Timing and Melody:
An Acoustic Study of Rhythmic Patterns of Chinese Dialects

by

Ya Li

B.Sc., Hunan University, 1989

M.A., University of Victoria, 2008

A Dissertation Submitted in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Linguistics

Timing and Melody:
An Acoustic Study of Rhythmic Patterns of Chinese Dialects

by

Ya Li
B.Sc., Hunan University, 1989
M.A., University of Victoria, 2008

Supervisory Committee

Dr. Hua Lin, Supervisor
(Department of Linguistics)

Dr. John H. Esling, Departmental Member
(Department of Linguistics)

Dr. Emmanuel Hérique, Non-Departmental Member
(Department of French)

Supervisory Committee

Dr. Hua Lin, Supervisor
(Department of Linguistics)

Dr. John H. Esling, Department Member
(Department of Linguistics)

Dr. Emmanuel Hérique, Non-Department Member
(Department of French)

ABSTRACT

Inspired by Lin and Huang's (2009) rhythmic study of Chinese dialects, this study examines speech rhythm of 21 Chinese dialects from three perspectives, timing, melody, and phonological structure. The 21 dialects belong to four major groups of Chinese and their respective sub-groups. The four major groups are Mandarin, Wu, Min, and Cantonese. Nine duration-based and four pitch-based metrics are used to quantify timing and melody, respectively. Four phonological structure-based metrics are used to explore the relationships between syllable structure and timing and between tone structure and melody. All the metrics are paired up according to five categories, duration-only, pitch-only, duration-pitch, duration-syllable, pitch-tone, and each pair is subject to a correlation analysis. Then timing and melody patterns of the Chinese dialects are determined by correlation patterns of relevant metric pairs.

The main findings of this study are as follows: 1) Timing and melody patterns of the Chinese dialects are far from homogenous across major groups and melody patterns are more distinct than timing patterns; 2) No single metric pair is able to quantify speech

rhythm consistently for all the Chinese dialects; nonetheless, pitch-based metric pairs fare better than duration-based ones; 3) Syllable-timedness and melodiousness are correlated positively for all the major groups except for Wu; 4) Phonological structure plays little role in shaping timing and melody patterns of most Chinese dialects.

The above findings are both expected and unexpected. They are expected in the sense that rhythmic perception invovles multiple acoustic cues, so it comes as no surprise that not all rhythmic metrics are successful in quantifying Chinese rhythm. They are unexpected for the reason that all the metrics are developed based more or less on phonological structure, yet the rhythmic patterns they reveal do not correspond to the structure affinity or group membership of the Chinese dialects. Overall, the findings suggest that pitch is a more import cue than duration to Chinese rhythm. As the first study of Chinese rhythm across multiple dialects and from different perspectives, this study not only lays a methodological foundation for future research but also contributes to our in-depth understanding of Chinese dialects.

Table of Contents

List of Tables

List of Figures

Appendix

Acknowledgments

As I am approaching the end of my long-winded PhD study, my feelings have gone from anxious, bitter-sweet, to grateful. I understand that without the support and even sacrifice from those around me, it is impossible for me to hang on to the end.

Firstly, I would like to thank you, Dr. Hua Lin. You have always been here for me, since the year you kindly took me in as your MA student. Your expertise in Chinese linguistics has helped kindle my passion in the similar area. You are also very kind, having always cared about me and my son. I feel a sense of warmth whenever around you. It is fair to say that without your strong support, I would not have achieved anything academic.

Secondly, I would like to thank you two, Dr. John Esling and Dr. Emmanuel Hérique. Dr. Esling, like Dr. Lin, you have always been supportive during all my years at UVic. Backed up by your encouragement and expertise in voice quality research, I dared to include voice quality as part of my research. Dr. Hérique, despite on my committee for just a year or so, you delayed no time to help me whenever I needed. I am also thankful to Dr. Sonya Bird and Dr. Martha McGinnis-Archibald. You two as supervisors of my candidacy projects have helped me to expand my linguistic knowledge in the areas of experimental phonetics and syntax.

A big thank-you goes to Dr. Xiaoxiang Chen from Hunan University. You offered me the strongest support when I was in China working on my dissertation project. I have been very blessed to have you as my best colleague-friend.

A special thank-you goes to Kellen Parker, a founding member of *phonemica.net (xiangyin yuan)*. You kindly allowed me to download and use the Chinese speech data. Also, I appreciate the support of the SSHRC Doctoral Scholarship, allowing me to complete my study without dealing with financial hardship.

Last, I am thankful to all of my wonderful colleagues and close friends who have been with me all these years. I can never express enough my heartfelt gratefulness for all of you. I just want you all know that your loving-kindness is what keeps me going and what makes me happy.

To my father

Love never fails (1 Corinthians 13:8).

Chapter 1
INTRODUCTION

This section provides an overview of the present research. Specifically, Section 1.1 provides crucial research context for the present work, and Section 1.2 introduces the research objects and relevant objectives. Section 1.3 provides definitions for a list of important terms used in the dissertation. Section 1.4 presents research questions and related hypotheses. Section 1.5 provides an outline of this dissertation.

1.1 Research context

More than half a century ago, Lloyd (1940) likened English and Spanish rhythm to two military terms, "Morse-code" and "machine-gun." Since then, researchers have embarked on a mission to find linguistically meaningful terms to classify rhythm. To date, after half a century's vigorous research, we have yet to find the best way to characterize rhythm. Nonetheless, we have gone through some important stages, each having helped to improve our understanding of speech rhythm. From isochrony-based (Pike, 1946; Abercrombie, 1967; Ladefoged, 1975) to phonology-based (Donegan & Stampe, 1983; Dauer, 1983&1987; Auer, 1993), from phonologically informed (Ramus, Nespor, & Mehler, 1999; Grabe & Low, 2002; Dellwo, 2006; White & Mattys, 2007) to perceptually informed (Dilley & McAuley, 2008; Kohler, 2009a, b; Cumming, 2010), these stages help us understand that what underlies the two simple yet fitting military metaphors is a linguistic structure deeply rooted in phonetic and phonological, especially prosodic, details such as syllable structure, timing, intensity, pitch, stress, tone, and intonation.

In the initial stage, Pike (1946) and Abercrombie (1967) proposed the stress- and syllable-timed rhythm to define the impressionistic Morse-code and machine-gun rhythm. The division between stress- and syllable-timing acknowledges the regularity in the recurrence of the two prosodic units, stress and syllable, or isochrony of stress and syllable intervals. Later, a third category of rhythm, mora-timing, was further proposed by Han (1962) and Ladefoged (1975). In a typical mora-timed language such as Japanese, it is successive morae rather than syllables (stressed or not) that are near-equal in duration. However, isochrony, whether based on stress, syllable, or mora, was quickly dismissed by the empirical evidence that stress intervals in so-called stress-timed languages is found no more regular or irregular than syllable intervals in syllable-timed languages (Dauer, 1983).

The second stage of research seeks to establish an association between phonological properties and speech rhythm, as how languages make rhythmic use of syllables or phonological words might hold the key to rhythmic classification (Dauer, 1983; Donegan & Stampe, 1983; Nespor &Vogel, 1986; Gil, 1986; Auer, 1993). Since it is found that stress-timed languages are usually associated with complex syllable structure, word-level stress, and vowel reduction in unstressed position; while syllable-timed languages with simple syllable structure, lexical stress or tone, and little vowel reduction, phonological structure was parameterized and then used to categorize different languages (Auer, 1993). Depending on which set of parameters each language has, the rhythmic class is assigned no longer between the initial dichotomy of syllable- versus stress-timing but rather along a continuum from mora- to syllable-, and to stress-timing. Later, this type of

phonologically based rhythmic research falls within the scope of prosodic typology (Jun, 2005; Maddieson, 2011; Schmid, 2012).

The use of segmental duration-based metrics to quantify rhythm marks the third stage in rhythmic research (Ramus et al., 1999, Grabe & Low, 2002, Dellwo, 2006, White & Mattys, 2007). At this stage, some phonological properties such as syllable structure and vowel reduction have been indirectly encoded in a series of duration-based acoustic measures such as %V (percentage of vowel duration over the total segment duration) and ΔC (the standard deviation of consonant intervals), so stress- and syllable-timed rhythm can be compared quantitatively in terms of these measures (Ramus et al., 1999). As a result, almost all the major world languages have been quantified using various duration-based metrics. However, according to Arvantini (2009), these phonologically informed durational metrics achieve only certain degree of success in quantifying speech rhythm: Only prototypical stress- and syllable- timed languages such as English and French can be distinguished reasonably well by these metrics. For other non-prototypical or rhythmically mixed or unclassified languages such as Greek and Korean, different metrics yield different classifications.

Due to limitations of durational metrics, perceptually informed rhythmic studies (Dilley & McAuley, 2008; Kohler, 2009a, b; Cumming, 2010) have been brought onto the scene. At this stage, studies of rhythmic grouping and patterns of prominence have been proliferated, as pitch, stress, and intonation, along with duration, are considered as essential components of speech rhythm. Later on, a number of pitch-based metrics, such as pitch height difference and pitch change rate, have been developed in order to quantify speech melody (Hirst, 2011, 2013) or intonation (Vicenik & Sundara, 2012). These

pitch-based metrics are used to complement existing duration-based metrics to make predictions on rhythmic timing and prominence patterns of different languages and dialects, which represents a renewed effort to classify rhythm. A main goal for researchers at this stage is still to find best possible measurable rhythmic parameters in speech production. To fulfill this goal, it is crucial to test out various rhythmic metrics, duration- or pitch-based, on both rhythmically diversified/matched languages/dialects.

Currently, the existing research literature is dominated by studies using duration-based metrics to quantify language rhythm: few studies seek to use duration-based metrics to quantify dialect rhythm, let alone to use pitch-based metrics to quantify dialect rhythm. This imbalance seen in the research literature gives rise to the present rhythmic research on Chinese dialects. This study aims at contributing to the current understanding of rhythmic production from not only phonological but also perceptual perspectives.

1.2 Research objects and objectives

The present research chooses four major Chinese dialect groups, Mandarin, Wu, Min, Cantonese (also referred to as Yue), along with some of their respective sub-dialects, as research objects. The choice of these Chinese dialects is based on the following considerations: 1) As a tonal language, Chinese makes a special case for pitch-based measures, because the use of lexical tone in almost every syllable may render them useless, or on the contrary, quite helpful in creating syllable-timed rhythm; Also, Chinese tone production, especially low tone production, is found to be associated with certain voice quality such as creakiness and breathiness (e.g., Chao, 1968; Rose, 1989; Cao & Maddieson, 1992; Davison, 1991; Belotel-Grenié & Grenié, 2004; Esposito, 2006; Lam & Yu, 2010). Since voice quality can affect segmental duration (Blankenship, 1997;

Gordon & Ladefoged, 2001; Gao, Hallé, Honda, Maeda, & Toda, 2011), it may in turn affect rhythm. Therefore, this study includes voice quality as part of its rhythm investigation, though it is not a main focus; 2) As a monosyllabic language, Chinese also makes a case for syllable-timed rhythm, and if so, it would be interesting to see how Chinese syllable structure is associated with syllable-timed rhythm; 3) Although Chinese writing is unified across China, Chinese dialects are far from homogeneous that some dialects such as Cantonese and Mandarin are mutually unintelligible (Tang, 2009). Therefore, investigating how Chinese dialects are related to one another from the perspective of rhythmic typology may shed light on Chinese dialectology; 4) The existing studies of Chinese rhythm are scanty: only a handful of them (Lin & Wang, 2005 & 2007; Lin & Huang, 2009; Mok, 2009) have used duration-based metrics to classify Chinese dialects. Therefore, Chinese rhythmic research is yet to be developed both in breadth and in depth.

Based on the above considerations, the present study aims to achieve three objectives: 1) to evaluate the usefulness of rhythmic measures, both duration- and pitch-based, in quantifying Chinese rhythm, 2) to identify the role of phonological structure in shaping Chinese rhythm, and 3) to enrich Chinese dialectological research from the rhythm perspective by extending Chinese rhythmic research from the major to sub- dialectal level.

1.3 Definitions of important terms

This section provides a brief definition for 21 important terms used in the dissertation and a detailed explanation of them will be provided along the way.

1) Prosody: an umbrella term for all the supra-segmental features occurring in speech. It includes three types of features, melodic (e.g., pitch, tone, and intonation), temporal

(duration, length, tempo, pause), and dynamic (e.g., loudness, stress, and voice quality) (Van Heuven & Sluijter, 1996). These features help to realize the prosodic structure and prominence relations within the structure (Jun, 2005).

2) Prosodic structure: utterances can be broken down into smaller speech units and form a hierarchical structure. In the prosodic hierarchy proposed by Selkirk (1986), for example, an utterance can be divided into phonological and intonation phrases, then into prosodic words and feet, and finally into stressed and unstressed syllables. Chinese prosodic structure is unique and will be discussed in Section 3.2.1.

3) Rhythm: the perceived regularity of prominent speech units (e.g., stressed and unstressed syllables, long and short syllables, and their combinations) (Crystal, 1985).

4) Duration-based rhythm: speech rhythm perceived from the temporal perspective. Perceptually, it is classified into syllable-timing (e.g. French), stress-timing (e.g. English), and mora-timing (e.g., Japanese). This study focuses on syllable-timedness of Chinese speech, specifically on how duration of certain speech units are shortened or lengthened in a patterned way by speakers of Chinese dialects.

5) Duration: the length of a speech unit. Here it refers to sonorant and inter-sonorant intervals. It is measured in milliseconds (ms). A sonorant interval (Son) includes a sonorant sound (a vowel, glide, nasal, or an approximant such as a lateral or rhotic sound) or some combination of sonorant sounds. An inter-sonorant (IS) interval includes an obstruent consonant sound (a stop, a fricative, or an affricate sound), some combination of obstruent sounds (a pause less than 10ms is included as part of its adjacent obstruent).

6) Duration-based metrics: a set of metrics used to quantify the timing difference between languages/dialects. They are originally designed to measure vocalic and consonantal

intervals directly from the speech signal, without any reference to syllables, words, or higher prosodic units (Harrington, Hoole, & Pouplier, 2013). An example of such a metric is %V (percentage of vocalic duration over the total vocalic and consonantal duration in a stretch of speech) and it is developed by Ramus et al. (1999). This study, however, measures sonorant and inter-sonorant intervals instead of vocalic and consonantal intervals. The metric %V hence becomes %Son (percentage of sonorant duration over the total sonorant and inter-sonorant duration) in this study.

7) Chinese syllables: they are traditionally divided into two parts: the initial and the final. The initial is usually a single consonant at the onset and the final is everything that follows (Lin, 2001a). Each Chinese character corresponds to a syllable, so Chinese is considered as phonologically monosyllabic.

8) Melody: pitch variations involving measurable physical parameters (Hirst & Di Cristo, 1998). A stretch of speech can be perceived as more or less melodious if there are larger/more or smaller/less pitch variations in the speech. Note that melody in this study does not refer to any specific prosodic levels. According to Hirst (2011), melody is related to pitch fluctuation over time, regardless of its linguistic source such as changes in tone, stress, and intonation; the more melodious a stretch of speech, the more the pitch fluctuates. In some literature reviewed in this dissertation (e.g., Vicenik & Sundara, 2012), the term intonation is used to refer to melody, but this study treats them separately because Chinese has tone instead of intonation as the major contributor to melody. Stricly speaking, intonation is a prosodic feature referring to the use of pitch variations to convey discoursal meaning and to mark phrases (Gussenhoven, 2004). For example, a falling intonation at the end of a sentence signals a statement or the complete status of the

sentence (Fox, 2000). It constitutes a level (intonation phrase) in the prosodic hierarchy proposed by Selkirk (1986). Hence, 'melody' is a better term than 'intonation' when describing pich variation at the phonetic level.

9) Pitch-based rhythm: speech rhythm perceived from the melodic perspective. This study focuses on melody patterns, specifically on whether pitch variations are patterned for Chinese dialects and if so, how they are patterned.

10) Pitch: the perceptual correlate of f0 (Fundamental frequency), the frequency at which the vocal folds vibrate.

11) Pitch-based metrics: a set of metrics used to quantify the melody difference between languages/dialects. An example of such a metric is PE (pitch excursion or the difference between the highest and lowest pitch on a local pitch contour). It measures pitch variations directly from the speech signal, without any reference to tone, stress, and intonation.

12) Chinese tones: Chinese as a tonal language uses pitch variations to signal meaning differences. Each Chinese tone is associated with a Chinese syllable and a specific pitch pattern. Chinese tones contrast either by the direction of pitch variation (level, falling, or rising), by the amount of pitch variation (high, mid, or low), or by both. Sometimes, a tone can be produced with certain voice quality (e.g., creakiness or breathiness).

13) Voice quality: speech characteristics related to the voice source, the volume velocity airflow through the glottis during phonation (Gobl & N íChasaide, 1992). Phonation is the status of the glottis. Different states of the glottis give rise to different types of voice quality or phonation types, and three common ones are modal, breathy, and creaky voice.

14) Voice source metrics: a set of metrics used to quantify voice quality in terms of creaky, modal, and breathy voice. An example of such a metric is H1*-H2* (the difference between the spectral magnitudes of the first two source harmonics; * means the corresponding magnitudes are corrected for the effect of the first and second formants, F1 and F2) (Iseli, Shue, & Alwan, 2007).

15) Chinese dialects: regional varieties of the Chinese language. It is traditionally classified into seven major groups, Mandarin, Wu, Min, Kejia (Hakka), Yue (Cantonese), Xiang, and Gan (Lin, 2001a). Each major dialect group can be further classified into sub-dialect groups (sub-groups) according to the geographical regions they are spoken.

17) Standard Chinese: also called standard Mandarin or *Putonghua* in mainland China and *Guoyu* in Taiwan. It is developed from the Mandarin dialect spoken in Beijing and used in all the major media systems and school teaching in China (Lin, 2001a).

1.4 Research questions and hypotheses

The current study focuses on the following research questions, each associating with a hypothesis. The rationale for the hypothesis is also explained.

Question 1: How do duration- and pitch-based metrics fare in quantifying Chinese rhythm at major and sub- dialectal levels?

Hypothesis 1: Duration-based metrics fare better than pitch-based metrics in quantifying Chinese rhythm at the major dialectal level but neither of them fares well at the sub-dialectal level.

Rationale:

According to Auer (1993), if a language already exploits pitch for lexical distinction, it is unlikely for pitch to take the extra functional load for rhythmic distinction. If this is the

case, then duration rather than pitch will be used as a major cue to Chinese rhythm. Specifically, Chinese should be more distinct durationally than melodiously at the major dialectal level but not at the sub-dialectal level, because dialects in sub-groups are more likely to pattern rhythmically together in terms of both duration and pitch due to their similar phonological structure.

Also, how well the metrics fare means how consistent the metrics are in showing the same/similar rhythmic patterns for the same major/sub- group of dialects: the more consistent they are, the better they fare. For example, if majority of duration-based metrics show that Mandarin is more syllable-timed than Cantonese and majority of pitch-based metrics show that Mandarin is less melodious than Cantonese, then both duration- and pitch-based metrics are considered faring well; If majority of duration-based metrics agree on the pattern that Mandarin is more syllable-timed than Cantonese but majority of pitch-based metrics can not agree on the pattern that Mandarin is less melodious than Cantonese, then the former are considered faring better than the latter.

Question 2: How are duration-based timing and pitch-based melody patterns related?

Hypothesis 2: There is a positive correlation between syllable-timedness and melodiousness.

Rationale:

Since duration and pitch both contribute to rhythmic perception, it is expected that there exists some correlation between the duration- and pitch-based rhythmic patterns. According to Auer (1993), syllable-timing is often associated with tone and stress-timing with complex syllable structure. Dialects with complex syllable structure are expected to have complex tone structure and be more stress-timed or less syllable-timed. Complex

tone structure often means more tones and more tones result in more restricted pitch variation. Since more restricted pitch variation means less melodiousness, syllable-timedness is positively correlated with melodiousness.

Question 3: How are syllable and tone structures and Chinese rhythm related at major and sub- dialectal levels?

Hypothesis 3: Structural complexity is negatively correlated with syllable-timedness and melodiousness at the major but not sub- dialectal level.

Rationale:

Based on the previous research, syllable-timing is predictable from the complexity of syllable structure (e.g., Auer, 1993). All the Chinese dialects are monosyllabic, but they do vary in syllable type and segmental composition. It can be assumed that dialects with a complex syllable structure will have more syllable types and segments and be less syllable-timed than dialects with a simple syllable structure. However, this correlation may not hold true for Chinese sub-dialects, as there is not much difference among their syllable type and segmental composition.

Major Chinese dialect groups also vary in tone structure, so it can be assumed that dialects with a complex tone structure will have more tone types and be less melodious than dialects with a simple tone structure. However, this correlation may not hold true for sub-dialects, as there is not much difference among their tone types.

1.5 Outline of the dissertation

Chapter 1 provides research context, objectives, and hypotheses. Chapter 2 reviews relevant research literature. Chapter 3 describes phonological/prosodic characteristics of major Chinese dialect groups. Chapter 4 explains the research method and data analysis

process. Chapter 5 presents results and discusses findings for individual major dialect groups. Chapter 6 provides a general discussion of research hypotheses and summary of the most important findings. Chapter 7 concludes the research by pointing out research contributions, limitations, and future directions.

Chapter 2
LITERATURE REVIEW

As overviewed in Chapter 1, half a century's research work has improved our understanding of speech rhythm from various perspectives. This section provides a closer look at the current research field: Section 2.1 clarifies key theoretical concepts to be used in the present research and Section 2.2 reviews most influential rhythmic studies with a focus on their use of quantitative research methods to classify rhythm.

2.1 Defining rhythm

According to Fraisse (1982), rhythm is difficult to define, because it involves several variables fusing together: duration, intensity, pitch, and pause can all induce the perception of regularity in ordered elements (e.g., speech sounds in the present context). A commonly assumed definition of rhythm refers to the recurring patterns of prominence over time, triggered by physical properties such as vowel and consonant duration (Kohler, 2009b, p6). However, a later definition sees rhythm as recurring effects of temporal control over larger domains than syllable and stress foot. The larger domain is created by pitch grouping, so pitch is also a key player in rhythmic perception, as languages may differ in "the rhythmic control of speech by coordinating short-term vocal tract opening-closing gestures differently with more long-term pitch grouping" (Kohler, 2009b, p. 9).

If timing and pitch are two defining characteristics of rhythm, we need to understand how they contribute to rhythmic perception. Section 2.1.1 and Section 2.1.2 respectively present phonological and perceptual bases of speech rhythm.

2.1.1 Phonological basis of rhythm

Speech timing is traditionally classified into two main categories, syllable-timing and stress-timing (Pike, 1946; Abercrombie, 1967). When comparing syllable- and stress-timed languages, one cannot help noticing some correlations between the type of timing a language has and its phonological properties. For example, Yoruba and French are both considered syllable-timed and Arabic and English stress-timed languages (Abercrombie, 1967). A noticeable difference between the former two and the latter two is in the absence/presence of lexical stress or accent. Also, Yoruba as a prototypical syllable-timed tonal language has the simplest CV structure and no vowel reduction, while English as a prototypical stress-timed language has the most complex syllable structure, allowing maximally three or four consonants as syllable onset or coda. Note that syllable onset and coda are called 'shell' by Auer (1993). Also, there is always vowel reduction in un-stressed or non-accented syllables. French and Arabic, on the other hand, have moderately complex syllable structures, falling on a rhythmic continuum between the two prototypes, with French leaning towards the syllable-timed due to the absence of stress and Arabic towards the stress-timed due to the presence of stress. Table 2.1 lists the aforementioned phonological differences between the four languages (adapted from Auer, 1993, p. 9):

Table 2.1 Phonological difference between syllable- and stress-timed languages

	Yoruba	French	Arabic	English
Vowel reduction in non-accented syllables	no	marginal	no	yes
Maximal syllable shell* complexity	CV	CC…CC	CCVCC	CCC…CCCC
Tone	yes	no	no	no
Accent	no	no	yes	yes

*Syllable shell: onset and coda

Syllable- and stress-timing were once referred to as syllable- and word-rhythm by Donegan and Stampe (1983), as they consider that perception of timing is not restricted to syllable duration and foot length but characterized by a host of linguistic features. The syllable- and word-based rhythm is distinguished by phonotactic and prosodic differences in such as syllable structure and word/phrasal accent among different languages. A similar view is held by Dauer (1983). Dauer (1983) studied the inter-stress intervals of some stress-timed and syllable-timed languages and found no regular intervals in either type of languages. She then noted that English has varied syllable types and also long vowels in stressed syllables, and that these structural properties tend to reinforce the impression of stress timing in English. In general, the difference between stress and syllable-timing is a result of differences in phonological properties such as syllable structure, vowel reduction, and the phonetic realization of stress (Dauer, 1983).

Another relevant view is advanced by Auer (1993) when he realized that many phonological properties in a wide range of languages have either one of the two levels of

prosodic units, syllable and phonological word, as their domain. In other words, most phonological processes take place either in the syllable or in the phonological word. As a result, languages are distinguished not by whether or not word accent (stress) plays a central role in rhythmic classification but by which prosodic domain is the most relevant to phonological processes. Particularly, Auer (1993) proposed a phonological typology model based on a set of phonological parameters, such as shell (onset & coda) complexity, syllable or word related processes/phonotactics, tone, accent (stress), and vowel reduction. These parameters help to establish two prototypes, syllable and word languages. All the languages then are compared with the two prototypes and placed on a continuum with the two prototypes at the two ends. Table 2.2 associates the aforementioned parameters with different language types, including two opposing prototypes (adapted from Auer, 1993, p. 94).

Table 2.2 List of languages and their phonological parameters

	Processes	Vowel reduction	Accent	Tone	Shell complexity
Prototype	S			S	L
Yoruba	S			S	L
Mandarin	S	(+)	+	(S)	L
French	S/W	(+)			H
Arabic	W	(+)	+		H
English	W	+	+		H
Prototype	W	+	+	(W)	H

S = Syllable-related; W = Word-related; S/W = both or neither of them; ( ) = marginal or restricted; + = presence; L(ow) = CG…C or less; H(igh) = CC…CC or more; M(id) = in between L and H; blanks = not applicable or absence.

Syllable-related processes are also called syllable structure enhancing processes, which involve creating simple structure through phonological processes such as vowel epenthesis and assimilation/re-syllabification both within and across word boundaries. Word-related processes, on the other hand, are called syllable structure destroying processes, which involve creating complex structure through processes such as vowel deletion and consonant-consonant assimilation. In the table, no languages match the two prototypes on all the phonological parameters. Accent, supposedly the central parameter for stress-timing, occurs in Mandarin as well, because Mandarin toneless syllables are unstressed (Lin, 1999 & 2001b). Also, Mandarin is marked as restricted in the tone parameter, because it permits toneless syllables. French, in contrast, has highly complex syllable shell but the absence of word-level accent makes it more like the syllable prototype.

Schiering, Bickel, and Hildebrandt (2012) questioned Auer's (1993) use of the word domain dominance to characterize stress-timed languages through a typological study of a database containing 58 languages from Indo-European, Austro-Asiatic, and Sino-Tibetan families. They coded phonological processes in terms of their word-level absence/presence and then calculated a frequency value for the number of word domain reference. Then they compared these languages' frequency distribution with word-rhythm predicted by Auer (1993). The results show that stress-timed languages show only slightly more frequent word domain reference in Indo-European and Sino-Tibetan families than in Austro-Asiatic family. Also, stress-timed languages show slightly less frequent word domain reference than non-stress-timed languages in Austro-Asiatic

family, indicating that the word-level dominance cannot be used to predict stress-timing across language families.

Overall, however, it is possible to distinguish some languages based on syllable- and word-level phonological properties. If this is the case, then rhythmic classification, whether dubbed as syllable- and word-rhythm or syllable- and stress-timing, should also be possible from the phonological perspective.

2.1.2 Perceptual basis of rhythm

Since stress is a major perceptual cue to prominence-based rhythm and pitch is a major acoustic cue to stress, it is necessary to understand how stress and ultimately how pitch influences rhythm. The following sections review previous understanding of stress-induced rhythm. Specifically, Section 2.1.2.1 describes the relationship between stress and rhythm and Section 2.2.2.2 introduces different types of prominence-based rhythm. Section 2.2.2.3 describes the interaction between timing and stress or influence of pitch on duration perception.

2.1.2.1 Stress and rhythm

The relationship between stress and rhythm has long been studied within the metrical framework (Liberman & Prince, 1977; Selkirk, 1986; Hayes, 1995). According to Arvaniti (2009), rhythm is represented by meter, and it is the alternation of strong and weak stress. Meter-based rhythm renders syllable duration irrelevant or at most secondary to stress (Benadon, 2014). Based on the metrical stress theory proposed by Hayes (1995), the meter-based rhythm has a number of general characteristics: (1) stress is hierarchical: sequences of rhythmic beats (a point in time) have multiple levels of strength; (2) rhythm obeys a tendency to even spacing at all intervals of repetition; (3) rhythm obeys a law of

downward implication (i.e., culminativity): a beat on a high layer also serves as a beat on all lower layer; (4) there is no stress assimilation under the assumption that stress is not a feature but the linguistic manifestation of rhythmic structure (Liberman, 1975; Liberman & Prince, 1977). These characteristics determine that stress patterns contribute significantly to the perception of rhythm.

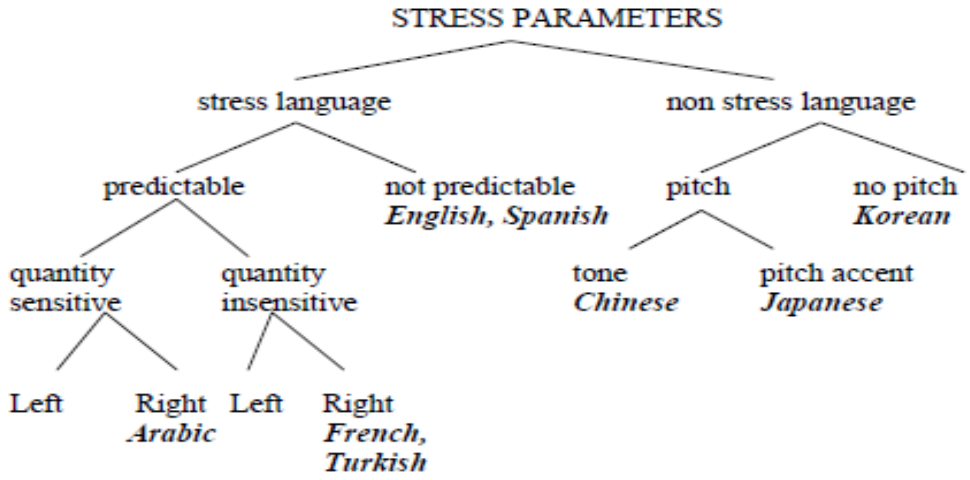
Note that the term 'stress' is often used in a general sense equivalent to the term 'prominence.' According to Kohler (2009a), stress in a strict sense refers only to lexical stress, and it is part of the phonology of a word. Prominence, on the other hand, refers to relative syllable salience in an utterance, and it is syllable- not word-oriented. Stress at the lexical level and prominence at the syllabic level are often treated as equivalent, because stress is usually manifested through a syllable in a word by acoustic parameters such as f0, duration, and intensity.

There are also two terms to be distinguished from lexical stress: one is 'pitch accent' and the other is 'tone.' Pitch accent refers to prominence of a mora/syllable manifested by pitch height (High or Low f0), and tone refers to prominence of a syllable/word manifested through pitch shape, be it level, rising, falling, or combinations of rising and falling. Also, both terms are word-oriented like stress, yet unlike stress, which does not regularly contribute to word meaning, both pitch accent and tone are used to convey word meaning. For example, Japanese words *h<u>a</u>shi* ('chopsticks') and *hash<u>i</u> (*'bridge'), when spoken, differ only by the position of pitch accent in each word: In the first word, the higher pitch falls on the first mora but in the latter word, it is on the second mora (see underlined letters in the two words). Chinese words *mā* ('Mom') and *mǎ* ('horse'), when spoken, differ only by the tone shape each word carries: The first word carries a high

level pitch but the second word carries a low falling-rising pitch. Japanese and Chinese show how pitch can be used in addition to contributing to lexical stress.

Altmann (2006) classified languages based on how they employ lexical stress, pitch accent, and tone or by stress parameters: Stress languages all employ word-level stress but stress assignment in a word varies from language to language: Arabic, for example, regularly has stress on the utterance-final heavy syllable. Non-stress languages either use pitch to signal word meaning or do not use it at all. Based on how they use pitch, non-stress languages can be further divided into tone, pitch accent, and no-pitch languages: Chinese, for example, uses different tones with a syllable to convey different meanings. Note that a syllable combined with a tone is considered as a lexical morpheme in Chinese (associated with one Chinese character) and Chinese words mostly contain two lexical morphemes (associated with two Chinese characters; Lin, 2001a). Figure 2.1 summarizes the above prominence-based language classification (copied from Altmann, 2006, p. 38):

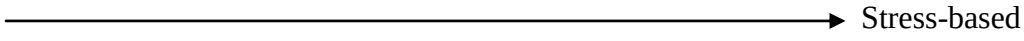
Figure 2.1 Classification of world languages based on stress parameters



Japanese, Chinese, and Korean are classified as non stress language, implying that they cannot be stress-timed languages. However, if stress is understood as prominence in a

general sense, then all the non-stress languages can be stress-timed or, perhaps a better term, prominence-timed, because the three types of pitch variations, pitch accent, tone, and lexical stress, can all contribute to prominence, be it on a stressed syllable in an English word, a high-pitched mora in a Japanese word, or a high-toned syllable in Chinese, as long as the mora/syllable is relatively salient compared with adjacent moras/syllables. The view is consistent with what Dauer (1983, p. 1) claimed: any language can be considered more or less stress-based, depending on how large a role stress plays in a language, and rhythmic grouping takes place even in the so-called syllable-timed languages. In Dauer's (1983) study, stress is found to recur at a rate of 1.9 to 2.3 per second for most languages, suggesting a universal temporal organization in language. Dauer (1983) placed some languages along the dimension of stress-based rhythm instead of stress-timed rhythm, illustrated as follows (p. 60):

Figure 2.2 Illustration of stress-based rhythmic continuum

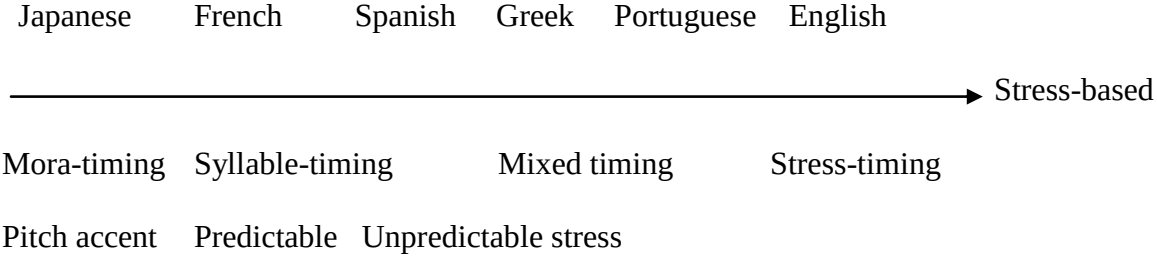


Japanese French Spanish Greek Portuguese English

Japanese as a typical mora-timed language is placed at the beginning of the stress-based continuum and English as a typical stress-timed language is placed at the end of the continuum. Next along the continuum toward the end are French and Spanish, the typical syllable-timed languages, followed successively by Greek and Portuguese, the mixed timing type (Horton & Arvaniti, 2013; Frota & Vigário, 2001; Cruz, 2013). On this continuum, Japanese has pitch accent falling on a mora, resulting in mora prominence (Venditti, 2005); French has neither pitch accent nor lexical stress, but it does have predictable stress at the phrasal level, which can be found on the final syllable in the

Parisian variety and on the penultimate syllable in the Swiss variety (Avanzi, Obin, Bardiaux, & Bordal, 2012); Spanish, despite the absence of lexical stress, has word-level stress on the final syllable in consonant-final non-verb words and on the penultimate syllable in vowel-final non-verb words (Aske, 1990; Bakovic, 2014); Greek has complex stress at all lexical, word, and phrasal levels (Botinis, 1989). As for Portuguese, it is similar to Spanish in terms of stress pattern, but according to Frota and Vigário (2001), Portuguese, especially the European variety, has a tendency toward stress-timing, because of some phonological processes such as vowel reduction similar to those in English. English has stress but its placement is not fixed in one position, so stress in English is considered non-predictable (Altmann, 2006). Note that in Figure 2.1, Spanish is identified with English rather than with French in terms of stress type, because its stress is not determined by syllable position alone but by word type. However, Spanish is identified with French rather than English in terms of rhythmic class (Pike, 1946; Dauer, 1983), so there is a slight mismatch between Altmann's (2006) stress parameters and previous rhythmic classification. Nonetheless, there is a general correspondence between timing- and stress-based rhythmic classifications, as illustrated on the Dauer's (1983) stress-based continuum below (slightly adapted by the author).

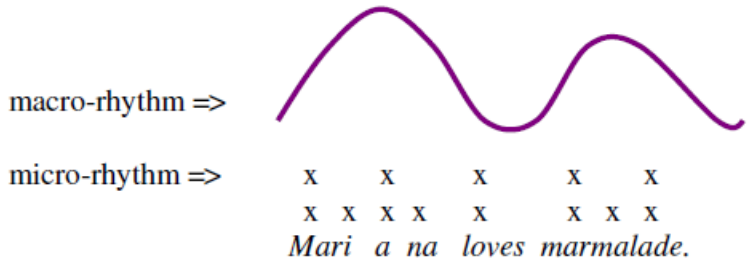
Figure 2.3 Correspondence between timing- and stress-based rhythmic classifications



2.1.2.2 Micro-rhythm versus macro-rhythm

In the prosodic typology proposed by Jun (2005), pitch-based rhythm is distinguished at the micro- and macro-level. Micro-rhythm is word prosody (lexical tone, pitch accent, and stress) based, so it can be associated with speech timing-based rhythm or syllable-, stress-, and mora-timing. Macro-rhythm refers to perceived regularity in pitch movement at the phrasal level (within an intonation phrase), whether or not it results from stress, lexical/ phrasal/boundary tone, or both. Figure 2.4 from Jun (2012, p. 2) illustrates the two types of rhythm.

Figure 2.4 Illustration of micro- and macro-rhythm



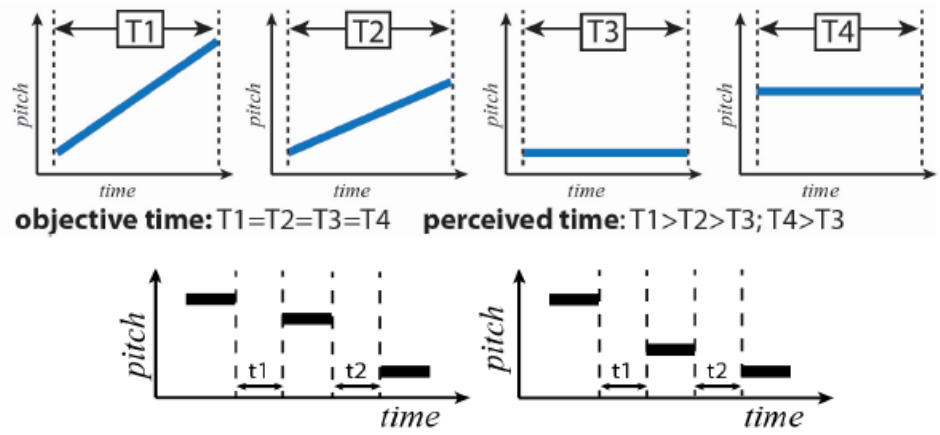
According to Jun (2012), there are three criteria to capture macro-rhythm: low/high pitch alternation, similarity and regularity of sub-tonal units which can be larger or smaller than a word. Strong macro-rhythm is associated with a pitch contour having alternating low and high pitch (LH or HL) and similar sub-tonal units occurring regularly (LH-LH-LH… or HL-HL-HL…). By these criteria, syllable- or mora-timed languages tend to have a strong macro rhythm, because they can freely use pitch instead of duration variation to cue word prominence. In contrast, stress-timed languages tend to have a weak macro-rhythm, because they can only use duration freely instead of pitch variation to cue word prominence. However, there are some exceptions. For example, a tonal language like Chinese is not likely to have a strong macro-rhythm, because its lexical

tones determine that Chinese phrases will not have regular pitch patterns due to varying tones they may contain. In general, macro-rhythm can be used to describe rhythmic patterns not easily characterized by stress (Jun, 2011).

2.1.2.3 Interaction between duration and pitch

The interaction between timing and pitch-based prominence has long been recognized in a so-called auditory kappa effect in speech, which describes the distortion of timing caused by pitch variations (Cohen, Hansel, & Sylvester, 1954; Yoblick & Salvendy, 1970). For example, vowels carrying contour or high level tones are perceived longer than vowels carrying level or low level tones (Lehiste, 1976; Wang, Lehiste, Chuang, & Darnovsky, 1976; Pisoni, 1976; Shigeno, 1986; Henry & McAuley, 2009; Yu, 2010). Mackenzie (2007) attributes the kappa effect to the assumption that contour changes may be more salient than pitch interval changes. A perceived lengthening effect of dynamic f0 on vowels is also found among listeners with different language backgrounds (Cumming 2011), though it is still questionable as to whether or not this effect is language universal or specific based on Lehnert-LeHouilier's (2007) study. Brugos and Barnes (2012) in their study of prosodic grouping also found that tones closer in pitch (height if they are level tones or shape if they are rising or falling tones) are also perceived as closer in time, and listeners tend to group syllables close in pitch more than in time. Figure 2.5 illustrates two cases of the auditory kappa effect (copied from Brugos & Barnes, 2014, p. 388):

Figure 2.5 Illustration of the auditory kappa effect



The upper figure shows that the rising tone is heard longer than level pitch (T1> T2> T3), the higher rising tone longer than lower rising tone (T1> T2), and the high level pitch longer than low level tone (T4> T3). The lower figure shows that t1 (t: silent intervals between two level tones) is perceived shorter than t2 (t1< t2) on the left but longer (t1> t2) on the right, despite that t1 and t2 are the same in length. The perceived differences are caused by the fact that the difference is smaller between the first and middle pitch levels on the left than between the middle and last pitch levels on the right.

There is also the tau effect, which refers to the distortion of pitch caused by timing variations. In a study by Henry, McAuley, and Zaleha (2009), for example, variations in timing introduce systematic distortions in perceived pitch, so the tau and kappa effects lead Henry et al. (2009) to the conclusion that pitch and timing relations are fundamentally interdependent in perception. Some researchers have explored the perceptual role of both duration and f0 in the creation of rhythm (Dilley & McAuley, 2008; Kohler, 2009a; Cumming, 2010, 2011) and made some progress in finding the relationship between duration and pitch in perceived rhythm.

Dilley and McAuley (2008) tested whether or not the last two words in a sequence were perceived as a compound depends on the f0/duration patterns of their preceding words. They found that participants' perception of prosodic constituency was based on both f0 and duration cues. For example, when the 8-syllable sequence "channel dizzy foot note book worm' is assigned the pitch sequence "HL HL HL H L H" and the syllable duration is held constant, listeners tend to perceive the last two syllables as two separate words 'book worm'. That is to say, a falling pitch contour helps to create a monosyllabic context. On the other hand, when the 8-syllable sequence is assigned the pitch sequence "LH LH L H L H", listeners tend to perceive the last two syllables as a disyllabic compound 'bookworm' instead. Therefore, a rising pitch contour helps to create a disyllabic context. The same effect can be found in two other conditions: when f0 patterns are held constant but duration patterns are manipulated to create either mono- or di-syllabic contexts and when both f0 and duration patterns are manipulated to create either mono- or di-syllabic contexts. Furthermore, the simultaneous manipulation of f0 and duration has the strongest effect on the final two words perceived as one word while the manipulation of duration alone condition has the smallest effect. In other words, pitch plays a larger role than duration in creating perceptual grouping.

Kohler (2009a) conducted another type of experiment on prominence perception of the disyllable *baba*. Specifically, he systematically changed f0 (specifically, the direction of the pitch movement), duration, and acoustic energy across the bisyllable and then asked sixteen German listeners to judge whether the first or second syllable was more prominent. What he found is that f0 is a more powerful cue than the other two, because changing f0 on the second *ba* from falling to rising-falling produces a steep change from

first-syllable to second syllable prominence during the identification task. On the other hand, changing duration and acoustic energy respectively from long-short and strong-weak to short-long and weak-strong also produces a shift of prominence from the first to second syllable, but not as steep as changing f0. Furthermore, if f0 is rising-falling, prominence perception cannot be counteracted by changes of duration and acoustic energy. In other words, f0 carries greater weight than the other two for prominent-syllable perception.

Based on an extensive survey of previous rhythm research, Kohler (2009b) concluded that rhythm was related to recurring timing patterns of at least four acoustic parameters, f0, duration, energy, and spectral dynamics. Particularly, he pointed out that f0 was more powerful as a chunking device than the other three but has barely been included in rhythm research, and rhythm should be related to global temporal patterns rather than a linear succession of local metrics restricted to segmental durations.

Cumming (2010, 2011) also explored the effect of f0 and duration on perceived rhythm. Two crucial findings are that 1) pitch and duration are interdependent cues in the perception of rhythmic grouping and 2) the relative weighting of tonal and durational cues depends on listeners' native language. In her 2010 study, Cumming used a judgment task to investigate how rising pitch and increased duration affect rhythmic grouping of syllables. Specifically, she manipulated f0 and duration in the second and/or third syllable in a 5-syllable sequence and then asked native speakers of Swiss German, Swiss French, and French (36, 38, & 36 participants, respectively) and bilingual speakers of Swiss German and Swiss French (20 participants) to judge whether the 5-syllable sequence has a 3+2 or 2+3 grouping. The results show that the two cues are significantly

more effective when heard simultaneously than either one of them is heard. Whether one cue is more effective than the other, however, depends on the listeners' language background. For Swiss German listeners, a sufficiently large f0 rise (=30Hz) has more influence than an increase in duration (1.5 times longer) on rhythmic grouping, but for Swiss French and French listeners, increased syllable duration has a greater perceptual weight than rising pitch (f0). Native language influence on rhythmic grouping is more evident in bilingual responses: bilinguals responded like Swiss German when listening to Swiss German stimuli but like Swiss French when listening to French.

Cumming (2011) went on to investigate how f0 and duration influence the perceived rhythm of sentences. She asked native speakers of Swiss German, Swiss French, and French (47, 50, 48 participants, respectively) to judge whether a given sentence has the most natural-sounding rhythm. The given sentence is manipulated in terms of duration and f0, with changing duration and f0 on a certain syllable. The main finding is that f0 and duration, when one of them is in the normal range and another made deviant from the normal range, can influence the listeners' perception of rhythmicality of the sentence. In other words, the duration and f0 excursion of prominent syllables both have to be non-deviant for speech to be judged rhythmic.

As for which cue is weightier in the judgment, the finding is that it depends on the listeners' native language. Swiss French and French listeners are sensitive both to tonal and durational manipulations but Swiss German listeners are more sensitive to durational than to tonal manipulations. There seems to be a contradiction between this finding and the earlier one, where it is pitch rather than duration that is a more effective cue to Swiss German (Cumming, 2010). Cumming (2011) explained it away based on the fact that the

two studies examine different aspects of rhythmic perception: the 2010 study requires the listeners to locate the boundary between two rhythmic groups and the duration and f0 manipulations do not deviate from a normal production, but the 2011 study requires the listeners to judge the rhythmic naturalness of a sentence with multiple rhythmic groups and the duration and pitch manipulations are deviant. Swiss German has vowel and consonant length contrasts, so the native speakers are sensitive to deviant segmental duration in the sentence. In grouping words, if both cues are in their respective normal ranges, the listeners naturally give more attention to the cues that vary more, as segment duration is constrained by phonological length contrasts in Swiss German and the less constrained cue for Swiss German is pitch. Consequently, a sufficiently large pitch rise (30Hz) wins out in the judgment of rhythmic grouping.

Note that Cumming (2011) did not give the listeners the definition of rhythm before they performed judgment tasks, yet they were able to complete the judgment tasks systematically based on pitch and duration properties. A post-test survey shows that all the listeners have their own ideas about what rhythm is like and their responses reflect the complexity of rhythm: intuition, fluency of speech, intonation, word groups, length of vowels/syllables, pauses, speech style, speed/timing/tempo, stress/accentuation, sentential meaning/structure can all serve as criteria. However, some criteria are more commonly used than others and differ from group to group: The Swiss German listeners tend to base their judgment on feeling and intuition (or how they personally would say it), stress/accentuation (both duration and pitch related), and vowel/syllable length (duration related), whereas the Swiss French and French listeners on speed (duration related), intonation (pitch related), and stress/accentuation. The Swiss German group's preference

for the durational cue in syllables is consistent with the finding from the preceding perceptual study (2011), again reflecting structural properties of the Swiss German language. Note that stress/accentuation is a common criterion shared by all three language groups, further suggesting that both pitch and duration are important in rhythm perception. In general, native language properties determine which cues, tonal or durational, the listeners attend more to when making a rhythmic judgment.

Cumming (2011) also explored the perceptual influence of f0 and duration on each other and found that f0 movement would increase the perceived duration. Also, it is found that the listeners perceive vowels with a dynamic f0 as longer than those with a level f0. Therefore, duration alone may not be able to capture rhythmic patterns the listeners actually perceive.

In fact, Arvaniti (2009) emphasized that speech rhythm as a perceptual phenomenon should be distinguished from speech timing: while the latter is related to durational patterns of speech events, the former has to do with the abstract pattern of periodicities which is extracted by listeners from durations and other acoustic cues such as f0, intensity, and spectral dynamics. Therefore, it is necessary for rhythmic research to focus more on rhythmic grouping or patterns of prominence than on durational variability of segments (vowels or consonants) alone.

Rhythmic grouping, according to Cumming (2011), is a cover term for the grouping of prosodic units such as syllables, metrical feet, and phrases to form rhythmic perception. It is based on the relative prominence of certain prosodic units as manifested through such acoustic cues as duration (perceived as length), f0 (perceived as pitch), and intensity (perceived as loudness). Rhythm is hence considered as the perceived regularity induced

by the grouping of these prominences (Cumming, 2011). All of the acoustic cues can contribute to the perception of rhythm, yet how much each cue contributes varies from language to language (Cumming, 2011). That is to say, rhythmic grouping is based not only on duration (length perceptually) but also on f0 (pitch perceptually), intensity or amplitude (loudness perceptually), and other acoustic cues (e.g., spectral dynamics). Therefore, it is important for every rhythmic study of different languages to adopt a psychological understanding of rhythm.

Table 2.3 summarizes three of the aforementioned prominence-based rhythmic studies according to the acoustic cues they used and the main findings.

Table 2.3 Summary of three prominence-based rhythmic studies

Prominence-based rhythmic studies	Acoustic parameters manipulated	Main finding
Dilley & McAuley, 2008	f0, duration	Distal f0 and duration patterns affect the perceptual grouping of locus syllables.
Kohler, 2009a,b	f0, duration, amplitude	F0 is a more powerful cue than the other two for prominence perception.
Cumming, 2010, 2011	f0, duration	Rising f0 and increased duration are significantly more effective than either one of them when heard simultaneously.

