CROSS-REGIONAL WORD DURATION PATTERNS IN MANDARIN

BY

DI WU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Associate Professor Chilin Shih, Chair
Associate Professor Ryan Shosted
Professor Jerome Packard
Associate Professor Torrey Loucks
Assistant Professor Xun Yan

Abstract

Duration contrasts can convey many types of information, including language background, word structure, word frequency, speech genre, intention, and emotions. An understanding of duration lays the foundation for many aspects of speech technology since duration plays a major role in speech production and perception. This dissertation explores the duration patterns of Mandarin words among three Mandarin dialectal regions—Beijing, Taiwan, and Malaysia.

This dissertation brings diverse methodologies on speech data collection, annotation, and corpus construction to investigate linguistic pattern. Three speech production studies are conducted to explore the duration patterns of words with different length and internal structures.

These studies reveal the general duration patterns of Mandarin Words. First of all, all the multi-syllabic words demonstrate the disyllabic long-short metrical form. Second, linguistic factors—syllable structure, positions (syllable position, word position, and sentence position), word frequency, word category, word internal structure, particle attachment, speech rate of sentence have significant effects on syllable duration. Thirdly, social factor—region interacts with multiple linguistic factors (word structure, syllable position, and particle attachment) and plays an important role in duration prediction.

Quantitative data from these studies reveal that there are regional differences in rhythmic contrast among different Mandarin speaking regions. Beijing Mandarin speakers are more sensitive to the length change of linguistic unit and show stronger rhythmic contrast than speakers from Taiwan and Malaysia Mandarins. The results also display that Malaysia

Mandarin speakers show the similar rhythmic pattern as Beijing Mandarin speakers.

The investigation of duration patterns in this dissertation provides a detailed description of word duration in Mandarin. This dissertation also provides the foundation for further research on duration pattern related super-segmental feature. A comprehensive understanding of duration pattern with linguistic and social factors is helpful to improve the quality of durational models used in speech technology.

To my family

Acknowledgments

I would like to thank everyone who made my Ph.D. Journey possible and successful with consistent and patient care and encouragement.

First and foremost, I would like to thank my academic advisor, Dr. Chilin Shih, for her invaluable support and encouragement during the path of my doctoral life. She has been a great mentor with a sharp vision of research and a strong sense of social responsibility. She trained me as a linguist with enthusiasm.

I would also like to thank my other committee members: Ryan Shosted, Jerome Packard, Torrey Loucks, and Xun Yan. They provided valuable insights and feedbacks to guide me in thinking about the analysis of data from different perspectives. I am also grateful to all my other professors at UIUC, especially Dr. Marina Terkourafi.

My gratitude also to my professors in Beijing and Taiwan: Liming Zhao (THU), Kuang-yu Chang (NTHU), Zhiyun Cao (BLCU), Rixin Zhao (BLCU), Hongyin Nie (CASS), and Xiaoping Yao (BFSU). They introduced me to linguistics and inspired me to pursue my passion for phonetics and phonology in every possible way. Without their support and encouragement, it would not have been possible for me to become a linguist.

I am fortunate to be supported by many researchers and friends. Dr. Jingsong Zhang at BLCU generously provided research facilities for my data collection. Li-hsin, Ning, Chung-Yu Chen, You Li, Yuyun Lei, and Yan Sun assisted me in stimuli evaluation and annotation. Li-Wen Weng, Chung-Yu Chen and Shuju Shi offered great help with the participant enrollment. Without their support, this work would have been impossible.

I have much benefited from colleagues and friends in the Department of Linguistics

and the Phonetics and Phonology Lab. Many thanks to: Dora Lu, Chen-huei Wu, Yung-Hsiang Chang, Li-hsin Ning, John Jang, Nicole Wong, Tim Mahrt, Sarah Johnson, Marissa Goldrich and Suyeon Im. There are so many precious moments we discussed ideas and designs, supported each other's experiments and celebrated achievements in the lab.

I am also grateful to my friends Yiyi Wang, Taishan Zhu, and Yanan Liu who have always been my side in the past two years.

Finally, my deepest gratitude goes to my husband, Andrew Hinderliter, and my parents who have always been supportive in all respects.

