–53.

Feng, Renzhao. 1998. The Hongkongnese: Who are the Hongkongnese? *Chinese
Sociology and Anthropology.* 30(3). 37-44.

Ferguson, Daniel. 1993. The Shāndōng Highwayman: Mechanisms of inclusion
and resistance and the predication of Cantonese identity through

Fujisaki, H. 1981. Dynamic characteristics of voice fundamental frequency in
speech and singing: Acoustical analysis and physiological interpretations.


*Proceedings of the 17th International Congress of the Phonetic Sciences. Hong Kong.* 1754-1757.

Schellenberg, Murray. 2012a. Does language determine music in tone languages? 
*Ethnomusicology.* 52(2). 266-278.


Siu, Helen F. 1996. Remade in Hong Kong: Weaving into the Chinese cultural tapestry. In Liu, Tao Tao and Faure, David, (eds.) *Unity and Diversity: Local Cultures and Identities in China.* Hong Kong: Hong Kong University Press. 177-197.