2.2 Quantifying rhythm

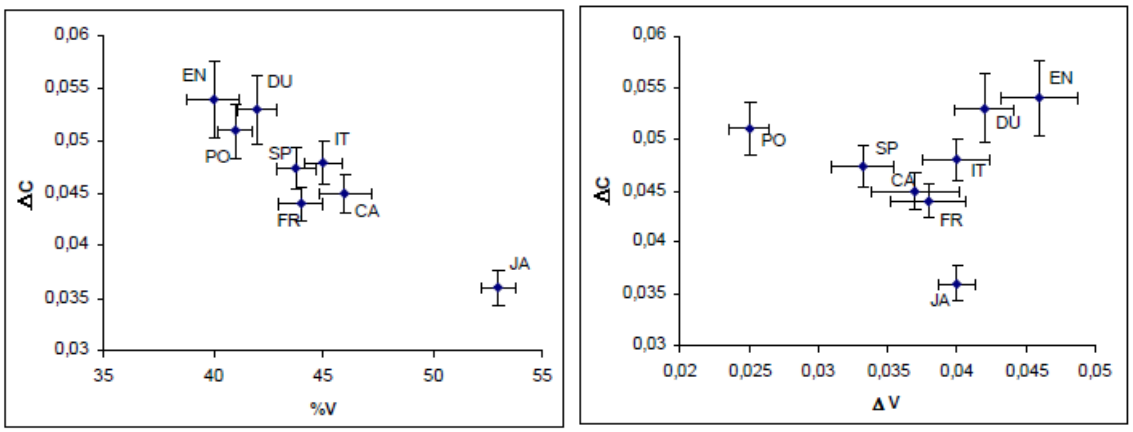
Despite that early empirical research (e.g., Faure, Hirst, & Chafcouloff, 1980; Wenk & Wioland, 1982) found evidence of isochrony neither in stress- nor in syllable-timed languages, some researchers persisted in seeking further empirical evidence (e.g., Ramus et al., 1999; Grabe and Low; 2002), and with the advance of computer technology, they developed more complicated rhythmic metrics to replace previous simple calculations of stress intervals/syllable duration. As a result, they are finally able to find some evidence for the traditional classification of speech rhythm. Section 2.2.1 and Section 2.2.2 respectively introduce duration- and pitch-based metrics used to study speech timing and melody. Section 2.2.3 introduces voice source metrics used to study voice quality. Section 2.2.4 and 2.2.5 respectively review previous rhythmic studies on dialects of non-tonal languages and of Chinese.

2.2.1 Duration-based metrics

Duration-based metrics are used to reveal patterns of speech timing. Instead of measuring stress intervals or syllable duration directly, the recent metrics measure consonant and vowel intervals and use them to infer timing patterns. For example, Ramus et al. (1999) first proposed a set of consonant and vowel duration-based metrics to quantify rhythm: %V (percentage of duration taken up by vocalic intervals in a sentence), ΔV (standard deviation of the duration of vocalic intervals in a sentence), and ΔC (standard deviation of the duration of consonantal intervals in a sentence). The rationale for adopting these metrics is that stress-timed languages, unlike syllable-timed ones, allow vowel reduction and complex consonant clusters, so vowel duration and consonant duration should vary more in the former than in the latter. This hypothesis is largely supported by the results:

When %V-ΔC and ΔV-ΔC are paired respectively to serve as the x-y axis by which languages are plotted, stress-timed languages such as English, syllable-timed languages such as French, and mora-timed languages such as Japanese are clustered in separate rhythmic groups as expected. Figure 2.6 below illustrates the findings (copied from Ramus, 2002, p. 1 & p. 3):

Figure 2.6 Rhythmic classification based on %V-ΔC and ΔV-ΔC

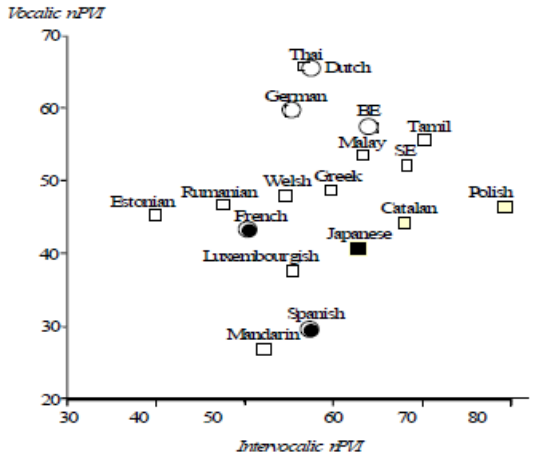


In the figure on the left, English (EN) has a smaller %V but larger ΔC than French (FR), indicating that English is more stressed timed or less syllable-timed than French. Japanese (JA), on the other hand, has much smaller ΔC but larger %V than the rest of the languages examined, indicating that it has a distinct rhythmic pattern. The figure on the right shows a similar result, as English, French, and Japanese are separated nicely from one another.

Grabe and Low (2002) developed the second set of duration-based rhythmic metrics: nPVI (normalized Pairwise Variability Index) and rPVI (raw Pairwise Variability Index) based on the same rationale Ramus et al. (1999) used. These two metrics make pairwise comparisons of successive vocalic and intervocalic intervals, respectively. The difference between rPVI and nPVI is that the latter is normalized for speech rate while

the former is not. The normalization process is used to reduce the influence of speech rate on vowel duration, since speech rate is found to correlate directly with vowel duration (Gay, 1978). When vocalic nPVI and intervocalic rPVI serve respectively as the x- and y-axis by which languages are plotted, stress-timed English, syllable-timed French, and mora-timed Japanese are also able to separate nicely, as illustrated in Figure 2.7 (copied from Grabe & Low, 2002, p. 7):

Figure 2.7 Rhythmic classification based on rPVI-nPVI



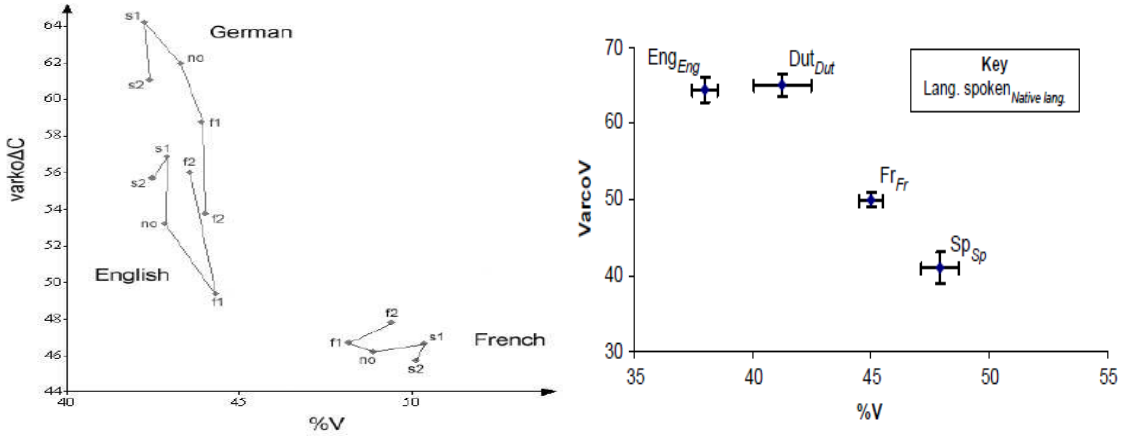
Stress-timed British English (BE) is located in the upper right region with relatively large nPVI and rPVI, indicating that it has more varied vocalic and intervocalic intervals. Syllable-timed Spanish is in the lower region towards the left with relatively small nPVI and rPVI, meaning that it has less varied vocalic and intervocalic intervals. Mora-timed Japanese is in the central region towards the right with moderate nPVI and rPVI, indicating that it has moderate variation in vowel and consonant duration. Mandarin is close to Spanish with smaller nPVI and rPVI, suggesting that it has relatively stable vocalic and intervocalic intervals, so it can be classified as syllable-timed.

Dellwo (2006) introduced an additional metric, VarcoΔC (variation coefficient of ΔC), the deviation value of ΔC corrected for speech rate. When VarcoΔC serves as the y-axis

in place of ΔC in the earlier %V-ΔC pair, stress-timed English and syllable-timed French are able to separate better. White and Mattys (2007) further proposed the use of VarcoV (variation coefficient of ΔV) in addition to VarcoΔC (also written as VarcoC).

When %V and VarcoV serve respectively as the x- and y-axis, stress-timed English and syllable-timed Spanish can also separate very well. Figure 2.8 illustrates the above findings (Left: copied from Dellwo, 2006, p. 5; Right: copied from White & Mattys, 2007, p. 517):

Figure 2.8 Rhythmic classification based on %V-varcoΔC and %V-varcoV



In the figure on the left, Stress-timed English and German are on the left side ranging from middle to high depending on speech rate (s1: very slow; s2: slow; no: normal; f1: fast; f2:fastest possible): Slower speech rate tends to induce more stress-timed rhythm, as the associated %V is smaller but varcoΔC is larger. In contrast, syllable-timed French is on the lower right side regardless of speech rate, but slower speech rate tends to induce more syllable-timed rhythm, as the associated %V is larger but varcoΔC is smaller.

In the figure on the right, stress-timed English and Dutch are on the upper left region and syllable-timed French and Spanish are on the lower right region, respectively indicating that stress-timing is related to smaller vowel percentage but greater speech

rate-induced variation in vocalic intervals while syllable-timing to larger vowel percentage but smaller speech-rate-induced variation in vocalic intervals. In other words, vowel duration is more sensitive to speech rate in stress-timed than in syllable-timed languages.

As for which pair of rhythmic metrics is most successful in classifying rhythm, Loukina, Kochanski, Rosner, Keane, and Shih (2009) and Loukina, Kochanski, Shih, Keane, and Watson (2011) tested all the metrics on a large corpus of speech involving five languages including Mandarin and found no single pair performing well in separating all the languages: it takes at least three metrics to separate all five languages. Also, within-language variability in these metrics is found larger than or comparable to that between languages. They also investigated the influence of speech rate on rhythmic measures and found that metrics already taking speech rate into consideration fared better than those without. The Mandarin-related results show that the correct rate for the %V-ΔC pair is 64% and the rPVI-C and nPVI-V pair is 70% (Loukina, et al., 2009).

Table 2.4 summarizes four of the aforementioned duration-based rhythmic studies according to the duration-based metrics they developed, the languages they classified, and how well the languages are classified by the metrics they used.

Table 2.4 Summary of four duration-based rhythmic studies

Duration-based rhythmic studies	Duration-based metrics	Languages compared	how well languages are separated
Ramus et al., 1999	%V, ΔV, ΔC	8 languages: English, French, Japanese, etc.	Yes among stress-, syllable-, and mora-timed
Grabe & Low, 2002	nPVI, rPVI	18 languages: English, Spanish, Japanese, Greek, Mandarin, etc.	Yes between stress- and syllable-timed; No for mora-timed and unclassified or mixed
Dellwo, 2006	VarcoΔC	German, English, French	Yes between stress- and syllable-timed
White & Mattys, 2007	VarcoΔV	English, Dutch, Spanish, French	Yes between stress- and syllable-timed

All the duration-based metrics introduced above involve measuring vowel and consonant or intervocalic intervals, but Dellwo, Fourcin, and Abberton (2007) and Fourcin and Dellwo (2009) later explored the use of different intervals, voiced and voiceless intervals, to replace vocalic and consonantal intervals and the results show equally well rhythmic classification. The division of voicing and non-voicing is both methodologically and perceptually more advantageous: the data segmentation process can be fully automated because it relies solely on acoustic information not on phonological knowledge. The latter requires human intervention, so it is error prone and labor consuming; on the other

hand, as previous research shows that even infants, who have yet to acquire language, are able to distinguish between different rhythms (Nazzi, Bertoncini, & Mehler, 1998; Nazzi & Ramus, 2003), it is reasonable to assume that naïve listeners make heavy use of voiced or voiceless signals to distinguish rhythm.

Another methodological improvement is also seen in a study by Galves, Garcia, Duarte, and Galves (2002). A rough measure of sonority directly from speech signals is developed instead of the measure of vocalic and consonantal intervals. It assigns values close to 0 for regions with obstruent signals and 1 for regions with sonorant signals in the spectrogram. Then two measures, *S* (sample mean of a sonority function) and δS (How large the variation is in the region with high obstruency), are then formulated to replace %V and ΔC. The two sonority-based metrics not only correlate with %V and ΔC quite well but also eliminate the need for segmentation of vowels and consonants.

Last, a particular way of quantifying rhythm is implemented in Stojanović's (2013) study of the effect of phonotactics on rhythmic classification of 21 different languages. Specifically, she developed a set of phonotactic metrics to capture the levels of the phonotactic complexity involving consonant-cluster size, sonority relations, and word length. The phonotactic length-based metrics parallel with actual duration-based metrics such as %V and ΔC. For example, an onset has a length value of 3 if it contains three segments and 2 if it contains two segments, regardless of the actual duration of the onset. The results show that the two types of metrics are highly correlated, so, for example, where syllable structure is simple, both the phonotactic length- and actual duration-based ΔC are small. It is also found that word-final consonant clusters are a defining factor for rhythmic class. In other words, languages with similar word-final consonant clusters are

also rhythmically similar. These findings suggest that speech rhythm may be quantified based on phonotactics.

Despite various duration-based metrics, there are always some languages evading all kinds of classifications. Korean, for example, is known to be difficult to classify, because of the conflicting results in the past (Lee, Jin, Seong, Jung, & Lee, 1994; Cho, 2004; Mok & Lee, 2008; Arvaniti, 2009, 2012). According to Mok and Lee (2008), Korean has a simple syllable structure similar to a syllable-timed language, but its strong final lengthening and frequent use of taps are more like a stress-timed language. In addition, the choice of speech materials such as designed versus uncontrolled speech and difference in elicitation methods such as read versus spontaneous speech both have a bearing on the final results. Therefore, mixed-timed languages like Korean call for non-duration-based rhythmic quantification.

2.2.2 Pitch-based metrics

If language rhythm cannot be reliably characterized by durational variability in consonants and vowels alone, it is necessary to examine other factors influencing speech rhythm. In fact, researchers have already made some successful attempts in using pitch to quantify melody patterns and in turn using melody patterns to infer speech rhythm (Vicenik & Sundara, 2008; Hirst, 2013).

In Vicenik and Sundara's (2008) study of closely related languages, American English (AE) and German (G), and closely related dialects, American and Australia English (AuE), both duration and f0 are used to determine how these languages and dialects are different from one another. Their study includes two parts: Experiment 1, an acoustic study of hundreds of sentences read in American English, German, and Australian

English; and Experiment 2, a perceptual study of whether subjects could discriminate between these prosody-similar languages/dialects using only prosodic cues (i.e., duration/f0).

In Experiment 1, both duration- and pitch-based measurements are taken from the recorded sentences to reveal the duration- and pitch-based prominence patterns of the three languages/dialects. The duration measurements include almost all the previous duration-based metrics with the one important change: instead of vocalic/consonantal intervals, the metrics use sonorant/obstruent (Son/Obs) intervals. Hence the previous duration-based metrics, %V, ΔV, ΔC, rPVI/nPVI for C and V, VarcoC, and VarcoV respectively become %Son, ΔSon, ΔObs, rPVI/nPVI for Obs and Son, VarcoObs, and VarcoSon, instead. The pitch-based measurements include minimum, maximum, and mean pitch (f0), the number of pitch rise, average rise height, and average slope (the rate of f0 change) for sonorant segments in each sentence. These measurements are used to identify and compare pitch-based intonation patterns between American English and German and between American and Australian English.

According to Vicenik and Sundara (2013), the measure of pitch rise is important in their study because stressed syllables in these languages/dialects are often marked with a high tone following a shallow or steep rise, represented by H* or L+H* in the Tobi transcription system developed by Beckman and Pierrehumbert (1986). Also, a basic assumption here is that a language using a shallow pitch rise more frequently should have a lower average slope than a language using a steep rise more frequently. In other words, the higher the average pitch rises, the steeper the average slope is.

Next, the duration and f0 data were subjected to a stepwise binary logistic regression analysis in order to obtain classification scores for two language/dialect pairs under three different conditions, pitch only (using only pitch cues), duration only (using only duration cues), and pitch and duration combined. It is found that in the pitch only condition, American English and German can be classified correctly 86.3% of the time, more than in the duration only condition (70.2%). In the duration and pitch combined condition, the classification score increases to 89.2%. The same can be said between American and Australian English: in the pitch only condition, these two dialects can be classified correctly 78.8% of the time, more than in the duration only condition (76.6%), but in the pitch and duration combined condition, the classification score increases to 87.8%. These findings indicate that duration and pitch are acoustically distinct even between closely related languages and dialects and pitch may play a more important role than duration in language/dialect discrimination.

The subsequent perceptual experiment conducted by Vicenik and Sundara (2013) basically supports the above findings. When native speakers of the two language/dialect pairs were tested on their response to stimuli, which are utterances with segment information removed but prosodic information kept, they were able to distinguish between their native language/dialect and rhythmically similar language/dialect based either on duration information alone, on pitch information alone, or on both, though the performance level under all three conditions is just slightly above chance.

In addition, results from individual duration- and pitch-based measurement also reveal some patterns relevant to the present study but not explicitly discussed by Vicenik and Sundara (2013; based on the data table provided on p. 300). First, T-test comparisons

show that four duration-based metrics (%Son, nPVI_Son, varcoObs, & varcoSon) are non-significant between American and Australian English (AmE & AuE), while only one (nPVI_Obs) is non-significant between American English and German (G). Second, three pitch-based metrics (minf0, meanf0, & slope) are non-significant between American and Australian English, while only one (minf0) is non-significant between American English and German. In other words, the duration and pitch difference is larger between languages than between dialects of the same language.

Last, there are noticeable corrections among duration- and pitch-based results as well: First, there is a highly positive correlation between pitch rise height and pitch slope for all three languages/dialects; second, there is a highly positive correlation between %Son and varcoSon but negative correlation between %Son and varcoObs for all three languages/dialects; Third, there is a highly positive correlation between %Son and pitch rise slope for G and AmE/AuE.

The positive correlation between rise height and slope (see the upward trend line in Figure 2.9) means that the higher and faster the pitch rises, the more intonational (or more pitch fluctuation) the sentence becomes. German has smaller rise height and slope than both American and Australian English, indicating that it is less intonational than the latter two English dialects. The positive correlation between %Son and varco_Son and the negative correlation between %Son and varcoObs (see the upward trend line in Figure 2.10 & the downward trend line in Figure 2.11) mean the larger %Son, the larger varcoSon and the smaller varcoObs. If varcoSon is equivalent to varcoV and %Son to %V, then based on White and Mattys (2007), varcoSon should form an inverse relationship with %Son, just as varcoV does with %V. Here the relationship between

varcoSon and %Son is reversed. A possible explanation is that varcoSon is not a good indicator of stress-timing: larger varcoSon does not necessarily mean more stress-timing. The inverse relationship between %Son and varcoObs, on the other hand, is consistent with the one between %V and varcoC. German has smaller %Son but larger varcoObs than both American and Australian English, indicating that it is more stress-timed than the latter two English dialects.

The positive correlation between %Son and pitch rise height (see the upward trend line in Figure 2.12) means the larger the %Son and the rise height, the less stressed-timed and more intonational. German has smaller %Son but larger rise height than both American and Australian English, indicating that it is more stressed-timed but less intonational than the latter two English dialects. In other words, stress-timing is inversely related to intonation-ness.

Figure 2.9 Illustration of correlation between rise height and slope of pitch

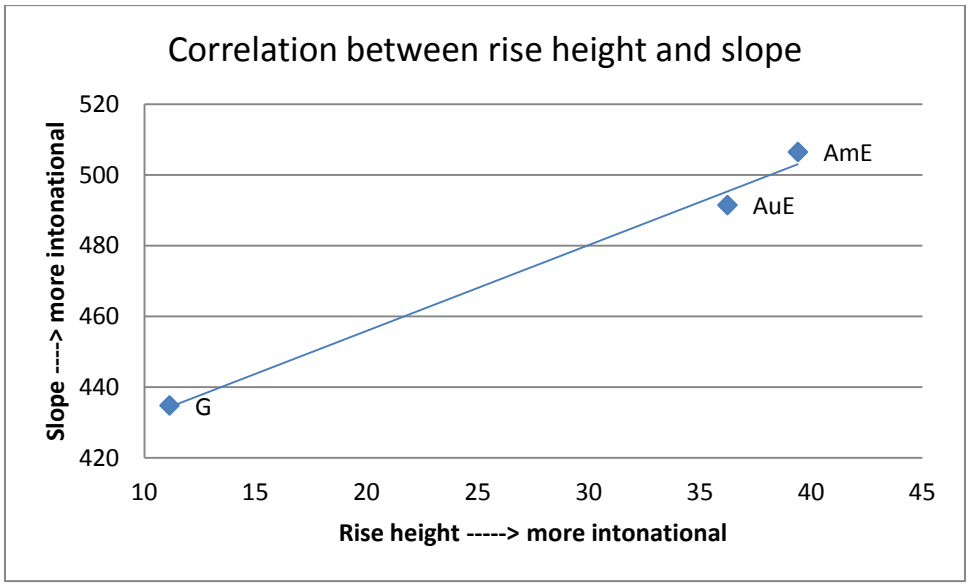


Figure 2.10 Illustration of correlation between %Son and varcoSon

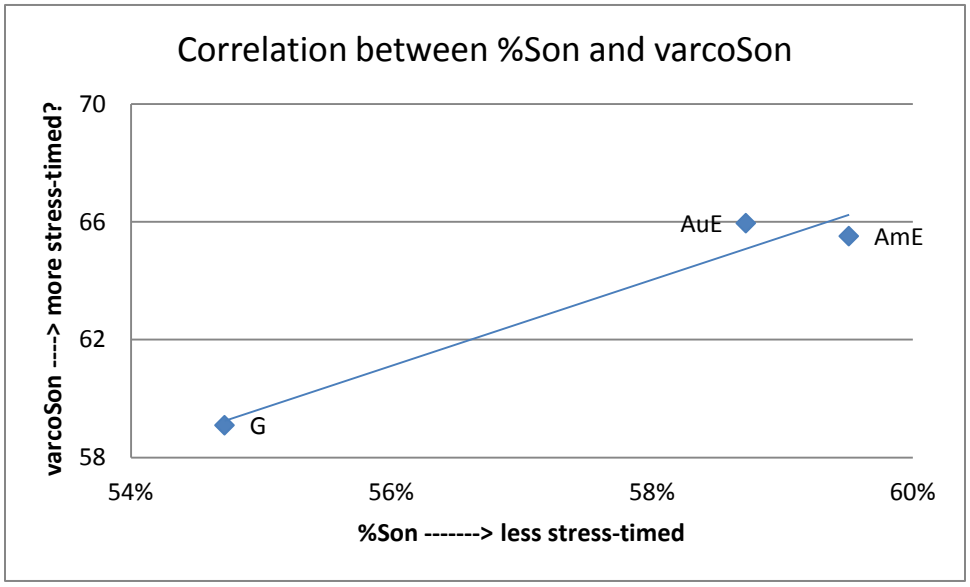


Figure 2.11 Illustration of correlation between %Son and varcoObs

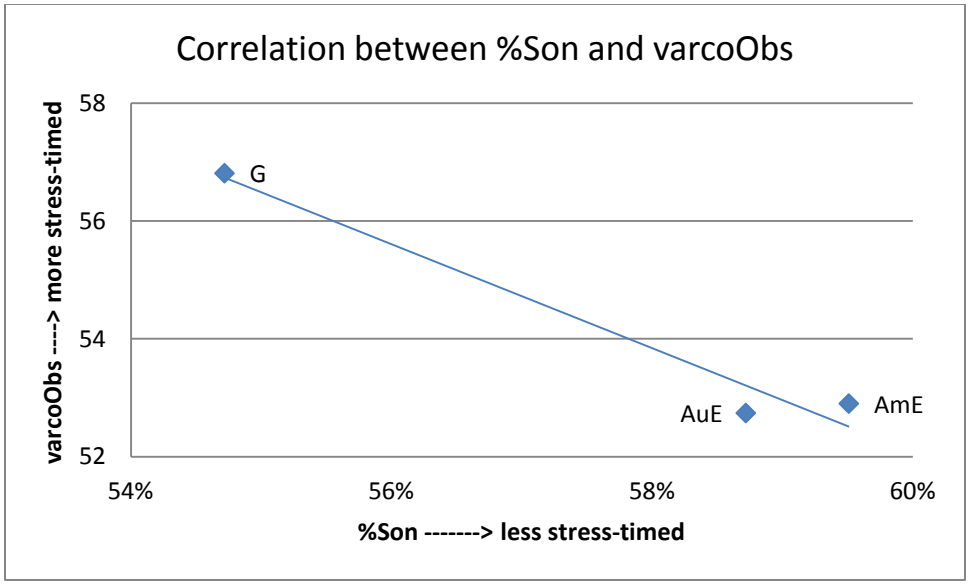
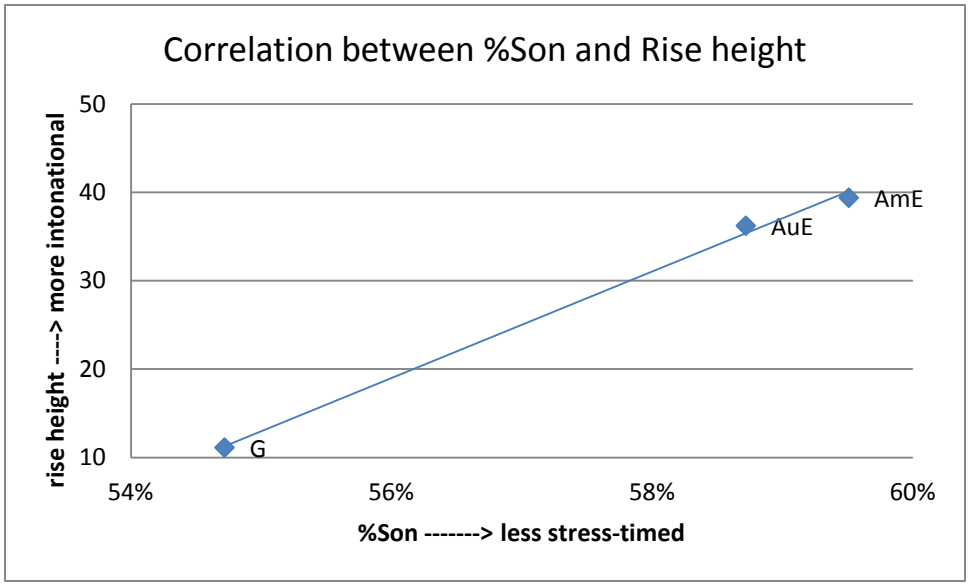


Figure 2.12 Illustration of correlation between %Son and pitch rise height

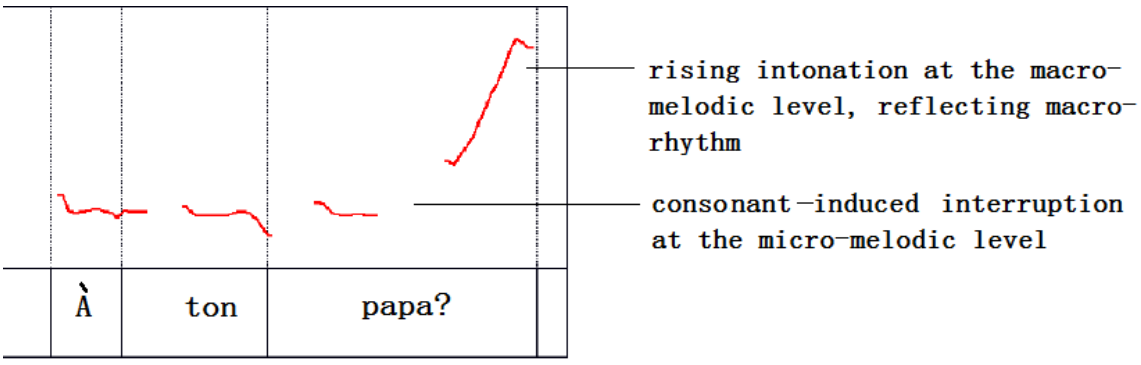


Recall that stress-timing is associated with weak macro-rhythm and syllable-timing with strong macro-rhythm (Jun, 2005). If stress-timing is negatively related to intonation-ness, then intonation-ness is positively correlated with strong macro-rhythm and in turn syllable-timing.

Another way to understand pitch-based rhythm is from Hirst (2013). He developed a set of so-called melody metrics for quantifying prosodic typology. He defined pitch excursion and pitch slope in terms of octave interval and rise- and fall-slope and then calculated the mean and standard deviation of these melody metrics for British English, French, and Mandarin Chinese speech samples. The results show a clear difference between the former two languages and Chinese, as the former two have larger mean and standard deviation for both interval and slope measures, indicating that pitch movement in Mandarin Chinese is more dramatic, faster, variable than in English and French or simply Mandarin Chinese sound more melodic or melodious than the two languages.

Note that Hirst (2013) used the term ‘melody’ instead of ‘intonation’ used in Vicenik and Sundara (2013), because he preferred ‘intonation’ used with an abstract phonological system occurring in a prosodic hierarchy and ‘melody’ used to mean pitch movement over time (Hirst, 2011). Hirst’s (2011) study also distinguishes micro-melody from macro-melody. Micro-melody refers to pitch discontinuity or fluctuations caused by the disturbing airflow in consonant production. Macro-melody refers to accentuation and intonation induced pitch movements. Therefore, they are not the same as micro- and macro-rhythm defined by Jun (2005). When pitch-based metrics are applied to speech, they operate at the lowest micro-melodic level, but able to reflect the influence from macro-melody and in turn from both micro- and macro-rhythm. In other words, micro-melody is segment-related but macro-melody is prosody-related, and it is formed based on both micro-rhythm and macro-rhythm. What pitch-based metrics capture is the accumulated influence of these levels on pitch movement, as shown in Figure 2.13 (adapted from Hirst, 2011, p. 64):

Figure 2.13 Illustration of micro- and macro-melody



In this phrase “À ton papa?” (‘To your Daddy?’), the stops [t, p] introduce gaps into the otherwise continuous pitch contour, reflecting micro-melody. Then at the end of the phrase, the rising intonation is introduced into the otherwise falling pitch contour,

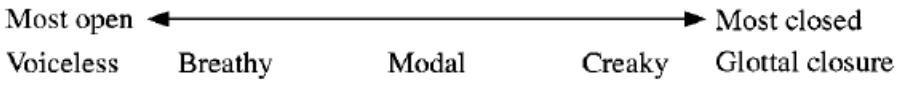
reflecting a macro-melody. The macro-melody is also a reflection of macro-rhythm, and it signals a question. No micro-rhythm is clearly shown here as all the words are short, carrying no lexical tone, pitch accent, or stress.

The relationships among melodiousness (Hirst, 2011), intonation-ness (Vicenik & Sundara, 2013), macrorhythmicality (Jun, 2005), and stressed-timedness can be described as follows: intonation-ness or melodiousness is positively correlated with syllable-timing and strong macro-rhythm and negatively with stress-timing and weak macro-rhythm.

2.2.3 Voice source metrics

Speech sounds can have certain accompanying voice quality used either contrastively or as a perception-enhancing cue. Voice quality is traditionally viewed as a function of the glottis, involving different degrees of glottal constriction during phonation, ranging from the most open (for voiceless sounds) to the most constricted (heard as a glottal stop), with breathy, modal, and creaky voice in between, as shown below (reproduced from Gordon & Ladefoged, 2001):

Figure 2.14 Continuum of phonation types



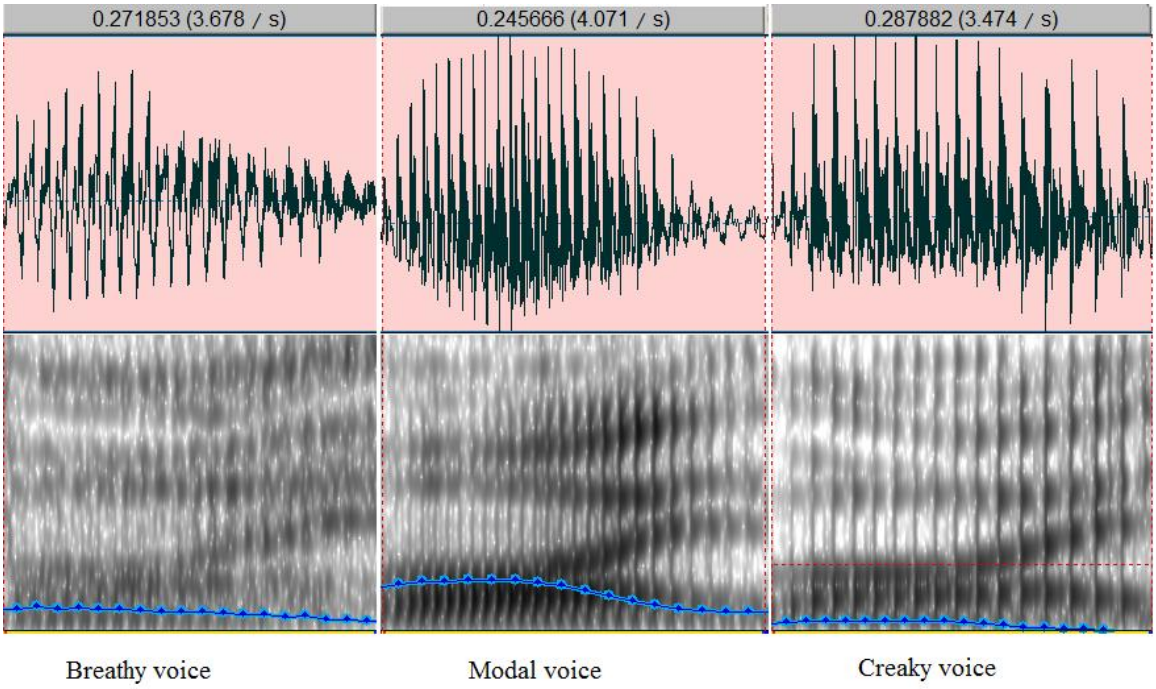
Later articulation models such as Esling's (2005) laryngeal articulator model, Edmondson and Esling's (2006) valves-of-the-throat model, and Moisik's (2013) epilarynx (the supraglottal part of the larynx) model acknowledge the major role the whole larynx plays in producing different voice qualities. Taking the complicated vowel/tone systems of four Tibeto-Burman languages (Jingpho, Hani, Wa, & Yi) spoken

in Southwest China for example, Maddieson and Ladefoged (1985) initially found that there is a phonation-related tense-lax distinction in these languages' vowels. Further studies of Bai (also a Tibeto-Burman language spoken in Southwest China) and Yi vowels and tones using a laryngoscopy reveal that what lies behind the tense and lax distinction is a strikingly complex laryngeal constrictor mechanism rather than simply the glottal state difference (Edmondson, Esling, Ziwo, Harris, & Li, 2001).

According to Scott and Esling (2011), all the above phonation types can be nicely incorporated into these later developed models: breathy, slack, or modal voice has a natural affinity with the unconstricted epilaryngeal tube and lowered larynx while creaky or tense voice with the constricted epilaryngeal tube and raised larynx.

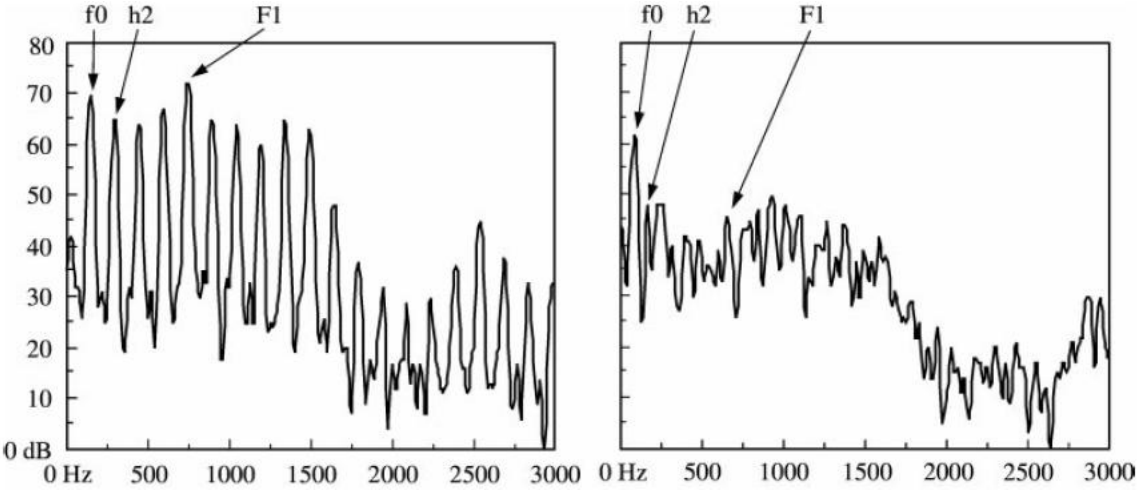
Acoustically, modal voice refers to normal speaking voice characterized by clearly defined, regularly spaced glottal pulses on the waveform display. In contrast, breathy voice involves a jagged waveform and blurry pulses (due to increased spectral noise) while creaky voice involves irregularly and widely spaced pulses (resulting in low fundamental frequency or f0). The three types of phonation are acoustically manifested in the waveforms (upper) and spectrograms (lower) in Figure 2.15. Note that the associated sound is /oi/, and it was produced by Dr. Esling and acquired through Bird and Wang (2009).

Figure 2.15 Acoustic manifestation of three phonation types

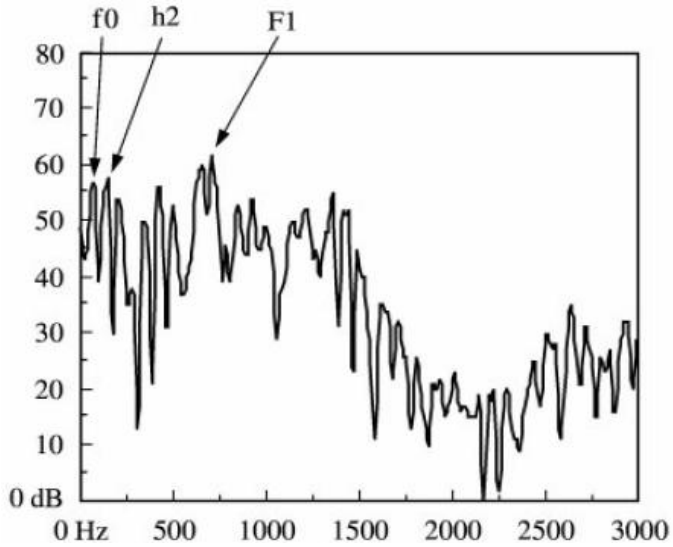


A comparison of the three spectrograms shows that breathy and creaky voices, compared to modal voice, exhibit reduced acoustic energy (or intensity; bars appear less dark in the spectrogram). Their energy differences can be captured by spectral tilt, the degree to which intensity drops off as frequency increases (Gordon & Ladefoged, 2001). There are different ways to quantify spectral tilt and two of common ones mentioned by Gordon and Ladefoged (2001) are comparing the amplitude of f0 (= H1) to that of the second harmonic (H2) or to that of the first formant (A1), as illustrated on the spectrum display in Figure 2.16 (reproduced from Gordon & Ladefoged, 2001):

Figure 2.16 Spectra for the three phonation types



Modal voice Breathy voice



Creaky voice

For modal voice, the difference in intensity between f0 and h2 (H1-H2) or F1 (H1-A1) is relatively small, forming a non-steep spectral tilt (see the upper-left figure). For breathy voice, intensity drops greatly from f0 to h2 or F1, resulting in a steeply falling spectral tilt (H1-H2 & H1-A1 > 0; see the upper-right figure). For creaky voice, the opposite is true: intensity rises from f0 to h2 and F1, resulting in a relatively steeply rising spectral tilt (H1-H2 & H1-A1 < 0; see the lower figure).

Since different types of voice quality have different acoustic properties, a number of acoustic measures have been adopted to quantify voice quality of different sounds. The above spectral tilt measures (H1-H2 & H1-A1), for example, have helped to distinguish modal, creaky, breathy vowels in Jalapa Mazatec (a language spoken in Mexico; Silverman, Blankenship, Kirk, & Ladefoged, 1995), modal and breathy vowels in Hmong (a Southeast Asian language; Huffman, 1987), and breathy and modal nasals in Tsonga (Traill & Jackson, 1988).

Improved versions of the spectral tilt metrics also occur in numerous voice quality studies (e.g., Hanson & Chuang, 1999; Iseli, et al., 2007; Keating, Kuang, Esposito, Garellek, & Khan, 2012). These later developed metrics are marked with asterisks (e.g., H1*-H2* and H1*-A3*), meaning that the metrics are corrected for the effects of formants. For example, different types of vowels have different F1 values, which may affect the actual amplitude of the harmonics (Hanson & Chuang, 1999). In addition, voice quality characteristics vary with speakers' age and sex, but the corrected metrics can capture the speaker difference. For example, H1*-H2* is found to correlate with f0 for low-pitched speakers and with F1 for high-pitched speakers (Iseli, et al., 2007). Overall, the spectral tile metrics, especially to H1*-H2*, appear to be among the most frequently used and the most effective measures of voice quality across languages (Keating & Esposito, 2006; Kuang, 2011).

There are also other voice source metrics used to quantify voice quality, some of which are CPP (Cepstral Peak Prominence), Subharmonic-to-Harmonic Ratio (SHR), and Energy. CPP and SHR are two ways of measuring aperiodicity of glottal pulses. Energy,

on the other hand, helps to detect the decrease of overall acoustic intensity (Shue, Keating, Vicenik, & Yu, 2011).

Since different metrics capture different aspects of voice quality characteristics, they will not be all successful in distinguishing voice quality. Most likely, if the whole larynx is engaged in the production, the resulting voice quality characteristics will be too complicated for the existing acoustic metrics to measure.

2.2.4 Rhythmic studies of dialects

Previous rhythmic research mainly involves classifying rhythmically similar or different languages, only a handful focusing on different varieties of the same languages, but most of which are non-tonal languages. Leemann, Dellwo, Kolly, & Schmid's (2012) study of eight Swiss German dialects and Schmid's (2012) study of nine Italo-Romance dialects are two of such research. German is considered as a stress-timed language and Italian a syllable-timed language. The two studies show that there are significant differences in timing patterns across the different dialects of the same language.

In Leemann et al.'s (2012) study, the eight duration-based metrics, %V, ΔC, ΔV, varcoC, varcoV, rPVI-C, nPVI-C, nPVI-V, are used and the speech data contain four sentences of six speakers from each of the eight Swiss German dialects. The eight dialects include four Alpine dialects and four Midland dialects, each of which can be further divided into two Eastern and two Western varieties. Statistical tests, ANOVA and t-test, are performed respectively on vocalic based measures, %V, ΔV, varcoV, nPVI-V, and consonantal based metrics, ΔC, varcoC, rPVI-C, nPVI-C, across the eight dialects, between Eastern and Western groups, and between Alpine and Midland groups. The results show that the vocalic based metrics are major cross-dialectal discriminators

between Alpine and Midland groups: Most of them show less variability for the Alpine group, which tends to retain the full vowel in unstressed syllable position, than for the Midland group. The consonantal based metrics, however, do not fare as well as the vocalic based ones in differentiating dialect groups, except that PVI-based metrics show significant differences between the Eastern and Western groups.

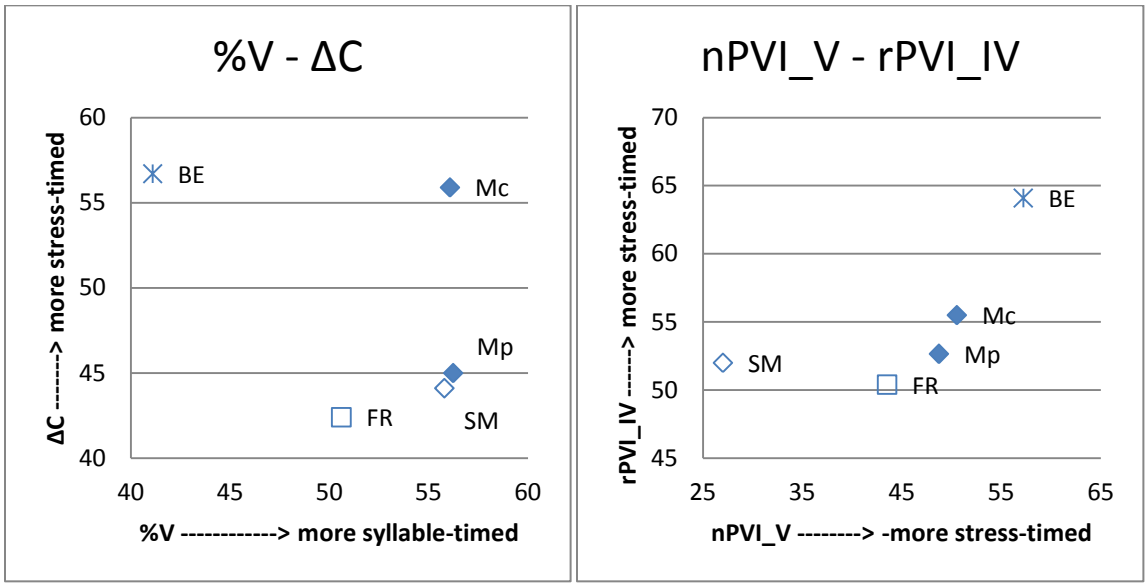
In Schmid's (2012) study, the nine dialects differ in terms of phonotactic complexity; for example, Pisan and Friulian respectively have the fewest and most syllable types. Pisan and Friulian represent typical syllable- and word-based rhythm based on Auer (1993). Six metrics, %V, ΔC, varcoC, ΔV, nPVI-V, and rPVI-C are used to analyze the speech data containing ten utterances from each of the nine dialects. The results show that Friulian stands out as the most stress-timed and Pisan the most syllable-timed. Generally, %V is well correlated with syllable complexity, ΔC and varcoC with complexity of consonant clusters, and nPVI-V with vowel reduction patterns.

2.2.5 Rhythmic studies of Chinese

As a tonal language, Chinese is considered as syllable-timed from the phonological point of view (Auer, 1993), but an experimental study by Cao (2004) finds no evidence of isochrony in Beijing Mandarin. Another study by Grabe and Low (2002) uses the duration-based nPVI and rPVI metrics to place Singapore Mandarin into the syllable-timing camp. Lin and Wang (2007) adopted a similar methodology and confirmed the syllable-timedness of standard Mandarin for the first time using the four metrics, %V, ΔC, intervocalic rPVI (rPVI_IV), and vocalic nPVI (nPVI_V). They studied two speech styles: interview-like conversation and passage-reading. A comparison of their results with those from Grabe and Low (2002) is illustrated in Figure 2.17 (based on the data

tables from Lin & Wang, 2007, p. 134; Grabe & Low, 2002, p. 12 & p. 14). The figure shows how Mandarin in the conversation style (Mc) and in passage-reading style (Mp) in Lin and Wang's (2007) study fit together with Singapore Mandarin (SM), French (FR), and British English (BE) in Grabe and Low's study (2002). The left figure plots the five types of speech by the metric pair %V-ΔC and the right one by nPVI_V-rPVI_IV. The assumption is the larger the %V and the smaller the ΔC, the more syllable-timed the speech and the larger the nPVI_V and rPVI_IV, the more stress-timed the speech.

Figure 2.17 Comparison of duration-based results among Mandarin, British English, and French



The left figure shows that Mp and SM are closer to syllable-timed FR than stress-timed EN, while Mc has a mixed pattern, as it has a larger %V comparable to Mp and SM but a larger ΔC comparable to stress-timed BE. The right figure shows that both Mc and Mp are identified with FR but SM and BE are far from the rest and even further away from each other by the nPVI measurement. The large discrepancy occurring between SM and standard Mandarin on the right is expected: Mandarin spoken in Singapore is heavily

influenced by Min Chinese (Zhou, 2006), so its rhythmic pattern is bound to deviate from that of standard Mandarin in some way.

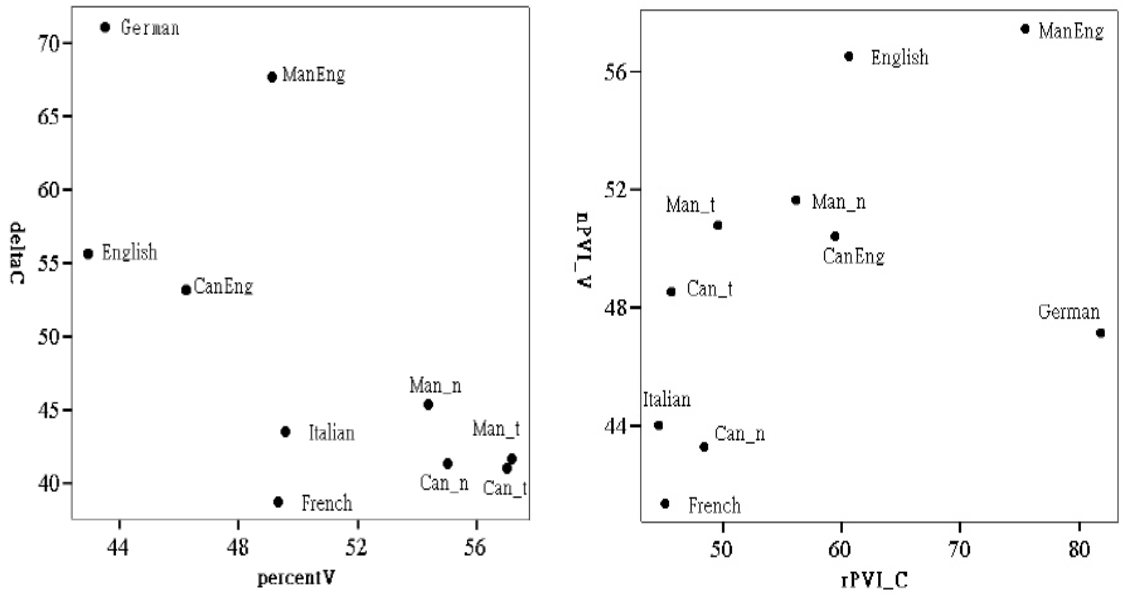
A comparison from the two speaking styles shows that almost all the measurements are larger for conversation than for passage-reading, indicating that more durational variations occur in casual speech than in formal speech. Particularly, Mc is much larger than Mp by ΔC, indicating that the casual style involves much more consonantal variation than the formal style.

Lin and Huang (2009) also performed the same kind of duration-based study on four varieties of Chinese, Mandarin (M), Shanghainese (S), Cantonese (C) and Taiwanese (T). Mandarin refers to standard Mandarin (based on a variety spoken in Beijing), Shanghainese is a sub-dialect of Wu, and Taiwanese a sub-dialect of Min. The results show that %V ranges from 58% to 52% in the following order C > M > S > T with Cantonese having the highest score while Taiwanese the lowest; ΔC ranges from 8.03 to 5.04 in this order T > S > C > M with Taiwanese having the highest score while Mandarin the lowest; nPVI ranges from 72.03 to 49.01 in this order S > T > C > M with Shanghainese having the highest score while Mandarin the lowest; rPVI ranges from 67.80 to 54.07 in this order T > S > C > M with Taiwanese having the highest score while Mandarin the lowest. These results indicate that there can be considerable variations in the duration results among even closely related dialects. Since %V and ΔC are negatively correlated and nPVI and rPVI positively correlated, T and S are in reverse order with C and M by the former two metrics, and T and S comes before C and M by the latter two metrics. These orders mean that Taiwanese and Shanghainese are more stress-timed than Cantonese and Mandarin or the latter two dialects are more syllable-timed

than the former two sub-dialects. For each duo, which dialect is more syllable-timed or stress-timed is not conclusive; for example, Taiwanese ranks lower than Shanghainese by nPVI but higher by rPVI. Also, Cantonese ranks higher on %V but higher on nPVI at the same time. These conflicting orders suggest that Taiwanese and Shanghainese may have a comparable degree of syllable-timedness, so do Cantonese and Mandarin.

Mok (2009) conducted a similar duration-based study on Beijing Mandarin and Cantonese rhythm and compared the results with those from two stressed timed languages, German and British English, and two syllable-timed languages, French and Italian (Dellwo, Aschenberner, Dancovicova, Steiner, & Wagner, 2004). As shown in Figure 2.18, Beijing Mandarin (Man) and Cantonese (Can) share syllable-timed rhythm with French and Italian on all four metrics (i.e., %V, ΔC, rPVI_C, & nPVI_V), but Cantonese has a stronger degree of syllable-timedness (i.e., larger %V & smaller ΔC, rPVI_C, & nPVI_V) than Beijing Mandarin. Mok (2009) attributes the different degrees of syllable-timing between Cantonese and Mandarin rhythm to the absence of lexical stress in Cantonese but not in Beijing Mandarin. Also, the results from two different speaking styles, read speech (reading a story from a script; marked as _n) and semi-spontaneous speech (repeating the story without the script; _t), indicate that the style difference does affect speech rhythm but the direction of the influence is yet to be determined, as all four metrics give inconsistent results as to which speaking style renders Mandarin and Cantonese more syllable-timed.

Figure 2.18 Comparison of duration-based results among Mandarin, French, Italian, British English, and German



In general, the above duration-based rhythmic studies on Chinese, though scanty, agree that Chinese dialects are syllable-timed, despite the varying degree of syllable-timedness. There is yet to be a single rhythmic study based on pitch, so the current study can fill the research gap by revealing the melody pattern of Chinese dialects and their relative degree of melodiousness.

Chapter 3
CHINESE PHONOLOGICAL TYPOLOGY AND PROSODIC PHONOLOGY

Chinese as a monosyllabic tonal language has unique prosodic features that may affect its rhythm. Also, Chinese phonological characteristics vary from dialect to dialect, so their influence on rhythm may vary as well. This chapter provides an overview of Chinese dialects in terms of their phonological typology (Section 3.1) and prosodic phonology (Section 3.2), followed by a summary (Section 3.3).

3.1 Chinese phonological typology

This section introduces how Chinese dialects are classified (Section 3.1.1) and what phonological characteristics each dialect has (Section 3.1.2).

3.1.1 Major Chinese dialect groups

According to Lin (2001a), Chinese is traditionally classified into seven major dialect groups, Mandarin, Wu, Gan, Xiang, Min, Kejia (i.e., Hakka), and Yue (i.e., Cantonese). These dialect groups are unevenly distributed within China: Mandarin has the largest coverage, ranging from North, Central, West, to far Northwest China (Xinjiang). Note that in the far Southwest China includes Xizang (Tibet) and Qinghai, Tibetan rather than Chinese is spoken. The remaining six groups cover East, South, and Southeast China. Specifically, Wu Chinese mainly covers East China, including Shanghai, Zhejiang, the southern part of Jiangsu and Anhui. Gan and Xiang Chinese mainly cover the central areas in South China, including Hunan and Jiangxi, respectively. Min Chinese mainly

covers the coastal areas in South and Southeast China, including Fujian, Taiwan, Hainan, and some part of Guangdong (Canton) and Guangxi. Hakka has the widest distribution in South and Southeast China. It is typically spoken in the eastern and northern parts of Canton, but almost every province in the southern areas has Hakka speakers. Cantonese is best known overseas just like Mandarin, and it mainly covers the central and western parts of Canton, Hong Kong, and Macau.

Each major dialect groups can be further classified into some sub-dialect groups. Taking Mandarin as an example, it can be divided into three sub-dialect groups, Northern, Eastern, and Southwestern. Northern Mandarin mainly covers Beijing, Tianjin, the Central-Plain region (Hebei, Henan, Shandong, Shanxi, Gansu, Shaanxi, Xinjiang), and Northeast China (Liaoning, Jilin, Heilongjiang). Some researchers propose that the Mandarin dialects spoken in the Central-Plain region should break away from Northern Mandarin as a separate sub-dialect, called Central-Plain Mandarin.

Eastern Mandarin mainly covers Anhui and the northern part of Jiangsu and Southwestern Mandarin mainly covers Hubei, Sichuan, Yunnan, Guizhou, northern parts of Hunan and Guangxi. Table 3.1 summarizes the seven major groups and their respective sub-groups listed in *Chinese Dialect Lexicon* (1989), along with 20 representative dialects going by the names of cities and provinces where they are spoken.

Table 3.1 List of Chinese dialects and their geographical distributions

No.	Major group	Sub-group	No.	Representative Dialect by city name	Representative Dialect by province name
1	Mandarin	Northern	1	Beijing	n/a (not applicable)
			2	Jinan	Shandong
		Central-Plain	3	Xi'an	Shaanxi
			4	Taiyuan	Shanxi
		Southwestern	5	Wuhan	Hubei
			6	Chengdu	Sichuan
		Eastern	7	Hefei	Anhui
			8	Yangzhou	Jiangsu
2	Wu		9	Suzhou	Jiangsu
			10	Wenzhou	Zhejiang
3	Xiang	Eastern	11	Changsha	Hunan
		Central	12	Shuangfeng	Hunan
4	Gan		13	Nanchang	Jiangxi
5	Hakka (Kejia)		14	Meixian	Canton
6	Cantonese (Yue)	Central	15	Guangzhou	Canton
		Southwestern	16	Yangjiang	Canton
7	Min	Southern	17	Xiamen (Amoy)	Fujian
			18	Chaozhou	Canton
		Eastern	19	Fuzhou	Fujian
		Northern	20	Jian'ou	Fujian

The following Chinese dialect map illustrates the distributions of the seven major groups and three sub-groups of Mandarin throughout China (copied from the *Wikimedia* website).

Figure 3.1 Chinese dialect map

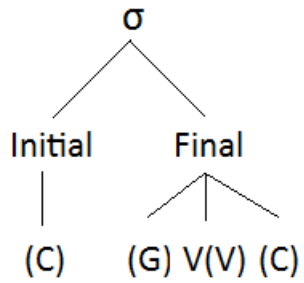
3.1.2 Phonological characteristics of Chinese dialects

All the Chinese dialects are monosyllabic and tonal, since each Chinese character is associated with a single syllable and each syllable carries a tone. Section 3.1.2.1 describes the syllable structure and segmental composition of Chinese dialects and Section 3.1.2.2 describes the tone structure of Chinese dialects. Section 3.1.3 focuses on voice quality of Chinese sounds and tones.

3.1.2.1 Syllable structure

A Chinese syllable is traditionally considered to have a two-part structure: Initial and Final. The initial part is optional: when absent, it is called zero-initial, but once occupied, it must be filled with a single consonant (C), be it sonorant (nasal or lateral) or obstruent (plosive, fricative, affricate). The final part is always filled with a vowel (V: monophthong or VV: diphthong) or in some case, a syllabic consonant (treated as V here). To make the final a little more complex, the vowel can optionally take with it a pre-vocalic glide (G), a post-vocalic consonant (nasal, rhotic, plosive), or both. Figure 3.2 illustrates such a structure.

Figure 3.2 Chinese syllable structure



Different combinations of initial and final segments (consonants, vowels, and glides) yield 16 attested syllable types across Chinese dialects. Most types of syllables are

acceptable in Mandarin except the ones with the (G)VVC Final. Syllables with the (G)VVC Final generally occur in Min. Table 3.2 lists all of them and the syllable samples (in IPA symbols) and the dialects to which they belong.

Table 3.2 Attested Chinese syllable types

No.	Zero-Initial	Syllable Sample	Dialect of Occurrence	No.	Initial-Final	Syllable Sample	Dialect of Occurrence
1	V	[i]	Mandarin	9	CV	[ta]	Mandarin
2	VV	[ai]	Mandarin	10	CGV	[lwo]	Mandarin
3	VC	[in]	Mandarin	11	CVV	[tai]	Mandarin
4	VVC	[ouʔ]	Min	12	CVC	[tan]	Mandarin
5	GV	[ja]	Mandarin	13	CVVC	[touʔ]	Min
6	GVC	[jin]	Mandarin	14	CGVC	[twan]	Mandarin
7	GVV	[wei]	Mandarin	15	CGVV	[twei]	Mandarin
8	GVVC	[jouʔ]	Min	16	CGVVC	[tjouʔ]	Min

Table 3.3 summarizes the 20 representative dialects listed in *Chinese Dialect Lexicon* (1989), along with the number (#) of syllable types they have and the syllable types they are missing.

Table 3.3 List of Chinese dialects and count of their syllable types

No.	Major group	Sub-group	No.	Representative dialect	# of syllable types	Missing syllable type
1	Mandarin	Northern	1	Beijing	12	(CG)VVC
			2	Jinan	12	(CG)VVC
		Central-Plain	3	Xi'an	12	(CG)VVC
			4	Taiyuan	12	(CG)VVC
		Southwestern	5	Wuhan	12	(CG)VVC
			6	Chengdu	12	(CG)VVC
		Eastern	7	Hefei	10	(C)VVC&(C)GVV(C)
			8	Yangzhou	12	(CG)VVC
2	Wu		9	Suzhou	10	(C)VVC&(C)GVV(C)
			10	Wenzhou	12	(CG)VVC
3	Xiang	Eastern	11	Changsha	12	(CG)VVC
		Central	12	Shuangfeng	10	(C)VVC&(C)GVV(C)
4	Gan		13	Nanchang	12	(CG)VVC
5	Hakka		14	Meixian	12	(CG)VVC
6	Cantonese	Central	15	Guangzhou	16*	No
		Southwestern	16	Yangjiang	16*	No
7	Min	Southern	17	Amoy	16	No
			18	Chaozhou	16	No
		Eastern	19	Fuzhou	16	No
		Northern	20	Jian'ou	16	No

* Cantonese does not have diphthong vowels followed by C, but it has long vowels followed by C (e.g., [a:k]), so it is still treated here to have the VVC Finals.

Most dialects have 12 syllable types with the four complicated (CG)VVC types missing. Three dialects, Hefei Mandarin, Suzhou Wu, and Shuangfeng Xiang, have the fewest syllable types, with six complicated (C)VVC and (C)GVV(C) types missing. All the Cantonese and Min dialects have the most syllable types, with all 16 possible (Initial)-Final combinations.

Although many types of segments such as labials, nasals, velars, glides, principal vowels are shared among Chinese major dialect groups, they make up very different combinations of initials and finals in different major groups. Table 3.4 summarizes the 20 representative dialects listed in *Chinese Dialect Lexicon* (1989), along with the number (#) of initials and finals, Fin:Ini, and sumFI. Fin:Ini is the ratio of finals to initials, and sumFI is the total number of initials and finals. Note that zero initial are counted as one type of initials. The two parameters, Fin:Ini and sumFI, will be used to reflect the complexity of Chinese syllable structure. Larger Fin:Ini and sumFI mean more complex syllable structure and less syllable-timedness. So far no studies seem to have used them in Chinese rhythmic studies, so their usefulness is to be tested in this study.

Table 3.4 List of Chinese dialects and count of their initials and finals

No.	Major group	Sub-group	No.	Representative dialect	# of Fin	# of Ini	Fin:Ini	sumFI
1	Mandarin	Northern	1	Beijing	40	22	1.82	62
			2	Jinan	38	24	1.58	62
		Central-Plain	3	Xi'an	40	27	1.48	67
			4	Taiyuan	36	21	1.71	57
		Southwestern	5	Wuhan	37	19	1.95	56
			6	Chengdu	36	20	1.80	56
		Eastern	7	Hefei	41	21	1.95	62
			8	Yangzhou	47	17	2.76	64
2	Wu		9	Suzhou	49	28*	1.75	77
			10	Wenzhou	34	29*	1.17	63
3	Xiang	Eastern	11	Changsha	38	20	1.90	58
		Central	12	Shuangfeng	33	28	1.18	61
4	Gan		13	Nanchang	65	19	3.42	84
5	Hakka		14	Meixian	76	18	4.22	94
6	Cantonese	Central	15	Guangzhou	68	18*	3.78	86
		Southwestern	16	Yangjiang	61	19*	3.21	80
7	Min	Southern	17	Amoy	76	17	4.47	93
			18	Chaozhou	85	18	4.72	103
		Eastern	19	Fuzhou	48	15	3.20	63
		Northern	20	Jian'ou	34	15	2.27	49

*In these dialects, glides such as /j/ or /w/ are considered as part of initials rather than finals.

Gan, Hakka, Cantonese, and Southern Min dialects have the more finals than initials (Fin:Ini > 3) and the larger sum of initials and finals (≥80) than the rest of the dialects. If syllable-timedness is correlated negatively with the complexity of syllable structure, then dialects with large Fin:Ini (i.e., more finals than initials) and large sumFI may be less syllable-timed because of more complex syllable structure. Gan, Hakka, Cantonese and

Southern Min dialects are hence likely to be among the least syllable-timed due to having large Fin:Ini and sumFI.

3.1.2.2 Tone structure

Chinese is also characterized by its use of lexical tones to denote meaning. Almost every Chinese syllable is associated with a tone. Tones are distinct in terms of both overall tone height (tonal register) and individual tone shape (tonal contour). Tonal register refers to a phonological feature of tones which Yip (1992) divides into two main types: high and low, respectively representing "the pitch of the voice at the upper and lower range" (Yip, 1992, p. 245). Tone shape refers to different combinations of high (H), mid (M), and low (L) pitch levels within a tone (Lin, 1989). There are three types of tone shape: level, contour, and short. Level tones are made of at least two identical pitch levels, contour tones of different pitch levels, and short tones of just one pitch level, be it H, M, or L. For example, Mandarin Chinese has four lexical tones, T1, T2, T3, and T4. T1 has a high level pitch contour, T2 a high rising one, T3 a low falling-rising one, and T4 a high falling one, and they can be represented as HHH, MHH, LMH, and HML, respectively (Lin, 1989). Note that in Middle Chinese, tones are categorized into *ping* ('level'), *shang* ('rising'), *qu* ('departing'), and *ru* ('entering'), corresponding originally to level, rising, falling, and short tones, respectively. Short tones are commonly associated with syllables with a consonant coda (usually a stop); for example, Yip, a Cantonese surname read as /jip/, carries a high, short entering tone, H. Within each category, there is also a division between *yin* and *yang*, corresponding originally to voiceless and voiced consonant initials, respectively. Today, the actual tone categories and tone height and shape within each category vary from dialect to dialect. Mandarin,

for example, does not have entering tones any more with the loss of stop codas. While Beijing Mandarin *yin-ping* tone (T1) is a high-level tone, but Tianjin Mandarin *yin-ping* tone becomes a low-falling tone. Therefore, Middle Chinese tone categorization can no longer reflect actual tone type.

Another common way to describe Chinese tones is to use numerical values. Chao (1930) devised a 5-level numerical scale to represent pitch height: H = 4 or 5, M = 3, and L = 1 or 2. Based on this numerical scale, Mandarin T1, T2, T3, and T4 have pitch values of 55, 35, 214, and 51, respectively.

Traditionally, T1, T2, and T4 are considered as a high register tone and T3 as a low register tone. Note, however, that Wang et al. (2012) considered T2 a low register tone as they found T2 tended to occur within the lower pitch range in continuous speech.

Mandarin also has a so-called neutral tone (T0). The neutral tone does not have a fixed tone shape or tonal letters assigned but can be considered as phonologically short and low (Lin, 2001b, 2006). It is often associated with the de-accented (or unstressed) syllable in a di-syllabic word (corresponding to two Chinese characters) or with a functional word (often corresponding to one Chinese character), so it is not considered as a lexical tone. Table 3.5 summarizes the five Mandarin tones, their respective features, and common representations:

Table 3.5 Five Mandarin tones

Tone name	T1	T2	T3	T4	T0
Tonal register	High	High	Low	High	Low
Tone shape	Level	Rising	Falling-Rising	Falling	Short
Numerical value	55	35	213	51	-
Pinyin* representation	mā	m á	mǎ	m à	ma
Number representation	ma1	ma2	ma3	ma4	ma0
Character example	妈	麻	马	骂	吗
English gloss	Mom	hemp	horse	scold	question marker

* Pinyin is a partially phonetic spelling system used to transcribe Mandarin Chinese sounds. Hereafter, it will be used to represent Chinese syllables and words.

Similar to phonological changes of sounds, tones are subject to changes when interacting with adjacent tones. This phenomenon is called "tone sandhi." Mandarin Chinese, for example, is known for its "third tone sandhi" process. It is the change of T3 into T2 when it is followed by another T3. A most-cited example is the greeting word *ni3hao3* ('Hello'; literally 'you good'). Each syllable in this word carries T3 when in isolation, but when they are together, the first T3 on *ni* changes into T2. As a result, the final reading of this word is *ni2hao3*. Note that a tone sandhi process occurs not only across word boundaries but also within a word and its application to a word/phrase with more than three T3s in a row is of not only phonological but also syntactic relevance.

Table 3.6 summarizes the 20 representative dialects listed in *Chinese Dialect Lexicon* (1989), along with the number (#) of high and low register tones, HT:LT, and sumT. HT:LT is the ratio of high to low register tones, and sumT is the total number of tones. The two parameters HT:LT and sumT will be used to reflect the complexity of Chinese

tone structure. Larger HT:LT and smaller sumT mean less complex tone structure and more melodiousness. Again, so far no studies seem to have used them in Chinese rhythmic studies, so their usefulness is to be tested in this study. Note that listed are all lexical tones, so the neutral tone is excluded. Also, each register may include level, contour, and short tones. To avoid possible disagreement on which tone is of high or low register in a dialect, this study uses the following criteria to count high and low register tones:

HT:

1) Any level or contour tone starting with H pitch; e.g., pitch values = 55, 44, 41, 52.

2) Any level or contour tone starting with M, going through H or M if applicable, and ending with H or M pitch; e.g., pitch values = 33, 35, 343.

3) Any short tone with H or M pitch; e.g., pitch values = 5, 4, 3.

LT:

1) Any level or contour tone starting with L pitch; e.g., pitch values = 13, 24.

2) Any level or contour tone starting with M, going through L if applicable and ending with L or M pitch; e.g., pitch values = 323, 32, 31, 312.

3) Any short tone with L pitch; e.g., pitch values = 2, 1.

Table 3.6 List of Chinese dialects and count of their high and low tones

No.	Major group	Sub-group	No.	Representative dialect	# of HT	# of LT	HT:LT	sumT
1	Mandarin	Northern	1	Beijing	3	1	3.00	4
			2	Jinan	2	2	1.00	4
		Central-Plain	3	Xi'an	2	2	1.00	4
			4	Taiyuan	3	2	1.50	5
		Southwestern	5	Wuhan	3	1	3.00	4
			6	Chengdu	2	2	1.00	4
		Eastern	7	Hefei	3	2	1.50	5
			8	Yangzhou	4	1	4.00	5
2	Wu		9	Suzhou	4	3	1.33	7
			10	Wenzhou	4	4	1.00	8
3	Xiang	Eastern	11	Changsha	3	3	1.00	6
		Central	12	Shuangfeng	3	2	1.50	5
4	Gan		13	Nanchang	3	4	0.75	7
5	Hakka		14	Meixian	3	3	1.00	6
6	Cantonese	Central	15	Guangzhou	5	4	1.25	9
		Southwestern	16	Yangjiang	5	4	1.25	9
7	Min	Southern	17	Amoy	4	3	1.33	7
			18	Chaozhou	5	3	1.67	8
		Eastern	19	Fuzhou	3	4	0.75	7
		Northern	20	Jian'ou	3	3	1.00	6

Among all the Chinese dialects listed, Mandarin tends to have the fewest lexical tones (4 or 5) and Cantonese has the most (9). In terms of HT:LT, Yangzhou, Wuhan, Beijing Mandarin rank top three with the largest ratio of HT to LT (≥3.00). If more tones (larger sumT) in a dialect mean less melodiousness because of smaller pitch excursion and larger HT:LT means more melodiousness because of more larger pitch excursion, then Mandarin dialects may be more melodious than dialects from the rest of the major groups due to having fewer tones (smaller sumT) and larger HT:LT.

3.1.3 Voice quality of Chinese sounds and tones

A number of studies (e.g., Rose, 1989; Cao & Maddieson, 1992; Davison, 1991; Belotel-Grenié & Grenié, 2004; Pan, 2005; Esposito, 2006; Lam & Yu, 2010; Gao et al., 2011; Keating et al., 2012) examined the voice quality of the four major groups of Chinese dialects, Mandarin, Cantonese, Wu, and Min. For Mandarin Chinese, the main finding is that creaky voice is often associated with the low-dipping tone, T3 (Davison, 1991; Belotel-Grenié & Grenié, 2004; Keating et al., 2012), and it is even more so with the non-lexical neutral tone, T0 (Belotel-Grenié & Grenié, 2004). Given that the neutral tone is often low in pitch, it is no surprise that it shares with the low-pitched third tone in creakiness.

Similar to Mandarin, Cantonese often has creakiness on its lowest tone, T4, and creakiness on T4 (a low-falling tone) is an important perceptual cue used by many speakers to distinguish T4 from the low level tone, T6 (Lam & Yu, 2010).

Min Chinese represents a little more complicated case with voice quality: Pan (2005) found some creakiness in the low-falling tone of Taiwan Min (a Southern Min dialect), but Esposito (2006) found that there is breathiness rather than creakiness associated with the low falling-rising tone in Fuzhou Min (an East Min dialect).

Wu Chinese has the most complex voice quality associated with its segmental/tonal production. Based on Rose (1989), Zhenhai Wu (a Wu dialect spoken in Ningbo, a city to the south of Shanghai) has three unique phonation types ranging from whisper, whispery voice (to be distinguished from breathy), to growl (or harsh, a variety of growl). The three phonation types occur on different types of Yang syllables. Note that Yang syllables here refer to syllables carrying Yang tones T3, T4, T6, which all start with a low pitch. In contrast, all the Yin tones, T1, T2, T5, start with a mid to high pitch.

In Cao and Maddieson's (1992) study, Shanghai Wu was also found to have a certain degree of breathiness associated with syllables with T3 and T5 (low rising tones, also called Yang tones). A relatively recent study of Shanghai Wu by Gao et al. (2011) found further that slackness (or a moderate degree of breathiness) in Yang tones also affects the duration of syllable initials: consonant initials with Yang tones are significantly shorter than those with Yin tones, suggesting that Yang tones tend to shorten the duration of consonants, whether obstruent or sonorant, in Shanghai Wu. Note that Gao et al. (2011) may be the first to investigate the influence of voice quality on Chinese segmental duration, though this aspect is not the main focus of their study. For languages other than Chinese, Blankenship (1997) investigated the timing of non-modal phonation in vowels and found that non-modal vowels have longer duration in languages where non-modality is contrastive on vowels than where it is not. Note that this study will not examine how voice quality affects segmental duration in each Chinese dialect; instead, it will present a big picture on whether or not voice quality is associated with duration-based rhythm.

Note also that for low tone production, different and even opposite phonation types can be observed across Chinese dialects. An explanation was offered by Moisik, Lin, and Esling (2014), who conducted a simultaneous laryngoscopy and laryngeal ultrasound study of Mandarin tones. They found that speakers employed two strategies to produce Mandarin T3 (a low falling-rising tone): one is by lowering the larynx and the other is by raising the larynx with laryngeal constriction. The former strategy gives the tone a modal or breathy quality and the latter gives it a creaky quality. Therefore, speakers of different dialects may choose to use one or the other strategy to produce low tones. If one strategy is consistently used, then the associated voice quality can be used to enhance the low tone

perception, as in the case of Mandarin, or contrast with another tone, as in the case of Bai and Yi (Edmondson, et al., 2005; Esling, 2005; Edmondson & Esling, 2006).

In general, voice quality can affect segmental/tonal duration, which in turn may contribute to rhythmic perception. Therefore, this study is interested in exploring its role in Chinese rhythmic patterning.

3.2 Chinese prosodic phonology

This section provides an overview of Chinese prosodic systems from three perspectives: structure (Section 3.2.1), stress-induced pitch variation (Section 3.2.2), and dialectical difference (Section 3.2.2).

3.2.1 Chinese prosodic structures

Due to the use of lexical tones as additional meaning units, Chinese is traditionally understood as having a unique two-layer prosodic system, or the "small ripples riding on big waves", a result of "an algebraic sum of the two kinds of waves" (Chao, 1968, p. 39). The small ripple refers to the local pitch contour of a lexical tone or a tonal contour. The big wave refers to the "underlying intonation skeleton" (Cao, 2004), the global pitch contour of a syntactically defined sentence or an intonation contour. Cao (2004) finds that the local pitch contour is kept relatively stable to preserve the tonal identity while the global pitch contour, the overall pitch height represented by intonation register, varies greatly to reflect intonation. Here intonation register is a phonological concept, representing "the level on which a lexical tone is actually realized in an utterance" (Yip, 1992, p. 225). It can be phonetically equivalent to the mid pitch range. She also recognized two types of effects of intonation on tones: one is all tones raise or lower as a

whole, the other is the pitch range expansion with high tones rising and low tones lowering.

In order to reveal how "small ripples" are riding on "large waves" or the interaction between tones and intonation in further details, Cao (2004) proposed a three-layer prosodic hierarchy for Chinese, prosodic word (PW), prosodic phrase (PP), and intonation phrase (IP). The speech units of spoken Chinese are basically bi- or tri-syllabic PW. PP is syntactically defined phrase or clause. IP contains one or more PPs and is syntactically defined sentence (Cao, 2004). Based on this three-layer structure, a sentence can be represented as follows (based on Cao, 2004, p. 34):

Figure 3.3 Three-layer Chinese prosodic structure

IP: { IP}

PP: { PP1}{ PP2}{ PP3}{ PP4}

PW: ω1 ω2 ω3 ω4 ω5 ω6 ω7 ω8 ω9 ω10 ω11

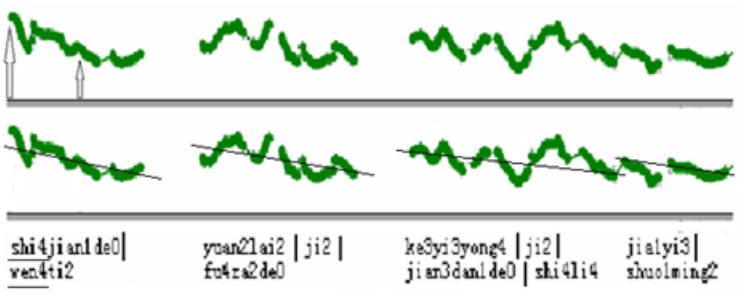
shijiande | wenti, | yuanlai | **ji** | fuzade, | keyiyong | **ji** | jiandande | shili | **jiayi** | **shuoming**

"The problems *in the world, originally* **extremely** *complex, can* ***be*** **explain*ed*** *using* **extremely** *simple* instances" *

*The same type of font is used for Chinese syllables and English words with the same meaning.

This sentence includes eleven prosodic words, four prosodic phrases, and one intonation phrase. The prosodic words are represented by local pitch contours and the prosodic phrases and intonation phrase by global pitch contours. Since the sentence is declarative, its intonation should be falling toward the end. Figure 3.4 illustrates how local tonal contours are integrated into global pitch contours in this sentence (adapted from Cao, 2004, p. 34).

Figure 3.4 Illustration of local and global pitch contours



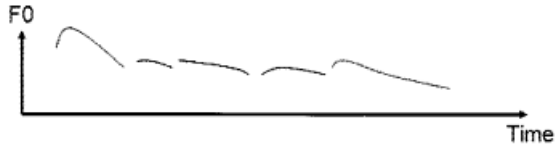
The first row is the actual pitch contour of the sentence. The four downward going lines in the second row show the global pitch movement within the four prosodic phrases. All the prosodic phrases start with a higher intonation register and ending with a lower intonation register, which reflects the general tendency for the pitch level of an intonation unit (PP here) to decline with time. Note that the intonation register in the fourth prosodic phrase ***jiayi*** **shuoming** (***'be*** **explain*ed'***) is lower than the preceding three phrases, likely due to the fact that the phrase is short and at the sentence-final position, hence subject to the pitch declination effect both at the PP and at the IP level. In other words, there is no need to reset the intonation register to the initial pitch level as the sentence is soon trailing off.

Figure 3.4 also reveals the way for "small ripples" to ride on "big waves", and it is by "retaining local lexical distinction by keeping their basic contour patterns, while carrying the information of global intonation through the variation of pitch register" (Cao, 2004, p. 34). Specifically speaking, it is by superposing both local and global pitch contours as an algebraic sum. Take the first syllable *shi4* ('world') and the fourth syllable wen4 ('question') as an example. The long arrow points to the start of the first syllable *shi4*, and the short arrow points to the start of the fourth syllable wen4. Since both syllables carry the same falling T4 (hence *shi4* and wen4), theoretically speaking, they should both

show an identical falling pitch contour, but here the first T4 starts with a much higher pitch and with a much steeper slope than the fourth T4, reflecting the influence of the global pitch contour. Specifically speaking, the intonation register of the first syllable is high and that of the fourth syllable is low; as a result, the first T4 shows a much higher overall pitch than the fourth T4.

The above interpretation of Chinese prosody has faced some challenge in recent years. Tseng (2006) suggested that tonal and intonational information is not the most significant contributor to speech prosody because neither tonal nor syntactic specifications are always realized in connected speech. In his study of Mandarin Chinese spontaneous speech, only 36% of lexical tones are found to have corresponding phonetic realizations and only 20% of declarative sentences have declining pitch contours and 45% have final fall only. The remaining 35% have unclassifiable contour patterns (Tseng, 2006). These results suggest that lexical prosody and intonation are not the only players in speech prosody. Tseng (2006) hence assumed that there is higher-level discourse information contributing to speech prosody, and he calls it prosodic Phrase Grouping (PG). This new level is added above the PP layer to reflect prosody at the discourse level. Note that it is not equivalent to the IP-layer shown in the aforementioned three-layer hierarchy, as it can include one or several IPs, optional breath groups, discourse markers, and prosodic fillers. The new PG-layer encodes cross-phrase associations at the discourse level through three positions: PG-initial, PG-medial, and PG-final. The schematic illustration of the new layer is presented in Figure 3.5 (copied from Tseng, 2006, p. 63):

Figure 3.5 Illustration of the new PG-layer



The five lines represent the global trajectory of F0 contours of a PG containing five prosodic phrases with declarative intonation. The PG is featured by a pitch register high in the beginning phrase (PG-initial), moderate in the middle three phrases (PG-medial), and low in the final phrase (PG-final). Tseng (2006) used two experiments to prove that this extra PG layer is essential to predicting actual speech prosody. A PG containing three prosodic phrases is used in the experiments. In the two experiments, he respectively teased out the phrase component (smoother global pitch variations over time) and the accent component (more drastic local F0 variations over time) from the speech prosody. In the first experiment, he made two kinds of predictions, one with and the other without taking the PG-effect into consideration, and compared them with the actual prosody output (the global pitch contour). The results show that the PP (Prosodic Phrase) layer can only predict about 40% of the prosody output while the PG layer predicts an additional 25% or so. The two layers make a total 65% of contributions to the correct prediction. If the whole PG-prosody hierarchy is considered, then the remaining 35% of contributions should come from the levels lower than the PP layer.

In the second experiment, Tseng (2006) studied the contributions to the local pitch contour from the lower syllabic (syllable type and tone type), lexical (prosody word), and prosodic phrasal levels. The results show that the syllabic level contributes about 20% to the correct prediction, the lexical level contributes an additional 1%, and the PP level contributes another additional 5%. Note that 20% of the contributions at the syllabic

level are from taking both tone type and syllable type into consideration. If only tone type is considered, then the accuracy rate for the correct prediction drops to only 12.5%. In total, the three levels make about 26% of contributions.

Based on the results from the two experiments, there are still 9% (=35%-26%) of contributions unaccounted for. Some likely contributors are stress, focus, and emphasis. According to Tseng (2006), they can all be treated as subsequent add-ons to PG.

3.2.2 Stress-induced pitch variation

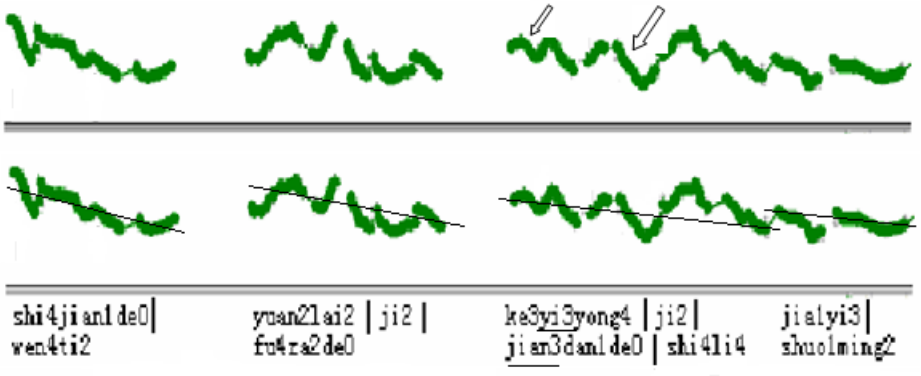
Contrastive stress is not regularly employed in Chinese, but it is generally agreed that Chinese has non-contrastive word stress (Lin, 2001b; Duanmu, 2004). Recognizing the fact that a disyllabic unit plays an important role in Chinese prosody, Lin (2001b) proposed that Mandarin disyllabic words have domain-initial stress. For high/contour toned disyllables, the stress on the first syllable can be achieved by adjusting the height difference between the two syllables, but for two T3 syllables, the height adjustment is restricted by their low-tone identity; that is, the room for the phonetic adjustment of their tonal height is limited. Thus, the T3 sandhi rule has to be in place to make sure one syllable is more prominent than the other. In other words, T3 sandhi is required by prosodic stress assignment.

Duanmu (2005) proposed a similar view but replaced prosodic stress with syntactic stress. According to Duanmu (2005), Chinese has an abstract stress system, which manifests itself through tone sandhi processes at both lexical and phrasal level. The abstract stress is syntactically determined, falling on the syntactic non-head of a word/phrase. If this is the case, then non-head stress-induced pitch variations can be

predicted from tone sandhi patterns. Since tone sandhi is mostly syntactically determined, tone sandhi patterns make stable building blocks for a global pitch contour (Wu, 1990).

Regardless of its nature, stress or accent is usually handled in two ways cross-linguistically, and they are pitch range expansion and pitch register elevating. Chinese as a tonal language nonetheless uses a special way to handle stress-induced pitch variations: when a syllable is accentuated, its tonal register will become higher if it is originally high and lower if it is originally low (Cao, 2004). Take the pitch contour of the aforementioned sentence as an example (adapted from Cao, 2004, p. 34):

Figure 3.6 Illustration of stress-induced pitch variation



Based on Cao (2004), the smaller arrow in the third phrase points to the falling part of the pitch contour of the second syllable *yi3* in the word *ke3yi3* ('may') and the bigger arrow points to the falling part of the pitch contour of the fifth syllable *jian3* in the word *jian3dan1de0* ('simple'). Here the second syllable *yi3* is unstressed and the fifth syllable *jian3* is stressed, but their tonal context is similar: The syllable *ke3* preceding *yi3* carries the low T3, but is showing up as the rising T2 due to the third tone sandhi effect, so the second syllable *yi3* is actually preceded by a T2 rather than a T3 syllable. The fifth syllable *jian3* is also preceded by a T2 syllable, **ji2** ('extremely'). Moreover, the second syllable *yi3* is followed by the T4 syllable *yong4* ('use') and the fifth syllable *jian3* by the

T1 syllable *dan1*. In other words, both *yi3* and *jian3* are surrounded by tones with a high tonal register, be it T2, T4, or T1.

Although both the syllables carry the same low falling-rising T3 and share a similar tonal context, the stressed *jian3* has a much lower tonal register than the unstressed *yi3*. Also, the pitch range of the stressed *jian3* is more expanded than that of unstressed *yi3*, as it has a much longer slope. Note that the global intonation contour (the third downward going line in the second row) may have some lowering effect on the tonal register of the fifth syllable *jian3*, but it does not necessarily expand the pitch range, so the long falling contour of *jian3* must come mainly from the local accentuation.

To sum up, tone sandhi and stress work together from different phonological levels to exert their influence on a tonal contour.

3.2.3 Prosodic differences in dialects

The prosodic systems identified in the previous research are all based on standard or Beijing Mandarin. This section reviews some studies on dialectal influence on tonal register and intonation (Section 3.2.3.1) and dialectal differences in tone sandhi patterns (Section 3.2.3.2).

3.2.3.1 Dialectal influence on tonal register and intonation

In order to find how dialectical variations affect tonal register and the intonation declination pattern, Huang and Fon (2008) studied two sub-dialects of Taiwan Mandarin, Northern and Central. The Northern sub-dialect is considered as standard Taiwan Mandarin (*Guoyu*, 'national language') and the Central sub-dialect is said to be heavily influenced by Min, as 80% of the population in Taiwan speaks a variety of the Southern Min dialect, also called Taiwanese (Cheng, 1985). Taiwan Mandarin itself is also

slightly different from standard Mandarin spoken in Mainland China (*Putonghua*, 'common speech', standard Chinese) in terms of the four lexical tones, as it has a lower tonal register than the latter, which is also likely to be an influence from the Southern Min spoken in Taiwan or Taiwanese. Taiwanese has five lexical tones, four of which have a lower tonal register than their Mandarin counterparts. Table 3.7 compares the lexical tones in terms of their numerical values across Taiwan Mandarin (Fon & Chiang, 1999), Putonghua, and Taiwanese (Cheng, 1997):

Table 3.7 Comparison of lexical tones in *Guoyu, Putonghua,* and Taiwanese

Tone name	T1	T2	T3	T4	T5	T6	T7
Guoyu (standard Mandarin spoken in Taiwan)	44	323	312	42	-	-	-
Putonghua (standard Mandarin spoken in China)	55	35	214	51	-	-	-
Taiwanese (Southern Min spoken in Taiwan)	44	12	53	21	22	3	5

The comparison of the three dialects shows that *Putonghua* has the highest overall pitch, followed by Taiwan Mandarin and Taiwanese. Huang and Fon (2008) then predicted that the tonal register of Taiwanese will have more influence on the Central than on the Northern Mandarin dialects in Taiwan; specifically, the tonal register will be lower and the intonation contour will be less steep for the Central than for the Northern Mandarin.

Huang and Fon (2008) measured the pitch of a T1 syllable in different sentential positions, initial, medial, and final. The results show that the Northern dialect has a

higher T1 register than the Central dialect in all three positions both for male and for female speakers, but the register difference is more prominent for male than for female speakers. The intonation declination contour, on the other hand, is steeper at the medial and final positions for Northern than for Central male speakers and less steeper at all three positions for Northern than for Central female speakers. The results from male speakers are consistent with the predictions and can be related to the tonal register difference between the two sub-dialects. As for why male speakers show a larger degree of dialectal differences than female speakers, Huang and Fon (2011) attributed it to the gender difference in conforming to the social norm; that is to say, women are more likely to conform to standard linguistic forms than men. A further study by Huang, Wu, and Fon (2012) reveals that highly proficient speakers of Taiwanese have a lower tonal register and a narrower pitch range in Taiwan Mandarin than less proficient speakers of Taiwanese.

Since standard Mandarin is the official language in China, all dialect speakers, especially younger speakers, can more or less speak standard Mandarin, so Chinese dialects they speak may be subject to influence from standard Mandarin. If this is the case, then it is expected to see two dialects belonging to different major groups but geographically close to each other may have a pitch pattern more similar to two dialects belonging to the same major group but geographically far apart.

3.2.3.2 Dialectal differences in tone sandhi patterns

As mentioned in Section 3.2.2, tone sandhi patterns are important contributors to speech prosody, because they reflect tone-stress-syntax interactions at every syntactic and prosodic level. Duanmu (2004) compared three tone sandhi patterns in standard

Mandarin, Wu, and Min Chinese. In standard Mandarin, the most discussed third tone sandhi, for example, has an effect of neutralizing the difference between T3 and T2 (Duanmu, 2004) or changes a low tone into a high tone due to physiological constraints (Lin, 1996). In Shanghai Wu, tone sandhi spreads the tone on first syllable to the entire sandhi domain and removes the tone on the rest of syllables in the domain. In Min spoken in Xiamen, every syllable is associated with two tones, final and non-final, corresponding to two syllable positions: domain final or in isolation and non-final. Note that the Min tone alternation strictly speaking is not tone sandhi as it is syllable position related rather than a result of tone-tone interaction (Duanmu, 2005). Figure 3.7 illustrates the three types of tone sandhi (adapted from Duanmu, 2005, p. 2 & p. 3):

Figure 3.7 Illustration of three types of tone sandhi

(1)Standard Mandarin:	T3 T3 → T2 T3		T2 T3
	/mai ma/	Cf:	/mai ma/
	buy horse		bury horse
(2)Shanghai Wu:	HL-HL → H-L		
	/se pe/		
	three cups		
(3)Xiamen Min:	Final: 24		Non-final: 22
	/p^he <u>we/</u>	Cf:	/<u>we</u> tua/
	leather shoe		shoe lace

Chen (2000) described Tianjin Mandarin tone sandhi as a process of changing two identical tones or different contour tones but with the same pitch height adjacent into a contour tone.

Since Chinese dialects employ various types of changes to the original tonal patterns, their tone structure may not be a predictor to the pitch-based rhythm in these Chinese dialects and other dialects with similar or different tone sandhi patterns.

3.3 Summary

Chinese dialects vary greatly in terms of syllable and tone structures, so their syllable and tone structures may correlate with duration- and pitch-based rhythmic patterns. However, voice quality characteristics associated with syllables and tones may render them not good predictors of rhythm. Moreover, the interference from tone sandhi at every level of prosodic structure, complicated by stress and intonation, may render pitch-based Chinese rhythm unpredictable.

Chapter 4
METHODOLOGY

4.1 Speech materials

The speech data include 21 audio recordings downloaded from the *phonemica.net* website under authorization. The content of these recordings is mostly about personal stories or descriptions of one's hometown casually told by male speakers. Male speakers were chosen because there are more male than female speakers contributing to the website. There are three main advantages for adopting the speech material from the website: First, the speech has the naturalness similar to the conversational style of speech; Second, it is easier to find various dialect speakers online than offline, as the speakers contributing to the website are from all over the country; Third, all the speakers seem to have grown up and been still living in the area where the dialects are used, so they have daily exposure to their own dialects. Note that their age is between eighteen and thirty, so their speech may be slightly different from the older generation.

The recordings are divided into four major groups. Group 1 includes eight recordings representative of Mandarin Chinese spoken in northern and southwestern parts of China; Group 2 includes four recordings representative of Northern and Southern Wu spoken in East China; Group 3 includes five recordings representative of East, Central, and Southern Min spoken in Southeast China; Group 4 includes four recordings representative of Guangzhou and Hong Kong Cantonese spoken in South China. Details of the four groups (including nine sub-groups) and the associated dialects are listed in Table 4.1.

Table 4.1 Summary of the speech data of 21 Chinese dialects

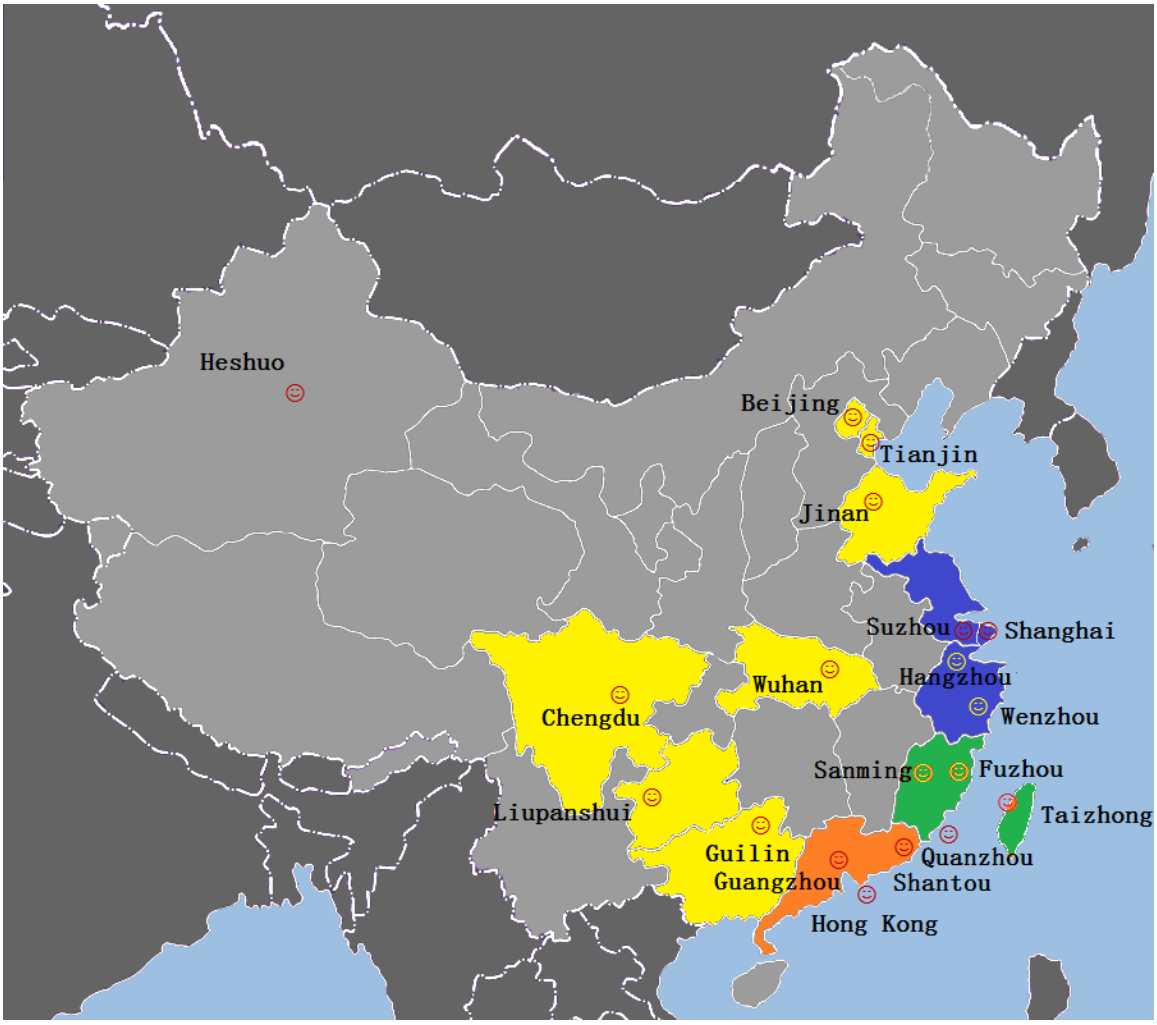
Group	Major group	Sub-group	No.	Dialect by speaker's hometown	Dialect by province/region	Speaker's age
1	Mandarin	Northern	1	Beijing	North	19
			2	Tianjin	North	19
			3	Jinan	Shandong	29
			4	Heshuo	Xinjiang	20
		South-western	5	Wuhan	Hubei	26
			6	Chengdu	Sichuan	19
			7	Liupanshui	Guizhou	25
			8	Guilin	Guangxi	23
2	Wu	Northern	9	Hangzhou	Zhejiang	24
			10	Shanghai	East	26
			11	Suzhou	Jiangsu	20
		Southern	12	Wenzhou	Zhejiang	29
3	Min	Southern	13	Quanzhou	Fujian	24
			14	Taizhong	Taiwan	24
			15	Shantou	Guangdong	24
		Eastern	16	Fuzhou	Fujian	25
		Central	17	Sanming	Fujian	24
4	Cantonese	Guangzhou (GZ)	18	GZ1	Guangdong	26
			19	GZ2	Guangdong	20
		Hong Kong (HK)	20	HK1	South	21
			21	HK2	South	20

Each recording is named after the speaker's hometown as reported by the speaker. Note that in Group 1, the Heshuo dialect (No. 4) is reported as standard Mandarin by the speaker. The auditory perception confirms that it is standard Mandarin but with a little Xinjiang accent. In Group 4, four Cantonese speakers are from just two cities,

Guangzhou and Hong Kong, so their speech is differentiated by both city name and speaker number as GZ1, GZ2, HK1, and HK2 (see Figure 4.1 for the geographical distribution of the 21 dialects as reported by their speakers at the end of this section).

Since Group 1, 2, 3, and 4, representative of Mandarin, Wu, Min, and Cantonese, have more than one speech recordings involved, they can be used to answer the questions of whether or not phonologically similar dialects are also rhythmically similar and how well duration- and pitch-based metrics can differentiate speech at the sub-dialectal level or by different speakers of the same dialect.

Figure 4.1 Geographical distribution of the 21 Chinese dialects



Note that the yellow areas include seven Mandarin dialects excluding the dialect, Heshuo, by a Mandarin speaker from West China; the blue areas include four Wu dialects; the red area includes three Min dialects (inside the Fujian province); and the orange areas include four Cantonese and two Min (one in the Cantonese province & one in Taiwan) dialects.

4.2 Metrics

In order to reveal how duration, pitch, voice quality, and phonological structures are related to rhythm, this study uses four kinds of metrics, duration-, pitch-, voice source-, and phonological structure-based metrics to analyze speech data.

4.2.1 Duration-based metrics

There are nine duration-based metrics extracted from each speech recording: SumD, s_rate, %Son, ΔSon, varco_Son, nPVI_Son, ΔIS, varcoIS, and rPVI_IS. The specific meaning of each metric, followed by further explanations, is listed in Table 4.2.

Table 4.2 Nine duration-based metrics

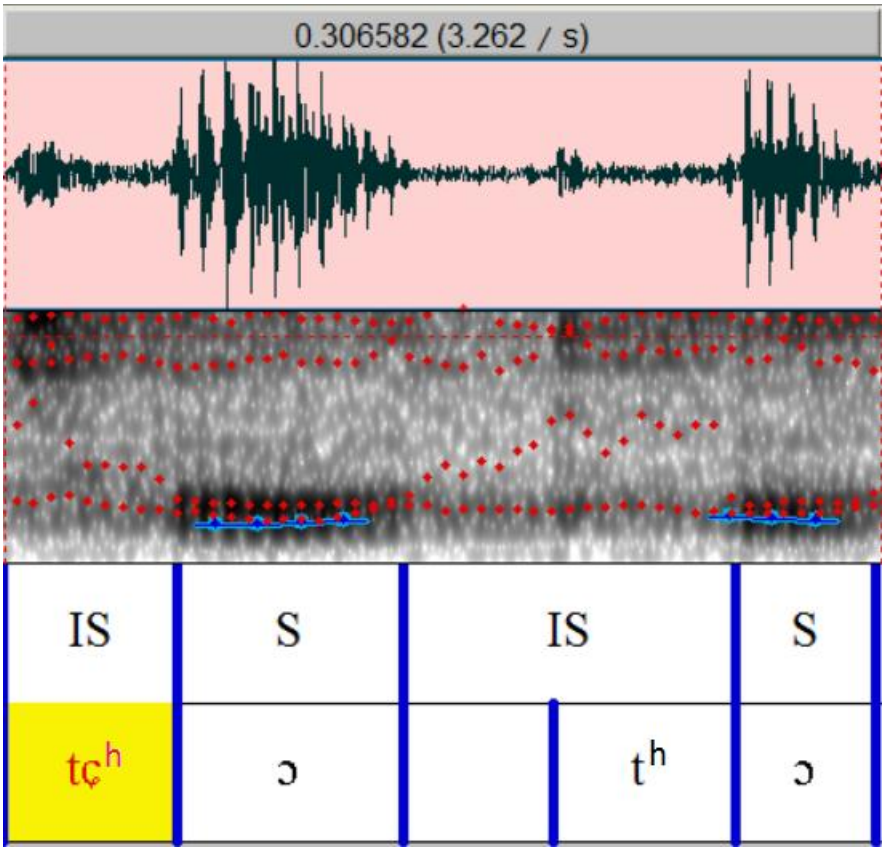
1	sumD	Sum Duration of sonorant and inter-sonorant intervals (i.e., sum of sonorant and inter-sonorant duration)
2	s_rate	Speech Rate, number of sonorant intervals per second
3	%Son	Percentage of Sonorant duration over sumD
4	ΔSon	Standard deviation of Sonorant duration
5	varcoSon	variation coefficient of ΔSon
6	nPVI_Son	normalized Pairwise Variability Index for Sonorant intervals
7	ΔIS	Standard deviation of Inter-Sonorant duration
8	varcoIS	variation coefficient of ΔIS
9	rPVI_IS	raw Pairwise Variability Index for Inter-Sonorant intervals

The first metric SumD is calculated by summing up the duration of all the sonorant and inter-sonorant intervals. Similar to Vicenik and Sundara (2013), sonorant intervals are measured here instead of vowel intervals. Since sonorants are main pitch carriers, measuring duration of sonorant intervals not only provides a good indication of prominence patterns but also eliminates the difficulty to segment sonorant consonants such as glides. Take a glide [j] as an example, its formant transition into a vowel can be rather gradual, so it is better to treat it as part of a sonorant interval.

Unlike Vicenik and Sundara (2013), inter-sonorant intervals are measured here instead of obstruent intervals. Measuring inter-sonorant intervals helps avoid the difficulty in separating stop consonants from short pauses. Taking an unreleased stop coda [pˀ] (a sound in Cantonese) as an example, the duration of its closure portion can be hard to measure if there is a following pause. In this case, it is treated as part of an inter-sonorant interval if the interval is shorter than 10ms. Therefore, inter-sonorant intervals include not only obstruents but also pauses shorter than 10ms. Pauses longer than 10ms, usually occurring between intonation phrases, are excluded from inter-sonorant intervals.

Sonorant and inter-sonorant intervals are analogous to vocalic and inter-vocalic intervals defined by Grabe and Low (2002): In the spectrogram, the former is the stretch of signal between sonorant onset and offset, acoustically indicated by formants, regardless of the number of sonorants (vowels/glides/nasals/ approximants) included; the latter refers to the stretch of signal between sonorant offset and onset, regardless of the number of obstruents (stops/fricatives/affricates) included. Figure 4.2 illustrates sonorant and inter-sonorant intervals in the spectrogram.

Figure 4.2 Illustration of sonorant and inter-sonorant intervals



The first IS (Inter-Sonorant interval) includes the aspirated affricate /tɕʰ/). The first S (Sonorant interval) includes the vowel /ɔ/. The second IS includes a brief background noise (a less than 10ms pause between syllables) followed by the aspirated stop /tʰ/. The second S includes the vowel /ɔ/. Note that during each S, the visible pitch contour (the short dotted line) does not run through exactly from the beginning to end of the interval. This is because pitch at some section cannot be defined by the Praat, but the missing pitch section is still considered as part of the sonorant, as the segmentation is based on formants (the dotted line) not on pitch. The undefined pitch will be ignored by pitch-based measurement, so it does not affect the final results.

As for the second metric s_rate, it is traditionally calculated as the number of syllables per second, but syllable-counting is hardly accurate whether manually or automatically, so s_rate is calculated here as the number of sonorant intervals per second.

Of the remaining seven metrics, %Son, ΔSon, and ΔIS are parallel to %V, ΔV, and ΔC in Ramus et al.'s (1999) study, rPVI_IS and nPVI_Son to inter-vocalic rPVI and vocalic nPVI in Grabe and Low's (2002) study, and varcoSon and varcoIS to VarcoΔC in Dellwo's (2006) and VarcoΔV in White and Mattys' (2007) study. The correspondence between the present sonorant (Son)- and inter-sonorant (IS)-based and previous vowel (V)- and consonant(C)-based metrics is listed in Table .

Table 4.3 Comparison of present and previous duration-based metrics

No.	Present Son- & IS-based	Previous V- & C-based	Previous study
1	sumD (Sonorant & Inter-Sonorant duration)	sumD (Vowel & Consonant or Vocalic & Inter-vocalic duration)	Ramus et al. (1999) Grabe & Low (2002)
2	s_rate (#sonorant intervals/s)	s_rate (#syllables/s)	conventional
3	%Son	%V	Ramus et al. (1999)
4	ΔSon	ΔV	Ramus et al. (1999)
5	varcoSon	VarcoΔV	White & Mattys (2007)
6	nPVI_Son	Vocalic nPVI	Grabe & Low (2002)
7	ΔIS	ΔC	Ramus et al. (1999)
8	varcoIS	VarcoΔC	Dellwo (2006)
9	rPVI_IS	Inter-vocalic rPVI	Grabe & Low (2002)

4.2.2 Pitch-based metrics

There are four pitch-based metrics extracted from each speech recording: meanPE, ΔPE, meanPS, and ΔPS. These metrics are used to show pitch-based prominence patterns.

The specific meaning of each metric is listed in Table 4.4, followed by further explanations.

Table 4.4 Four pitch-based metrics

1	meanPE	Pitch Excursion size: the log difference between maxf0 and minf0 averaged over all the pitch contours in a stretch of speech
2	ΔPE	Standard deviation of Pitch Excursion size
3	meanPS	Pitch Slope: the rate of f0 change averaged over all the pitch contours in a stretch of speech
4	ΔPS	Standard deviation of Pitch Slope

PE (Pitch Excursion) involves a pitch rise or a pitch fall on a pitch contour (i.e., local rise or local fall). The PE-based metrics meanPE and ΔPE are derived from the measures of maxf0 and minf0 across all pitch contours. Vicenik and Sundara (2013) only used local rise alone as a metric (i.e., PR) and defined it as any minimum (minf0) followed by the closest maximum (maxf0) on the same pitch contour. Analogously, pitch fall can be defined as any maximum followed by the closest minimum on the same pitch contour. The use of pitch excursion here instead of pitch rise can capture pitch change in both directions. Specifically, PE is calculated in terms of the difference between the log values of maxf0 and minf0, as shown in the following formula:

$$\text{PE} = 12 \times (\log_2 maxf0 - \log_2 minf0)$$

According to Hirst (2011), the simple difference between maxf0 and minf0 (maxf0 - minf0) has no meaning from a perceptual point of view, so it needs to be converted to a perceptually meaningful value. By using the above formula, the f0 unit can be converted

from Hertz (Hz) into Semitones (sts). Semitones correspond well to our relative perception of pitch. Hence sts instead of Hz is adopted here as the unit of PE.

PS (Pitch Slope) is measured as instantaneous rates of f0 change at a fixed interval (10ms, i.e., the f0 sample rate) during each sonorant interval. Because all sonorant intervals are longer than 10ms, there will be multiple PSs during each pitch contour. The PS-based metrics meanPS and ΔPS are derived from the measures of PS across all pitch contours. The unit of PS is Semitones per second (sts/s).

4.2.3 Phonological structure metrics

Chinese syllabic structures are measured by Fin:Ini and sumFI and tonal structures are by HT:LT and sumT. Specific values are listed for the 21 Chinese dialects in Table 4. (see Appendix 1-5 for their inventories of initials, finals, and tones on which the calculations are based).

Table 4.5 Results from the four phonological structure metrics (all the dialects)

No.	Major group	Sub-group	No.	Dialect	Syllable-based		Tone-based	
					Fin:Ini	sumFI	HT:LT	sumT
1	Mandarin	Northern	1	Beijing	1.82	62	3	4
			2	Tianjin	1.54	61	1	4
			3	Jinan	1.58	62	1	4
			4	Heshuo	1.77	61	3	4
		Southwestern	5	Wuhan	1.95	56	3	4
			6	Chengdu	1.8	56	1	4
			7	Liupanshui	1.68	51	1	4
			8	Guilin	1.94	53	1	4
2	Wu	Northern	9	Hangzhou	1.3	69	0.75	7
			10	Shanghai	1.54	71	1.5	5
			11	Suzhou	1.81	76	1.33	7
		Southern	12	Wenzhou	1.17	63	1	8
3	Cantonese	Guangzhou (GZ)	13	GZ1	3.78	86	1.25	9
			14	GZ2	3.78	86	1.25	9
		Hong Kong (HK)	15	HK1	3.89	88	1.25	9
			16	HK2	3.89	88	1.25	9
4	Min	Southern	17	Quanzhou	6.21	101	1.33	7
			18	Taizhong	5.67	100	1.33	7
			19	Shantou	4.28	95	1.67	8
		Eastern	20	Fuzhou	3.13	62	0.75	7
		Central	21	Sanming	2.18	54	1	6

In order to find out how phonological structures affect duration- and pitch-based rhythm, the syllable-based Fin:Ini and sumFI values will be used to correlate with the results from duration-based metrics and the tone-based HT:LT and sumT values with the results from pitch-based metrics. Note that these values change with the inventory change and the

inventory changes with time and also depends on how they are investigated. To ensure their accuracy, the author would use the latest and most authoritative documentation possible (see Appendix 5 for all the data sources).

4.2.4 Voice source metrics

As mentioned in Section 2.2.3, the most effective voice quality metric is H1*-H2*, which measures the spectral tilt of the first two harmonics while having the influence of the vowel formants (F1 & F2) corrected (using Iseli, et al.'s algorithm, 2007). Breathy voice should have a larger positive value of H1*-H2* than modal voice and in turn than creaky voice, which should have a negative value of H1*-H2*.

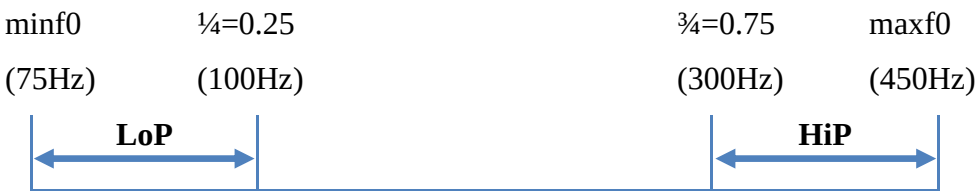
Another metric used in this study is CPP (Cepstral Peak Prominence), which measures the aperiodicity of breathy voice. According to Hillenbrand and Houde (1996), a highly periodic speech signal will have a clear harmonic structure and thus a high prominent peak on the cepstrum (an inverse spectrum), but a non-periodic signal occurring in breathy voice renders the harmonic structure of breathy voice unclear and hence breathy voice has a weak cepstral peak. CPP is also categorized as a type of harmonic-to-noise measure to emphasize the influence of noise (the breathy component) on harmonics (Keating & Garellek, 2015). In Esposito's (2006) study, CPP is found to be quite successful in detecting the breathy phonation for a variety of languages including Fuzhou Min. Note that creaky voice may also have a lower cepstral peak than a modal voice if it is aperiodic (Kuang, 2011; Keating & Garellek, 2015); however, the next paragraph will show that the object to be measured excludes non measurable aperiodic creaky voice, so CPP is used mainly to detect breathiness in speech signals.

Note that in this study, the two types of metrics are not used to measure individual segments and tones in each dialect, as it is difficult to compare their voice quality across different dialects, given that voice quality may occur in different types of segments and/or tones in different dialects. Alternatively, voice quality is measured directly from high- and low-pitched (HiP & LoP) speech signals, regardless of segmental and tonal types with which they are associated. These signals are able to cover the highest and lowest tones in each dialect. Also, high-pitched signals are assumed to have modal voice quality and used to contrast with low-pitched speech signals, which are assumed to have either breathy or creaky voice quality. Note that certain aperiodic signals such as those occurring in creaky voice may not have a measurable pitch, so they can not be the object of the measurement.

The high- and low-pitched signals are identified based on the maxf0 (pitch ceiling) and minf0 (pitch floor) algorithm developed by Hirst (2011). According to Hirst (2011), the maxf0 and minf0 used for f0 estimation in speech analysis software such as Praat have default values set to 75Hz and 600Hz, respectively. These values can not reflect individual speakers' pitch range and often cause octave errors. To solve this issue, Hirst (2011) recommended a two-pass method. The first pass uses the minf0 and maxf0 values, 60Hz and 700Hz, to calculate f0, followed by the second pass, which uses the values of q1 and q3 (the first and third quartiles of the f0 distribution) to calculate the pitch floor and pitch ceiling. In the first pass, the range of minf0 and maxf0 (60-700Hz) is larger than the default one (75-600Hz) so as to capture highly expressive voices. In the second pass, q1 and q3 are used to provide robust estimates of the f0 dispersion. Then pitch floor and ceiling can be determined by multiplying q1 with 0.75 and q3 with 1.5,

respectively. For example, if q1 is 100Hz, then minf0 is 75Hz (=100x0.75); if q3 is 300Hz, then maxf0 is 450Hz (=300x1.5). Now that pitch floor and ceiling are determined, it is possible to determine low- and high-pitched signals: any signals with a pitch range between minf0 and q1 are considered low-pitched and between q3 and maxf0 high-pitched. Figure 4.3 illustrates how low- and high-pitched signals are identified:

Figure 4.3 Illustration of LoP and HiP ranges



A comparison of the voice source measures of high- and low-pitched signals helps to reveal whether voice quality is merely a reflection of individual speakers' voice identity or a phonetic cue accompanying certain tonal production. Also, the H1*-H2* and CPP data will be used to compare with duration- and pitch-based data to see if voice quality contributes to rhythm.

Table 4. summarizes the two voice source metrics and their meanings:

Table 4.6 Two voice source metrics

1	H1*-H2*	The amplitude difference of the first two Harmonics, corrected to avoid the influence of vowel formants F1 and F2
2	CPP	Cepstral Peak Prominence

4.2.5 Summary

There are a total 19 metrics used in this study, nine of which are duration-based, four pitch-based, four phonological structure metrics, and two voice source metrics (see Table

4.7). The nine duration-based metrics are further grouped into four Son-based (%Son, ΔSon, varcoSon, & nPVI_Son), three IS-based (ΔIS, varcoIS, & rPVI_IS), and two extras (sumD & s_rate). The four pitch-based metrics are further grouped into two PE-based (meanPE & ΔPE) and two PS-based (meanPS & ΔPS). The four phonological structure metrics are also grouped into two syllable structure metrics (Fin:Ini & sumFI) and two tone structure metrics (HT:LT & sumT). The two voice source metrics are H1*-H2* (measure of spectral tilt) and CPP (measure of noise).

Table 4.7 Summary of 19 metrics from 4 metric types

Metric type (4)	Sub-type (7)	Individual metric (17)			
Duration-based (9)	Son-based (4)	%Son	ΔSon	varcoSon	nPVI_Son
	IS-based (3)	ΔIS	varcoIS	rPVI_IS	
	Extra (2)	sumD	s_rate		
Pitch-based (4)	PE-based (2)	meanPE	ΔPE		
	PS-based (2)	meanPS	ΔPS		
Phonological structure (4)	Syllable structure (2)	Fin:Ini	sumFI		
	Tone structure (2)	HT:LT	sumT		
Voice source (2)	Spectral tilt (1)	H1*-H2*			
	Noise (1)	CPP			

4.3 **Analysis:**

4.3.1 Acoustic analysis

Acoustic analysis of the speech data in terms of duration-, pitch-, and voice source-based measures was performed automatically with manual adjustment according to the five steps below:

(1) A Praat script developed by Hirst (2011) was used to determine the pitch floor and ceiling (f0 detection range required for acoustic analysis) of each speaker. The purpose of this step is to minimize octave jump occurring during pitch calculation.

(2) The Praat script Prosgram2.9f developed by Mertens (2012) was used to segment speech data into sonorant and inter-sonorant intervals.

(3) A Praat script was used to calculate %Son, ΔSon, ΔIS, varcoSon, and varcoIS and a Python script was used to calculate nPVI_Son and rPVI_IS. Both scripts were developed by Yoon (2008).

(4) The Praat script ProsodyPro5.3.2 developed by Xu (2013) was used to calculate PE and PS. Then Microsoft Excel was used to calculate meanPE, meanPS, ΔPE, and ΔPS.

(5) The VoiceSauce software (running on Matlab) developed by Shue et al. (2011) is used to calculate H1*-H2* and CPP values of high- and low-pitched signals.

4.3.2 Statistical analysis

The results of all the duration-, pitch- and structure-based measurement were subject to the correlation analysis so as to find out how duration-, pitch-, voice-source-, and structure-based metrics are related to one another. The results take the form of the correlation coefficient (cc) values. All the cc values fall between -1 and +1 inclusive. A positive cc value indicates that the associated two metrics tend to move together in one direction, whereas a negative cc value indicates that that two metrics tend to move together in opposite direction. If two metrics are unrelated, then their absolute cc value is smaller than 0.5 (i.e., $|cc| < 0.5$ or $-0.5 < cc < 0.5$). If an absolute cc value is larger than

0.7, then it can be said that the two metrics are highly correlated. Also, the correlation is considered low if the absolute cc value is between 0.5 and 0.6 and moderate if it is between 0.6 and 0.7.

To sum up, the methodology adopted here is different from previous studies in five aspects: 1) the speech material is natural rather than designed; 2) the duration-based metrics use inter-sonorant intervals instead of obstruent or consonant intervals; 3) the pitch-based metrics use the perceptually meaningful unit semitones (sts) instead of Hertz (Hz); 4) the f0 detection setting is individualized according to speakers' pitch range; 4) the voice source-based metrics measure high- and low-pitched speech signals instead of individual segments/tones. It is hoped that all these metrics will help us to understand rhythm from a better perspective, or at least from a different perspective.

4.4 Predictions

In order to show how duration-, pitch-, and structure-based metrics in forming rhythmic patterns for different dialects, all the metrics except the two voice-source-based ones (to be discussed separately) are paired up and subjected to the correlation analysis. The results of the correlation analysis are used to reveal duration- and pitch-based rhythmic patterns and correlation patterns among duration-, pitch-, and phonology-based metrics. As shown in Table 4.8, there are 57 pairs of metrics being used to predict the relative degrees of syllable-timedness and melodiousness for all the dialects. They are divided into five categories, five pairs in duration-only, two in pitch-only, 28 in duration-pitch, 14 in duration-syllable, and eight in pitch-tone categories (5+2+28+14+8=57). The five metric pairs in the duration-only category all occurred in the previous rhythmic studies (hereafter "conventional pairs"). The two metric pairs in the pitch-only category are

chosen based on Vicenik and Sundara (2012). The duration-pitch category includes all 28 possible metric pairs (= 7 duration-based metrics x 4 pitch-based metrics) and no previous studies have explicitly studied their relationships. Similarly, the duration-syllable category includes all 14 possible metric pairs (= 7 duration-based metrics x 2 syllable-based metrics) and the pitch-tone category includes all eight possible metric pairs (= 4 pitch-based metrics x 2 tone-based metrics).

The predictions for the five conventional metric pairs in the duration-only category are as follows: %Son is highly (H) and negatively (−) correlated with ΔIS, varcoSon, and varcoIS, ΔSon is highly (H) and positively (+) correlated with ΔIS, and rPVI_IS is also highly (H) and positively (+) correlated with nPVI_Son. Also, the larger the %Son is and the smaller the remaining six metrics are, the more syllable-timed the speech is.

The predictions for the two metric pairs in the pitch-only category are as follows: meanPE and meanPS are highly (H) and positively (+) correlated, so do ΔPE and ΔPS. Also, the larger the four metrics are, the more melodious the speech is.

The predictions for the 28 metric pairs in the duration-pitch category are as follows: only %Son is highly (H) and positively (+) correlated with the four pitch-based metrics, and all the remaining six duration-based metrics are highly (H) and negatively (−) correlated with the four pitch-based metrics. Also, the larger the %Son is, the smaller the remaining six duration-based metrics are, and the larger the four pitch-based metrics, the more syllable-timed and more melodious the speech is.

The predictions for the 14 metric pairs in the duration-syllable category are as follows: only %Son is highly (H) and negatively (−) correlated with the two syllable-based metrics, and all the remaining six duration-based metrics are highly (H) and positively (+)

correlated with the two syllable-based metrics. Also, the larger the %Son is, the smaller the remaining six duration-based and two syllable-based metrics are, the more syllable-timed the speech is.

The predictions for the eight metric pairs in the pitch-tone category are as follows: The four pitch-based metrics are highly (H) and positively (+) correlated with HT:LT but highly (H) and negatively (−) correlated with sumT. Also, the larger the four pitch-based metrics are, the larger HT:LT, and the smaller sumT, the more melodious the speech is.

Table 4.8 Predicted correlations for 57 metric pairs from 5 categories

No.	Category (# of metric pairs)	metric pair	Predicted correlation	Associated rhythmic pattern
1	Duration-only (5 pairs)	%Son - ΔIS %Son -varcoSon %Son -varcoIS	H−	larger %Son ↔ more syllable-timed
		ΔSon - ΔIS rPVI_IS - nPVI_Son	H+	
2	Pitch-only (2 pairs)	meanPE - meanPS ΔPE - ΔPS	H+	melodious
3	Duration-pitch (28 pairs)	%Son	H+	larger %Son ↔ more syllable-timed & more melodious
		ΔSon varcoSon nPVI_Son ΔIS varcoIS rPVI_IS - meanPE - meanPS - ΔPE - ΔPS	H−	
4	Duration-syllable (14 pairs)	%Son	H−	smaller Fin:Ini, smaller sumFI ↔ more syllable-timed
		ΔSon varcoSon nPVI_Son ΔIS varcoIS rPVI_IS - Fin:Ini - sumFI	H+	
5	Pitch- tone (8 pairs)	meanPE - meanPS - ΔPE - ΔPS - HT:LT sumT	H+ H−	larger HT:LT, smaller sumT ↔ more melodious

Since the values of the two syllable-based metrics Fin:Ini and sumFI and the two tone-based metrics HT:LT and sumT are known for all 21 Chinese dialects (see Table 4.6), the relative degree of syllable-timedness and melodiousness of these dialects can be somehow ordered based on the predictions in Table 4.8. According to these values, Fin:Ini ranges from 1.17 to 6.21, sumFI from 51 to 101, HT:LT from 0.75 to 3, and sumT from 4 to 9. Now it can be assumed that as each metric increases or decreases in value, the degree of syllable-timedness or melodiousness also increases or decreases for the associated dialect, depending on which metric is involved.

Table 4.9 lists detailed rankings of all 21 dialects in terms of the four structure-based measures. The first four columns show that as Fini:Ini and sumFI increases, the degree of syllable-timedness decreases. For example, the two Wu dialects Wenzhou and Hangzhou have a smaller Fin:Ini value (<1.5) than the rest of the Chinese dialects, so they are listed among the most syllable-timed. In contrast, the three Southern Min dialects Shantou, Taizhong, and Quanzhou are the least syllable-timed as they have a larger Fin:Ini value (> 4) than the rest of the Chinese dialects. Note that the two syllable structure-based measures do not rank all the major groups of dialects consistently: they agree only on the rankings of the four Cantonese and the three Southern Min dialects but not on the rankings of the four Wu, eight Mandarin, and two East and Central Min dialects. This internal inconsistency of the structure-based measures suggests that they may not both correlate well with rhythm and melody patterns.

The next four columns show that as HT:LT decreases and sumT increases, the degree of melodiousness decreases. For example, the three Mandarin dialects Beijing, Heshuo, and Wuhan are the most melodious as they all have the largest HT:LT value (= 3). On

the other hand, Fuzhou Min and Hangzhou Wu have the smallest HT:LT value (= 0.75), so they are listed among the least melodious.

Similarly, the two tone structure measures do not rank all the major groups of dialects consistently: they agree only on the rankings of the three Mandarin dialects Beijing, Heshuo, and Wuhan, recognizing that they are among the most melodious. For the rest of the dialects, however, the rankings are rather messy, suggesting that HT:LT and sumT cannot both be good indicators of melodiousness.

To summarize, predictions made in this section are all based on previous studies, and it is hoped that they can generalize to this study. For the phonological structure measures, they already show some inconsistency in predicting the relative syllable-timedness and melodiousness of all the dialects, but it is yet to be seen which one of them has the best predictive power.

Table 4.9 Relative syllable-timedness and melodiousness predicted by phonological structure metrics for all 21 Chinese dialects

Syllable-timedness↓	Fin: Ini↑	Syllable-timedness↓	sumFI↑	Melodious-ness↓	HT: LT↓	Melodious-ness↓	sumT↑
ΔWenzhou	1.17	◊Liupanshui	51	♦Beijing ♦Heshuo ◊Wuhan	3	♦Beijing ♦Tianjin ♦Jinan ♦Heshuo ◊Wuhan ◊Chengdu ◊Liupanshui ◊Guilin	4
▲Hangzhou	1.27	◊Guilin	53				
♦Tianjin ▲Shanghai	1.54	□Sanming	54	▲Shanghai	1.5		
		◊Wuhan ◊Chengdu	56	■Quanzhou ■Taizhong ▲Suzhou	1.33		
♦Jinan	1.58	♦Tianjin ♦Heshuo	61	●GZ1 ●GZ2 ○HK1 ○HK2	1.25	▲Shanghai	5
◊Liupanshui	1.68					□Sanming	6
♦Heshuo	1.77	♦Beijing ♦Jinan □Fuzhou	62			▲Hangzhou ▲Suzhou ■Quanzhou ■Taizhong □Fuzhou	7
◊Chengdu ▲Suzhou	1.8 1.81						
♦Beijing	1.82	ΔWenzhou	63	♦Tianjin ♦Jinan ◊Liupanshui ◊Chengdu ◊Guilin ΔWenzhou □Sanming ■Shantou	1		
◊Guilin	1.94	▲Hangzhou	69				
◊Wuhan	1.95	▲Shanghai	71				
□Sanming	2.18	▲Suzhou	76			ΔWenzhou ■Shantou	8
□Fuzhou	3.13						
●GZ1 ●GZ2	3.78	●GZ1 ●GZ2	86			●GZ1 ●GZ2 ○HK1 ○HK2	9
○HK1 ○HK2	3.89	○HK1 ○HK2	88				
■Shantou	4.28	■Shantou	95				
■Taizhong	5.67	■Taizhong	100	□Fuzhou ▲Hangzhou	0.75		
■Quanzhou	6.21	■Quanzhou	101				

*↑: ascending order; ↓: descending order; ♦/◊Northern/Southwestern Mandarin; ▲/Δ Northern/Southern Wu; ■/□Southern/Eastern & Central Min; ●/○ GZ/HK Cantonese

Chapter 5
RESULTS

This chapter reports the results obtained from duration-based, pitch-based, phonological structure, and voice source measures of the Chinese dialect speech data (see Appendix 1-11 for individual values of all the measures). Section 5.1 and Section 5.2 respectively present correlation results in the duration-only and pitch-only categories to show duration-based timing and pitch-based melody patterns of four major dialect groups. Section 5.4 presents voice source results to show voice quality patterns of four major groups. Section 5.3 presents correlation results to show how phonological structure, syllable-timedness, and melodiousness are correlated in four major groups. Section 5.5 presents results to show how timing and melody are respectively patterned and correlated with phonological structure for all 21 Chinese dialects regardless of their group membership. Note that voice quality does not seem to be a major player in shaping rhythmic patterns of the Chinese dialects, so the correlation between voice quality and rhythm is not pursued further but will be briefly discussed in Chapter 6.

5.1 Duration-based timing patterns

Correlation results in the duration-only category and the associated timing patterns are presented successively for Mandarin, Wu, Min, and Cantonese in Section 5.1.1, 5.1.2, 5.1.3, and 5.1.4. Section 5.1.5 summarizes the results.

5.1.1 Mandarin

Correlation results for Mandarin are listed inTable 5.1.

Table 5.1 Correlation results in the duration-only category (Mandarin)

	Column	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	
	Metric	*sumD*	*s_rate*	*%Son*	*ΔSon*	*varco-Son*	*nPVI_Son*	*ΔIS*	*varco-IS*	*rPVI_IS*	*Row*
	sumD	1									*1*
	s_rate	0.061	1								*2*
Son_based	*%Son*	-0.603	-0.544	1							*3*
	ΔSon	0.017	<u>-0.910</u>	0.370	1						*4*
	varcoSon	-0.380	-0.171	0.255	0.162	1					*5*
	nPVI_Son	0.132	0.122	-0.175	0.093	-0.588	1				*6*
IS-based	*ΔIS*	0.618	0.151	<u>-0.860</u>	0.087	-0.340	0.315	1			*7*
	varcoIS	-0.579	0.129	0.162	-0.128	0.657	-0.447	-0.183	1		*8*
	rPVI_IS	0.613	0.155	-0.851	0.044	-0.477	0.324	<u>0.980</u>	-0.291	1	*9*

In Column 1 and 2 (C1 & C2), the highest correlation occurs between s_rate and ΔSon (cc = -0.910, underlined in C2). The negative correlations indicate that faster speech tends to have less variability in sonorant duration. As for sumD, it is not highly correlated with all the Son- and IS-based metrics (|cc| < 0.7; see C1), indicating that speech length does not contribute much to timing.

Of the five conventional duration-based metric pairs, only the pair %Son-ΔIS shows a high correlation (|cc| = |-0.860| > 0.7, underlined in C3). The negative correlation between %Son and ΔIS is predicted and it means that dialects with a longer portion of sonorant duration and less variability in inter-sonorant duration tend to be more syllable-timed.

Of the 16 unpredicted duration-based metric pairs (C3-C9, except the 7 pairs with cc =1 and the 5 predicted pairs), the pair ΔIS-rPVI_IS has the highest correlation (cc = 0.980, underlined in C7). The positive correlation between ΔIS and rPVI_IS means that dialects with less variability in inter-sonorant duration tend to be more syllable-timed.

This correlation is not predicted but can be expected because the two metrics represent two different ways of measuring the same inter-sonorant duration, and therefore they should vary together in the same direction. In other words, all the IS-based metrics, whether it is calculated by ΔIS, by varcoIS, or by rPVI_IS, are expected to be small for syllable-timed languages.

In order to reveal the rhythmic patterns of the seven Mandarin dialects in further detail, Figure 5.1 and Figure 5.2 respectively use ΔIS-rPVI_IS and %Son-ΔIS as the x-y axis to plot all seven Mandarin dialects along the syllable-timedness continuum in the form of a scatter chart. The first pair has the highest correlation among all the pairs and the second has the highest correlation among the five predicted pairs. The selection of the two pairs is based on the assumption that they can best represent duration-based metrics to provide answers to the given research questions. The negative correlation is represented by the downward trend line and the positive correlation by the upward trend line. Along the upward trend line, the degree of syllable-timedness can be clearly ordered for the four northern dialects and the four Southwestern dialects respectively as follows: Beijing > Heshuo > Jinan > Tianjin and Wuhan > Chengdu > Guilin > Liupanshui. Along the downward trend line, the degree of syllable-timedness is similarly ordered for most Mandarin dialects, except that the places of Guilin and Chengdu are switched.

A comparison of the two sub-dialect groups shows a tendency for the Southwestern dialects to be more syllable-timed than northern dialects except Beijing, but since the eight Mandarin dialects do not form two separate clusters according to their respective membership in Northern and Southwestern sub-groups, it can be said that duration-based metrics can not distinguish Mandarin rhythm at the sub-dialectal level.

Figure 5.1 Timing pattern based on ΔIS-rPVI_IS (Mandarin)

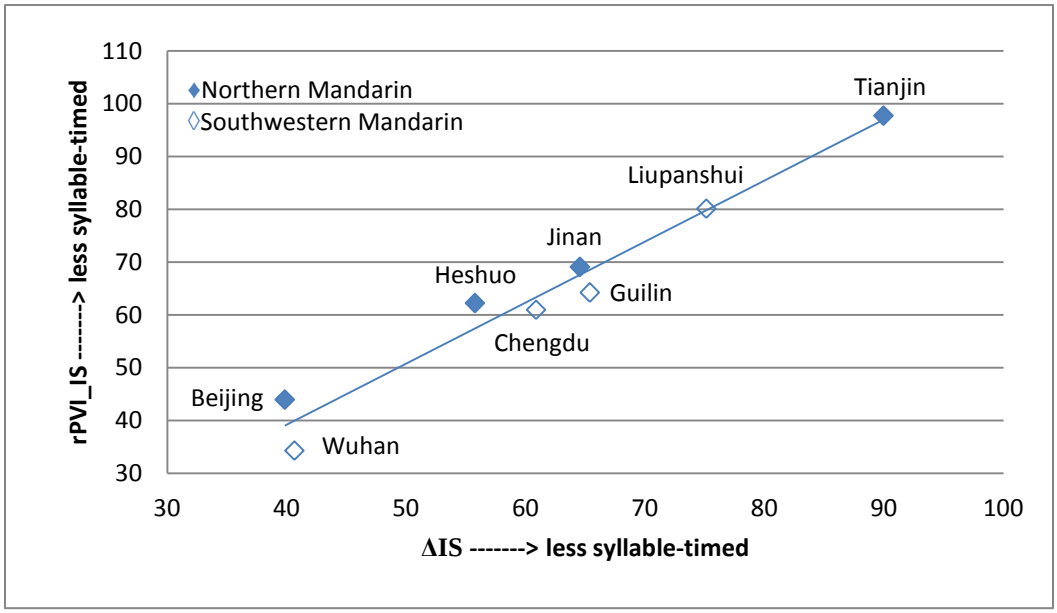
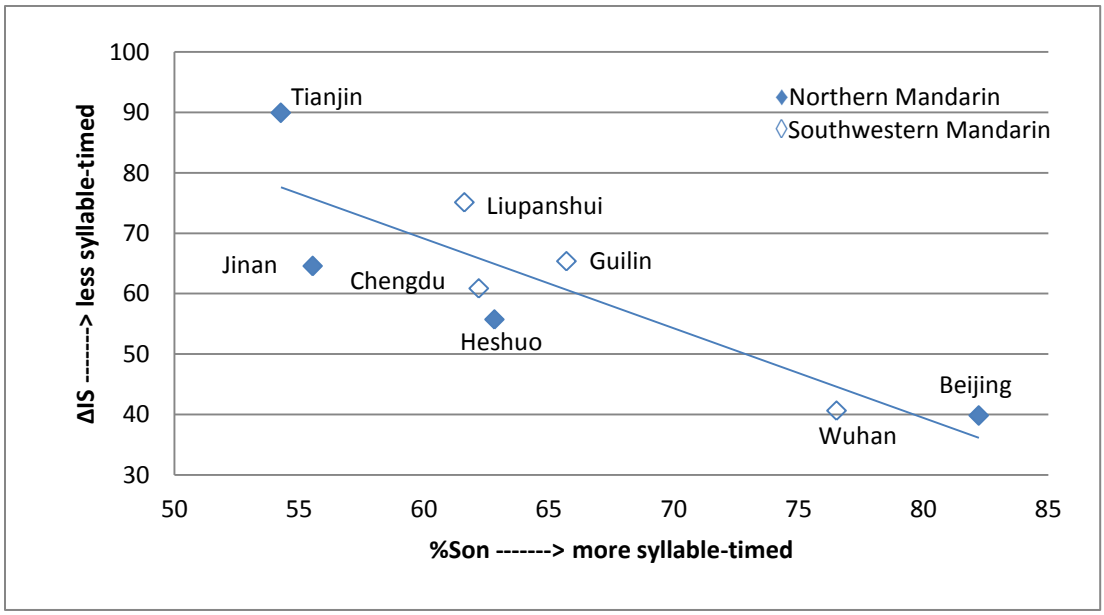


Figure 5.2 Timing pattern based on %Son-ΔIS (Mandarin)



5.1.2 Wu

Correlation results for Wu are listed in Table 5.2.

Table 5.2 Correlation results in the duration-only category (Wu)

	Column	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	
	Metric	*sumD*	*s_rate*	*%Son*	*ΔSon*	*varco Son*	*nPVI_ Son*	*ΔIS*	*varco IS*	*rPVI_ IS*	*Row*
	sumD	1									*1*
	s_rate	-0.531	1								*2*
Son_based	*%Son*	0.986	-0.403	1							*3*
	ΔSon	0.787	-0.842	0.735	1						*4*
	varcoSon	0.890	-0.857	0.808	0.909	1					*5*
	nPVI_Son	0.327	-0.399	0.362	0.718	0.366	1				*6*
	ΔIS	-0.697	-0.096	-0.727	-0.107	-0.394	0.335	1			*7*
IS-based	*varcoIS*	-0.400	-0.549	-0.507	0.202	0.043	0.274	0.857	1		*8*
	rPVI_IS	-0.888	0.105	-0.924	-0.426	-0.601	-0.035	0.928	0.772	1	*9*

In C1 and C2, the highest correlations occur respectively between sumD and %Son (cc = 0.986; underlined in C1-R3) and between s-rate and varcoSon (cc = -0.857; underlined in C2-R5). The positive correlation between sumD and %Son and the negative correlation between s-rate and varcoSon indicate that longer speech tends to have a larger portion of sonorant duration and faster speech tends to have smaller variability in sonorant duration. The latter pattern can be expected as faster speech tends to shorten vowel duration and hence reduce the range of variation in sonorant duration.

Of the five conventional duration-based metric pairs, three pairs, %Son-varcoSon, %Son-ΔIS, and %Son-varcoIS, are correlated and the pair %Son-varcoSon has the highest correlation among the three pairs (cc = 0.808, underlined in C3). The positive correlation between %Son and varcoSon, however, is in conflict with the prediction that %Son and varcoSon are inversely related.

Of the 16 unpredicted duration-based metric pairs, seven pairs are correlated and the highest correlation occurs between ΔIS and rPVI_IS (cc = 0.928, underlined in C7-R9). The positive correlation is expected and it means that dialects with less variability in inter-sonorant duration, whether measured by ΔIS or rPVI_IS, tend to be more syllable-timed.

Figure 5.3 and Figure 5.4 respectively use the metric pair ΔIS-rPVI_IS and %Son- ΔIS as the x-y axis to plot all the dialects along the syllable-timedness continuum. The first pair has the highest correlation among all the pairs and the second pair has the second highest correlation among the five predicted pairs. Note that %Son-varcoSon as the highest correlated pair among the five predicted pairs is not usable here, because its positive correlation gives conflicting timing patterns.

Along the upward trend line, the degree of syllable-timedness can be clearly ordered for all four Wu dialects as follows: Hangzhou > Suzhou > Wenzhou > Shanghai. Along the downward trend line, the order is a little different: Hangzhou > Suzhou, Wenzhou > Shanghai. Suzhou and Wenzhou are put into the same group as %Son-ΔIS cannot differentiate them. In general, the two metric pairs can well distinguish the relative syllable-timedness for the three northern Wu dialects, Hangzhou, Suzhou, and Shanghai.

Figure 5.3 Timing pattern based on ΔIS-rPVI_IS (Wu)

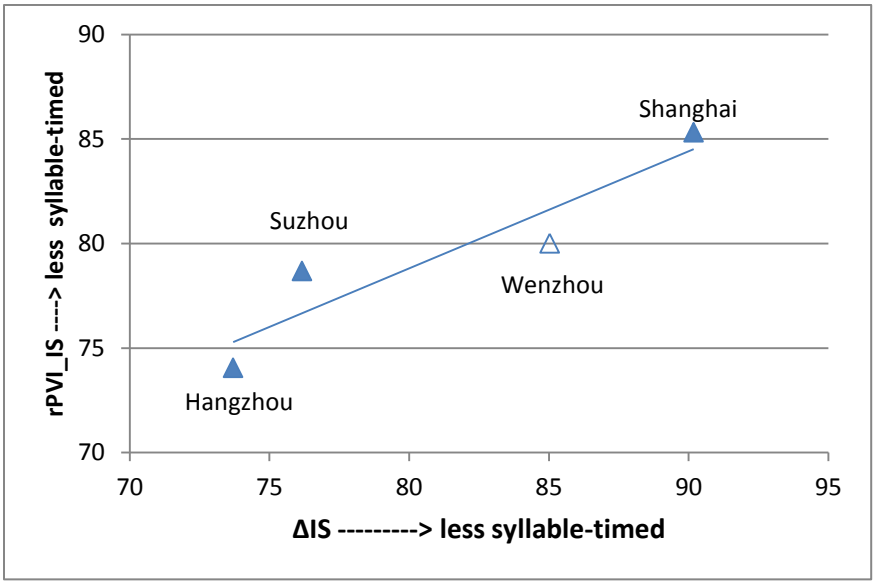
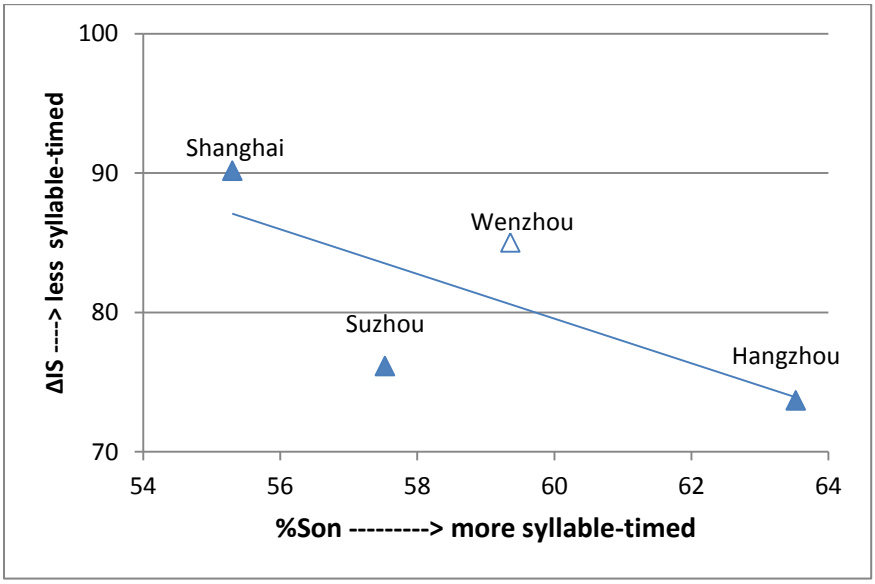


Figure 5.4 Timing pattern based on %Son-ΔIS (Wu)



5.1.3 Min

Correlation results for Min are listed in Table 5.3.

Table 5.3 Correlation results in the duration-only category (Min)

	Column	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	
	Metric	*sumD*	*s_rate*	*%Son*	*ΔSon*	*varco Son*	*nPVI_ Son*	*ΔIS*	*varco IS*	*rPVI_ IS*	*Row*
	sumD	1									*1*
	s_rate	-0.346	1								*2*
Son_based	*%Son*	0.146	-0.485	1							*3*
	ΔSon	-0.332	-0.699	0.637	1						*4*
	varcoSon	0.248	-0.958	0.703	0.805	1					*5*
	nPVI_Son	0.529	-0.897	0.707	0.551	0.932	1				*6*
IS-based	*ΔIS*	0.786	-0.754	0.259	0.090	0.656	0.838	1			*7*
	varcoIS	0.226	-0.842	-0.058	0.453	0.669	0.573	0.660	1		*8*
	rPVI_IS	0.536	-0.933	0.206	0.430	0.798	0.796	0.850	0.917	1	*9*

In C1 and C2, the highest correlations occur between sumD and ΔIS (cc = 0.786; underlined in C1-R7) and between s_rate and varcoSon (cc = -0.958; underlined in C2-R5), respectively. The positive correlation between sumD and ΔIS and the negative correlation between s_rate and varcoSon indicate that longer speech tends to have more variability in inter-sonorant duration and faster speech tends to have smaller variability in sonorant duration.

Of the five conventional duration-based metric pairs, two pairs, %Son-varcoSon and rPVI_IS-nPVI_Son, are correlated and the pair rPVI_IS-nPVI_Son has a higher correlation (cc = 0.796, underlined in C6). Of the 16 unpredicted duration-based metric pairs, 14 pairs are correlated, and the highest correlation occurs between varcoSon and nPVI_Son (cc = 0.932, underlined in C5-R6). This correlation is not predicted but can be expected because the two metrics represent two different ways of measuring the same sonorant duration, and they should vary together in the same direction. In other words, all the Son-based metrics except %Son, whether it is calculated by ΔSon, by varcoSon, or by nPVI_Son, are expected to be small for syllable-timed languages.

Together, there are 16 correlated metric pairs and they all have positive cc values (see C3-8 for those with cc > 0.5). Note that the positive correlation even holds for %Son-varcoSon, %Son-ΔSon, and %Son-nPVI_Son (see C3-R4-6), despite that negative correlations are predicted for them.

Figure 5.5 and Figure 5.6 respectively use two pairs of metrics, varcoSon-nPVI_Son and rPVI_IS-nPVI_Son, as the x-y axis to plot all the dialects along the syllable-timedness continuum. The first pair has the highest correlation among all the correlated pairs and the second pair has the highest correlation among the five predicted pairs. Along the upward trend line in the first figure, the degree of syllable-timedness can be ordered for the five Min dialects by four groups: Quanzhou > Shantou > Taizhong, Sanming > Fuzhou. The relative syllable-timedness of Taizhong and Sanming in the third group is to be determined. In the second figure, the trend line can only order the dialects by two groups: Quanzhou, Shantou > Taizhong, Fuzhou, Sanming. A comparison of the three dialect groups shows that Eastern and Central Min dialects tend to be more syllable-timed than the three Southern Min dialects, but in general, the two duration-based metric pairs cannot distinguish the relative syllable-timedness of Min dialects consistently.

Figure 5.5 Timing pattern based on varcoSon-nPVI_Son (Min)

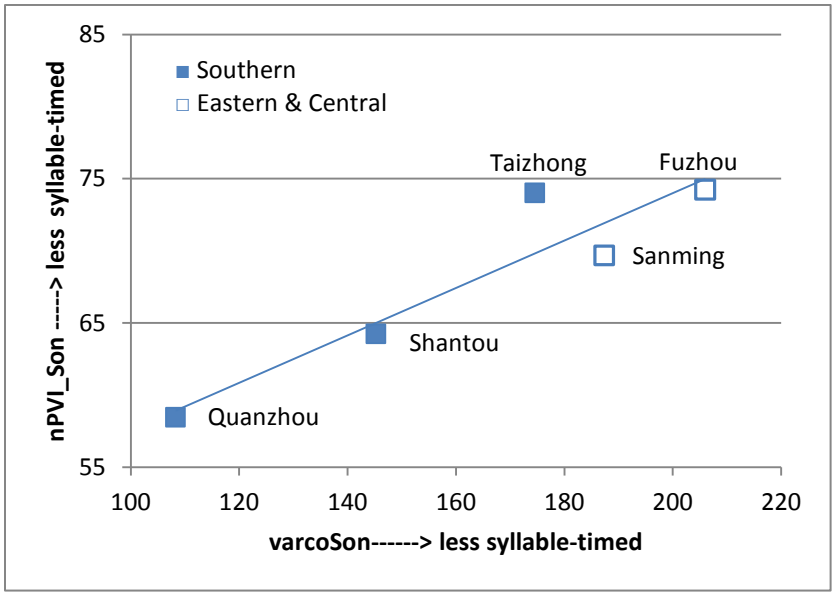
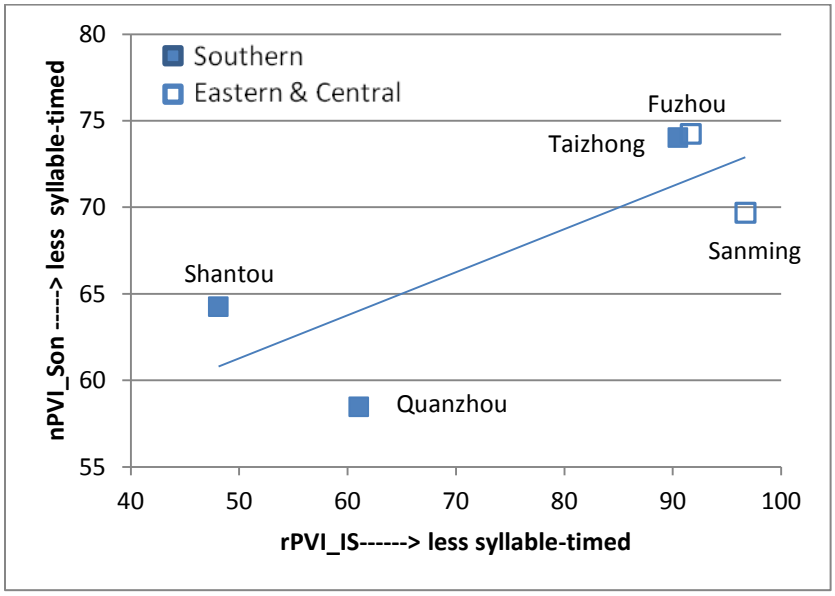


Figure 5.6 Timing pattern based on rPVI_IS-nPVI_Son (Min)



5.1.4 Cantonese

Correlation results for Cantonese are listed in Table 5.4.

Table 5.4 Correlation results in the duration-only category (Cantonese)

	Column	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	
	Metric	*sumD*	*s_rate*	*%Son*	*ΔSon*	*varco Son*	*nPVI_ Son*	*ΔIS*	*varco IS*	*rPVI_ IS*	*Row*
	sumD	1									*1*
	s_rate	-0.306	1								*2*
Son_based	*%Son*	0.529	-0.131	1							*3*
	ΔSon	-0.164	-0.447	-0.053	1						*4*
	varcoSon	0.481	-0.740	0.527	0.037	1					*5*
	nPVI_Son	0.515	-0.491	0.817	0.092	0.889	1				*6*
IS-based	*ΔIS*	0.416	-0.141	0.234	-0.684	0.644	0.490	1			*7*
	varcoIS	-0.207	-0.508	-0.659	0.162	0.287	-0.120	0.261	1		*8*
	rPVI_IS	-0.411	-0.199	-0.932	0.194	-0.202	-0.568	-0.094	0.877	1	*9*

In C1 and C2, the highest and the only high-level correlation occurs between s_rate and varcoSon (cc = -0.740; underlined in C2-R5). The negative correlation between s_rate and varcoSon indicates that faster speech tends to have smaller variability in sonorant duration.

Of the five conventional duration-based metric pairs, four are correlated but none of them are highly correlated (|cc| < 0.7), and the highest correlation occurs between ΔSon and ΔIS (cc = -0.684; underlined in C4-R7). The negative correlation is not predicted and it gives conflicting timing patterns. Of the 16 unpredicted duration-based metric pairs, only four pairs are highly correlated, and the highest correlation occurs between %Son and rPVI_IS (cc = -0.932; underlined in C3). Again, this correlation is not predicted but can be expected, as rPVI_IS is similar to ΔIS, both measuring inter-sonorant variability. If %Son and ΔIS should correlate negatively, %Son and rPVI_IS should, too.

Figure 5.7 and 5.8 respectively use two pairs of metrics, %Son-rPVI_IS and %Son-varcoIS, as the x-y axis to plot all the dialects along the syllable-timedness continuum. The first pair has the highest correlation among all the pairs. The second pair has the

second highest correlation among the five correlated pairs (cc = -0.659; underlined in C3-R8). Note that ΔSon-ΔIS as the highest correlated pairs among the five predicted pairs is not usable here, because its negaitive correlation gives conflicting timing patterns.

Along the downward trend line in the first figure, the degree of syllable-timedness can be ordered for the four Cantonese dialects in three groups: GZ2 < GZ1, HK2 < HK1. The relative syllable-timedness of GZ1 and HK2 is to be determined, but within each sub-group, the order is clear: GZ2< GZ1 and HK2 < HK1. Along the upward trend line in the second figure, the order is clear within sub-groups but not across sub-groups, meaning that duration-based measures cannot distinguish the relative syllable-timedness consistently across Cantonese sub-groups.

Figure 5.7 Timing pattern based on %Son-rPVI_IS (Cantonese)

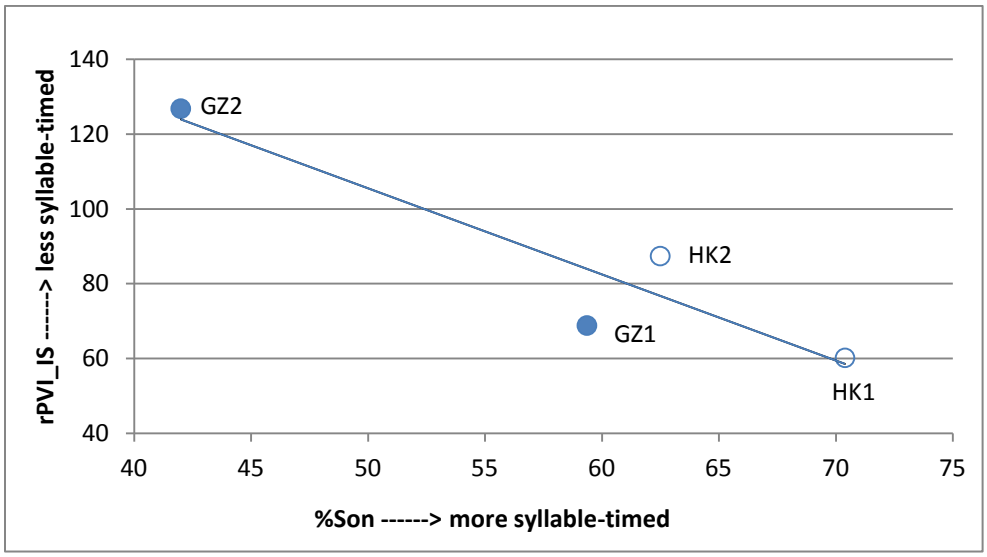
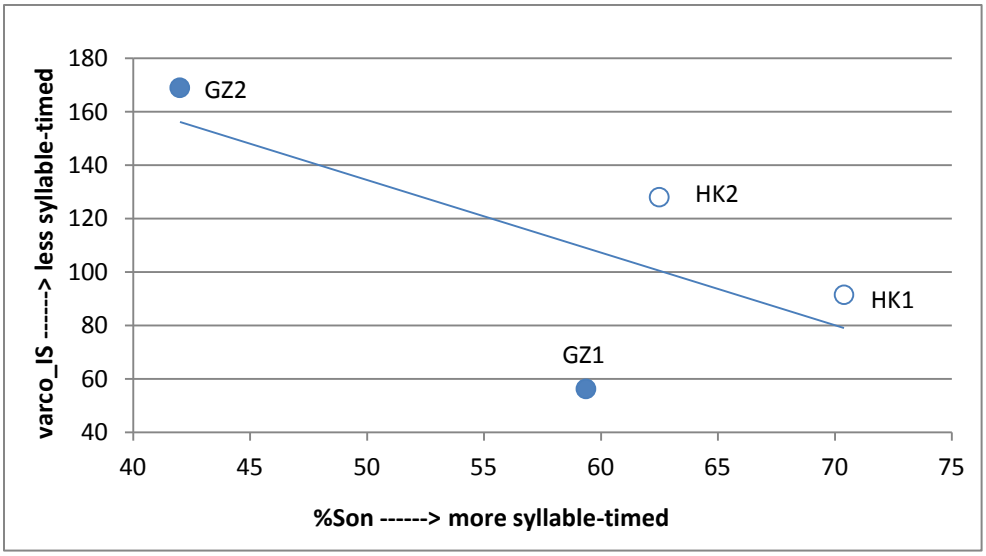


Figure 5.8 Timing pattern based on %Son-varcoIS (Cantonese)



5.1.5 Summary

Duration-based rhythmic patterns vary greatly across the four major groups of Chinese dialects. As shown in Table 5.5, almost half of the correlations are not significant (|cc| < 0.5; marked as ns). Ideally, the five duration-based metric pairs used in previous studies (call them conventional pairs for shorthand) should all have a high correlation (i.e., |cc| > 0.7, marked as H; see C1). Also, the direction of the correlation should be positive (+) for the two pairs ΔSon-ΔIS and rPVI_IS-nPVI_Son and negative (−) for the three %Son-based pairs %Son-ΔIS, %Son-varcoIS, and %Son-varcoSon. The actual results, however, are very different from what is predicted.

For each major dialect group, none of the highest correlation occurs in the five conventional pairs (see underlined for the highest correlation in each group). Nonetheless, %Son-varcoSon occurs the most in correlated pairs (3 times in total; see R3), and the first metric %Son is also frequently seen in other correlated pairs, so it may be the best metric in revealing the relative syllable-timedness for all the Chinese dialects.

As for the actual direction of correlation of the five conventional pairs, only %Son and ΔIS are correlated negatively as predicted across the dialect groups who have them correlated (i.e., Mandarin and Wu). All the remaining pairs have directions contrary to what is predicted for at least one group. It seems that duration variability measures alone cannot predict syllable-timing.

Table 5.5 Summary of correlation results in the duration-only category (four major groups)

Row	Duration-based metric pairs	Predicted correlation type	Mandarin	Wu	Min	Cantonese
1	%Son-ΔIS	H−	H−	H−	n/c	n/c
2	%Son-varcoIS	H−	n/c	L−	n/c	M−
3	%Son-varcoSon	H−	n/c	H+	H+	L+
4	ΔSon-ΔIS	H+	n/c	n/c	n/c	M−
5	rPVI_IS-nPVI_Son	H+	n/c	n/c	H+	L−
6	ΔIS-rPVI_IS	-	H+	H+	-	-
7	varcoSon-nPVI_Son	-	-	-	H+	H+
8	%Son-rPVI_IS	-	-	-	-	H−
	Column	*1*	*2*	*3*	*4*	*5*

H: Highly correlated (|cc|> 0.7); M: Moderately correlated (0.6 <|cc|<0.7); L: Lowly correlated (0.5< |cc| < 0.6). - : Not applicable; n/c: Not correlated

The sumD- and s_rate-based results show that not a single metric is consistently correlated with sumD or s_rate across all four major groups (see Table 5.6). Nonetheless, there is a tendency for s_rate to correlate with varcoSon consistently (see R10): s_rate-varcoSon occurs in three major groups except Mandarin and all have a highly negative correlation (see R10-C5).

Also, there are six out of fourteen pairs occurring three times in three dialects and they are sumD-%Son, sumD-varcoSon, sumD-rPVI_Son, s_rate-ΔSon, s_rate-varcoSon, and s_rate varcoIS (see C5). The frequent occurrence of Son-based metrics in these pairs indicates that a longer and faster speech tends to influence sonorant duration, though the direction of the influence is not predictable. This pattern can be expected, however, after all, since sonorancy is the most important perceptual cue.

As for the influence of sumD and s_rate on individual dialects, Min suffers the most, as it has 9 correlated pairs (see C3) and five out of seven pairs involving s_rate have a high correlation (see R8-R14-C3). Wu, on the other hand, suffers badly from sumD, as four out of seven pairs involving sumD have a high correlation.

Generally speaking, duration-based rhythmic patterns are just as heterogeneous as dialects themselves. Not a single pair of metrics is able to correlate as well as predicted and even if they correlate well, they may not be able to distinguish the relative syllable-timedness consistently for all the Chinese dialects. Worse yet, most of them are neither correlated nor correlated in the expected direction.

Table 5.6 Summary of sumD- and s_rate-based correlation results (four major groups)

Row	Correlated metric pairs (14)		Mandarin (6)	Wu (8)	Min (9)	Cantonese (4)	# of occurrence in each group
1	sumD-	%Son	M−	H+	n/c	M+	3
2		ΔSon	n/c	H+	n/c	n/c	1
3		varcoSon	n/c	H+	n/c	n/c	1
4		nPVI_Son	n/c	n/c	M+	M+	2 (M+)
5		ΔIS	M+	M−	H+	n/c	3
6		varcoIS	M−	n/c	n/c	n/c	1
7		rPVI_IS	M+	H−	M+	n/c	3
8	s_rate-	%Son	M−	n/c	n/c	n/c	1
9		ΔSon	H−	H−	M−	n/c	3
10		varcoSon	n/c	H−	H−	H−	3 (H−)
11		nPVI_Son	n/c	ns	H−	n/c	1
12		ΔIS	n/c	n/c	H−	n/c	1
13		varcoIS	n/c	M−	H−	M−	3
14		rPVI_IS	n/c	n/c	H−	n/c	1
	Column		*1*	*2*	*3*	*4*	*5*

5.2 Pitch-based melody patterns

Correlation results in the pitch-only category and the associated melody patterns are presented successively for Mandarin, Wu, Min, and Cantonese in Section 5.2.1, 5.2.2, 5.2.3, and 5.2.4. Section 5.2.5 summarizes the results.

5.2.1 Mandarin

Correlation results for Mandarin are listed in Table 5.7.

Table 5.7 Correlation results in the pitch-only category (Mandarin)

	Column	*1*	*2*	*3*	*4*
Row	*Metric*	*meanPE*	*ΔPE*	*meanPS*	*ΔPS*
1	meanPE	1			
2	ΔPE	0.917156	1		
3	meanPS	0.565157	0.636065	1	
4	ΔPS	0.452686	0.613892	0.940599	1

Both meanPE-meanPS and ΔPE-ΔPS are positively correlated as predicted, but the degrees of their correlations are respectively low and moderate (cc = 0.565157 & 0.613892 < 0.7, underlined in C1-R3 & C2-R4). The positive correlations are predicted and they mean that dialects with a larger pitch excursion and slope and larger variability in pitch excursion and slope tend to be more melodious.

In order to reveal the pitch-based rhythmic patterns of the seven Mandarin dialects in further detail, Figure 5.9 and Figure 5.10 respectively uses meanPE-meanPS and ΔPE-ΔPS as the x-y axis to plot all the dialects along the melodiousness continuum.

Along the upward trend line in both figures, Beijing shows up as the most melodious. The two Southwestern dialects Chengdu and Guilin are among the least melodious.

Heshuo, Liupanshui, Tianjin, Jinan, and Wuhan are in between. The position of Tianjin shifts from close to Liupanshui to far away from it, causing timing inconsistency across the two sub-groups. The overall distribution of the eight Mandarin dialects does not show a clear division between Northern and Southwestern Mandarin groups, so pitch-based metrics cannot distinguish Mandarin rhythm well at the sub-dialectal level.

Figure 5.9 Melody pattern based on meanPE-meanPS (Mandarin)

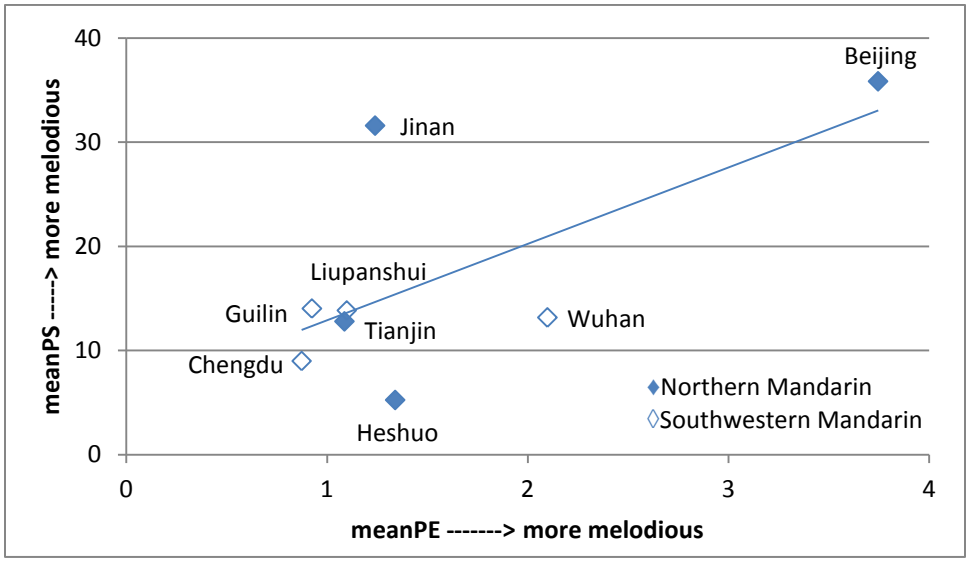
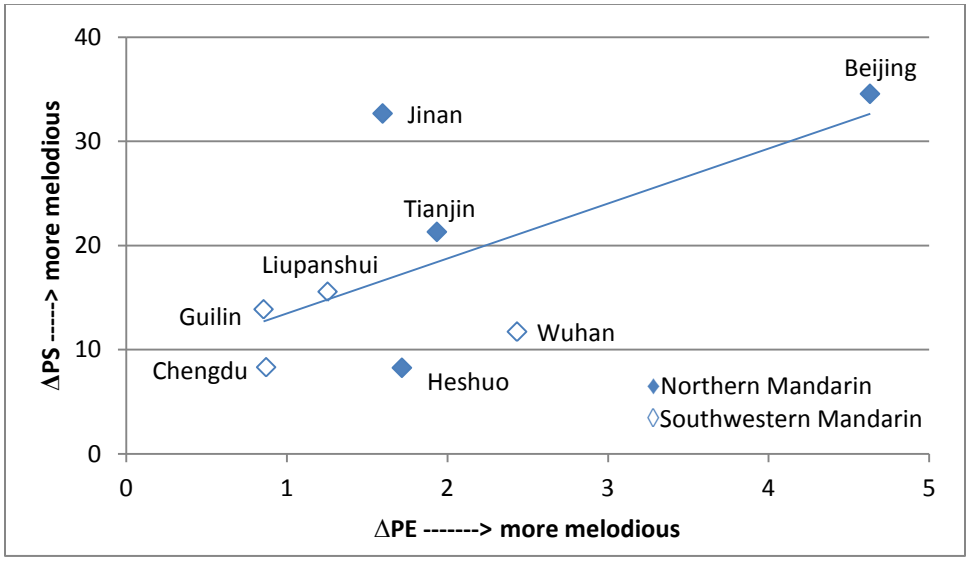


Figure 5.10 Melody pattern based on ΔPE-ΔPS (Mandarin)



5.2.2 Wu

Correlation results for Wu are listed in Table 5.8.

Table 5.8 Correlation results in the pitch-only category (Wu)

	Column	*1*	*2*	*3*	*4*
Row	*Metric*	*meanPE*	*ΔPE*	*meanPS*	*ΔPS*
1	meanPE	1			
2	ΔPE	0.81143	1		
3	meanPS	0.98537	0.869404	1	
4	ΔPS	0.985792	0.72654	0.969932	1

Both meanPE-meanPS and ΔPE-ΔPS pairs have highly positive correlations as predicted correlated (cc = 0.98537 & 0.72654, underlined in C1-R3 & C2-R4). As shown in Figure 5.11 and Figure 5.12, the degree of melodiousness can be ordered along the upward trend line for all four Wu dialects as follows: Hangzhou < Wenzhou < Suzhou < Shanghai. The two pitch-base metric pairs hence can distinguish the four Wu dialects consistently.

Figure 5.11 Melody pattern based on meanPE-meanPS (Wu)

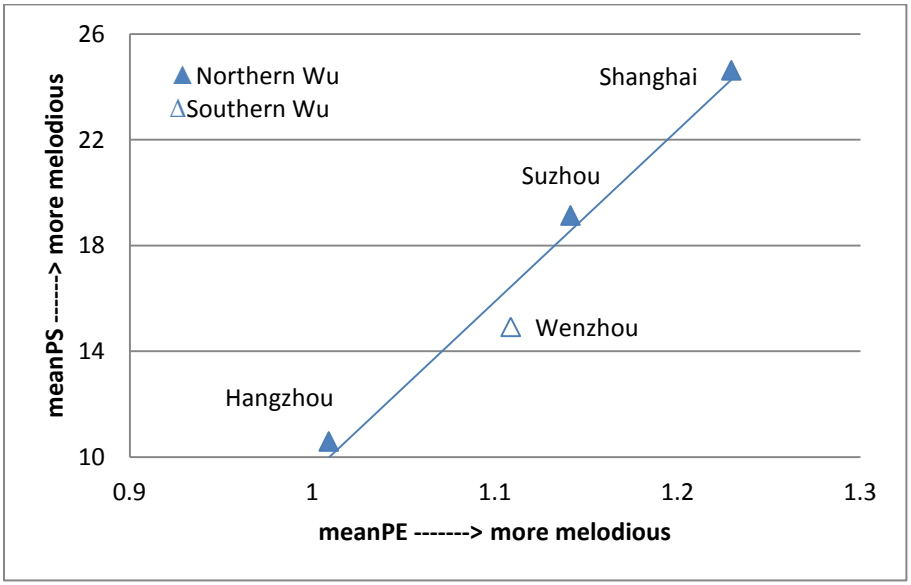
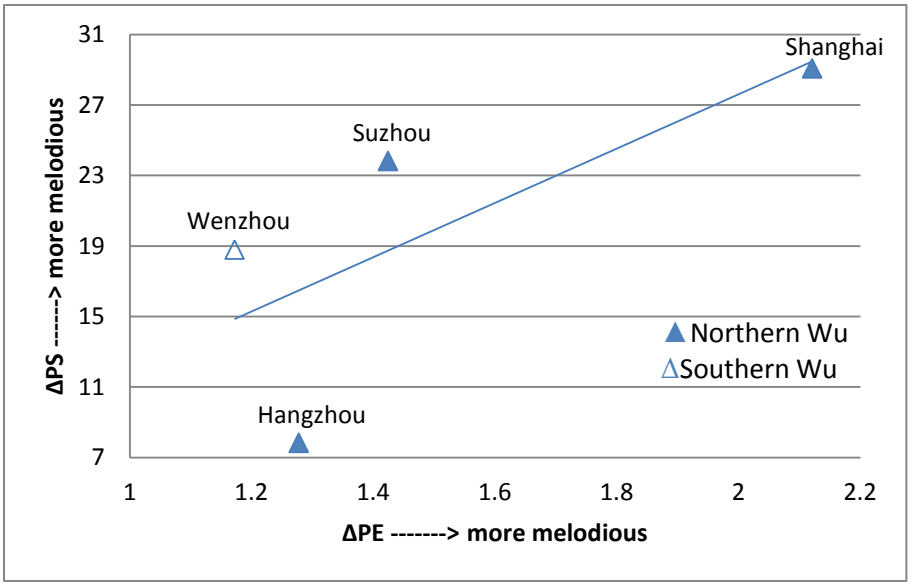


Figure 5.12 Melody pattern based on ΔPE-ΔPS (Wu)



5.2.3 Min

Correlation results for Min are listed in Table 5.9.

Table 5.9 Correlation results in the pitch-only category (Min)

	Column	*1*	*2*	*3*	*4*
Row	*Metric*	*meanPE*	*ΔPE*	*meanPS*	*ΔPS*
1	meanPE	1			
2	ΔPE	0.957534	1		
3	meanPS	0.966312	0.96887	1	
4	ΔPS	0.831442	0.830522	0.930691	1

Both meanPE-meanPS and ΔPE-ΔPS pairs have highly positive correlations as predicted (cc = 0.966312 & 0.830522, underlined in C1-R3 & C2-R4). As shown in Figure 5.13 and Figure 5.14, the Eastern dialect Fuzhou is the most melodious and the Central dialect Sanming is the least melodious along the upward trend line. For the three Southern Min dialects, the degree of melodiousness can be ordered along the upward trend line by two

groups as follows: Quanzhou < Taizhong, Shantou. In general, the four pitch-based metrics can distinguish the degree of melodiousness for Min dialects reasonably well.

Figure 5.13 Melody pattern based on meanPE-meanPS (Min)

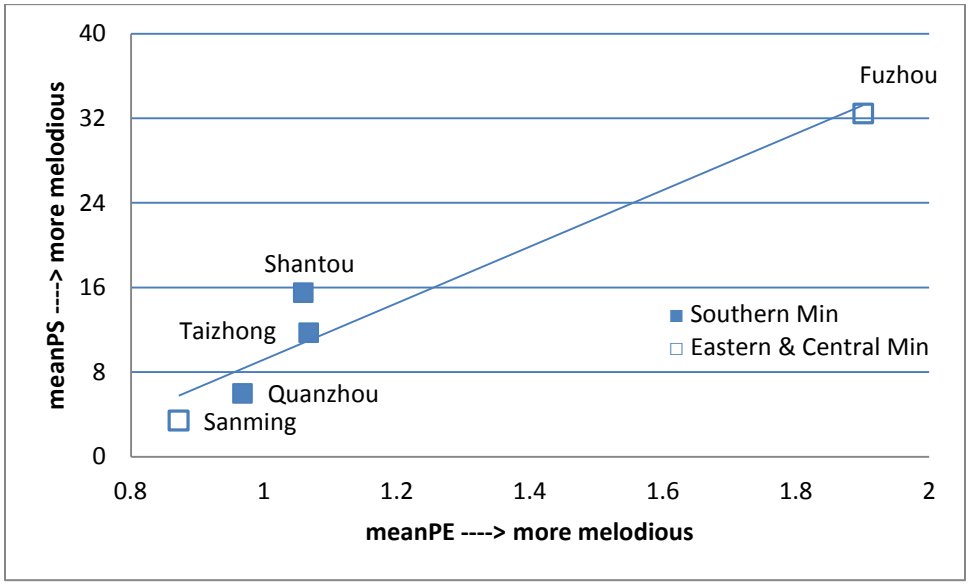
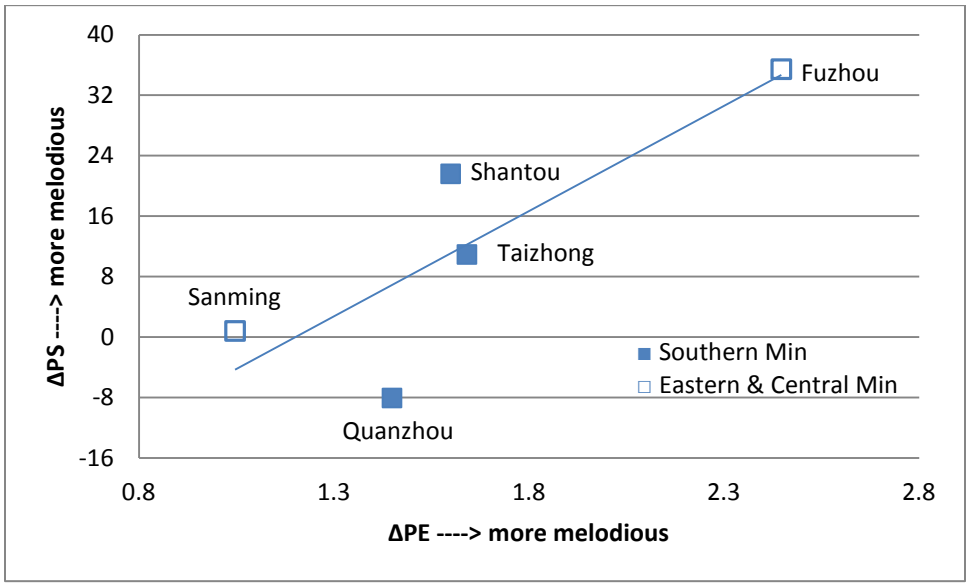


Figure 5.14 Melody pattern based on ΔPE-ΔPS (Min)



5.2.4 Cantonese

Correlation results for Cantonese are listed in Table 5.10.

Table 5.10 Correlation results in the pitch-only category (Cantonese)

	Column	*1*	*2*	*3*	*4*
Row	*Metric*	*meanPE*	*ΔPE*	*meanPS*	*ΔPS*
1	meanPE	1			
2	ΔPE	0.966991	1		
3	meanPS	0.982705	0.979946	1	
4	ΔPS	0.942265	0.981985	0.931455	1

Both meanPE-meanPS and ΔPE-ΔPS pairs have highly positive correlations as predicted (cc = 0.982705 & 0.981985, underlined in C1-R3 & C2-R4). As shown in Figure 5.15 and Figure 5.16, the degree of melodiousness is the largest for HK2, as it is at the top of the upward trend line. For the remaining three Cantonese dialects GZ1, GZ2, and HK1, they cannot be consistently ordered by the two metric pairs. Within the HK group, HK1 can be identified as less melodious than HK2 by both metric pairs, while within the GZ group, the relative melodiousness of GZ1 and GZ2 cannot be consistently distinguished. Overall, pitch-based metrics is not all successful in distinguishing the melodiousness of the Cantonese dialects within sub-groups and across sub-groups.

Figure 5.15 Melody pattern based on meanPE-meanPS (Cantonese)

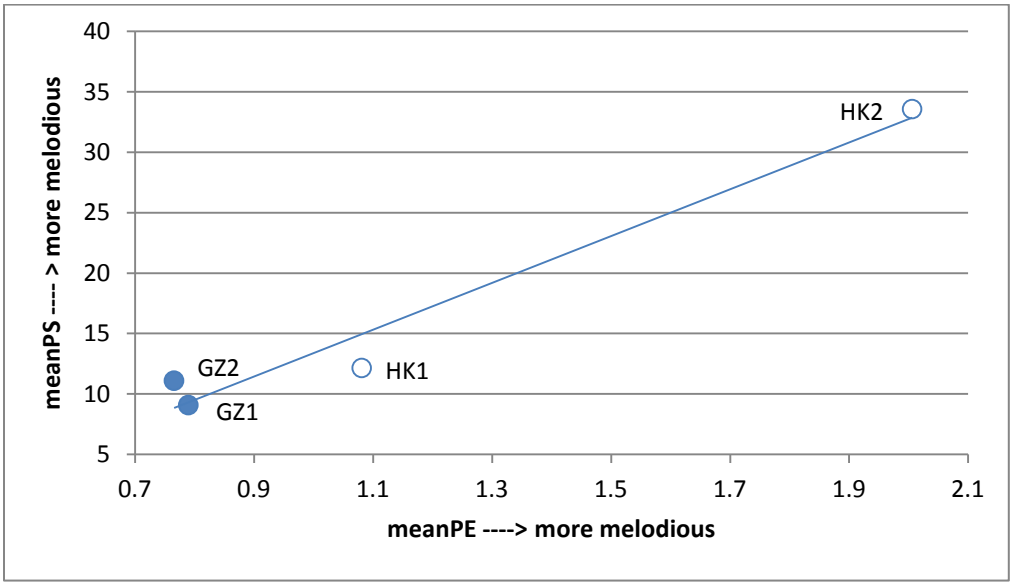
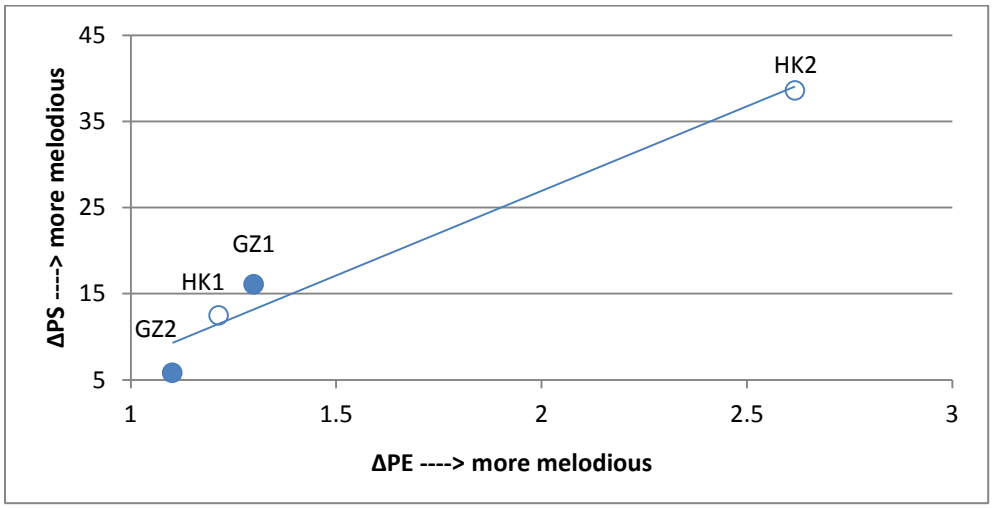


Figure 5.16 Melody pattern based on ΔPE-ΔPS (Cantonese)



5.2.5 Summary

Pitch-based melody patterns are basically consistent across all four major groups. As shown in Table 5.11, all the pitch-based metrics are positively correlated as predicted (C2-5), despite that not all of them are highly correlated. Only Wu, Min, and Cantonese have the two pitch-based metric pairs highly correlated as predicted (see C3-5). Also, the

highest correlation occurs between meanPE and meanPS for all four major groups except for Mandarin (underlined in R1). Despite some inconsistencies, the pitch-based metrics do reasonably well in distinguishing the relative melodiousness for the four major groups.

Table 5.11 Summary of correlation results in the pitch-only category (four major groups)

Row	Pitch-based metric pairs	Predicted correlation type	Mandarin	Wu	Min	Cantonese
1	meanPE-meanPS	H+	L+	H+	H+	H+
2	ΔPE-ΔPS	H+	M+	H+	H+	H+
	Column	*1*	*2*	*3*	*4*	*5*

5.3 Correlations among syllable-timing, melody, and phonological structure

The results reported in Section 5.1 and 5.2 have shown timing and melody patterns of four major Chinese dialect groups. This section presents further results to show if timing and melody patterns are related to each other as well as to phonological structure in each group and if so, how they are related. Correlation patterns between syllable-timedness and melodiousness, between syllable-timedness and syllable structure, and between melodiousness and tone structure are presented in Section 5.3.1, 5.3.2, and 5.3.3, respectively. Section 5.3.4 summarizes all the correlation patterns and key findings.

5.3.1 Correlation between syllable-timedness and melodiousness

Correlation results in the duration-pitch category are presented successively for Mandarin, Wu, Min, and Cantonese.

5.3.1.1 Mandarin

Correlation results for Mandarin are listed in Table 5.12.

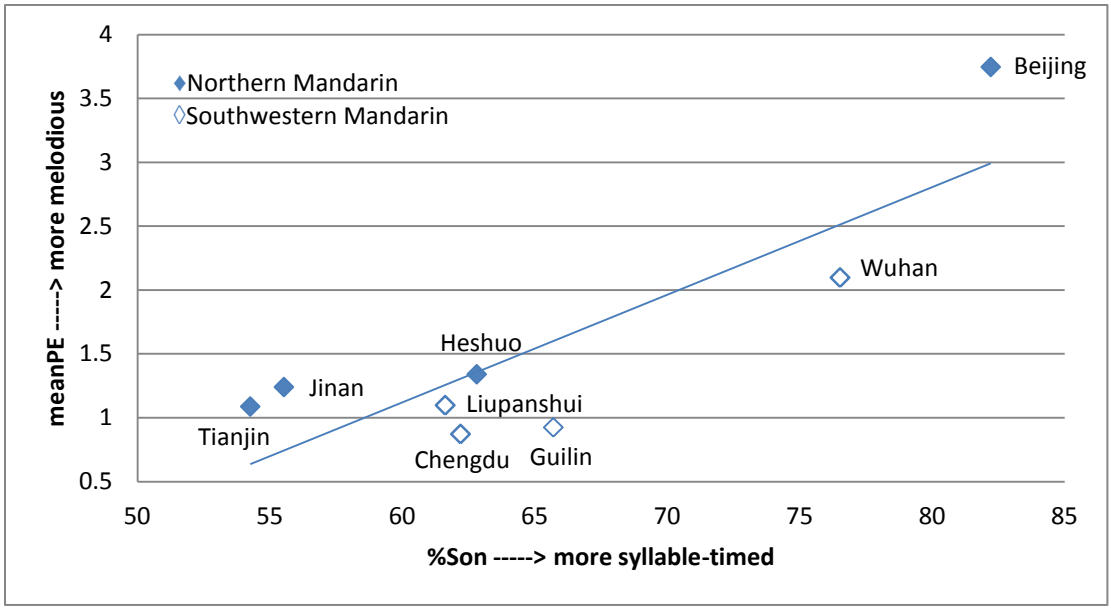
Table 5.12 Correlation results in the duration-pitch category (Mandarin)

Duration-based \ *Pitch-based*		*PE-based*		*PS-based*		
		meanPE	*ΔPE*	*meanPS*	*ΔPS*	*Row*
Son-based	*%Son*	0.845766	0.728198	0.331689	0.130747	*1*
	ΔSon	0.322776	0.375161	-0.05315	0.021174	*2*
	varcoSon	-0.03837	-0.09688	-0.32323	-0.43923	*3*
	nPVI_Son	-0.09149	-0.0477	0.200699	0.293907	*4*
IS-based	*ΔIS*	-0.70774	-0.56878	-0.30331	-0.06225	*5*
	varcoIS	-0.08846	-0.17661	-0.18673	-0.33099	*6*
	rPVI_IS	-0.61822	-0.46305	-0.22552	0.033286	*7*
	Column	*1*	*2*	*3*	*4*	

There are only five pairs correlated, all of which occur between PE- and duration-based metrics (see underlined in C1 & C2). The highest correlation occurs between %Son and meanPE (cc = 0.845766, underlined in R1-C1). The positive correlation is predicted and it means that dialects with a larger degree of syllable-timedness tend to be more melodious as well.

As an illustration, Figure 5.17 uses %Son-meanPE as the x-y axis to plot all the dialects along the syllable-timedness and melodiousness dimensions. Along the upward trend line, both the degrees of syllable-timedness and melodiousness can be ordered for the eight Mandarin dialects in four groups as follows: Tianjin, Jinan > Liupanshui, Chengdu, Heshuo, Guilin > Wuhan > Beijing. A comparison of the two sub-dialect groups shows a tendency for Southwestern dialects to be less melodious than northern dialects, as the former are all below the trend line but the latter on or above it.

Figure 5.17 Correlation pattern based on %Son-meanPE (Mandarin)



5.3.1.2 Wu

Correlation results for Wu are listed in Table 5.13.

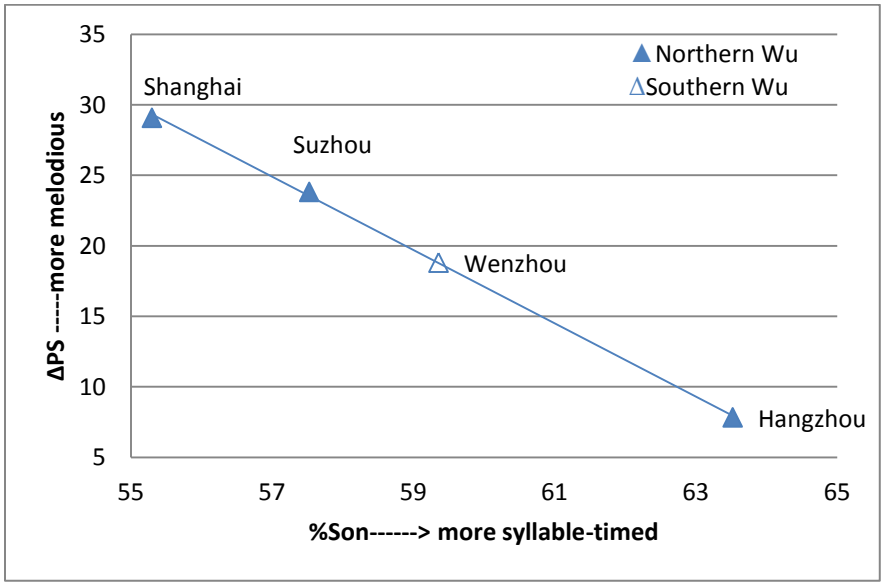
Table 5.13 Correlation results in the duration-pitch category (Wu)

Duration-based \ *Pitch-based*		*PE-based*		*PS-based*		
		meanPE	*ΔPE*	*meanPS*	*ΔPS*	*Row*
Son-based	*%Son*	-0.98972	-0.74033	-0.97389	**-0.99968**	*1*
	ΔSon	-0.63497	-0.2697	-0.65425	-0.75105	*2*
	varcoSon	-0.72091	-0.21635	-0.66989	-0.8219	*3*
	nPVI_Son	-0.29599	-0.34852	-0.42767	-0.3731	*4*
IS-based	*ΔIS*	0.793165	0.646521	0.708163	0.713292	*5*
	varcoIS	0.623946	0.805357	0.604204	0.484924	*6*
	rPVI_IS	0.962315	0.794408	0.918087	0.915487	*7*
	Column	*1*	*2*	*3*	*4*	

All 28 pairs but seven pairs (underlined) are correlated. The highest correlation occurs between %Son and ΔPS (cc = -0.99968, bolded in C4-R1). The negative correlation is not predicted as it means that dialects with a larger degree of syllable-timedness tend to be less melodious rather than more melodious.

Figure 5.18 uses %Son-ΔPS as the x-y axis to plot all the dialects along the syllable-timedness and melodiousness dimensions. Along the upward trend line, the degree of syllable-timedness increases and melodiousness decreases clearly from Shanghai, Suzhou, Wenzhou, to Hangzhou.

Figure 5.18 Correlation pattern based on %Son-ΔPS (Wu)



5.3.1.3 Min

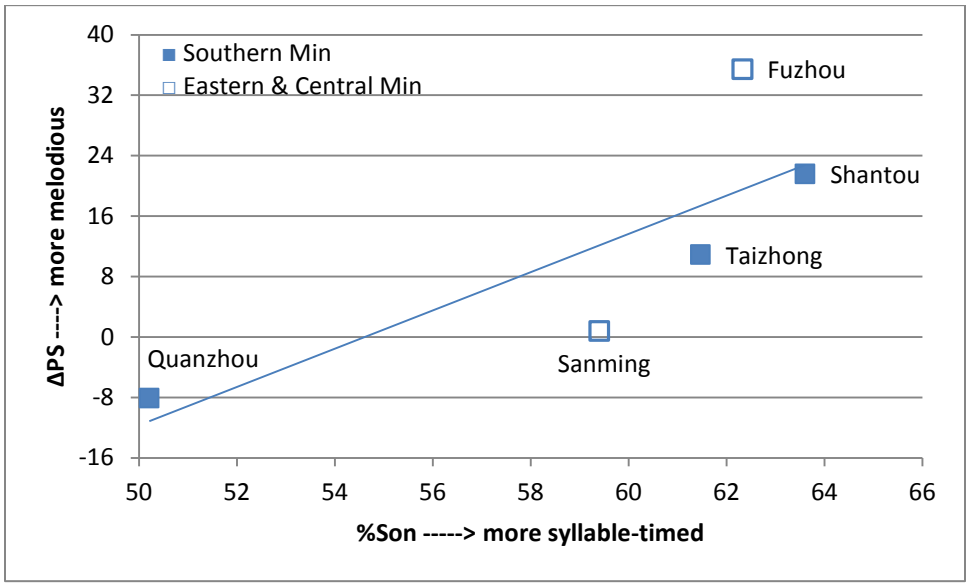
Correlation results for Min are listed in Table 5.14.

Table 5.14 Correlation results in the duration-pitch category (Min)

Duration-based \ *Pitch-based*		*PE-based*		*PS-based*		
		meanPE	*ΔPE*	*meanPS*	*ΔPS*	*Row*
Son-based	*%Son*	0.37312	0.36056	0.526626	0.793464	*1*
	ΔSon	0.531811	0.316863	0.520328	0.644112	*2*
	varcoSon	0.566003	0.401462	0.52514	0.629132	*3*
	nPVI_Son	0.515355	0.435602	0.498606	0.59687	*4*
IS-based	*ΔIS*	0.324725	0.296056	0.230216	0.194658	*5*
	varcoIS	0.369437	0.15091	0.152055	0.028135	*6*
	rPVI_IS	0.274848	0.098675	0.116604	0.102017	*7*
	Column	*1*	*2*	*3*	*4*	

There are ten pairs correlated, all of which occur between Son- and pitch-based metrics (see underlined). The highest and also the only high correlation occurs between %Son and ΔPS (cc = 0.793464 > 0.5; underlined in R1-C4). The positive correlation is predicted. Figure 5.19 uses %Son-ΔPS as the x-y axis to plot all the dialects along the syllable-timedness and melodiousness dimensions. Along the upward trend lines, the degrees of syllable-timedness and melodiousness can both be ordered for the five Min dialects in four groups: Quanzhou < Sanming < Taizhong < Shantou, Fuzhou. A comparison of the sub-dialect groups does not show any clear pattern.

Figure 5.19 Correlation pattern based on %Son-ΔPS (Min)



5.3.1.4 Cantonese

Correlation results for Cantonese are listed in Table 5.15.

Table 5.15 Correlation results in the duration-pitch category (Cantonese)

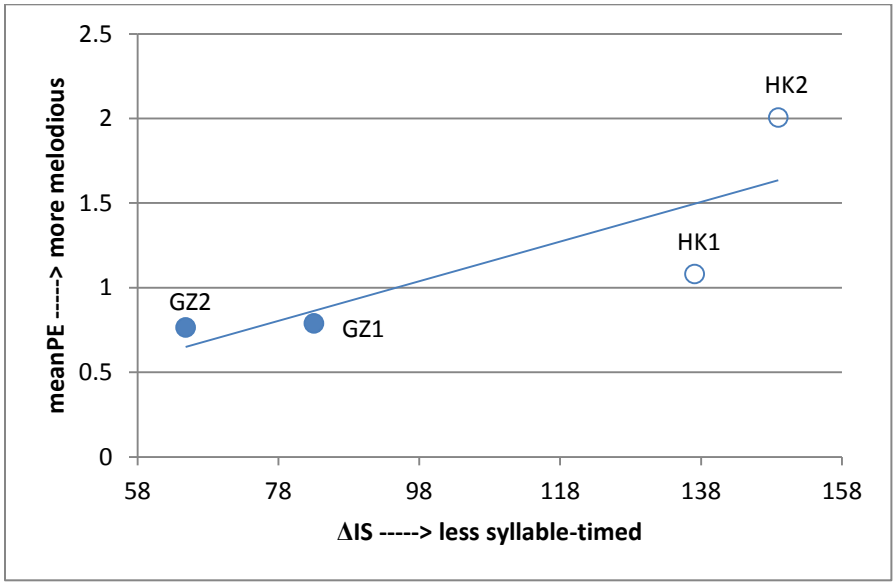
Duration-based \ *Pitch-based*		*PE-based*		*PS-based*		
		meanPE	*ΔPE*	*meanPS*	*ΔPS*	*Row*
Son-based	*%Son*	0.590323	0.764318	0.629137	0.818425	*1*
	ΔSon	-0.69164	-0.54922	-0.70404	-0.41873	*2*
	varcoSon	0.37952	0.404195	0.258441	0.568819	*3*
	nPVI_Son	0.444459	0.47262	0.329682	0.629855	*4*
IS-based	*ΔIS*	0.822248	0.683503	0.704695	0.720918	*5*
	varcoIS	0.226753	0.208805	0.077963	0.383992	*6*
	rPVI_IS	-0.25999	-0.31502	-0.41885	-0.14015	*7*
	Column	*1*	*2*	*3*	*4*	

There are 13 pairs correlated (see underlined). The highest correlation occurs between ΔIS and meanPE (cc = 0.822248; underlined in R1-C4), followed by the one

between %Son and ΔPS (cc = 0.818425; underlined in R1-C4). Both the correlations are positive as predicted.

Figure 5.20 uses ΔIS-meanPE as the x-y axis to plot all the dialects along the syllable-timedness and melodiousness dimensions. In the two figures, the correlation patterns are inconsistent and even conflicting both within and across sub-groups: in the first figure, GZ1 and GZ2 are more syllable-timed but less melodious than HK1 and HK2; while in the second figure, GZ1 becomes less syllable-timed than HK2 but more melodious than HK1. Despite some correlations found between duration- and pitch-based metrics, timing and melody patterns cannot be consistently shown for Cantonese.

Figure 5.20 Correlation patterns based on ΔIS-meanPE (Cantonese)



5.3.2 Correlation between syllable-timing and syllable structure

Correlation results in the duration-syllable category and the associated correlation patterns in the form of scatter charts are presented successively for Mandarin, Wu, and Min in Section 5.3.2.1, 5.3.2.2, and 5.3.2.3. Note that no correlation analysis was

performed for Cantonese data, as the four Cantonese dialects are very similar in syllable structure.

5.3.2.1 Mandarin

Correlation results for Mandarin are listed in Table 5.16.

Table 5.16 Correlation results in the duration-syllable category (Mandarin)

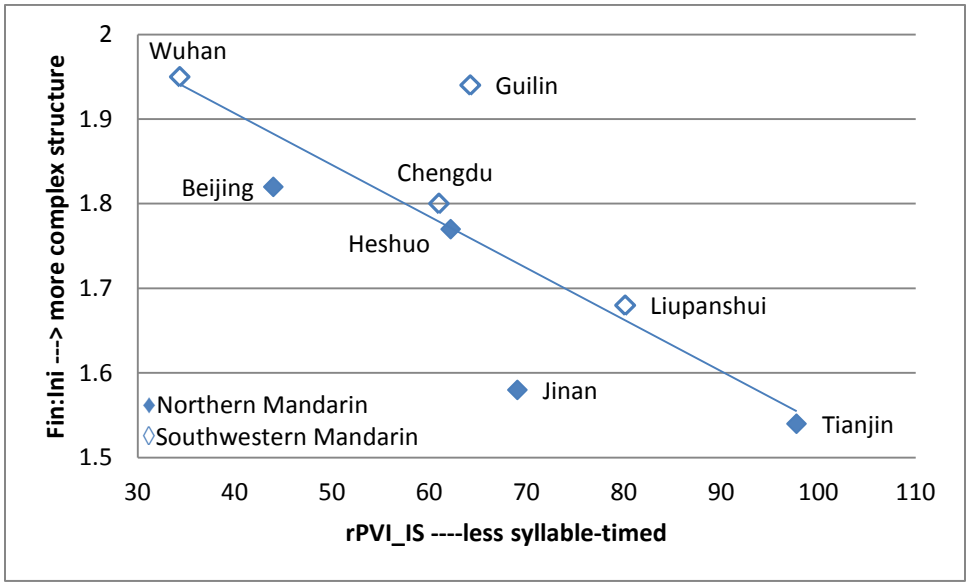
Syllable-based / *Duration-based*		*Fin:Ini*	*sumFI*	*Row*
Son-based	*%Son*	0.733312	-0.0163	*1*
	ΔSon	0.164954	-0.23037	*2*
	varcoSon	0.426646	-0.24789	*3*
	nPVI_Son	-0.03602	0.15458	*4*
IS-based	*ΔIS*	-0.70703	-0.15357	*5*
	varcoIS	0.318394	-0.32958	*6*
	rPVI_IS	-0.79197	-0.02043	*7*
	Column	*1*	*2*	

There are only three pairs correlated, all of which occur between Fin:Ini and duration-based metrics(see underlined). No correlation between sumD and syllable-timing. The highest correlation occurs between rPVI_IS and Fin:Ini (cc = -0.79197, underlined in R7-C1). The negative correlation is not predicted as it means that dialects with a larger Fin:Ini tend to have a smaller rPVI_IS and hence be more rather than less syllable-timed.

Figure 5.21 uses rPVI_IS-Fin:Ini as the x-y axis to plot all the dialects along the syllable structure complexity and syllable-timedness dimensions. For the four Northern Mandarin dialects, the degree of syllable-timedness decreases as Fin:Ini decreases from Beijing, Heshuo, Jinan, to Tianjin. For the four Southwestern Mandarin dialects except for Guilin, the degree of syllable-timedness decreases as Fin:Ini decreases from Wuhan,

Chengdu, to Liupanshui. Guilin has a little smaller Fin:Ini but is much less syllable-timed than Wuhan. A comparison of Northern and Southwestern Mandarin dialects indicates that Southwestern Mandarin dialects tend to have a larger Fin:Ini and larger degree of syllable-timedness than Northern Mandarin dialects except Beijing. In general, dialects with a larger number of finals than initials tend to be more syllable-timed.

Figure 5.21 Correlation pattern based on rPVI_IS-Fin:Ini (Mandarin)



5.3.2.2 Wu

Correlation results for Wu are listed in Table 5.17.

Table 5.17 Correlation results in the duration-syllable category (Wu)

Duration-based \ *Syllable-based*		*Fin:Ini*	*sumFI*	*Row*
Son-based	*%Son*	-0.55321	-0.34916	*1*
	ΔSon	-0.85286	-0.67392	*2*
	varcoSon	-0.56789	-0.31217	*3*
	nPVI_Son	-0.9716	-0.99715	*4*
IS-based	*ΔIS*	-0.15197	-0.32681	*5*
	varcoIS	-0.19311	-0.22826	*6*
	rPVI_IS	0.226199	0.038578	*7*
	Column	*1*	*2*	

There are six pairs correlated (see underlined). The highest correlation occurs nPVI_Son and sumFI (cc = -0.99715; underlined in R4-C2), followed by the one between nPVI_Son and Fin:Ini (cc = -0.9716; underlined in R4-C1). The negative cc value here is not predicted because it means that dialects with smaller sumFI and Fin:Ini tend to be less rather than more syllable-timed.

Figure 5.22 and Figure 5.23 respectively use nPVI_Son-sumFI and nPVI_Son-Fin:Ini as the x-y axis to plot all the dialects along the syllable structure complexity and syllable-timedness dimensions. Since the two figures consistently show that the degree of syllable-timedness decreases from Suzhou, Shanghai, Hangzhou, to Wenzhou as sumFI and Fin:Ini decreases, syllable structure may be a good indicator of syllable-timedness for Wu.

Figure 5.22 Correlation pattern based on nPVI_Son-sumFI (Wu)

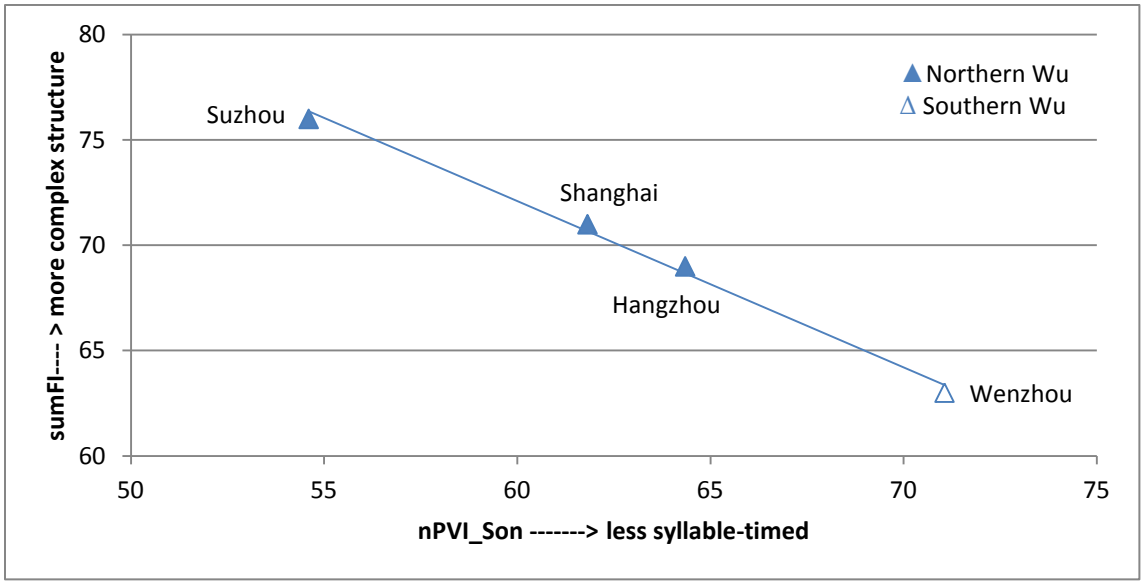
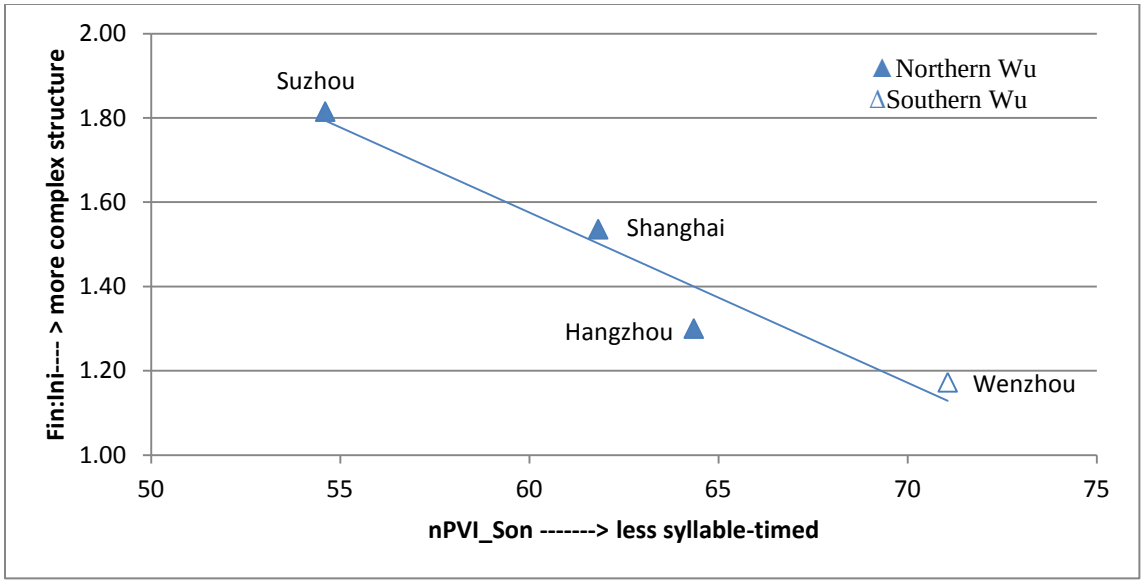


Figure 5.23 Correlation pattern based on nPVI_Son-Fin:Ini (Wu)



5.3.2.3 Min

Correlation results for Min are listed in Table 5.18.

Table 5.18 Correlation results in the duration-syllable category (Min)

Duration-based \ *Syllable-based*		*Fin:Ini*	*sumFI*	*Row*
Son-based	*%Son*	-0.50407	-0.29141	*1*
	ΔSon	-0.94877	-0.88489	*2*
	varcoSon	-0.73567	-0.75331	*3*
	nPVI_Son	-0.4549	-0.48072	*4*
IS-based	*ΔIS*	-0.04603	-0.20755	*5*
	varcoIS	-0.50604	-0.75021	*6*
	rPVI_IS	-0.48163	-0.6564	*7*
	Column	*1*	*2*	

There are eight pairs correlated (see underlined). The highest correlation occurs between ΔSon and Fin:Ini (cc = -0.94877, underlined in R2-C1), followed by the one between ΔSon and sumFI (cc = -0.88489, underlined in R2-C2) . The negative correlation means that dialects with larger Fin:Ini and sumFI (hence more complex syllable structure) tend to be more syllable-timed and this pattern is opposite to what is predicted.

Figure 5.24 and 5.25 respectively use ΔSon-Fin:Ini and ΔSon-sumFI as the x-y axis to plot all the dialects along the syllable structure complexity and syllable-timedness dimensions. For the three Southern Min dialects, the degree of syllable-timedness decreases as Fin:Ini and sumFI decreases from Quanzhou, Taizhong, to Shantou, though their sumFI differences are small. All the Southern Min dialects have larger Fin:Ini and sumFI and are more syllable-timed than the Central and Eastern Min dialects. The relative syllable-timedness of the Central Min dialect Sanming and the Eastern Min dialect Fuzhou, on the other hand, shows the opposite pattern: Sanming has less Fin:Ini

and sumFI than Fuzhou and is more syllable-timed. On the whole, the correlation pattern between syllable-timing and syllable structure is different for Min at the sub-dialectal level.

Figure 5.24 Correlation pattern based on ΔSon-Fin:Ini (Min)

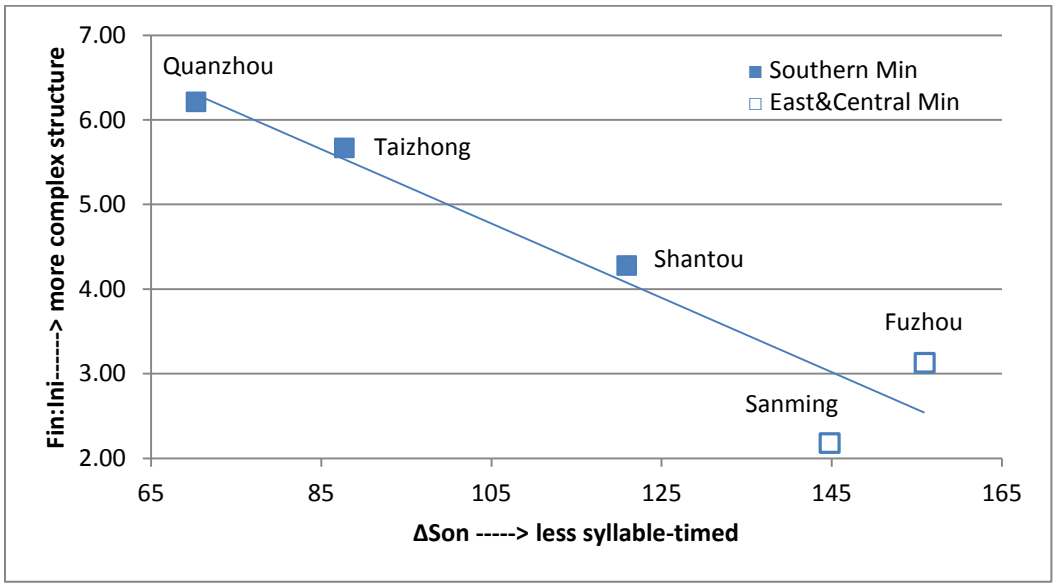
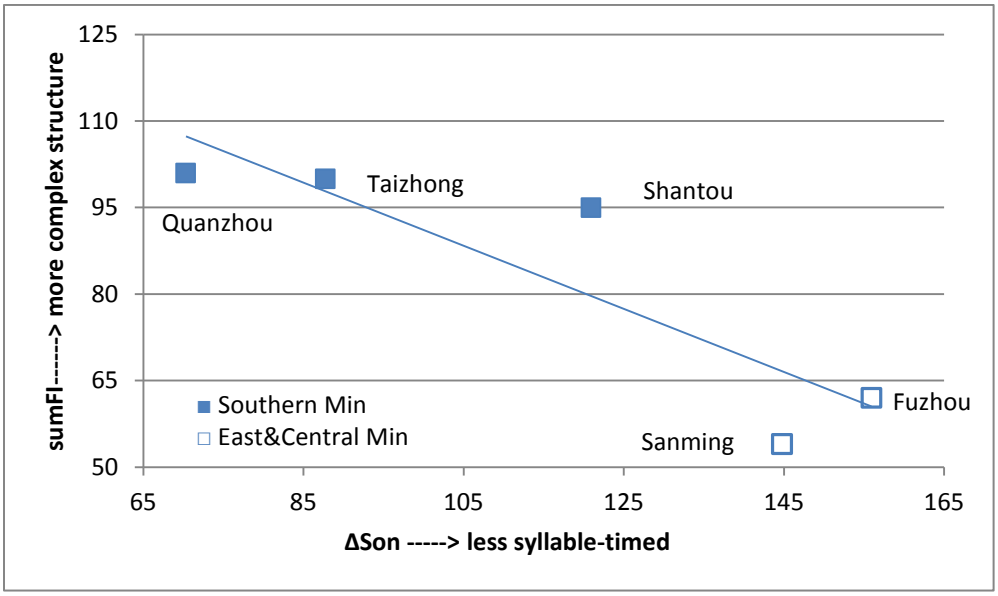


Figure 5.25 Correlation pattern based on ΔSon-sumFI (Min)



5.3.3 Correlation between melody and tone structure

Correlation results in the pitch-tone category and the associated correlation patterns are presented successively for Mandarin, Wu, and Min in Section 5.3.3.1, 5.3.3.2, and 5.3.3.3. Note that no correlation analysis was performed for Cantonese data, as the four Cantonese dialects are very similar in tone structure.

5.3.3.1 Mandarin

Correlation results for Mandarin are listed in Table 5.19.

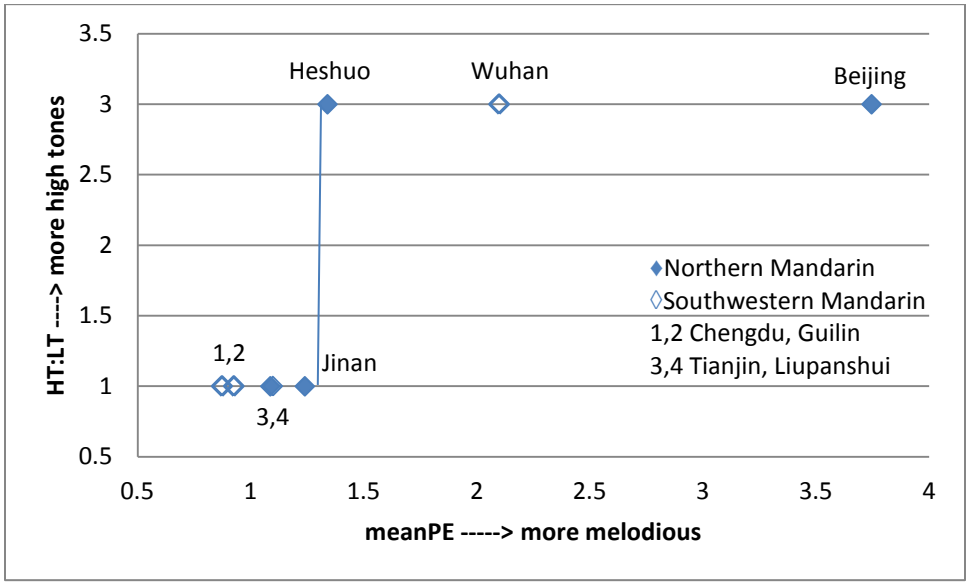
Table 5.19 Correlation results in the pitch-tone category (Mandarin)

Pitch-based / *Tone-based*	*PE-based*		*PS-based*		
	meanPE	*ΔPE*	*meanPS*	*ΔPS*	*Row*
HT:LT	0.723442	0.688676	0.087481	-0.0085	*1*
Column	*1*	*2*	*3*	*4*	

Only two pairs are correlated (see underlined). The only high correlation occurs between meanPE and HT:LT (cc = 0.723442 > 0.7; underlined in C1-R1). Note that the correlation analysis is not applicable to sumT, as all the Mandarin dialects have the same number of tones (sumT = 4). The positive correlation is predicted and it means that dialects with a larger HT:LT tend to be more melodious. Figure 5.26 uses meanPE-HT:LT as the x-y axis to plot all the dialects along the tone structure complexity and melodiousness dimensions. Chengdu, Guilin, Liupanshui, Tianjin, and Jinan have the same low HT:LT (= 1) and similarly small degrees of melodiousness. In contrast, Heshuo, Wuhan, and Beijing have the same high HT:LT (= 3), but Heshuo is close to Jinan in melodiousness (see the vertical line). A comparison of Southwestern and

Northern Mandarin does not show any particular melody pattern at the sub-dialectal level.

Figure 5.26 Correlation pattern based on meanPE-HT:LT (Mandarin)



5.3.3.2 Wu

Correlation results for Wu are listed in Table 5.20.

Table 5.20 Correlation results in the pitch-tone category (Wu)

Pitch-based / *Tone-based*	*PE-based*		*PS-based*		
	meanPE	*ΔPE*	*meanPS*	*ΔPS*	*Row*
HT:LT	0.96413	0.794755	<u>0.984715</u>	0.977294	*1*
sumT	-0.66378	<u>-0.97357</u>	-0.75204	-0.56924	*2*
Column	*1*	*2*	*3*	*4*	

All eight pairs are correlated (|cc| > 0.5), six of which are also highly correlated (cc > 0.7; see R1 & R2-C2-3). The highest correlation occurs between meanPS and HT:LT (cc = 0.984715; underlined in C3-R1). The positive correlation is predicted and it means that dialects with more high than low tones tend to be more melodious. The correlation between ΔPE and sumT is also high (cc = -0.97357; underlined in C2-R2). The negative

correlation is also predicted and it means that dialects with more tones tend to be less melodious.

Figure 5.27 and Figure 5.28 respectively use meanPS-HT:LT and ΔPE-sumT as the x-y axis to plot all the dialects along the tone structure complexity and melodiousness dimensions. The degree of melodiousness increases as HT:LT increases from Hangzhou, Wenzhou, Suzhou, to Shanghai. The degree of melodiousness increases as sumT decreases from Wenzhou, Suzhou, to Shanghai. Hangzhou has the same number of tones as Suzhou but is a little less melodious than Suzhou. Overall, both HT:LT and sumT are good indicators of the relative melodiousness for Wu dialects.

Figure 5.27 Correlation pattern based on meanPS-HT:LT (Wu)

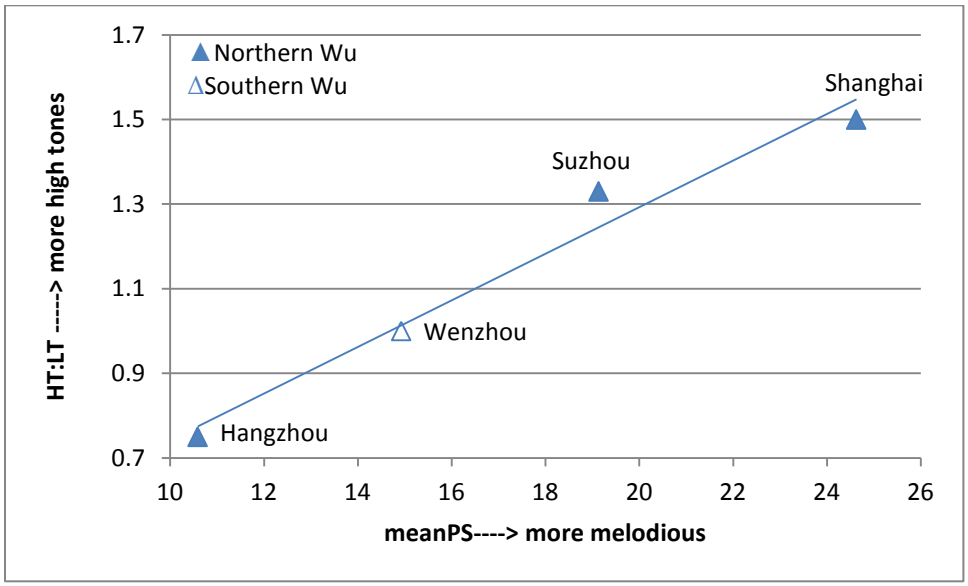
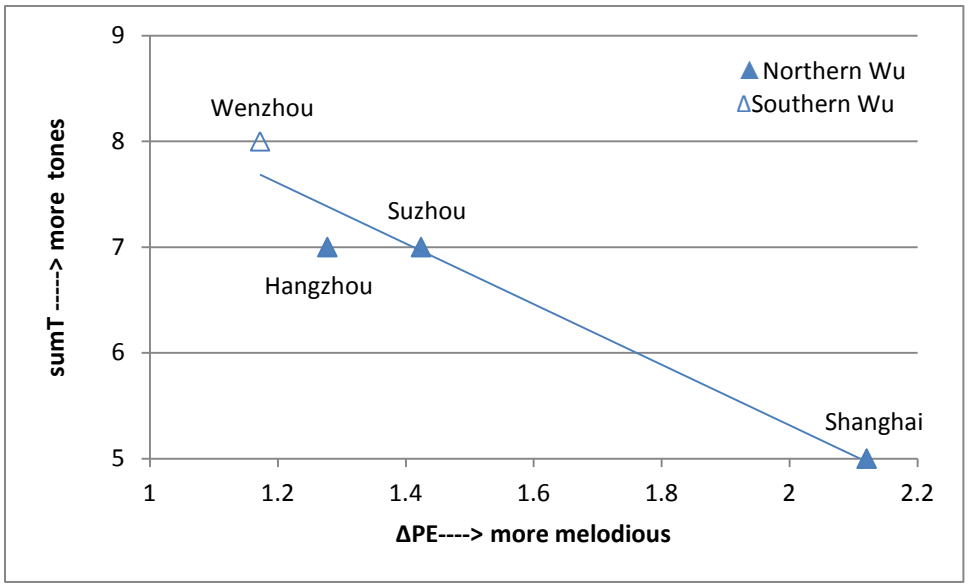


Figure 5.28 Correlation pattern based on ΔPE-sumT (Wu)



5.3.3.3 Min

Correlation results for Min are listed in Table 5.21.

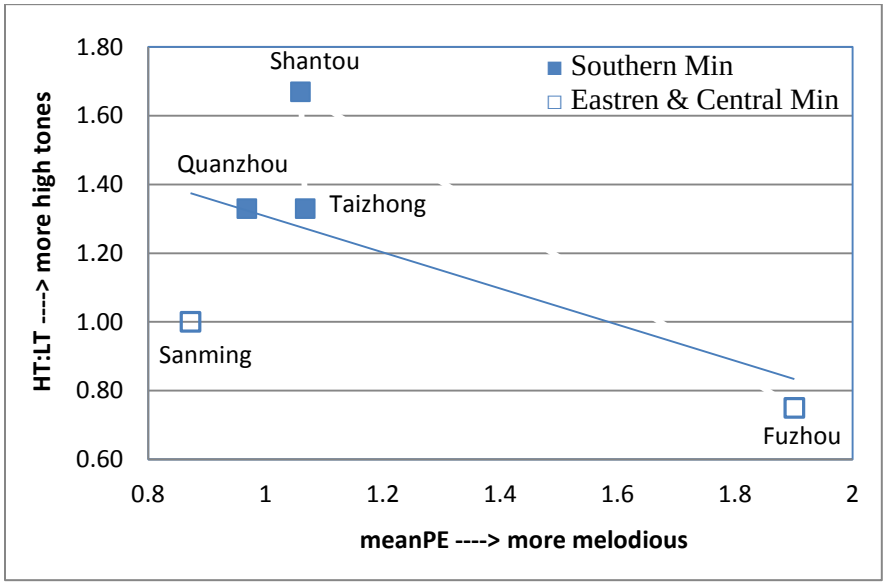
Table 5.21 Correlation results in the pitch-tone category (Min)

Pitch-based / *Tone-based*	*PE-based*		*PS-based*		
	meanPE	*ΔPE*	*meanPS*	*ΔPS*	*Row*
HT:LT	-0.61866	-0.40012	-0.42132	-0.27243	*1*
sumT	0.160312	0.383553	0.374078	0.429452	*2*
Column	*1*	*2*	*3*	*4*	

Only a moderate, negative correlation occurs between meanPE and HT:LT (cc = -0.61866; underlined in R1-C1). The negative correlation is not predicted as it means that dialects with a larger HT:LT tend to be less rather than more melodious. Figure 5.29 uses meanPE-HT:LT as the x-y axis to plot all the dialects along the tone structure complexity

and melodiousness dimensions. No particular melody pattern is shown in the figure, except that Fuzhou has the lowest HT:LT but largest degree of melodiousness.

Figure 5.29 Correlation pattern based on meanPE-HT:LT (Min)



5.3.4 Summary

Various correlation patterns emerge for the four major Chinese dialect groups. The most noticeable pattern is that not many metric pairs are correlated as predicted; it is often the case that either the correlations are not as high as predicted, or the direction of their correlations is opposite to what is predicted, or both.

Table 5.22 summarizes the number of correlated pairs and all the metrics pairs with the highest correlation in the same correlation category for all the applicable major dialect groups. The actual number of correlated pairs in each category in relation to the predicted number of correlated pairs (Actual/Predicted#; see C1) shows how well all the metric pairs correlate in each major group: Wu has the most correlated pairs in relation to the predicted pairs (37/50), followed successively by Cantonese (13/28), Min (20/50), and Mandarin (10/46). Wu is also the only major group with all eight metric pairs in the

pitch-tone category positively correlated as predicted (see C4-R4), except that not all of the correlations are as high as predicted.

In the duration-pitch and pitch-tone categories (see C2 & C4), %Son and HT:LT occurs in all the highest correlated pairs, suggesting that they are good indicators of syllable-timedness and melodiousness, respectively. In addition, meanPE occurs the most (3 times; see R2-C2, R2-C4, & R6-C4) among the highest correlated pairs in the syllable-pitch and pitch-tone categories, suggesting that it is the best indicator of melodiousness. In the duration-syllable category, the best performer is Fin:Ini, as it occurs the most (twice; see R2-C3 & R6-C3).

Generally, not a single correlation pattern is shared among all four major groups and not a single metric pair has the highest correlation across all four major groups. Nonetheless, %Son, meanPE, Fin:Ini, and HT:LT perform better than the rest of the metrics in their respective correlation categories.

Table 5.22 Summary of correlation results in duration-pitch, duration-syllable, and pitch-tone categories (four major groups)

Major group	*Correlation category→*	*Duration-Pitch*	*Duration-Syllable*	*Pitch-Tone*	*Row*
Mandarin	Actual/ Predicted # (10/46)	5/28	3/14	2/4	*1*
	Pair with highest correlation	%Son-meanPE (H+)	rPVI_IS-Fin:Ini (H−)	meanPE-HT:LT (H+)	*2*
Wu	Actual/ Predicted # (37/50)	21/28	6/14	8/8	*3*
	Pair with highest correlation	%Son-ΔPS (H−)	nPVI_Son-sumFI (H−)	meanPS-HT:LT (H+)	*4*
Min	Actual/ Predicted # (20/50)	10/28	8/14	1/8	*5*
	Pair with highest correlation	%Son-ΔPS (H+)	ΔSon-Fin:Ini (H−)	meanPE-HT:LT (M−)	*6*
Cantonese	Actual/ Predicted # (13/28)	13/28	-	-	*7*
	Pair with highest correlation	%Son-ΔPE (H+)	-	-	*8*
Column	*1*	*2*	*3*	*4*	

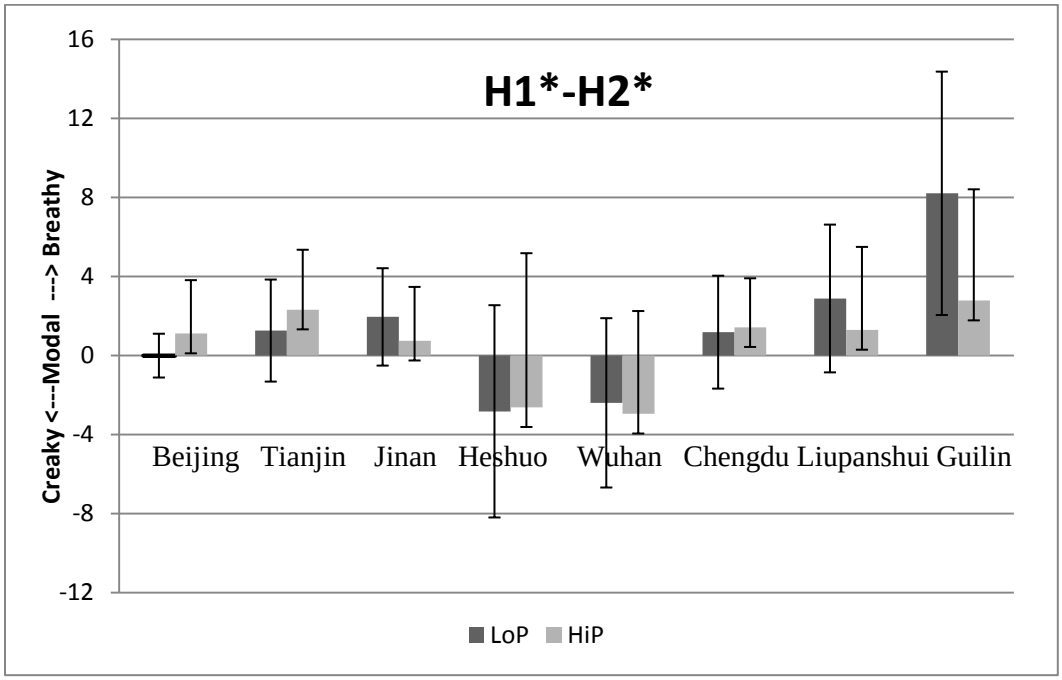
5.4 Voice source results

Voice source results are presented successively for Mandarin, Wu, Min, and Cantonese in Section 5.4.1, 5.4.2, 5.4.3, and 5.4.4. Section 5.4.5 summarizes the results.

5.4.1 Mandarin

H1*-H2* and CPP results for the eight Mandarin dialects are listed in Appendix 8. Based on H1*-H2* results, a comparison of voice quality is made between LoP and HiP across eight Mandarin dialects in the form of a column graph (see Figure 5.30).

Figure 5.30 Comparison of H1*-H2* between LoP and HiP (Mandarin)



Since a breathy voice has a positively larger H1*-H2* value while a creaky voice has a negatively larger H1*-H2* value than a modal voice, a tall column means a possible breathy voice if it is larger than 0dB (0 on the y-axis) and a possible creaky voice if it is below 0dB. Note that the word "possible" used here because the H1*-H2* difference fluctuates around 0dB for modal voice and it may take at least the 4dB difference for a voice to be distinguished between modal and breathy (Esposito, 2006). In general, the height difference between adjacent blue and red columns indicates the possible voice quality difference between LoP and HiP in a dialect. The error bar on each column is the standard deviation and the longer the bar, the larger the standard deviation.

The most noticeable pattern here is that the LoP production is rather breathier than the HiP production in Guilin, as the blue column is much higher than red one. It is very

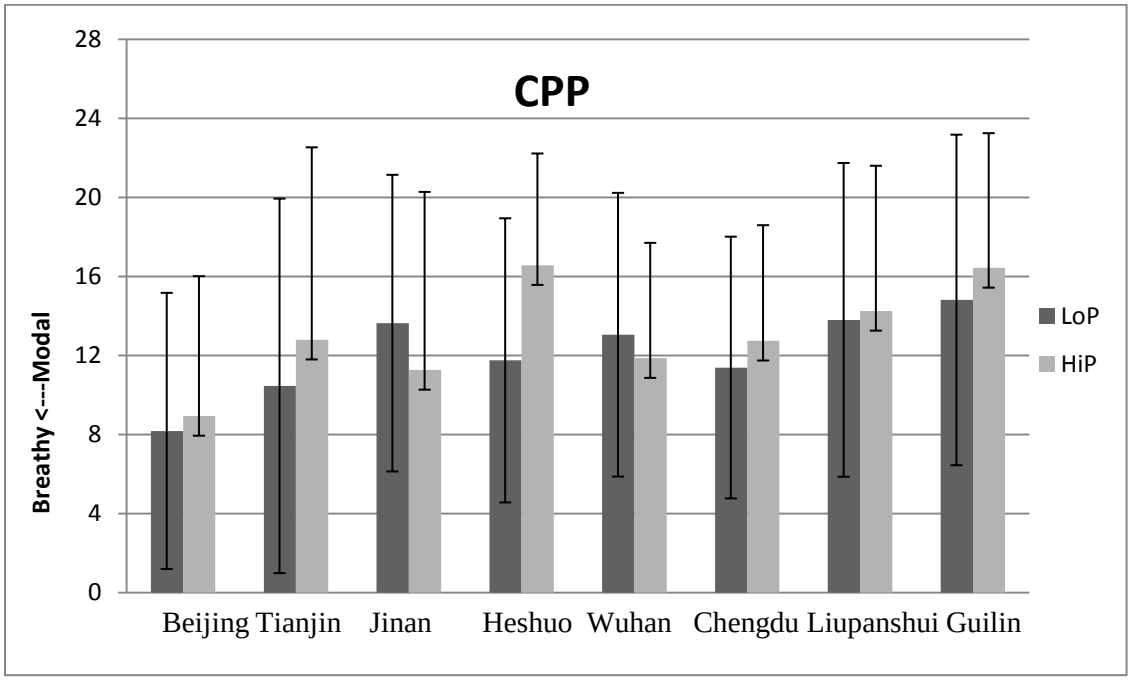
likely that breathiness is a phonetic feature associated with the Guilin speaker's low tone production.

Two other noticeable patterns are the slight creakiness of LoP in Beijing (see the shortest blue column below 0dB) and the overall creakiness in Heshuo and Wuhan (see the four middle columns below 0dB). The former pattern is in line with the previous finding that Mandarin low tone production tends to be creaky (Davison, 1991; Belotel-Grenié & Grenié, 2004; Keating et al., 2012; Belotel-Grenié & Grenié, 2004). The latter pattern, on the other hand, suggests that the two speakers in Heshuo and Wuhan may happen to have a creaky voice.

For the remaining four dialects, Tianjin, Jinan, Chengdu, and Liupanshui, LoP is breathier than HiP in the former two but not in the latter two dialects. However, it is also possible that both the LoP and HiP productions can be still considered modal for the four dialects, as all of their H1*-H2* differences are less than 2dB.

Based on CPP results, a comparison of voice quality is also made between LoP and HiP across eight Mandarin dialects in the form of a column graph (see Figure 5.31).

Figure 5.31 Comparison of CPP between LoP and HiP (Mandarin)



Since a breathy voice has a lower cepstral peak than a modal voice, the closer to 0dB a column, the breathier the voice is. The most noticeable pattern here is that LoP is rather breathier than HiP in Heshuo, as the blue column is much shorter than the red one.

For the rest of the dialects, the CPP difference between LoP and HiP is within or around ±2dB, suggesting that most LoP and HiP productions are close in voice quality.

A comparison of the two figures shows some discrepancies in the identification of voice quality: In the Heshuo dialect, for example, both the LoP and HiP production show similar voice quality (towards creakiness) by the H1*-H2* measure, whereas LoP is much breathier than HiP by the CPP measure. Guilin represents another case: its LoP is much breathier than HiP by the H1*-H2* measure but not so much by the CPP measure. A possible explanation for these discrepancies is suggested in Keating and Garellek

(2005): the same voice quality can have a variety of acoustic properties, but not all of them are present and measurable. What is most likely to happen in natural speech is that different voice qualities co-exist, so voice source measures may be biased towards the one they measure while the actual acoustic effect is from a combination of different acoustic properties. Therefore, the two voice source measures do not necessarily correlate, especially in uncontrolled speech.

5.4.2 Wu

H1*-H2* and CPP results for four Wu dialects are listed in Appendix 9. The results are illustrated in Figure 5.32 and Figure 5.33, respectively.

Figure 5.32 Comparison of H1*-H2* between LoP and HiP (Wu)

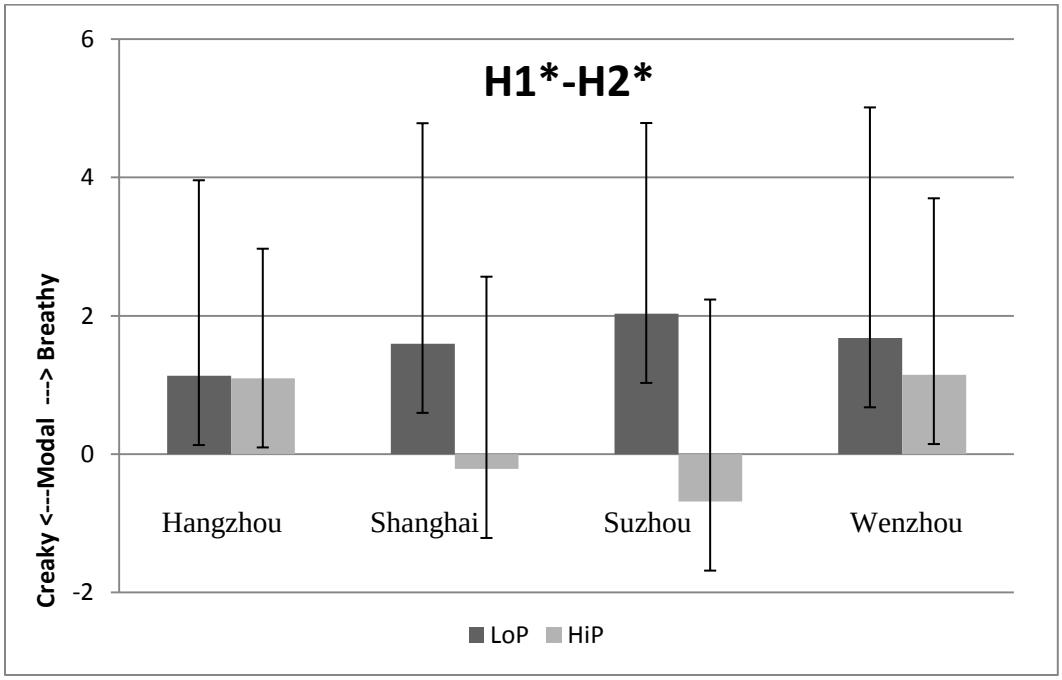
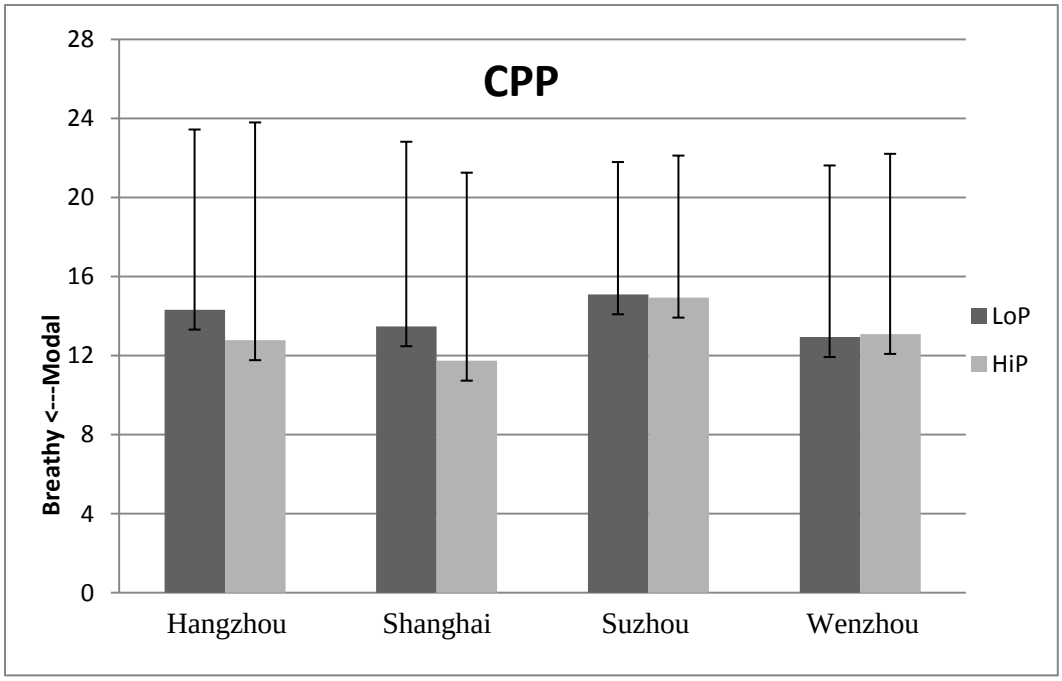


Figure 5.33 Comparison of CPP between LoP and HiP (Wu)



By H1*-H2*, the most noticeable pattern is that LoP is breathy and HiP is creaky in Suzhou and Shanghai, as the high blue columns are above 0dB and the red columns below 0dB. The breathiness of the LoP production found in Shanghai is in line with the previous finding that Shanghai low rising tone production tends to be breathy (Cao & Maddieson, 1992; Gao et al., 2011).

Wenzhou also has LoP breathier than HiP, as the blue column is taller than the red one. For Hangzhou, LoP and HiP are close in voice quality, as the two columns are about the same height.

By CPP, the most noticeable pattern here is that LoP is less breathy than HiP in Hangzhou and Shanghai, as the blue columns are higher than red ones. Suzhou and Wenzhou both have a CPP difference within $\pm$1dB, suggesting that their LoP and HiP productions have a similar voice quality.

A comparison of the two figures shows that the patterns for Shanghai seem to be opposite: LoP is breathier than HiP by the H1*-H2* but not by the CPP measure. A possible explanation is that the CPP measure may not serve as a good indicator of the voice quality difference for Shanghai, as its difference between LoP and HiP is rather small (< 2dB). Here the H1*-H2* measure seems to be more reliable, as a comparison of the standard deviations (see the vertical lines on top of the columns) shows that the standard deviation difference (ΔLoP – ΔHiP) for H1*-H2* is more than twice larger than for CPP (0.41dB vs. 0.17dB), meaning that H1*-H2* is more sensitive to the energy difference than CPP.

5.4.3 Min

H1*-H2* and CPP results for five Min dialects are listed in Appendix 10. The results are illustrated in Figure 5.34 and 5.35, respectively.

Figure 5.34 Comparison of H1*-H2* between LoP and HiP (Min)

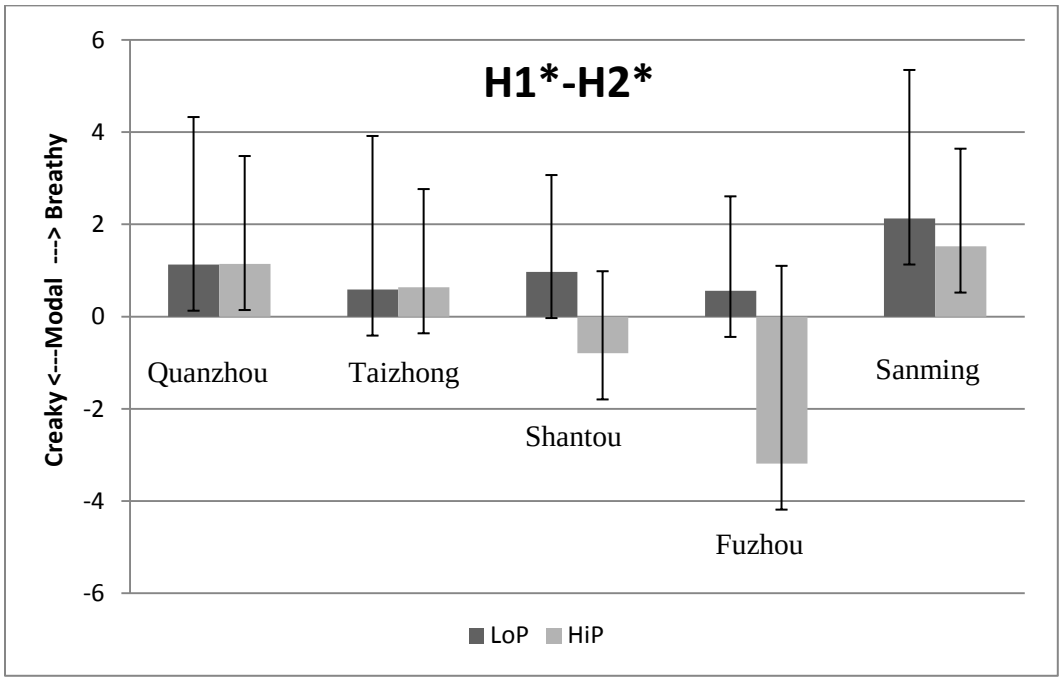
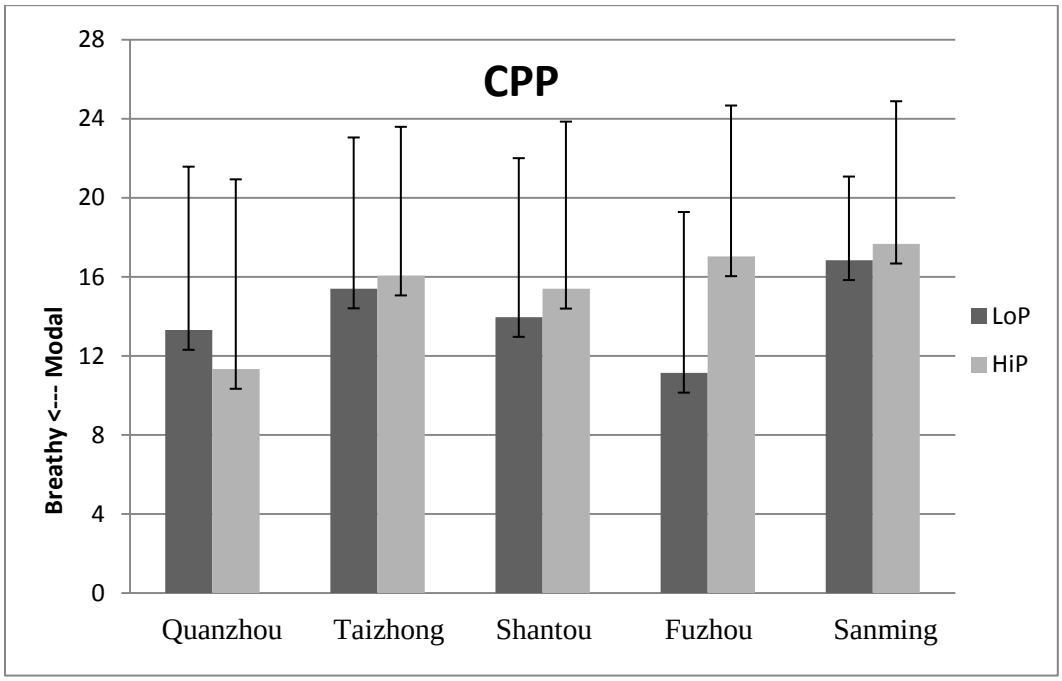


Figure 5.35 Comparison of CPP between LoP and HiP (Min)



By H1*-H2*, the most noticeable pattern is that HiP is rather creakier than LoP in Fuzhou, as shown by the long red column below 0dB and the short blue column above 0dB. The breathiness of the LoP production is in line with the previous finding that Fuzhou low tone production is breathy (Esposito, 2006). Note that creakiness in HiP here may actually mean tenseness in voice. Keating and Garellek (2015) also noted an instance of mid- or high-pitched tense voice, which shares some acoustic characteristics with low-pitched vocal fry (the typical creakiness perceived). In both voice types, the larynx is constricted, so it is not surprising that their energy patterns are similar.

Shantou shows a similar pattern: its HiP is also creakier than LoP. For Sanming, LoP is breathier than HiP. For Quanzhou and Taizhong, the H1*-H2* difference between is within ±1dB, suggesting that their LoP and HiP productions are close in voice quality.

Note that creakiness was found in Taiwan Min low tone production in Pan's (2005) study, but no evidence is shown here.

By CPP, the most noticeable pattern here is LoP is rather breathier than HiP in Fuzhou, as the blue column is much shorter than red one. This pattern is consistent with the pattern shown earlier: Fuzhou LoP production tends to be breathy, so the corresponding H1*-H2* has a positive value and CPP is small. Fuzhou HiP production, on the other hand, has a large CPP, which supports the idea that the type of creakiness in Fuzhou HiP production may actually be tense voice, as tense voice has a robust harmonic structure and hence a large CPP. An auditory examination of the Fuzhou speech confirmed the tenseness of Fuzhou HiP production.

Another noticeable pattern is that HiP is breathier than LoP in Quanzhou, but their CPP difference is within 2dB, so it may not be a reliable cue to the voice quality identification. As for the remaining three dialects, Taizhong, Shantou, and Sanming, their CPP difference between LoP and HiP is within or around $\pm$1dB, suggesting that their LoP and HiP productions are similar in voice quality.

5.4.4 Cantonese

H1*-H2* and CPP results for four Cantonese dialects are listed in Appendix 11. The results are illustrated in Figure 5.36 and 5.37, respectively.

Figure 5.36 Comparison of H1*-H2* between LoP and HiP (Cantonese)

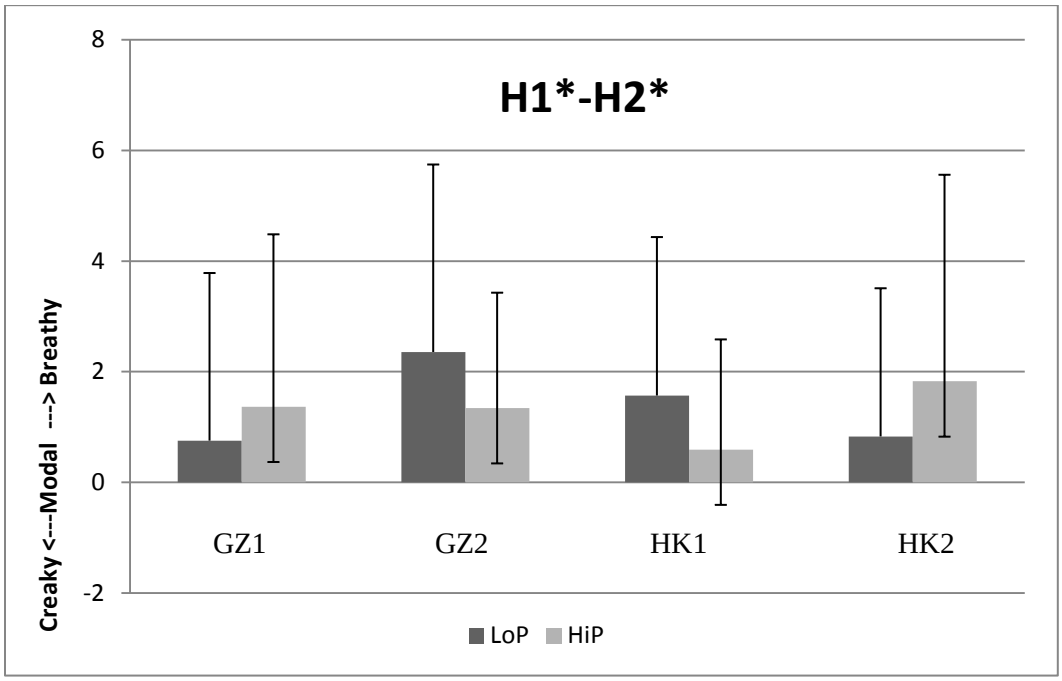
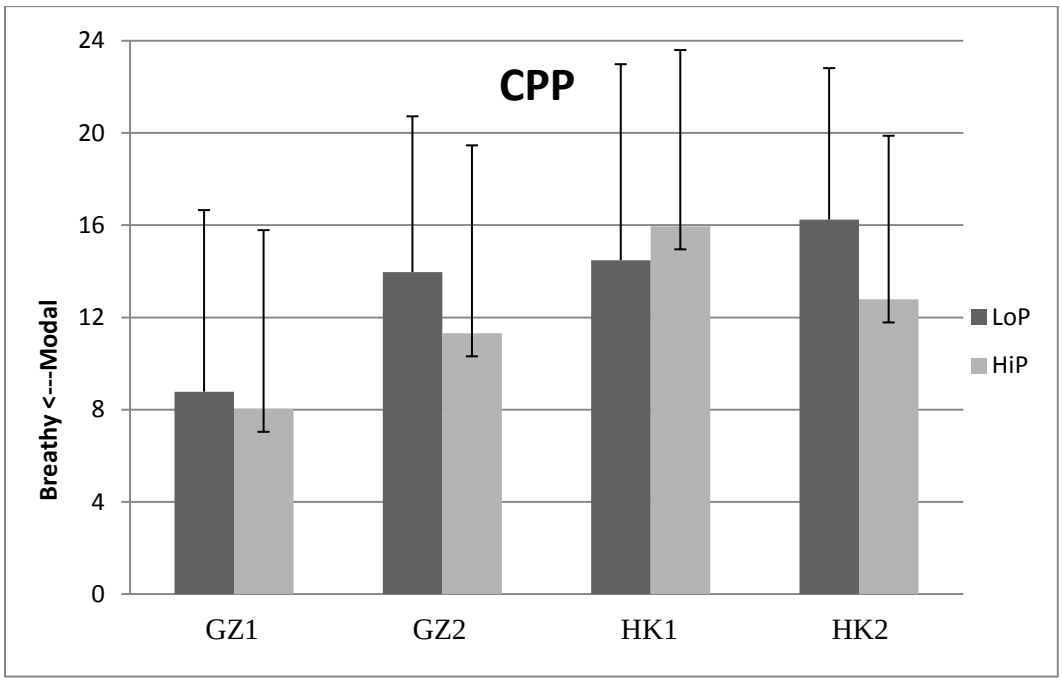


Figure 5.37 Comparison of CPP between LoP and HiP (Cantonese)



By H1*-H2*, all the Cantonese dialects show a H1*-H2* difference between LoP and Hi but not all of them have the same pattern. In GZ2 and HK1, LoP is breathier than HiP, whereas in GZ1 and HK2, LoP is less breathy than HiP. Note that creakiness was found to be a perceptual cue for Cantonese low falling tone (Lam &Yu, 2010), but no evidence of creakiness is shown in the LoP production here, as no red columns are below 0dB.

By CPP, both GZ2 and HK2 show that HiP is breathier than LoP, as the red columns are shorter than the blue ones. GZ1 shows a similar pattern, but the CPP difference is not as large as GZ2 and HK2. HK1 shows an opposite pattern: LoP is breathier than HiP, as the blue column is shorter than the red one.

A comparison of the two figures shows a conflict for GZ2: LoP is breathier than HiP by the H1*-H2* measure but less so by the CPP measure. Yet, both the Hi*-H2* and CPP difference between LoP and HiP is small (H1*-H2*: about 1dB; CPP: about 2.6dB), so it is hard to determine which measure better reflects the reality.

5.4.5 Summary

Section 5.4 shows that the LoP and HiP production can have different voice qualities in different dialects. Most noticeable patterns include the breathiness of LoP production in Guilin Mandarin and the tenseness of the HiP production in Fuzhou Min.

In addition, no consistent voice quality pattern is found either within a major group or across the four major groups. Part of the reason may be that the two voice source metrics H1*-H2* and CPP are not always consistent in determining voice quality. Sometimes, they even show conflicting voice quality patterns, as in the case of Shanghai Wu and GZ2 Cantonese. However, in these cases (also in most cases), the H1*-H2* and CPP difference between LoP and HiP is relatively small (within or around 2dB), so voice

quality variations may be largely idiosyncratic rather than systematic or phonological in most dialects.

5.5 Rhythmic patterns across all the Chinese dialects

The above four sections have presented rhythmic patterns respectively for four major Chinese dialect groups. This section presents correlation results across all 21 Chinese dialects. Section 5.5.1 and Section 5.5.2 respectively presents results in the duration- and pitch-only categories across all the dialects. Section 5.5.3 presents results in the duration-pitch, duration-syllable, and pitch-tone categories across all the dialects. Section 5.5.4 summarizes all the results and key findings.

5.5.1 Duration-based timing patterns

Correlation results in the duration-only category are listed in Table 5.23.

Table 5.23 Correlation results in the duration-only category (all the dialects)

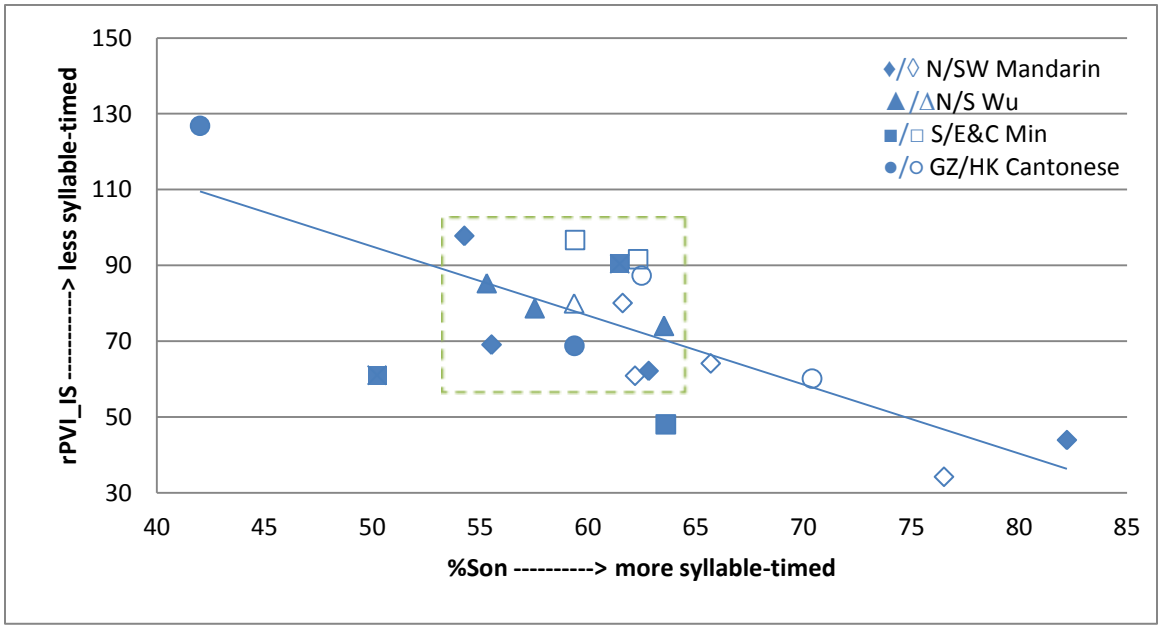
	Column	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	
	Metric	*sumD*	*s_rate*	*%Son*	*ΔSon*	*varco Son*	*nPVI_ Son*	*ΔIS*	*varco IS*	*rPVI_ IS*	*Row*
	sumD	1									*1*
	s_rate	-0.185	1								*2*
Son_based	*%Son*	-0.046	-0.319	1							*3*
	ΔSon	-0.199	-0.521	0.335	1						*4*
	varcoSon	0.152	-0.387	0.151	0.156	1					*5*
	nPVI_Son	0.408	-0.439	0.388	0.356	0.347	1				*6*
IS-based	*ΔIS*	0.590	-0.404	-0.294	-0.331	0.408	0.426	1			*7*
	varcoIS	-0.063	-0.246	-0.452	-0.032	0.600	0.015	0.510	1		*8*
	rPVI_IS	0.196	-0.332	-0.732	0.025	0.067	0.028	0.621	0.627	1	*9*

No high correlation occurs in C1 and C2, indicating that speech length and rate are not much related to timing across all 21 dialects. None of the five predicted duration-based metric pairs have the absolute cc value larger than 0.5, so all of them are not correlated.

Of all 16 unpredicted metric pairs, only five are correlated (see underlined) and the only high correlation occurs between %Son and rPVI_IS (cc = -0.732, underlined in C3-R9). The negative correlation is expected and it means that dialects with a larger portion of sonorant duration and smaller variability in inter-sonorant duration tend to be more syllable-timed.

Figure 5.38 uses %Son-rPVI_IS as the x-y axis to plot all the dialects along the syllable-timedness continuum. Most Chinese dialects have similar degrees of syllable-timedness based on %Son-rPVI_IS, as they clustered in a relatively small area (see the dotted box around the middle portion of the trend line). Cantonese and Mandarin dialects have the widest distribution along the downward trend line: Two Mandarin dialects (Beijing & Wuhan) are located at the bottom of the trend line and one Cantonese dialect (GZ2) is located above the top of the trend line. Also, all four major groups except Wu have different sub-dialects both inside and outside the small area, indicating that the degree of syllable-timedness varies greatly not only across major groups but also across sub-groups.

Figure 5.38 Timing pattern based on %Son-rPVI_IS (all the dialects)



5.5.2 Pitch-based melody patterns

Correlation results in the pitch-only category are listed inTable 5.23.

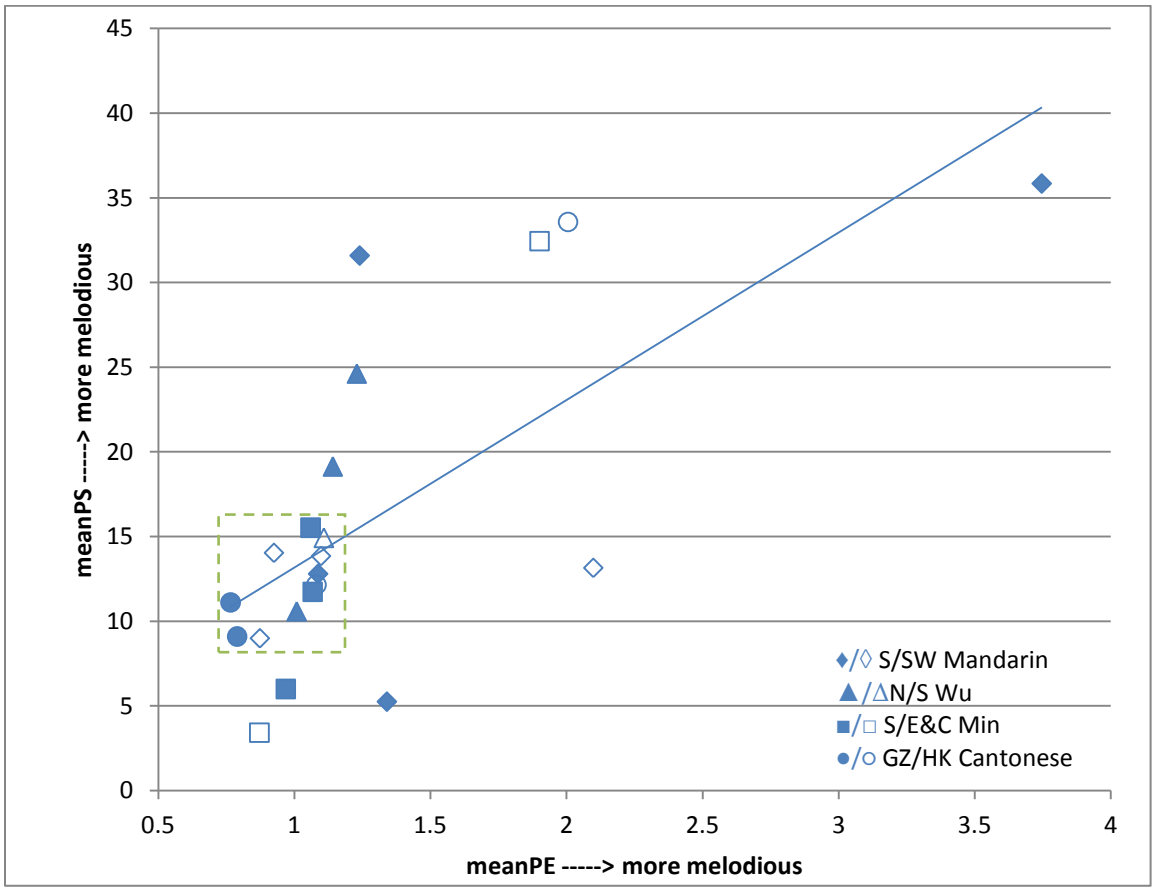
Table 5.24 Correlation results in the pitch-only category (all the dialects)

	Column	*1*	*2*	*3*	*4*
Row	*Metric*	*meanPE*	*ΔPE*	*meanPS*	*ΔPS*
1	meanPE	1			
2	ΔPE	0.957648	1		
3	meanPS	0.685446	0.701282	1	
4	ΔPS	0.559904	0.602305	0.92115	1

Both meanPE-meanPS and ΔPE-ΔPS have a positive correlation as predicted, but the degree of their correlation is only moderate (cc = 0.685446 & 0.602305; underlined in C1-R3 & C2-R4). Figure 5.39 uses meanPE-meanPS as the x-y axis to plot all the dialects along the melodiousness continuum. Over a half of Chinese dialects are clustered in a relatively small area (see the dotted box around the lower portion of the

trend line). All four major dialect groups and their sub-groups are present both inside and outside the box, meaning that melody patterns vary greatly across the Chinese dialects at both major and sub- dialectal levels.

Figure 5.39 Melody pattern based on meanPE-meanPS (all the dialects)



5.5.3 Correlations among syllable-timing, melody, and phonological structure

This section presents correlation results to show if timing and melody patterns are related to each other as well as to phonological structure across all the Chinese dialects. Correlation results in the duration-pitch, duration-syllable, and pitch-tone categories and the associated correlation patterns are presented in Section 5.5.3.1, 5.5.3.2, and 5.5.3.3, respectively.

5.5.3.1 Correlation between syllable-timedness and melodiousness

Correlation results in the duration-pitch category are listed in Table 5.25.

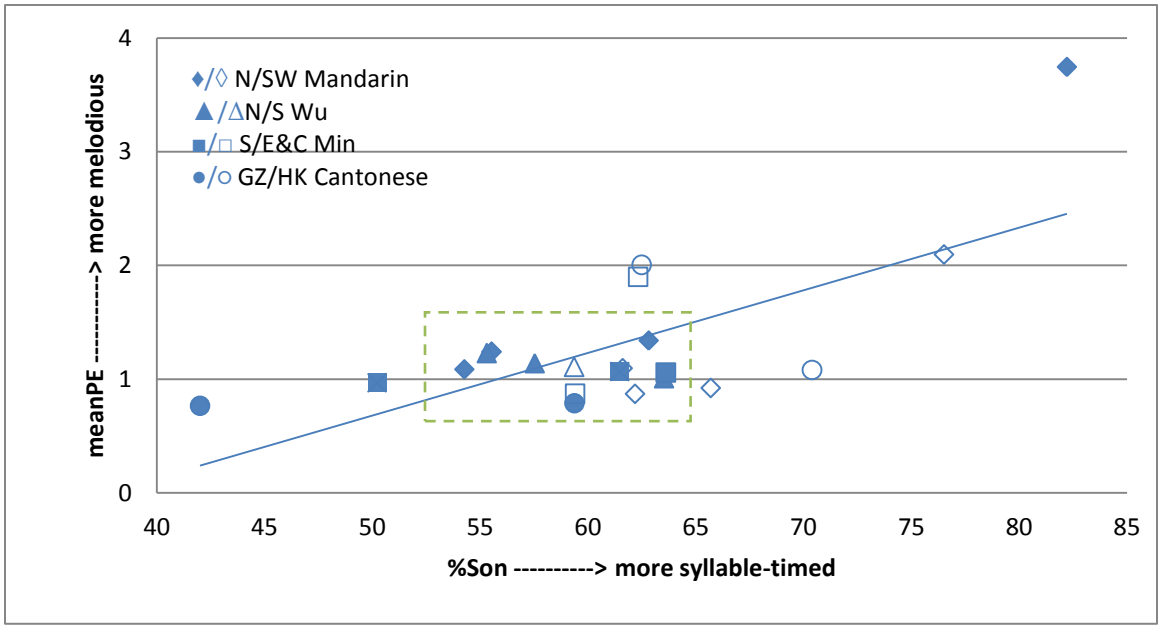
Table 5.25 Correlation results in the duration-pitch category (all the dialects)

Duration-based \ *Pitch-based*		*PE-based*		*PS-based*		*Row*
		meanPE	*ΔPE*	*meanPS*	*ΔPS*	
Son-based	*%Son*	<u>0.6912</u>	<u>0.55553</u>	0.269841	0.267442	*1*
	ΔSon	0.401303	0.376399	0.328108	0.39976	*2*
	varcoSon	-0.01864	-0.0552	0.0109	-0.02051	*3*
	nPVI_Son	0.204914	0.147602	0.25584	0.283131	*4*
IS-based	*ΔIS*	-0.31504	-0.26878	-0.07581	-0.07174	*5*
	varcoIS	-0.16937	-0.12736	0.037111	-0.09349	*6*
	rPVI_IS	-0.39176	-0.2884	-0.04071	-0.00921	*7*
	Column	*1*	*2*	*3*	*4*	

Only two pairs, %Son-meanPE and %Son-ΔPE, are positively correlated as predicted (see underlined). Yet, even the highest correlation is moderate (cc = 0.6912, see underlined in R1-C1).

Figure 5.40 illustrates how all the dialects are differentiated by %Son-meanPE: More than a half of Chinese dialects have degrees of syllable-timedness and melodiousness varying inside a relatively small area (see the dotted box around the middle portion of the trend line). Except Wu, all four major dialect groups and their sub-groups are present both inside and outside the box, meaning that both timing and melody patterns vary greatly across most of the Chinese dialects.

Figure 5.40 Correlation pattern based on %Son-meanPE (all the dialects)



5.5.3.2 Correlation between syllable-timedness and syllable structure

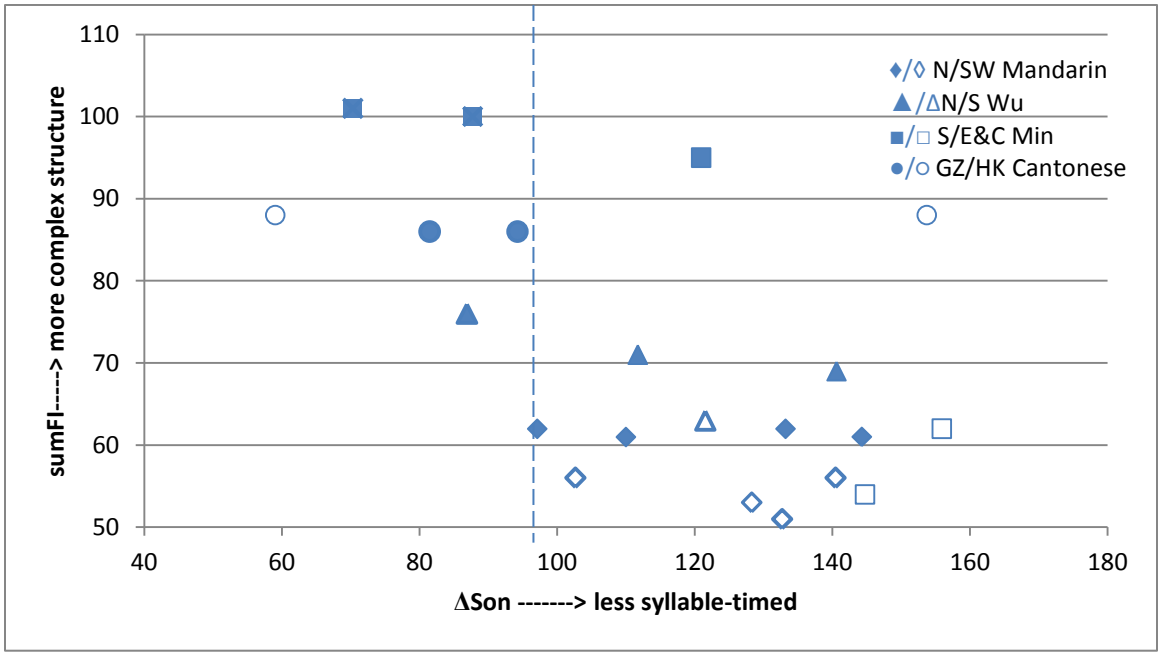
Correlation results in the duration-syllable category are listed in Table 5.26.

Table 5.26 Correlation results in the duration-syllable category (all the dialects)

Duration-based \ *Syllable-based*		*Fin:Ini*	*sumFI*	*Row*
Son-based	*%Son*	-0.22199	-0.29318	*1*
	ΔSon	-0.46571	-0.57103	*2*
	varcoSon	0.146448	0.149611	*3*
	nPVI_Son	0.045744	-0.11109	*4*
IS-based	*ΔIS*	0.342473	0.338067	*5*
	varcoIS	0.270136	0.314508	*6*
	rPVI_IS	0.066034	0.093267	*7*
	Column	*1*	*2*	

The only correlation occurs between ΔSon and sumFI (cc = -0.57103; underlined in C2-R2) and the degree of correlation is low (|cc| < 0.6). Figure 5.41 illustrates how all the Chinese dialects are differentiated by ΔSon-sumFI. Six dialects on the left side of the dotted vertical line have higher sumFI and smaller ΔSon than most dialects on the right side of the dotted vertical line. This pattern means that dialects with a complex syllable structure tend to have less variability in sonorant duration and in turn have a larger degree of syllable-timedness. This outcome is contrary to what has been predicted, the smaller the sumFI, the more syllable-timed the dialect or the larger the sumFI, the less syllable-timed the dialect. Notably, all eight Mandarin Chinese dialects are less syllable-timed (all on the right side of the line) than the three Cantonese, two Southern Min, and one Northern Wu dialects (on the left side of the line), despite that they have the fewest number of finals and initials.

Figure 5.41 Correlation pattern based on ΔSon-sumFI (all the dialects)



5.5.3.3 Correlation between melodiousness and tone structure

Correlation results in the pitch-tone category are listed in Table 5.27.

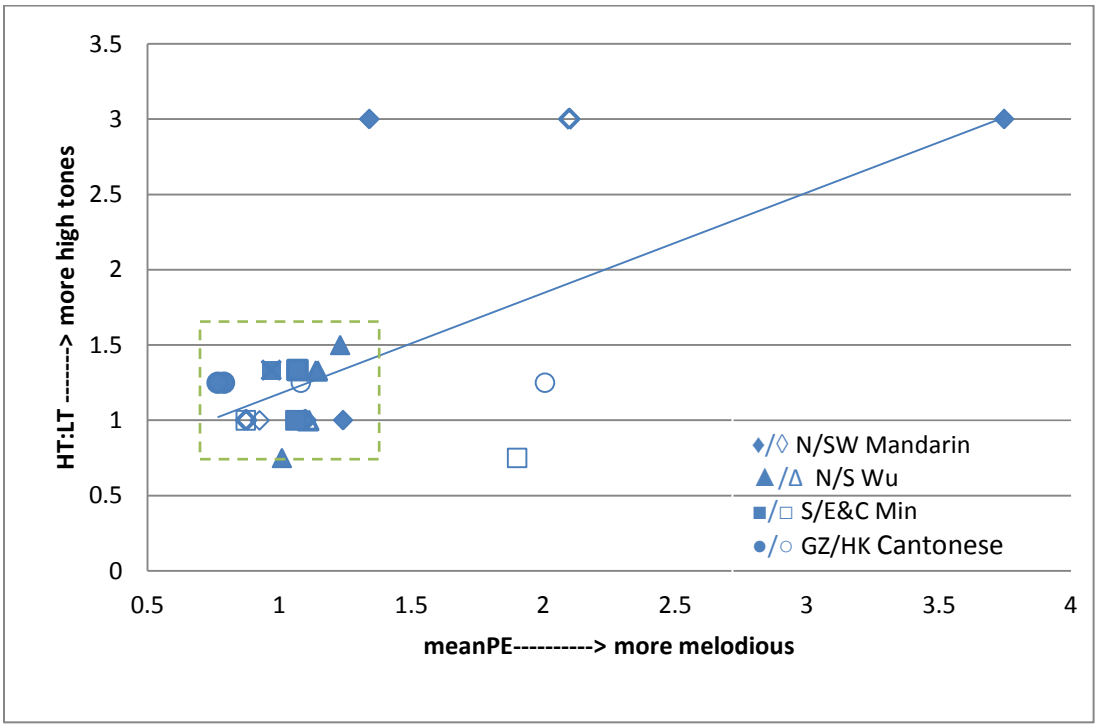
Table 5.27 Correlation results in the pitch-tone category (all the dialects)

Pitch-based / *Tone-based*	*PE-based*		*PS-based*		
	meanPE	*ΔPE*	*meanPS*	*ΔPS*	*Row*
HT:LT	0.637561	0.603945	0.068579	0.017187	*1*
sumT	-0.25524	-0.20715	-0.05594	-0.04155	*2*
Column	*1*	*2*	*3*	*4*	

There are only two moderate correlations occurring (see underlined). The highest correlation occurs between meanPE and HT:LT (cc = 0.637561; underlined in C1-R1). The positive correlation is predicted and it means that dialects with more high than low tones tend to be more melodious. Figure 5.42 illustrates how all the Chinese dialects are differentiated by meanPE-HT:LT. Most Chinese dialects have the degree of

melodiousness and the HT:LT value varying in a relatively small range (see the dotted box). The three Mandarin dialects outside the box have the same large HT:LT (=3), but they vary greatly in term of the degree of melodiousness (see the top three diamonds). Also, all four major groups except Wu have their sub-groups both inside and outside the box, indicating that the degree of melodiousness of these dialects can vary greatly, whether or not they have the same or similar tonal structure.

Figure 5.42 Correlation pattern based on meanPE-HT:LT (all the dialects)



5.5.4 Summary

When all the major dialect groups are examined individually, timing and melody patterns seem to vary greatly. When they are examined together, their timing and melody patterns vary, still greatly but not as great as expected, as most dialects are clustered in a relatively small area along the syllable-timedness, melody, and structure complexity dimensions.

On the other hand, almost all the metric pairs fail to distinguish all the Chinese dialects well: As shown in Table 5.28, the total number of correlated pairs is low (=12) compared to a total of 54 predicted pairs and there is only one highly correlated pair, %Son-rPVI_IS (see C3-R1). Despite that not a single metric performs consistently well, the metrics %Son, meanPE, and HT:LT seem to perform better than the rest in their respective correlation categories, as they occur the most in correlated pairs. That is to say, sonorant duration, pitch excursion, and the number of high tones can contribute to the relative syllable-timedness and melodiousness of Chinese dialects.

Based on these metrics, some general patterns emerge as predicted: first, sonorant duration and syllable-timedness are positively correlated; second, syllable-timedness and melodiousness are also positively correlated; third, high tone complexity and melodiousness are positively correlated. A pattern opposite to what is predicted also emerges: the number of finals and initials does not necessarily correlate with syllable-timedness negatively. On the whole, Chinese rhythmic patterns vary greatly from dialect to dialect, whether within a major or sub- group.

Table 5.28 Summary of correlation results in all 5 categories (all the dialects)

Row	*Correlation category*	*# of correlated metric pairs (12)*	*Correlated metric pairs*	*Correlation type*
1	*Duration-only*	5	%Son-rPVI_IS	H−
			varcoSon-varcoIS	M+
			rPVI_IS-varcoIS	M+
			ΔIS-rPVI_IS	M+
			ΔIS-varcoIS	M–
2	*Pitch-only*	2	meanPE-meanPS	M+
			ΔPE-ΔPS	M+
3	*Duration-pitch*	2	%Son-meanPE	M+
			%Son-ΔPE	L+
4	*Duration-syllable*	1	ΔSon-sumFI	L–
5	*Pitch-tone*	2	meanPE-HT:LT	M+
			ΔPE-HT:LT	M+
	Column	*1*	*2*	*3*

Chapter 6
DISCUSSION

The last chapter has reported all the duration-, pitch-, and structure-based results and shown that Chinese rhythmic patterns are homogeneous neither at the major nor at the sub- dialectal level. This chapter provides answers to the three research questions and evidence for or against the associated hypotheses in Section 6.1, 6.2, and 6.3, respectively. Section 6.4 summarizes the discussion and offers some implications for Chinese rhythm research.

6.1 Duration-based timing and pitch-based melody patterns

This study has used nine duration- and four pitch-based metrics to reveal timing and melody patterns of Chinese dialects. Given such amount of metrics used, there is a question of whether or not they can make a consistent rhythmic distinction among Chinese dialects and a further question is how many of them can if they can. These two questions are translated into the following two operational ones: 1) how many metric pairs are correlated as predicted and 2) how consistent they are in showing the relative syllable-timedness and melodiousness of individual dialects at major and sub- dialectal levels. The answers to these questions are used to address the first research question, how duration- and pitch-based metrics fare in quantifying Chinese rhythm at major and sub- dialectal levels.

Recall that the criterion for how well the metrics fare in quantifying rhythm (see Section 1.4) is that there must be majority of metric pairs capable of revealing a pattern

consistently. Here two specific measures, CP ratio and consistency range, are used to evaluate how well the metrics fare in quantifying rhythm.

The CP ratio is the number of correlated metric pairs in relation to the total number (5) of metric pairs as predicted, so it measures how many metric pairs are correlated as predicted. The total number of predicted pairs is five in the duration-only category (%Son-ΔIS, %Son-varcoSon, %Son-varcoIS, ΔSon-ΔIS, & rPVI_IS-nPVI_Son) and two in the pitch-only category (meanPE-meanPS & ΔPE-ΔPS). Note that some unpredicted pairs are also correlated but for the sake of simplicity, they will not be counted. Also, the predicted pairs involve all types of metrics, sufficiently representive of all the pairs in their respective categories.

The consistency range refers to how many sub-groups in a major group have their dialects consistently ordered by the two (if available) representative pairs (illustrated in Section 5.1 & 5.2). In other words, if the two pairs can give the same order of syllable-timedness or melodiousness for dialects within a sub-group of the four major dialect groups, then the consistency range is 1. By the same token, if two such sub-groups exist, the consistency range is 2. If no sub-groups have their dialects consistently ordered, then the consistency range is 0. Note that the two measures are not used to evaluate the correlation patterns for all the Chinese dialects shown in Section 5.5.1 and 5.5.2, because they are expected to do worse across major groups.

Now the initial criterion for how well the metrics fare in quantifying rhythm can turn into two specific ones: 1) The CP ratio must reach 2/3, meaning that at least two thirds (or majority) of metric pairs must be correlated; 2) the consistency range must equal 1 or above 1.

Table 6.1 summarizes the CP ratio and consistency range for the four major dialect groups Mandarin, Wu, Min, and Cantonese (cf. Table 5.5, Table 5.11, & Table 5.22).

Table 6.1 Comparison of CP ratio and consistency range in the duration- and pitch-only categories

Correlation category	*Data Range (# of dialects)*	*CP ratio (Actual/Predicted)*	*Consistency range*
Duration-only	Mandarin (8)	1/5	1
	Wu (4)	3/5	1
	Min (5)	2/5	0
	Cantonese (4)	<u>4/5</u>	2
Pitch-only	Mandarin (8)	2/2	2
	Wu (4)	2/2	2
	Min (5)	2/2	2
	Cantonese (4)	2/2	1

Duration-based metrics do not fare well for most dialect groups. Of the four major dialect groups, only Cantonese has over two-thirds of metric pairs are correlated (CP ratio >2/3; see underlined), and its top two most correlated pairs are able to order the dialects consistently for two sub-groups (the consistency range = 2). For the remaining three major dialect groups Mandarin, Wu, and Min, they all fail to reach the CP ratio of 2/3, and Min also fails the consistency range (=0 <1).

Pitch-based metrics fare far better than duration-based ones. In terms of the CP ratio, the two predicted metric pairs (meanPE-meanPS & ΔPE-ΔPS) are correlated for all the major dialect groups as well as across all the dialects (CP ratio = 1 > 2/3). In terms of consistency range, all four major dialect groups have at least one sub-group that can be consistently ordered.

As for how well the duration- and pitch-based metrics do in distinguishing between sub-groups of the same major group, the answer is no, because no two sub-groups can separate well on either the syllable-timedness or the melodiousness continuum.

Below are the answer to Question 1 and the discussion of the associated hypothesis:

Question 1: How do duration- and pitch-based metrics fare in quantifying Chinese rhythm at major and sub- dialectal levels?

Answer:

How duration- and pitch-based metrics fare depends on which major group or sub-group is involved. Judged by both CP ratio and consistency range, duration-based metrics fare poorly for all four major groups except for Cantonese, whereas pitch-based metrics fare well for all four groups especially for Wu, Min, and Cantonese. At the sub-dialectal level, no duration- and pitch-based metrics fare well, as most of them can not separate sub-groups by syllable-timedness or melodiousness.

Hypothesis 1: Duration-based metrics fare better than pitch-based metrics in quantifying Chinese rhythm at the major dialectal level (FALSE) but neither of them fares well at the sub-dialectal level (TRUE).

Discussion:

The first part is false because only the opposite is true. Pitch-based metrics fare better than duration-based ones at the major dialectal level, perhaps because pitch variations are restricted by tone types, and tone types are distinct across dialects. In other words, speakers cannot manipulate pitch too freely as it is restricted by available tone types. For segmental duration, however, except for Cantonese, which has long and short vowel distinction, its variation is not restricted, so more idiosyncratic variations can be involved

so as to fail duration-based metrics. The idiosyncratic aspect includes but does not limit to instances such as that speakers may manipulate segment intervals to achieve certain communicative effect. Hence, duration is subject more to influences of speech rate, content, and style than pitch. It has been shown that speech rate indeed has affected some measures (most notably, varcoSon) for some dialect groups (most notably, Min), so it is not surprising that more factors may come into play.

Note that the rationale for the first hypothesis provided in Section 1.4 may need to be qualified: when there are no confounding factors, Chinese rhythm may be more durationally than melodiously distinct at the major dialectal level. For example, if a speaker can use Mandarin and Min respectively to deliver a message in the same way, then the two dialects may have a rhythmic pattern more durationally than melodiously distinct.

The second part is true as expected. It means that rhythmic differences among sub-dialect groups, unlike those among different languages, are too small to be detected by either duration- or pitch-based metrics. Most sub-groups do not have distinct timing or melody patterns. In fact, even major groups cannot be differentiated: the patterns across all 21 Chinese dialects (see Section 5.5) show that most dialects, regardless of their group membership, fall into a small vicinity of syllable-timedness and melodiousness, suggesting that Chinese rhythm, after all, does not vary greatly as expected.

6.2 The relationship between duration- and pitch-based rhythm

There are good reasons to expect a relationship between duration- and pitch-based rhythmic patterns in Chinese: 1) both duration and pitch are cues to rhythm; 2) syllables as tone bearing units are inevitably associated with pitch.

In order to answer the second research question, "How duration-based timing and pitch-based melody patterns are related", the CP ratio and direction of correlation are summarized for the four major dialect groups in Table 6.2. The CP ratio here refers to the number of correlated metric pairs in relation to the total number (= 28) of metric pairs in the duration-pitch category. The direction of correlation refers to whether two metrics are positively (+) or negatively (−) correlated.

With respect to CP ratio, duration- and pitch-based metrics correlate well only for Wu, as Wu has over two-thirds of metric pairs (CP ratio = 21/28 > 2/3) correlated. The direction of correlation is negative for Wu but positive for the remaining three major groups.

Table 6.2 Comparison of CP ratio and direction of correlation in the duration-pitch category

Data Range (# of dialects)	*CP ratio*	*Direction of correlation*
Mandarin (8)	5/28	+
Wu (4)	21/28	−
Min (5)	10/28	+
Cantonese (4)	13/28	+

Below are the answer to Question 2 and the discussion of the associated hypothesis:

Question 2: How are duration-based timing and pitch-based melody patterns related?

Answer:

Syllable-timedness and melodiousness are more or less correlated for all 4 major groups and they correlate best for Wu.

Hypothesis 2: There is a positive correlation between syllable-timedness and melodiousness (FALSE for Wu).

Discussion:

For Mandarin, Min, and Cantonese dialects, a larger degree of syllable-timedness indeed means a larger degree of melodiousness. In the previous study by Hirst (2013), Mandarin is found to be more syllable-timed and also more melodious than English and French. The reason may be due to the fact that tonal languages tend to have less complex syllable structure and larger pitch variations than non-tonal languages. Among Chinese dialects, those with fewer tones tend to have less complex syllable structure and larger pitch variations. Therefore, the positive correlation between syllable-timedness and melodiousness is expected from the structural perspective.

Then, why are syllable-timedness and melodiousness in Wu negatively correlated? The reason may be related to the unique structure of Wu. Wu dialects are different from the other three major groups of dialects in that its complex syllable structure does not imply complex tone structure. For the four Wu dialects, Fin:Ini decreases from Suzhou, Shanghai, Hangzhou, to Wenzhou, but HT:LT increases from Hangzhou, Wenzhou, Suzhou, to Shanghai. The two orders are almost opposite, indicating that Wu dialects with more complex syllable structure (as indicated by larger Fin:Ini) tend to have less complex tone structure (as indicated by larger HT:LT). Hence, Shanghai Wu, for

example, despite having a relatively complex syllable structure, is very melodious due to its relatively simple tone structure. The next section will reinforce the structure-based reasoning above.

6.3 The influence of phonological structure on rhythm

Phonology-based rhythmic studies (Dauer, 1983; Auer, 1993) show that syllable-timing is predictable from syllable structure complexity. All the Chinese dialects are monosyllabic, so it is natural to assume that Chinese dialects may be close in syllable-timedness as they all have simple syllable structure compared to English. On the other hand, Chinese dialects do vary greatly in syllable type, specifically initials and finals, at the major dialectal level. Since syllable type also reflects syllable structure complexity, Chinese dialects are expected to vary in syllable-timedness. However, syllable structure may not affect syllable-timedness of sub-dialects, as they often have similar syllable types. The complexity of Chinese syllable structure is measured by Fin:Ini and sumFI, the larger the two measures, the more complex the syllable structure and the less syllable-timed the dialect.

Chinese dialects also vary greatly in tone type, especially high and low tones, at the major dialectal level. In the same vein, if tone structure complexity can predict melodiousness, then Chinese dialects are expected to vary in melodiousness. Again, tone structure may not affect melodiousness of sub-dialects, as they often have similar tone types. The complexity of Chinese tone structure, especially high tone structure, is measured by HT:LT and sumT, the larger HT:LT and smaller sumT, the less complex the tone structure and the more melodious the dialect.

The third research question is about how phonological structure affects speech rhythm of Chinese dialects, despite their shared mono-syllabicity. In order to answer this question, the CP ratio and direction of correlation are summarized respectively for the four major dialect groups in Table 6.3. The CP ratio here refers to the number of correlated metric pairs in relation to the total number of metric pairs predicted in their respective categories. The direction of correlation refers to whether two metrics in a metric pair are positively (+) or negatively correlated (–). The 'n/c' entry means no correlation is found (|cc| < 0.5) and 'n/a' means 'not applicable' (no correlation analysis is performed for the metric pairs).

Table 6.3 Comparison of CP ratio and direction of correlation in the duration-syllable and pitch-tone categories

Correlation category	*Duration-syllable*			*Pitch-tone*		
Data Range (# of dialects)	*CP ratio*	*Direction of correlation*		*CP ratio*	*Direction of correlation*	
		Fin:Ini	*sumFI*		*HT:LT*	*sumT*
Mandarin (8)	3/14	+	n/a	2/4	+	n/a
Wu (4)	6/14	+	+	8/8	+	−
Min (5)	8/14	+	+	1/8	−	n/c

With respect to syllable structure measures, no dialects have more than two-thirds of correlated metric pairs in the duration-syllable category, meaning that syllable structure may not be a good indicator of syllable-timing for the Chinese dialects. Also, all the correlations are positive instead of negative as predicted, meaning that complex syllable structure can increase rather than decrease syllable-timedness in Chinese dialects.

As for tone structure measures, only Wu has more than two-thirds of correlated metric pairs in the pitch-tone category (CP>2/3; see underlined), indicating that tone structure can contribute greatly to melody patterns in Wu. Also, HT:LT and melodiousness are positively correlated as predicted and sumT and melodiousness are negatively correlated as predicted.

The relationships between syllable structure and syllable-timedness and between tone structure and melodiousness can be also seen from comparisons between the relative syllable-timedness and melodiousness ranked by the four structure metrics and by the highest correlated metric pairs in the duration-only and pitch-only categories for Mandarin, Wu, and Min dialects (see Table 6.4 and Table 6.5). Note that the dialects in the same cell of the tables have either comparable or conflicting rankings in the degree of syllable-timedness/melodiousness, so their relative degrees of syllable-timedness/melodiousness are either comparable or to be further determined.

Table 6.4 Comparison of structure- and duration-based order of syllable-timedness

Major group	*Degree of syllable-timedness↓*		
	Order ranked by Fin:Ini↑	*Order ranked by sumFI↑*	*Duration-based order*
Mandarin	♦Tianjin	◊Liupanshui	♦Beijing ◊Wuhan
	♦Jinan	◊Guilin	♦Heshuo ◊Chengdu
	◊Liupanshui	◊Wuhan ◊Chengdu	◊Guilin ♦Jinan
	♦Heshuo		◊Liupanshui
	◊Chengdu	♦Tianjin ♦Heshuo	♦Tianjin
	♦Beijing		
	◊Guilin	♦Beijing ♦Jinan	
	◊Wuhan		
Wu	ΔWenzhou	ΔWenzhou	▲Hangzhou
	▲Hangzhou	▲Hangzhou	▲Suzhou
	▲Shanghai	▲Shanghai	ΔWenzhou
	▲Suzhou	▲Suzhou	▲Shanghai
Min	□Sanming	□Sanming	■Quanzhou
	□Fuzhou	□Fuzhou	■Shantou
	■Shantou	■Shantou	■Taizhong □Sanming
	■Taizhong	■Taizhong	□Fuzhou
	■Quanzhou	■Quanzhou	

*↑: ascending order; ↓: descending order

For the degree of syllable-timedness, the orders ranked by syllable structure metrics, Fin:Ini and sumFI, are different from each other and also from the one ranked by duration-based measures. In the case of Mandarin, the most noticeable pattern is that Tianjin has the second largest sumFI and ranks as the least syllable-timed while Beijing has the largest sumFI but ranked as the most syllable-timed. This pattern suggests that syllable structure can not reflect Mandarin timing well.

For the four Wu dialects, the three orders do not show any particular pattern, suggesting that syllable structure does not affect syllable-timing in Wu. For the five Min dialects, the two structure-based orders are the same, both of which are basically opposite to the duration-based order between the Southern and the remaining two sub-groups, East and Central. This pattern shows that syllable structure may have some influence on syllable-timing in Min at the sub-dialectal level.

Table 6.5 Comparison of structure- and pitch-based order of melodiousness

Major group	*Degree of melodiousness*↑		
	Order predicted by HT:LT↑	*Order predicted by sumT*↓	*Pitch-based order*
Mandarin	◊Chengdu ◊Guilin ◊Liupanshui ♦Tianjin ♦Jinan	(not applicable)	◊Chengdu
			◊Guilin
			◊Liupanshui ♦Heshuo
	◊Wuhan ♦Beijing ♦Heshuo		♦Tianjin ◊Wuhan ♦Jinan
			♦Beijing
Wu	▲Hangzhou	ΔWenzhou	▲Hangzhou
	ΔWenzhou	▲Hangzhou ▲Suzhou	ΔWenzhou
	▲Suzhou	▲Shanghai	▲Suzhou
	▲Shanghai		▲Shanghai
Min	□Fuzhou	■Shantou	□Sanming
	■Shantou □Sanming	□Fuzhou ■Quanzhou ■Taizhong	■Quanzhou
	■Quanzhou ■Taizhong		■Taizhong ■Shantou
		□Sanming	□Fuzhou

*↑: ascending order; ↓: descending order

For the degrees of melodiousness, the orders ranked by tone structure metrics, HT:LT and sumT, again are different from each other and also from the one ranked by pitch-based measures. In the case of Mandarin, the most noticeable pattern is that dialects with smaller HT:LT, regardless of their membership in sub-groups, tend to be less melodious. Since Southwestern dialects tend to have smaller HT:LT, they are less melodious. This pattern shows that tone structure can affect Mandarin melody.

For the four Wu dialects, there are no particular connections between sumT- and pitch-based orders, but HT:LT- and pitch-based orders are the same, suggesting that tone structure, especially high tone structure, can affect Wu melody. For the five Min dialects, the three orders do not show any particular pattern, suggesting that tone structure does not affect Min melody.

Note that for the four Cantonese dialects, since they all have close values for the four structure-based metrics, theoretically speaking, they should all have a comparable degree of syllable-timedness and melodiousness. However, both the duration-based order (see Figure 5.7) and the pitch-based order (see Figure 5.15) show vast differences among the four Cantonese dialects, GZ1, GZ2, HK1, and HK2. Therefore, syllable-timedness and melodiousness can vary greatly despite the shared syllable and tone structure.

In general, the relationship between phonological structure and rhythm is a complicated one: Syllable structure and timing are not necessarily correlated nor correlated in the direction of what is expected. Tone structure and melodiousness too are not necessarily correlated nor correlated to the same degree. Admittedly, phonological structure is not even homogeneous within the same major dialect group, so it is not surprising if they do not correlate well with rhythm across the Chinese dialects.

Below are the answer to Question 3 and the discussion of the associated hypothesis:

Question 3: How are syllable and tone structures and Chinese rhythm related at major and sub- dialectal levels?

Answer: Syllable and tone structures are not related to Chinese rhythm for most major dialect groups or their sub-groups and if they do relate, the relationship is not necessarily the same as predicted. Only for Wu dialects, tone structure correlates well with melodiousness as predicted.

Hypothesis 3: Structural complexity is negatively correlated with syllable-timedness and melodiousness at the major but not sub-dialectal level (FALSE).

Discussion:

The orders of syllable-timedness and melodiousness predicted by structure-based metrics are not consistent with the ones derived from duration- and pitch-based based metrics for most Chinese dialects. In other words, more complexity in syllable and tone structures does not necessarily mean less syllable-timedness and less melodiousness.

The lack of correlation between syllable structure- and duration-based metrics can be explained again by the fact that all the Chinese dialects after all are simple in syllable structure, so their relative syllable-timedness depends more on their idiosyncratic differences such as in speech rate and speech content than on their syllable structure differences. As for the tendency for dialects with more complex structure to be more syllable-timed, a close examination of the correlation patterns reveals that the highest correlated pairs always involve durational variability measures including rPVI_IS (for Mandarin), nPVI_Son (for Wu), and ΔSon (for Min & all 21 dialects), meaning that more finals tend to have less durational variability. This pattern is counter-intuitive at first, but

if we consider that more finals mean more restricted varaition in duration, just like more tones mean more restricted varation in pitch excursion size, then it is possible that the durational variability between different finals tends to be small.

The lack of correlation between tone structure- and pitch-based metrics, on the other hand, indicates that the tonal structure differences among major Chinese dialect groups are not as large as they appear to be so as to override the influences from their idiosyncratic differences. Another factor coming into play is perhaps the tone sandhi process. Min, for example, has a very complicated sandhi pattern, so that a high tone in running speech may change to a low tone or a low tone into a high one. If this is the case, then the use of HT:LT and sumT to predict the order of melodiousness becomes impossible for Min dialects. This is perhaps why they fare the worst in ranking the relative melodiousness for Min dialects.

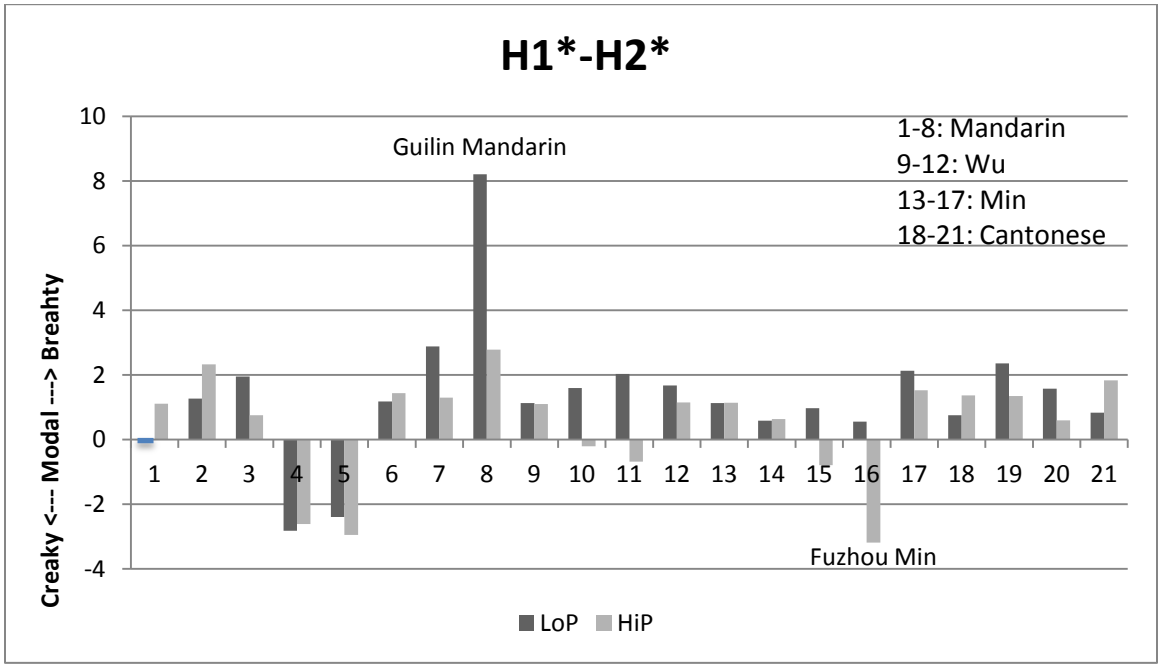
A last point worth mentioning is about the connection between Chinese rhythmic and dialectal classification. According to Tang (2009), Chinese dialectal classification is largely based on shared phonological characteristics, so it is natural to assume that each major dialect group may have its distinct rhythmic pattern since its sub-dialects must share some structural features that can shape such a pattern. This is true in a sense, as rhythmic patterns indeed vary across the four major dialect groups. However, this section just shows that their rhythmic variations are not a result of correlations among structure-, duration-, and pitch-based rhythm. Furthermore, when all the Chinese dialects are compared together in terms of their rhythmic patterns (cf. Section 5.5), it can be seen that most dialects, regardless of their group membership, are actually close in rhythm.

Therefore, there does not seem to be a strong connection between the two types of classifications.

6.4 The relationship between voice quality and rhythm

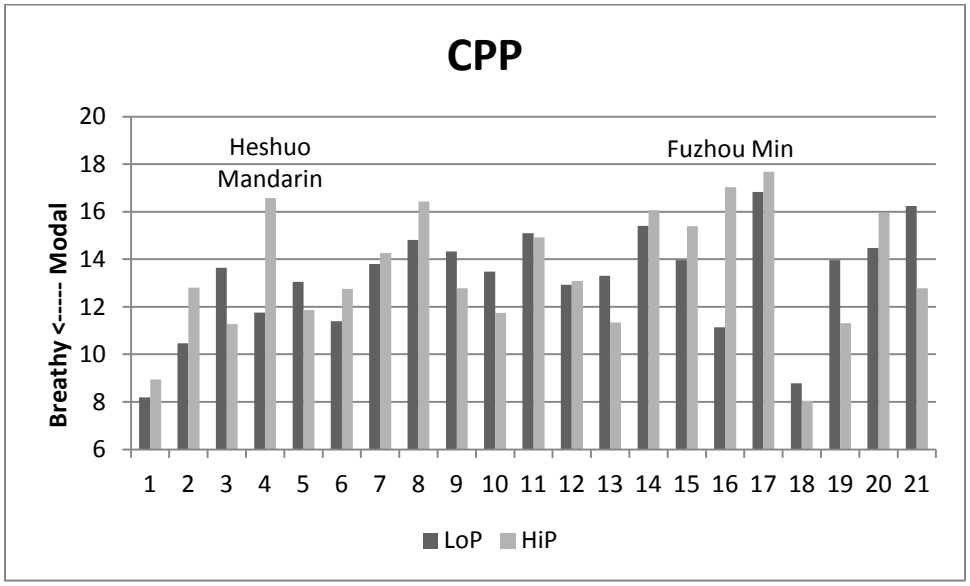
The voice source measures show that low- or high-pitched productions in most dialects are similar in voice quality. As shown in Figure 6.1, most dialects have an absolute H1*-H2* value less than or a little over 2dB, indicating that there is not much voice quality difference between LoP and HiP productions. The few exceptions include the breathiness of the LoP production in Guilin Mandarin (8) and the tenseness of the HiP production in Fuzhou Min (16).

Figure 6.1 Comparison of H1*-H2* between LoP and HiP (all the dialects)



In Figure 6.2, the CPP results show a somewhat different pattern: Heshuo Mandarin (4) and Fuzhou Min (16) have a much larger CPP difference between LoP and HiP productions. Specifically, Heshuo Mandarin LoP production is rather breathy and Fuzhou Min HiP production is rather tense.

Figure 6.2 Comparison of CPP between LoP and HiP (all the dialects)



The breathy quality of Guilin and Heshuo Mandarin LoP production is not unexpected: Although previous studies in Mandrin and Cantonese found that creakiness co-occurs often with low tones (Chao, 1968; Davison, 1991; Belotel-Grenié & Grenié, 2004; Lam & Yu, 2010), but Moisik, Lin, and Esling (2014) also found breathy T3 (a low tone) production in male speakers. Also, the tense quality of Fuzhou Min does not go aganst the reports on the connection between breathiness and low tone production in Wu and Min (Rose, 1989; Cao & Maddieson, 1992; Esposito, 2006; Gao, et al., 2011). The tense quality of HiP can signal the relatively lax quality of LoP in Fuzhou Min. If breathiness of LoP is for the purpose of enhancing perceptual salience of LoP, then tenseness of HiP is equally able to do so by making the sharp contrast between HiP and LoP. However, the question here is not about why there are all these differences in voice quality but rather about how these newly found voice qualities might affect rhythm.

A possible way for voice quality to affect rhythm was discussed by Gordon and Ladefoged (2001). They mentioned that non-modal phonation such as breathiness and creakiness is often associated with increased duration. If this is the case in Chinese, then it is possible that Guilin Mandarin and Fuzhou Min tend to have a longer portion of sonorant duration (a larger %Son) and therefore be more syllable-timed than other Chinese dialects with similar syllable and tone structures.

In the eight Mandarin dialects, the dialect closest to Guilin in syllable and tone structures is Chengdu, but Chengdu is more syllable-timed than Guilin by duration-based measures (see Table 6.4). Similarly, in the five Min dialects, the dialect closest to Fuzhou in syllable and tone structures is Sanming, but Sanming is also more syllable-timed than Fuzhou. Admittedly, these comparisons are too simplistic to yield meaningful results, and a separate study with strict experimental control is warranted. Nonetheless, the voice quality findings in this study can serve as a pointer to a possible direction in future research.

6.5 Summary

So far, all the research questions are answered and the associated hypotheses have been evaluated. These research questions and hypotheses have helped to explore Chinese rhythm and its relationship with phonological structure. None of the duration- and pitch-based metrics have succeeded in consistently ordering the relative syllable-timedness and melodiousness for all 21 dialects. However, pitch-based metrics have shown more success than duration-based ones at the major dialectal level. Most notably, Wu dialects can be quantified consistently by pitch-based metrics and their melodiousness correlates well with their high tone structures. Various correlation patterns are also found between

syllable-timedness and melodiousness at both major and sub- dialectal levels. When compared across all 21 Chinese dialects, neither duration- nor pitch-based rhythmic patterns reflect dialectal grouping. Most Chinese dialects, In fact, have similar degrees of syllable-timedness and melodiousness, regardless of their group membership.

The results of this study have some important implications for Chinese rhythm research. For Chinese, pitch rather than duration seems to play a major role in shaping rhythm, as pitch-based metrics fare better than duration-based ones and melody patterns are more distinct than timing patterns. Since pitch variation in Chinese dialects is mainly a result of tonal production, Chinese rhythmic research could focus on investigating the influence of tone on rhythm and seek to establish a formal connection between tone structure and melody. Also, inconsistency of duration-related correlation results suggests that Chinese timing patterns are more complicated than what current rhythmic metrics can capture. Given the uniqueness of Chinese as a monosyllabic and tonal language, previous notions of syllable- versus stress-timing may not be applicable to Chinese rhythm. Therefore, it is necessary for Chinese rhythm research to examine what timing really means for Chinese and then to devise a set of metrics capable of capturing Chinese rhythm.

Chapter 7
CONCLUSION

This dissertation concludes with a summary of the implications of the present study for rhytmic research in three areas: major contributions (Section 7.1), limitations (Section 7.2), and future directions (Section 7.3).

7.1 Major contributions

As mentioned in the introduction, previous research on Chinese rhythm focuses mostly on the two best known varieties of Chinese, namely, Mandarin and Cantonese and only on the timing perspective. The first systematic study of Chinese rhythm at the major dialectal level, to my knowledge, was the one conducted by Lin and Huang (2009). This study is an extension to Lin and Huang's (2009) study in terms of both breadth and depth, because it is the first systematic attempt to explore Chinese rhythmic patterns not only across multiple Chinese dialects but also from three perspectives, phonological structure, timing, and melody.

In addition, this study provides a methodological foundation for future research, as it is not only the first to use the PE- and PS-based metrics to compare melody patterns of Chinese dialects but also the first to establish the use of the four structure-based metrics, Fin:Ini, sumFI, HT:LT, and sumT, in quantifying the complexity level of Chinese syllable and tone structures. These novel metrics have shown some degrees of success in predicting duration- and pitch-based rhythm.

Lastly, this study contributes to our knowledge of Chinese rhythm by showing various timing- and melody patterns across Chinese dialect groups. It not only provides some

basis for Chinese rhythmic typology but also demonstrates the inadequacy of rhythmic measures in quantifying Chinese rhythm. By examining speech rhythm from both duration- and pitch-based perspectives, this study also reveals inherent relationships between timing and melody in Chinese dialects. By using phonological structure measures, this study further identifies the roles played by phonological structure. These roles, despite being insignificant in most cases, do help shape rhythmic patterns of some Chinese dialects. In general, this study as a piece of pioneering work helps to shed light on similar research in the future.

7.2 Limitations of the study

The main limitation of this study is data sourcing. Since all speech was uncontrolled natural speech, it is hard to match speech style and content across dialects. Every effort has been made by the researcher to select speech recordings with a similar genre. Generally, a stretch of speech was selected if it is about a personal story narrated with a usual rather than dramatic tone and uttered with certain spontaneity rather than read from a script (though the speaker may have rehearsed multiple times before making a formal recording). Note that the number of dialects for each major or sub- dialect group is also not matched, just for this very reason: it is hard to find even two stretches of acceptable speech in each dialect. Therefore, the rhythmic pattern of each dialect revealed in this study may change when more dialects and more controlled stretches of speech are involved. Nonetheless, the purpose of this study is not to find an accurate rhythmic pattern for each dialect but to provide some general pictures regarding how rhythm is associated with timing and melody patterns in Chinese dialects. This latter purpose is basically fulfilled given the above limitations in data sourcing.

Another limitation is related to the speech analysis process. Since any software programs will have some technical limitation, inaccurate analysis of speech may occur. The researcher has made every effort to ensure relatively accurate calculation of metrics. Some of these efforts include individualizing pitch range when calculating pitch-based metrics to avoid pitch errors inherent in the algorithm and manually adjusting inaccurate segmentation. Given the current technical means, it is not possible to eliminate all types of calculation inaccuracies. The possible remedy is to use a large amount of data to override the effect of calculation errors. This is just what this study has done. Even though there may be some errors occurring in the data analysis, some patterns still emerge and are informative for future research.

7.3 Future directions

As pioneering work, this study has much potential for further research. They can be expanded into different directions. First, it can be divided into a series of studies, each of which focuses on one major group of Chinese dialects with its sub-groups. In so doing, Chinese rhythmic typology can be established based on an extensive survey of Chinese rhythm at the dialectal level. Second, the study of Chinese rhythm can be extended to include different speech styles. As mentioned earlier, all the data in this study are casual personal narratives, so their rhythmic patterns may be different from those of broadcasting style or other styles. It would be interesting to see how style can influence Chinese rhythm. Third, this study can even become a second language study. As almost every Chinese speaker today is bilingual in terms of their ability to speak both the standard Mandarin and a regional variety of Chinese, a study of Mandarin rhythm in speech produced by non-native Mandarin speakers may provide some insight for

language acquisition in general. Also, the influence of voice quality on rhythm can be vigorously pursued, as little literature is available in this area. Last but not least, this study can also include a perception component. Rhythm by its definition is perceptual in nature, so it will be more enlightening to incorporate production and perception into rhythm research. By and large, this study opens up a host of possibilities for future research and it would be exciting to see if Chinese rhythm research can proliferate in these directions.

BIBLIOGRAPHY

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Altmann, H. (2006). *The perception and production of second language stress: a cross-linguistic experimental study*. Doctoral dissertation, University of Delaware.

Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica, 66(1-2),* 46-63.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40(3),* 351-373.

Aske, J. (1990). Disembodied rules versus patterns in the lexicon: testing the psychological reality of Spanish stress rules. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society 16* (pp. 30-45).

Auer, P. (1993). Is a Rhythm-based Typology Possible? A Study of the Role of Prosody in Phonological Typology. *KontRI Working Paper No. 21.* Universität, Fachgruppe Sprachwissenschaft.

Avanzi, M., Obin, N., Bardiaux, A., & Bordal, G. (2012). Speech Prosody of French Regional Varieties. In *Proceedings of Speech Prosody 2012* (pp. 603-606).

Baidupedia (Baidu baike). Retrieved multiple times in 2014 & 2015, from http://baike.baidu.com/

Bakovic, E. (2014). Exceptionality in Spanish stress. In Rafael, A. (Ed.). *The syllable in Romance languages: studies in honor of James W. Harris*. Nuñez Cedeño. Berlin & New York: Mouton de Gruyter.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology, 3(01),* 255-309.

Belotel-Grenié, A., & Grenié, M. (2004). The creaky voice phonation and the organisation of Chinese discourse. *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages,* Beijing, China.

Benadon, F. (2014). Metrical perception of trisyllabic speech rhythms. *Psychological research, 78(1),* 113-123.

Bird, S., & Wang, Q. (2009). *LING 380: Acoustic phonetics lab manual*. Department of Linguistics, University of Victoria.

Blankenship, B. (1997). *The time course of breathiness and laryngealization in vowels.* Doctoral dissertation, University of California, Los Angeles.

Botinis, A. (1989). *Stress and prosodic structure in Greek: a phonological, acoustic, physiological and perceptual study*. Lund: Lund University Press.

Brugos, A. & Barnes, J. (2012). Pitch trumps duration in a grouping perception task. *The 25th Annual CUNY Conference on Human Sentence Processing*, New York, NY.

Brugos, A., & Barnes, J. (2014). Effects of dynamic pitch and relative scaling on the perception of duration and prosodic grouping in American English. *Speech Prosody 7.* 388-392.

Cantonese lexicon (Jyutjyu samjam puici zifu). Retrieved multiple times in 2014 & 2015, from http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/

Cao, J. F., & Maddieson, I. (1992). An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics, 20,* 77-92.

Cao, J. F. (2004). Intonation structure of spoken Chinese: universality and characteristics. *Report of Phonetic Research* (pp. 31-38). Beijing: Institute of Linguistics, CASS.

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

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

Chen, M. Y. 2000. *Tone sandhi: patterns across Chinese dialects*. Cambridge University Press.

Cheng, R. L. (1985). A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language, 61(2),* 352-377.

Cheng, R. (1997). *Taiwanese and Mandarin structures and their developmental trends in Taiwan I: Taiwanese phonology and morphology*. Taipei: Yuan-Liou Publishing Co., Ltd.

Chinese dialect lexicon (Hanyu fangyin zihui). (1989). Department of Chinese Language and Literature, Beijing University. Beijing: Wenzi Gaige Press.

Cho, B. E. (2004). Issues concerning Korean learners of English: English education in Korea and some common difficulties of Korean students. *The East Asian Learner, 1(2),* 31-36.

Cohen, J., Hansel, C. E. M., & Sylvester, J. (1954). An experimental study of comparative judgements of time. *British Journal of Psychology. General Section, 45(2),* 108-114.

Cruz, M. (2013). *Prosodic variation in European Portuguese: phrasing, intonation and rhythm in Central-Southern varieties.* Doctoral dissertation, University of Lisbon.

Crystal, D. (1985). *A dictionary of linguistics and phonetics*. Oxford: Blackwell.

Cumming, R. E. (2010). *Speech rhythm: the language-specific integration of pitch and duration*. Doctoral dissertation, University of Cambridge.

Cumming, R. E. (2011). The interdependence of tonal and durational cues in the perception of rhythmic groups. *Phonetica, 67(4),* 219-242.

Database of the provincial government of Fujian. Retrieved multiple times in 2015, from http://www.fjsq.gov.cn/DSXZ.ASP

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*, 51-62.

Davison, D. S. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics, 78,* 50-57.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In Karnowski, P., & Szigeti, I. (Eds.). *Language and Language-Processing: Proceedings of the 38th Linguistic Colloquium* (pp. 231-241). Piliscsaba 2003. Frankfurt: Peter Lang.

Dellwo, V., Aschenberner, B., Dancovičová, J., Steiner, I., & Wagner, P. (2004). BonnTempo-corpus and BonnTempo tools: a database for the study of speech rhythm and rate. In *Proceedings of Interspeech 2004* (pp. 777-780). Jeju Island, Korea.

Dellwo, V., Fourcin, A., & Abberton, E. (2007). Rhythmical classification of languages based on voice parameters. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 1129-1132). Saarbrücken.

Donegan, P. J., & Stampe, D. (1983). Rhythm and the holistic organization of language structure. In *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax* (pp. 337-353). Chicago Linguistic Society.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word

segmentation and lexical processing. *Journal of Memory and Language, 59(3),* 294-311.

Duanmu, S. 2004. Left-headed feet and phrasal stress in Chinese. *Cahiers de Linguistique Asie Orientale, 33(1),* 65-103.

Duanmu, S. (2005). The tone-syntax interface in Chinese: some recent controversies. In *Proceedings of the Symposium "Cross-Linguistic Studies of Tonal Phenomena, Historical Development, Tone-Syntax Interface, and Descriptive Studies* (pp. 16-17).

Edmondson, J. A., Esling, J. H., Ziwo-Lama, Harris, J. G., & Li, S. N. (2001). The aryepiglottic folds and voice quality in the Yi and Bai languages: laryngoscopic case studies. *Mon-Khmer Studies, 31,* 83-100.

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies. *Phonology, 23(2),* 157-191.

Esling, J. H. (2005). There are no back vowels: the laryngeal articulator model. *The Canadian Journal of Linguistics, 50(1),* 13-44.

Esposito, C. M. (2006). *The effects of linguistic experience on the perception of phonation*. Doctoral dissertation, University of California, Los Angeles.

Faure, G., Hirst, D. J., & Chafcouloff, M. (1980). Rhythm in English: isochronism, pitch, and perceived stress. In Waught, L. R., & van Schooneveld, C. H. (Eds.). *The melody of language* (pp. 71-79). Baltimore, MD: University Park Press.

Fraisse, P. (1982). Rhythm and tempo. *The Psychology of Music, 1,* 149-180.

Fon, J., & Chiang, W. (1999). What does Chao have to say about tones? *Journal of Chinese Linguistics, 27(1),* 13-37.

Fourcin, A., & Dellwo, V. (2009). Rhythmic classification of languages based on voice timing. Department of Speech, Hearing, and Phonetic Sciences. UCL: London, UK.

Fox, A. (2000). *Prosodic features and prosodic structure: the phonology of suprasegmentals*. Oxford University Press.

Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus, 13(2),* 247-275.

Galves, A., Garcia, J., Duarte, D., & Galves, C. (2002). Sonority as a basis for rhythmic class discrimination. *Speech Prosody 2002*.

Gao, J., Hallé, P., Honda, K., Maeda, S., & Toda, M. (2011). Shanghai slack voice: acoustic and EPGG data. In *Proceedings of the 17th International Congress on Phonetic Sciences* (pp. 719-722).

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical society of America, 63(1),* 223-230.

Gil, D. (1986). A prosodic typology of language. *Folia Linguistica, 20(1-2),* 165-232.

Gobl, C., & NíChasaide, A. (1992). Acoustic characteristics of voice quality. *Speech Communication, 11(4),* 481-490.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics, 29(4),* 383-406.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology, 7,* 515-546.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Han, M. S. (1962). The feature of duration in Japanese. *onsei no kenkyuu (Phonetic Studies), 10,* 65-80.

Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America, 106(2),* 1064-1077.

Harrington, J., Hoole, P., & Pouplier, M. (2013). New directions in speech production. In Jones, M. J., & Knight, R.-A. (Eds.). *The Bloomsbury Companion to Phonetics* (pp. 242-259). A&C Black.

Hayes, B. (1995). *Metrical stress theory: principles and case studies*. University of Chicago Press.

Henry, M. & McAuley, J. (2009). Evaluation of an imputed pitch velocity model of the auditory kappa effect. *Journal of Experimental Psychology: Human Perception and Performance, 35(2),* 551–564.

Henry, M. J., McAuley, J. D., & Zaleha, M. (2009). Evaluation of an imputed pitch velocity model of the auditory tau effect. *Attention, Perception, and Psychophysics, 71(6),* 1399-1413.

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research, 39,* 311-321.

Hirst, D., & Di Cristo, A. (1998). *Intonation systems: a survey of twenty languages*. Cambridge University Press.

Hirst, D. (2011). The analysis by synthesis of speech melody: from data to models. *Journal of Speech Sciences, 1(1),* 55-83.

Hirst, D. (2013). Melody metrics for prosodic typology: comparing English, French, and Chinese. In *Proceedings of Interspeech 2013* (pp. 572-576).

Horton, R., & Arvaniti, A. (2013). Clusters and Classes in the Rhythm Metrics.

Huang, Y. H., & Fon, J. (2008). Dialectal variations in tonal register and declination pattern of Taiwan Mandarin. In *Proceedings of the 4th International Conference on Speech Prosody* (pp. 605-608).

Huang, Y. H., & Fon, J. (2011). Investigating the effect of Min on dialectal variations of Mandarin tonal realization. In *Proceedings of 17th ICPhS* (pp. 918-921). Hong Kong.

Huang, Y. H., Wu, E. C., & Fon, J. (2012). The effect of Min proficiency on production and perception of tones in Taiwan Mandarin. In *Proceedings of 6th International Conference on Speech Prosody* (pp. 22-25). Shanghai.

Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America, 81(2),* 495-504.

Iseli, M., Shue, Y. L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America, 121(4),* 2283-2295.

Jun, S.-A. (2005). Prosodic Typology. In Jun, S.-A. (Ed.). *Prosodic typology II: the phonology of intonation and phrasing*. Oxford University Press.

Jun, S. A. (2011). Tone-based macro-rhythm from the perspective of prosodic typology. *Journal of the Acoustical Society of America, 130(4),* 2471-2471.

Jun, S. A. (2012). Prosodic Typology Revisited: Adding Macro-Rhythm. *Speech Prosody 6,* Shanghai, China.

Keating, P. A., & Esposito, C. (2006). Linguistic voice quality. *The 11th Australasian International Conference on Speech Science and Technology*.

Keating, P., Kuang, J., Esposito, C., Garellek, M., & Khan, S. (2012). Multi-dimensional phonetic space for phonation contrasts. *Laboratory Phonology 13*.

Keating, P., & Garellek, M. 2015. Acoustic analysis of creaky voice. *LSA Annual meeting*.

Kohler, K. J. (2009a). The perception of prominence patterns. *Phonetica, 65(4),* 257-269.

Kohler, K. J. (2009b). Wither speech rhythm research? *Phonetica, 66(1-2),* 5-14.

Kuang, J. (2011). *Production and perception of the phonation contrast in Yi.* MA thesis, University of California, Los Angeles.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Lam, H. W., & Yu, K. M. (2010). The role of creaky voice quality in Cantonese tonal perception. *Journal of the Acoustical Society of America, 127(3),* 2023.

Lee, H. B., Jin, N. T., Seong, C. J., Jung, I. J., & Lee, S. M. (1994). An experimental phonetic study of speech rhythm in Standard Korean. *The 3rd International Conference on Spoken Language Processing*.

Leemann, A., Dellwo, V., Kolly, M. J., & Schmid, S. (2012). Rhythmic variability in Swiss German dialects. In *Proceedings of the 6th International Conference on Speech Prosody* (pp. 607-610).

Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics, 8,* 469-474.

Lehnert-LeHouillier, H. (2007). The influence of dynamic f0 on the perception of vowel duration: cross-linguistic evidence. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 757-760).

Liberman, M. (1975). *The intonational system of English*. Doctoral dissertation, Massachusetts Institute of Technology.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry 8,* 249-336.

Lin, H. (1996). *Mandarin tonology*. Taibei, Taiwan: Pyramid Press.

Lin, H. (1999). Putonghua qingsheng diaozhi de zonghe fenxi (A unified analysis of the neutral tone values in Mandarin). In Lu, S. N. (Ed). *Xiandai yuyinxue lunwenji (Papers on modern phonetics and phonology)* (pp. 175-183). Beijing, China: Golden City Press.

Lin, H. (2001a). *A grammar of Mandarin Chinese*. Lincom Europa.

Lin, H. (2001b). Stress and the distribution of the neutral tone in Mandarin. In Xu, D. B. (Ed.). *Chinese phonology in generative grammar* (pp. 139-161). San Diego: Academic Press.

Lin, H., & Wang, Q. (2005). Vowel quantity and consonant variance: A comparison between Chinese and English. In *Proceedings of "Between Stress and Tone."* Leiden, Holland.

Lin, H. (2006). Mandarin neutral tone as a phonologically low tone. *Journal of Chinese Language and Computing, 16(2),* 121-134.

Lin, H., & Wang, Q. (2007). Mandarin rhythm: an acoustic study. *Journal of Chinese Linguistics and Computing, 17(3),* 127-140.

Lin, H, & Huang, S-M. (2009). Rhythm in the Chinese dialects. *The Annual Canadian Linguistics Association Conference*, Ottawa.

Lloyd, J. (1940). *Speech signal in telephony*. London: Sir I. Pitman & Sons, Limited.

Loukina, A., Kochanski, G., Shih, C., Keane, E., & Watson, I. (2009). Rhythm measures with language-independent segmentation. In *Proceedings of Interspeech 2009* (pp. 1531-1534).

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C.-L. (2011). Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America*, *129*, 3258.

MacKenzie, N. (2007). *The kappa effect in pitch/time context*. Doctoral dissertation, Ohio State University.

Maddieson, I., & Ladefoged, P. (1985). "Tense" and "lax" in four minority languages of China. *UCLA Working Papers in Phonetics*, *60*, 59-83.

Maddieson, I. (2011). Typology of phonological systems. In Song, J. J. (Ed.). *The Oxford handbook of linguistic typology*. Oxford University Press.

Matthews, S., & Yip, V. (2013). *Cantonese: a comprehensive grammar*. New York: Routledge.

Mertens, P. (2012). *Prosogram v29f*. http://bach.arts.kuleuven.be/pmertens/prosogram/

Moisik, S. (2013). *The epilarynx in speech*. Doctoral dissertation. University of Victoria.

Moisik, S. R., & Esling, J. H. (2011). The 'whole larynx' approach to laryngeal features. In *Proceedings of ICPhS* (pp.1406-1409). Hong Kong.

Moisik, S. R., Lin, H., Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound. *Journal of the International Phonetic Association, 44,* 21-58.

Mok, P., & Lee, S. I. (2008). Korean speech rhythm using rhythmic measures. In *Proceedings of the 18th International Congress of Linguists.* Seoul, Korea.

Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics, 2,* 148-154.

Mok, P., & Wong, P. (2010). Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables. *Speech Prosody 2010.* Chicago.

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Experimental Psychology, 24(3),* 756-766.

Nazzi, T., Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41,* 233–243.

Nespor, M., & Vogel, I. 1986. *Prosodic phonology*. Dordrecht: Foris.

Pan, H. H. (2005). Voice quality of falling tones in Taiwan Min. In *Proceedings of Interspeech* (pp. 1401-1404).

Pike, K. L. (1946). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Pisoni, D. B. (1976). Fundamental frequency and perceived vowel duration. *Journal of the Acoustical Society of America, 59,* S39.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73(3),* 265-292.

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: perspectives. *Speech Prosody 2002*.

Rose, P. (1989). Phonetics and phonology of Yang tone; phonation types in Zhenhai. *Cahiers de Linguistique-Asie Orientale, 18 (2),* 229-245.

Schiering, R., Bickel, B., & Hildebrandt, K. A. (2012). Stress-timed = word-based?: testing a hypothesis in prosodic typology. *Language Typology and Universals, 65(2),* 157-168.

Schmid, S. (2012). Phonological typology, rhythm types, and the phonetics-phonology interface: a methodological overview and three case studies on Italo-Romance dialects. In *Methods in contemporary linguistics. A festschrift in honour of Iwar Werlen* (pp. 45-68). Mouton de Gruyter, Berlin/New York.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology, 3(01),* 371-405.

Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception and Psychophysics, 40,* 9-19.

Shue, Y.-L., P. Keating , C. Vicenik, K. Yu. (2011). VoiceSauce: a program for voice analysis. In *Proceedings of the ICPhS XVII* (pp. 1846-1849). http://www.phonetics.ucla.edu/voicesauce/

Silverman, D., Blankenship, B., Kirk, P., & Ladefoged, P. (1995). Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics, 37(1),* 70-88.

Stojanović, D. (2013). *Cross-linguistic comparison of rhythmic and phonotactic similarity.* Doctoral dissertation, University of Hawaii at Manoa.

Tang, C. (2009). *Mutual intelligibility of Chinese dialects: an experimental approach*. Doctoral dissertation, Leiden University.

Traill, A., & Jackson, M. (1988). Speaker variation and phonation type in Tsonga nasals. *Journal of Phonetics, 16 (4),* 385-400.

Tseng, (2006). Chapter 3 Prosodic analysis. In Lee, C-H., Li, H. Z., Lee, L-S., Wang, R. H., & Huo, Q. (Eds.). *Advances in Chinese spoken language processing* (pp. 57-76). Singapore: World Scientific Publishing.

Van Heuven, V. J. J. P., & Sluijter, A. M. (1996). Notes on the phonetics of word prosody. In Geodemans, R., van der Hulst, H., & Visch, E. (Eds.). *Stress patterns of the world, part 1: background* (pp. 233-269). The Hague: Holland Academic Graphics.

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In Jun, S.-A. (Ed.). *Prosodic typology II: the phonology of intonation and phrasing* (pp. 172-200). Oxford University Press.

Vicenik, C., & Sundara, M. (2008). The role of segmental and intonational cues in dialect discrimination. *Journal of the Acoustical Society of America, 123(5),* 3883.

Vicenik, C., & Sundara, M. (2013). The role of intonation in language and dialect discrimination by adults. *Journal of Phonetics, 41(5),* 297-306.

Wang, W. S. Y., Lehiste, I., Chuang, C. K., Darnovsky, N. (1976). Perception of vowel duration. *Journal of the Acoustical Society of America, 60,* S92.

Wang, Y. J., Jeong, D. Y., & Feng, J. L. 2012. The effect of intonation on Tone 2 and the register of Tone 2 in Mandarin: A preliminary study. *Tonal Aspects of Languages-Third International Symposium.*

Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10(2),* 193-216.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35(4),* 501-522.

Wikimedia. Retrieved multiple times in 2014 & 2015, from http://commons.wikimedia.org/wiki/

Wu, Z-J, (1990). "Can polysyllabic tone-sandhi patterns be the invariant units of intonation in spoken Standard Chinese? *International Conference on Spoken Language Processing 1990.* Kobe, Japan.

Xiao, Y. F. (2013). Ancient sounds and words in Liupanshui Chinese dialects. *Journal of Liupanshui Normal University, 25(4),* 74-79.

Xu, Y. (2013). *ProsodyPro5.3.2.* http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/

Yip, M. (1992). Tonal register in East Asian languages. In van der Hulst, H., & Snider, K. L. (Eds.). *The phonology of tone: the representation of tonal register* (pp. 245-268). Berlin: Mouton de Gruyter.

Yoblick, D. A., & Salvendy, G. (1970). Influence of frequency on the estimation of time for auditory, visual, and tactile modalities: the kappa effect. *Journal of Experimental Psychology, 86(2),* 157.

Yoon, T. J. (2008). *Rhythm_CSJ (a Praat script) & PVI (a Python script).* Acquired through personal communication.

Yu, A. C. L. (2010). Tonal effects on perceived vowel duration. *Laboratory Phonology 10*. Berlin: Mouton de Gruyter.

Zheng, W. (2008). *A study of phonological development in Taihu Wu (Taihu Pian Wuyu yinyun yanbian yanjiu).* Doctoral dissertation, University of California, Los Angeles.

Zhou, Q. H. (2006). The differences between Singapore Mandarin and the standard Mandarin and how to deal with the differences. Published in series on March 21-23 in *Singapore United Morning Newspaper*.

Appendix 1

Sound Inventory of the Eight Mandarin Dialects

The number of initials and finals of the eight Mandarin dialects:

Sub-group	*No.*	*Dialect*	*Fin*	*Ini*	*Fin:Ini*	*sumFI*
Northern	1	BJ (Beijing)	40	22	1.82	62
	2	TJ (Tianjin)	37	24	1.54	61
	3	JN (Jinan)	38	24	1.58	62
	4	HS (Heshuo)	39	22	1.77	61
Southwestern	5	WH (Wuhan)	37	19	1.95	56
	6	CD (Chengdu)	36	20	1.8	56
	7	LPS (Liupanshui)	32	19	1.68	51
	8	GL (Guilin)	35	18	1.94	53

Initials	1	2	3	4	5	6	7	8	9	10	11	12
BJ (22)	p	t	ts	tʂ	tɕ	k	l	x	Ø	s	m	
	p^h	t^h	ts^h	$tʂ^h$	$tɕ^h$	k^h	n	ɕ	ʐ	ʂ	f	
TJ (24)	p	t	ts	tʂ	tɕ	k	l	x	Ø	s	v	m
	p^h	t^h	ts^h	$tʂ^h$	$tɕ^h$	k^h	n	ɕ	ʐ	ʂ	f	ŋ
JN (24)	p	t	ts	tʂ	tɕ	k	l	x	Ø	s	m	ŋ
	p^h	t^h	ts^h	$tʂ^h$	$tɕ^h$	k^h	n	ɕ	ʐ	ʂ	f	ɳ
HS (22)	p	t	ts	tʂ	tɕ	k	l	x	Ø	s	m	
	p^h	t^h	ts^h	$tʂ^h$	$tɕ^h$	k^h	n	ɕ	ʐ	ʂ	f	
WH (19)	p	t	ts	tɕ	k	n	s	m	Ø			
	p^h	t^h	ts^h	$tɕ^h$	k^h	ŋ	ɕ	ɹ	f	x		
CD (20)	p	t	ts	tɕ	k	n	ɳ	s	m	Ø		
	p^h	t^h	ts^h	$tɕ^h$	k^h	ŋ	ɕ	z	f	x		
LPS (19)	p	t	ts	tɕ	k	ɕ	s	m	Ø			
	p^h	t^h	ts^h	$tɕ^h$	k^h	ŋ	z	f	x	l		
GL (18)	p	t	ts	tɕ	k	n	s	m	Ø			
	p^h	t^h	ts^h	$tɕ^h$	k^h	ŋ	ɕ	f	x			

Finals

V(VV)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
BJ	a	o	ɤ	i	u	y	ɿ	ʅ	ɚ	m̩						
(40)	ai	au	ou	ia	ie	ua	uo	ye	iau	iou	uai	uei	iai	yan		
	an	aŋ	ən	əŋ	in	iŋ	uŋ	yn	yŋ	iɛn	iaŋ	uan	uaŋ	uən	uəŋ	iuŋ
TJ	a	ɔ	ɤ	i	u	y	ɿ	ʅ	ɚ							
(37)	ai	au	ɤu	ia	iɛ	ua	uɔ	yœ	iau	iɤu	uai	uei	ei			
	an	aŋ	ən	əŋ	in	iŋ	uŋ	yn	yŋ	iɛn	yɛn	iaŋ	uaŋ	uən	uan	
JN	a	ɔ	ɛ	i	u	y	ɿ	ʅ	ɚ	ɤ						
(38)	ia	ie	iɛ	iɔ	ua	uɤ	uɛ	ye	ei	ou	iou	uei				
	aŋ	əŋ	iŋ	uŋ	iaŋ	iuŋ	uaŋ	uəŋ	æ̃	ẽ	iæ̃	iẽ	uẽ	uæ̃	yæ̃	yẽ
HS	a	o	ɤ	i	u	y	ɿ	ʅ	ɚ	m̩						
(39)	ai	au	ou	ia	ie	ua	uo	ye	iau	iou	uai	uei	yan			
	an	aŋ	ən	əŋ	in	iŋ	uŋ	yn	ioŋ	iɛn	iaŋ	uan	uaŋ	uən	uəŋ	iuŋ
WH	a	o	ɤ	i	u	y	ɿ	ɯ								
(37)	ai	au	ia	ie	ua	ue	yɛ	uɤ	io	ei	iau	iou	uai	uei	iɛi	ou
	an	aŋ	ən	in	yn	oŋ	iaŋ	uan	uaŋ	uən	joŋ	iɛn	yɛn			
CD	a	o	ɤ	i	u	y	ɿ	ɚ								
(36)	ai	au	ia	ie	ua	ue	ye	yo	ɛi	ei	iau	iəu	uai	uei	iɛi	
	an	aŋ	ən	in	yn	oŋ	iaŋ	uan	uaŋ	uən	yoŋ	yɛn	iɛn			
LPS	a	o	e	ɚ	i	u	ɿ									
(32)	ia	io	ie	ua	ue	ai	ei	au	əu	iau	iəu	iu	uai	uei		
	an	ən	aŋ	oŋ	in	ien	uan	uən	iɑŋ	ioŋ	uɑŋ					
GL	a	o	e	ə	i	u	y	ʅ								
(35)	ia	io	ie	ua	ye	yu	æi	əu	əi	iau	iəu	uæi	uəi	ua		
	ã	ən	ɑŋ	oŋ	ĩẽ	ũã	yẽ	in	un	yn	iɑŋ	ioŋ	uɑŋ			

Appendix 2

Sound Inventory of the Four Wu Dialects

The number of initials and finals of the four Wu dialects:

Sub-group	*No.*	*Dialect*	*Fin*	*Ini*	*Fin:Ini*	*sumFI*
Northern	1	HZ (Hangzhou)	39	30	1.3	69
	2	SH (Shanghai)	43	28	1.54	71
	3	SZ(Suzhou)	49	27	1.81	76
Southern	4	WZ (Wenzhou)	34	29	1.17	63

Initials	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
HZ(30)	p	t	ts	tɕ	k	b	d	dʑ	g	s	h	f	ɕ	l	Ø
	pʰ	tʰ	tsʰ	tɕʰ	kʰ	m	n	ɳ	ŋ	z	ɦ	v	ʑ	ʔ	ʊ
SH(28)	p	t	ts	tɕ	k	b	d	dʑ	g	s	h	f	ɕ	l	Ø
	pʰ	tʰ	tsʰ	tɕʰ	kʰ	m	n	ɳ	ŋ	z	ɦ	v	ʑ		
SZ (27)	p	t	ts	tɕ	k	b	d	dʑ	g	s	h	f	ɕ	l	Ø
	pʰ	tʰ	tsʰ	tɕʰ	kʰ	m	n	ɳ	ŋ	z	ɦ	v			
WZ (29)	p	t	ts	tɕ	k	b	d	dʑ	g	s	h	f	ɕ	l	Ø
	pʰ	tʰ	tsʰ	tɕʰ	kʰ	m	n	ɳ	ŋ	z	ɦ	v	dz	j	

Finals	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
HZ	ɑ	o	ʏ	ɔ	ɿ	ɪ	u	y	ɥ	ɛ	m̩	ŋ̍				
(39)	iɑ	iɔ	ʏo	eɪ	iɛ	uɛ	uo	uɑ	ou	ər	ueɪ					
	ɑŋ	iɑŋ	uɑŋ	ym	ən	in	uən	oŋ	ioŋ							
	ɐʔ	ɔʔ	iɔʔ	uɔʔ	uɐʔ	iɪʔ	yɪʔ									
SH	a	o	ʏ	ɔ	ɿ	ɪ	u	y	ø	ɛ	m̩	n̩	ŋ̍	aʔ	əʔ	oʔ
(43)	ia	iɔ	iʏ	yø	iɛ	uɛ	uø	yø	əl	iaʔ	uaʔ	iɪʔ	ioʔ	uəʔ	yɪʔ	
	ã	ɑ̃	ĩã	ũã	ĩɑ̃	ũɑ̃	ən	in	uən	yn	oŋ	ioŋ				
SZ	a	o	ʏ	ɥ	ɿ	i	u	y	ø	ɛ	ɐ	m̩	n̩	ŋ̍		
(49)	ia	io	iʏ	iø	iɪ	ua	uø	əu	əl	iæ						
	ã	ẽ	ɑ̃	ĩɑ̃	ũɑ̃	ĩã	ũã	ən	in	uən	yən	oŋ	ioŋ			
	aʔ	əʔ	oʔ	ɑʔ	iaʔ	iəʔ	ioʔ	uaʔ	uəʔ	yaʔ	yəʔ	iɑʔ				
WZ	a	o	ɛ	ɔ	ø	e	u	y	ɿ	ɛ	m̩	ŋ̍				
(34)	ia	iɛ	iai	iau	iəu	uɛ	uɔ	ua	uai	yɔ	yo	øy	əu	ai	ei	au
	aŋ	eŋ	oŋ	iaŋ	uaŋ	yoŋ										

Appendix 3

Sound Inventory of the Four Cantonese Dialects

The number of initials and finals of the two Cantonese dialects (four speakers, two speakers per dialect):

Sub-group	*No.*	*Dialect*	*Fin*	*Ini*	*Fin:Ini*	*sumFI*
GZ (Guangzhou)	1	GZ1	68	18	3.78	86
	2	GZ2	68	18	3.78	86
HK (Hong Kong)	3	HK1	70	18	3.89	88
	4	HK2	70	18	3.89	88

Initials	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GZ(18)	p	pʰ	t	tʰ	ʧ	ʧʰ	ʃ	k	kʰ	h	Ø	m	l	ŋ	n	j	w	f
HK(18)	p	pʰ	t	tʰ	ʧ	ʧʰ	ʃ	k	kʰ	h	Ø	m	l	ŋ	n	j	w	f
(2)								kw	kʰw									

Finals	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GZ	a	ɛ	i	ɔ	œ	u	y	m̩	ŋ̩									
(54)	ai	ɐi	au	ɐu	ei	-	iu	ɔi	ou	øy	ui							
(14)	ua	uai	uɐi	uɔ														
	am	an	aŋ	ɐm	ɐn	ɐŋ	-	ɛŋ	im	in	iŋ	ɔn	ɔŋ	œŋ	øn	un	ʊŋ	yn
		uan	uaŋ		uɐn	uɐŋ					uiŋ		uɔŋ					
	ap	at	ak	ɐp	ɐt	ɐk	ɛp	ɛk	ip	it	ik	ɔt	ɔk	œk	øt	ut	ʊk	yt
		uat	uɐt								uik						ʊak	
HK	a	ɛ	i	ɔ	œ	u	y	m̩	ŋ̩									
(56)	ai	ɐi	au	ɐu	ei	ɛu	iu	ɔi	ou	øy	ui							
	am	an	aŋ	ɐm	ɐn	ɐŋ	ɛm	ɛŋ	im	in	iŋ	ɔn	ɔŋ	œŋ	øn	un	ʊŋ	yn
	ap	at	ak	ɐp	ɐt	ɐk	ɛp	ɛk	ip	it	ik	ɔt	ɔk	œk	øt	ut	ʊk	yt

Note:

The way GZ and HK Cantonese count initials and finals is different: For GZ, /kw, kʰw/ (see the red font in HK Initials) are not counted as initials, so the number of initials is 18 in GZ but 20 in HK. As a trade off, GZ has 14 extra finals starting with /u/ (see the red font in GZ Finals) while HK does not have them, as /u/ is included as part of the two

extra initials /kw, k^hw/. Also, two extra finals /ɛu, ɛm/ (see the red font in HK Finals) are absent in GZ. As a result, GZ has 68 (=56-2+14) and HK has 56 finals. For ease of comparison, this study treats GZ and HK both having 18 initials and GZ having 68 while HK having 70 (=68+2) finals. Since the difference in Fin:Ini and sumFI between GZ and HK is fairly small, no comparison is made between syllable structure- and duration-based measures and between tone structure- and pitch-based measures across the four Cantonese dialects.

Appendix 4

Sound Inventory of the Five Min Dialects

The number of initials and finals of the five Min dialects:

Sub-group	*No.*	*Dialect*	*Fin*	*Ini*	*Fin:Ini*	*sumFI*
Southern	1	QZ (Quanzhou)	87	14	6.21	101
	2	TZ (Taizhong)	85	15	5.67	100
	3	ST (Shantou)	77	18	4.28	95
Eastern	4	FZ (Fuzhou)	47	15	3.13	62
Central	5	SM (Sanming)	37	17	2.18	54

Initials	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
QZ(14)	p	pʰ	t	tʰ	ts	tsʰ	s	k	kʰ	h	Ø	b/m	n/l	g/ŋ	-	-	-	-
TZ(15)	p	pʰ	t	tʰ	ts	tsʰ	s	k	kʰ	h	Ø	b/m	n/l	g/ŋ	-	-	-	dz
ST (18)	p	pʰ	t	tʰ	ts	tsʰ	s	k	kʰ	h	Ø	m	l	ŋ	n	b	g	dz
FZ (15)	p	pʰ	t	tʰ	ts	tsʰ	s	k	kʰ	h	Ø	m	l	ŋ	n	-	-	-
SM(17)	p	pʰ	t	tʰ	ts	tsʰ	s	k	kʰ	h	Ø	m	l	ŋ	ɡ̊	ɡ̊ʰ	ʃ	-

Finals-1

V(VV)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
QZ	a	i	u	o	e	ɔ	ə	ɯ								
(29)	ai	au	ia	io	iu	ua	ue	ui	iau	uai						
	ã	ɔ̃	ĩ	ãĩ	ãũ	ĩã	ĩũ	ũĩ	ũã	ĩãũ	ũãĩ					
TZ	a	i	u	e	ɔ	ɛ	o									
(30)	ai	au	ia	io	iu	ua	ue	ui	iau	uai						
	ã	ẽ	ɔ̃	ĩ	ãĩ	ãũ	ĩã	ĩõ	ĩũ	ũĩ	ũã	ĩãũ	ũãĩ			
ST	a	i	u	o	e	ɯ										
(34)	ai	au	ia	io	iu	ua	ue	ui	ou	oi	uai	iou				
	ã	ẽ	ɯ̃	ĩ	ãĩ	õĩ	ĩõ	ũã	ĩã	ĩẽ	ũẽ	ũãĩ	ũĩ	õũ	ãũ	ĩũ
FZ	a	i	u	y	ɔ	œ	ɛ									
(19)	ai	au	ia	ie	yɔ	ua	uɔ	øy	ɛu	ui	uai	iu				
SM	a	i	u	o	e	y	ɔ	ø	ɿ	ɯ						
(30)	au	aɯ	ia	iɔ	ie	iɯ	ue	ui	yo	ye	yi	iau				
	ã	õ	ĩã	ĩõ	ũã	ẽ	ỹã	ỹẽ								

Finals-2

(VVV)C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
QZ(58)	am	an	aŋ	əm	ɔŋ	im	in	iŋ	un	m̩	ŋ̩				
	iam	ian	iaŋ	iɔŋ	uan	uaŋ									
	aʔ	ɔʔ	əʔ	oʔ	eʔ	iʔ	uʔ	ɯʔ	mʔ	ŋʔ					
	ioʔ	iaʔ	uaʔ	ueʔ	auʔ	iuʔ	uiʔ	iauʔ							
	ãʔ	ɔ̃ʔ	ĩʔ	ẽʔ	ãĩʔ	ãũʔ	ĩũʔ	ũĩʔ	ĩãʔ	ĩãũʔ	ũãĩʔ				
	ap	at	ak	ɔk	ut	ip	it	iat	iap	iak	iɔk	uat			
TZ (55)	am	an	aŋ	ɔm	ɔŋ	im	in	iŋ	un	m̩	ŋ̩				
	ian	iaŋ	iam	uan	iɔŋ										
	aʔ	eʔ	iʔ	uʔ	ɔʔ	ɛʔ	oʔ	mʔ	ŋʔ						
	auʔ	iaʔ	iauʔ	ioʔ	iuʔ	uaʔ	ueʔ	iɔʔ							
	ãʔ	ɔ̃ʔ	ẽʔ	ĩʔ	ĩãʔ	ãũʔ	ĩãũʔ	ũãĩ ʔ							
	ap	at	ak	ɔk	ut	ik	ip	it	ɔp	iap	iat	iak	iɔk	uat	
ST (43)	am	aŋ	eŋ	ɔŋ	ɯŋ	uŋ	im	iŋ	ŋ̩						
	iɔŋ	iam	uam	iaŋ	uaŋ	ioŋ	ieŋ	ueŋ							
	aʔ	oʔ	eʔ	iʔ	ŋʔ	ioʔ	oiʔ	iaʔ	uaʔ	ueʔ	auʔ				
	ap	ak	ɔk	uk	ek	ip	ik	iap	uak	iak	iok	uap	ɯek	Iek	uek
FZ (28)	aŋ	uŋ	yŋ	iŋ	ŋ̩										
	øyŋ	iaŋ	ieŋ	ouŋ	uaŋ	uɔŋ	yɔŋ	eiŋ							
	aʔ	ɔʔ	œʔ	ɛʔ	iʔ	uʔ	yʔ								
	iaʔ	ieʔ	eiʔ	ouʔ	øyʔ	uɔʔ	yɔʔ	uaʔ							
SM (7)	am	aiŋ	iaiŋ	yaiŋ	iam	m̩	ŋ̩								

Appendix 5

Tonal Inventory of the 21 Chinese Dialects

No	*Major group*	*Dialect*	*Tone values*								*Tone count*	
			Ping		*Shang*		*Qu*		*Ru*		*HT:LT*	*sumT*
			Yin	*Yang*	*Yin*	*Yang*	*Yin*	*Yang*	*Yin*	*Yang*		
1	*Mandarin*	Beijing	55	35	214		51		-		3	4
2		Tianjin	21	45	24		53		-		1	4
3		Jinan	213	42	55		21		-		1	4
4		Heshuo	55	35	214		51		-		3	4
5		Wuhan	55	213	42		35		-		3	4
6		Chengdu	44	31	53		13		-		1	4
7		Liupanshui	45	31	51		24		-		1	4
8		Guilin	33	21	55		24		-		1	4
9	*Wu*	Hangzhou	323	212	51		334	113	55	12	0.75	7
10		Shanghai	53		-		34	23	55	12	1.5	5
11		Suzhou	55	13	51		31	513	5	3	1.33	7
12		Wenzhou	44	31	45	34	42	22	323	212	1	8
13	*Cantonese*	GZ1	55	21	35	23	33	22	5 33	2	1.25	9
14		GZ2	55	21	35	23	33	22	5 33	2	1.25	9
15		HK1	55	21	35	23	33	22	5 33	2	1.25	9
16		HK2	55	21	35	23	33	22	5 33	2	1.25	9
17	*Min*	Quanzhou	33	24	35	22	41		5	24	1.33	7
18		Taizhong	44	23	41		21	33	32	4	1.33	7
19		Shantou	33	55	52	35	213	22	32	5	1.67	8
20		Fuzhou	44	53	31		213	242	23	5	0.75	7
21		Sanming	553	41	21	213	33		12		1	6

Data source:

1. Beijing, Jinan, Wuhan, Chengdu (Mandarin), Wenzhou (Wu), Guangzhou (Cantonese): *Chinese dialect lexicon* (1989);
2. Tianjin, Heshuo, Guilin (Mandarin), Taizhong (Min): *Wikipedia* (accessed online, 2015).
3. Hangzhou, Shanghai, Suzhou (Wu): Zheng (2008).
4. Hong Kong: *Cantonese lexicon* (accessed online, 2015). Note that how many tones are in Cantonese is still arguable (Matthews & Yip, 2013). Some considers that Cantonese has six instead of nine tones, as the last three *Ru* tones, which always occur with unreleased final stops, can be viewed as allotones of the three

level tones (Mok & Wong, 2010). Some tones also have variants; for example, the *yin-ping* tone '55' can become '53.' This study assumes Hong Kong and Guangzhou Cantonese have the same set of tones.

5. Quanzhou, Fuzhou, Sanming (Min): *Database of the provincial government of Fujian* (accessed online, 2015)
6. Shantou (Min): *Wikipedia* & *Baidupedia* (accessed online, 2015).
7. Liupanshui (Mandarin): Xiao (2013) & Wikipedia.

Appendix 6

Duration-based Results for All 21 Chinese Dialects

No.	*Dialect*	*sumD (s)*	*s_rate (#Son/s)*	*%Son*	*ΔSon*	*varcoSon*	*nPVI_ Son*	*ΔIS*	*varcoIS*	*rPVI_ IS*
1	*Beijing*	14.5	3.37	82.2	133.2	53.5	66.4	39.8	66.3	43.9
2	*Tianjin*	90.2	3.43	54.2	144.3	90.4	67.5	89.9	63.6	97.7
3	*Jinan*	99.0	4.52	55.5	97.1	78.5	66.1	64.5	61.1	69.0
4	*Heshuo*	106.8	3.71	62.8	110.0	64.4	65.3	55.7	52.0	62.2
5	*Wuhan*	51.0	3.27	76.5	140.4	235.1	62.7	40.6	78.8	34.3
6	*Chengdu*	28.5	4.14	62.1	102.6	151.5	63.7	60.8	99.8	60.9
7	*Liupanshui*	154.8	3.47	61.6	132.6	74.4	63.5	75.1	65.4	80.1
8	*Guilin*	103.8	3.77	65.7	128.2	72.9	70.2	65.3	64.75	64.2
9	*Hangzhou*	139.3	3.4	63.5	140.6	186.7	64.4	73.7	116.3	74.1
10	*Shanghai*	76.4	3.5	55.3	111.7	157.0	61.8	90.2	138.8	85.3
11	*Suzhou*	86.4	4.1	57.5	86.9	144.0	54.6	76.2	108.2	78.7
12	*Wenzhou*	98.6	3.8	59.4	121.5	155.3	71.1	85.0	120.0	80.0
13	*Quanzhou*	70.2	4.65	50.2	70.3	108.2	58.4	63.4	113.2	61.0
14	*Taizhong*	162.1	3.52	61.4	87.7	174.5	74.0	139.8	119.0	90.4
15	*Shantou*	58.9	4.48	63.6	120.9	145.2	64.2	42.9	88.2	48.0
16	*Fuzhou*	66.7	3.04	62.3	155.9	205.9	74.2	111.7	136.7	91.6
17	*Sanming*	95.8	3.18	59.3	144.7	187.3	69.6	92.3	136.6	96.7
18	*GZ1*	76.0	3.8	59.4	94.2	59.7	58.6	64.8	56.3	68.8
19	*GZ2*	55.1	3.5	42.0	81.5	121.8	56.2	118.8	169.1	126.8
20	*HK1*	181.7	3.4	70.4	59.0	209.3	74.4	155.7	91.5	60.2
21	*HK2*	91.1	3.2	62.5	153.7	194.9	73.2	83.1	128.0	87.4

Appendix 7

Pitch-based Results for All 21 Chinese Dialects

No.	*Dialect name*	*meanPE (sts)*	*ΔPE*	*meanPS (sts/s)*	*ΔPS*
1	*Beijing*	3.746537	4.632931	35.85109	34.55461
2	*Tianjin*	1.087795	1.935499	12.80556	21.30174
3	*Jinan*	1.240496	1.597447	31.59672	32.66706
4	*Heshuo*	1.340532	1.717553	5.258887	8.253649
5	*Wuhan*	2.098075	2.433266	13.15911	11.7121
6	*Chengdu*	0.872878	0.869329	9.002992	8.311175
7	*Liupanshui*	1.097724	1.252802	13.85104	15.56078
8	*Guilin*	0.924552	0.854136	14.03796	13.87731
9	*Hangzhou*	1.009019	1.277722	10.58948	7.837629
10	*Shanghai*	1.229428	2.120963	24.62698	29.08113
11	*Suzhou*	1.141373	1.424258	19.13586	23.82785
12	*Wenzhou*	1.108645	1.171983	14.92256	18.80017
13	*Quanzhou*	0.96812	1.449164	5.997857	-8.08313
14	*Taizhong*	1.067618	1.641327	11.73646	10.91666
15	*Shantou*	1.059483	1.59964	15.52368	21.59264
16	*Fuzhou*	1.900826	2.446877	32.44386	35.44969
17	*Sanming*	0.871635	1.04596	3.420301	0.803632
18	*GZ1*	0.789584	1.299385	9.096341	16.08617
19	*GZ2*	0.765521	1.101195	11.11276	5.817532
20	*HK1*	1.080685	1.213953	12.15167	12.48453
21	*HK2*	2.005634	2.61684	33.57649	38.59972

Appendix 8

Voice Source Results (Mandarin)

Metric / *Dialect*	*H1*-H2* (dB)*		*CPP(dB)*	
	LoP	*HiP*	*LoP*	*HiP*
Beijing	-0.00576 (1.108051)	1.109814 (2.702054)	8.184592 (6.984615)	8.943212 (7.06917)
Tianjin	1.261915 (2.581441)	2.322902 (3.027052)	10.46428 (9.472624)	12.80113 (9.729713)
Jinan	1.952854 (2.462938)	0.748829 (2.722077)	13.63617 (7.504143)	11.27171 (9.004096)
Heshuo	-2.82453 (5.369183)	-2.61661 (7.792486)	11.75397 (7.188028)	16.57121 (5.647913)
Wuhan	-2.39504 (4.282868)	-2.94697 (5.196445)	13.05108 (7.175535)	11.86739 (5.832153)
Chengdu	1.181375 (2.855656)	1.431946 (2.474229)	11.3907 (6.621146)	12.74701 (5.844942)
Liupanshui	2.884488 (3.737436)	1.294205 (4.199326)	13.80246 (7.936603)	14.25754 (7.340987)
Guilin	8.205117 (6.156749)	2.777659 (5.630685)	14.81087 (8.361356)	16.43238 (6.814759)

*Numbers in ( ) are standard deviations.

Appendix 9

Voice Source Results (Wu)

Metric / *Dialect*	*H1*-H2* (dB)*		*CPP (dB)*	
	LoP	*HiP*	*LoP*	*HiP*
Hangzhou	1.132118 (2.826753)	1.097216 (1.872289)	14.32148 (9.122798)	12.77502 (11.02694)
Shanghai	1.596883 (3.186678)	-0.21117 (2.776239)	13.47983 (9.345096)	11.73881 (9.520906)
Suzhou	2.029516 (2.757058)	-0.68292 (2.917723)	15.096 (6.702532)	14.92579 (7.198822)
Wenzhou	1.677259 (3.334635)	1.146343 (2.551883)	12.93223 (8.693159)	13.09229 (9.121883)

*Numbers in ( ) are standard deviations.

Appendix 10

Voice Source Results (Min)

Metric / *Dialect*	*H1*-H2* (dB)*		*CPP (dB)*	
	LoP	*HiP*	*LoP*	*HiP*
Quanzhou	1.127228 (3.198697)	1.14215 (2.338499)	13.30922 (8.265055)	11.33697 (9.596526)
Taizhong	0.587537 (3.328145	0.63758 (2.127066)	15.40766 (7.643932)	16.06008 (7.529273)
Shantou	0.96733 (2.101215	-0.79674 (1.779938)	13.96547 (8.038107)	15.3943 (8.458757)
Fuzhou	0.558527 (2.048115	-3.18563 (4.285611)	11.14145 (8.137954)	17.03429 (7.632859)
Sanming	2.128504 (3.218136	1.520125 (2.119664)	16.83641 (4.236367)	17.67461 (7.208077)

*Numbers in ( ) are standard deviations.

Appendix 11

Voice Source Results (Cantonese)

Metric / *Dialect*	*H1*-H2* (dB)*		*CPP (dB)*	
	LoP	*HiP*	*LoP*	*HiP*
GZ1	0.755132 (3.028604)	1.369208 (3.113563)	8.778441 (7.872952)	8.036438 (7.74527)
GZ2	2.356224 (3.387907)	1.343615 (2.08437)	13.96491 (6.752247)	11.31026 (8.147967)
HK1	1.570393 (2.862836)	0.594175 (1.989886)	14.47792 (8.500203)	15.94977 (7.64206)
HK2	0.831139 (2.676736)	1.826596 (3.731749)	16.24111 (6.566589)	12.77926 (7.0961)

*Numbers in ( ) are standard deviations.