Table of Contents

List of Tables . . . . . **x**

List of Figures . . . . . **xi**

List of Abbreviations . . . . . **xiii**

Chapter 1 Introduction **1**

Chapter 2 Literature Review **4**
2.1 Regional Variation of Mandarin 4
2.2 Mandarin Words 6
2.3 Word Length 9
2.4 Durational Variation 10
2.4.1 Syllable and Syllable Structure 10
2.4.2 Articulation 13
2.4.3 Boundary and Position 14
2.4.4 Word Frequency 16
2.4.5 Prosodic Condition 16
2.4.6 Regional Variation 17
2.5 Research Approach 18
2.6 Research Questions 23
2.7 Summary 23

Chapter 3 Study 1: Duration Patterns of Monosyllabic and Disyllabic Words **24**
3.1 Introduction 24
3.2 Research Questions 25
3.3 Methodology 25
3.3.1 Participants 25
3.3.2 Materials 26
3.3.3 Procedure 27
3.3.4 Data Transcription 28
3.3.5 Corpus 30
3.4 Data Analysis and Result 32
3.4.1 Statistical Analysis 33

3.4.2 Syllable Duration . . . 35
3.4.3 Word Position . . . 36
3.4.4 Sentence Position . . . 37
3.4.5 Word Category . . . 42
3.4.6 Word Frequency . . . 44
3.4.7 Syllable Structure . . . 46
3.5 Discussion . . . 46
3.6 Summary of Study 1 . . . 49

Chapter 4 Study 2: Duration Patterns of Trisyllabic Words . . . 51
4.1 Branching Structures of Trisyllabic Words . . . 51
4.2 Research Question . . . 53
4.3 Methodology . . . 53
4.3.1 Materials . . . 54
4.3.2 Procedure . . . 55
4.3.3 Corpus . . . 56
4.4 Data Analysis and Result . . . 57
4.4.1 Statistical Analysis . . . 58
4.4.2 Result . . . 59
4.4.3 Syllable Duration . . . 61
4.4.4 Branching Structure . . . 62
4.4.5 *De* Attachment . . . 63
4.4.6 Word Frequency . . . 65
4.4.7 Syllable Structure . . . 66
4.5 Discussion . . . 67
4.5.1 Duration of Identical Character . . . 67
4.5.2 Regional Difference . . . 71
4.5.3 Duration Pattern . . . 73
4.6 Summary of Study 2 . . . 73

Chapter 5 Study 3: Duration Patterns of Quadrisyllabic idioms . . . 75
5.1 Introduction . . . 75
5.1.1 Idiom Length . . . 75
5.1.2 Idiom Structure . . . 76
5.1.3 Duplication of Character . . . 78
5.1.4 Content Character and Function Character . . . 79
5.2 Research Questions . . . 80
5.3 Methodology . . . 80
5.3.1 Materials . . . 80
5.3.2 Procedure . . . 81
5.3.3 Corpus . . . 82
5.4 Data Analysis and Result . . . 83
5.4.1 Statistical Analysis . . . 83
5.4.2 Result . . . 83

5.4.3 Syllable duration . . . 85
5.4.4 Syllable Position . . . 85
5.4.5 Idiom Structure . . . 89
5.4.6 Character Reduplication . . . 91
5.4.7 *De* Attachment . . . 95
5.4.8 Content Character and Function Character . . . 96
5.4.9 Syllable Structure . . . 96
5.5 Discussion . . . 98
5.6 Summary of Study 3 . . . 101

Chapter 6 General Discussion . . . 102
6.1 Rhythmic Pattern . . . 102
6.2 Word Frequency . . . 103
6.3 Individual Difference . . . 105
6.4 Gender Effect . . . 108
6.5 Further Research . . . 109
6.5.1 Tonal variation . . . 109
6.5.2 Mandarin Suffixes . . . 110

Chapter 7 Conclusion . . . 112

References . . . 114

Appendix A: Materials (stories) in Study 1 . . . 126

Appendix B: Sentences in Study 2 . . . 128

Appendix C: Sentences in Study 3 . . . 130

Appendix D: Idioms . . . 139

List of Tables

2.1 Chinese word length 10
2.2 Duration rules of MITalk system (based on the Chapter 9 in *From text to speech: The MITalk system* (Allen et al., 1987)). 20

3.1 Subject information 25
3.2 Story material word counts 27
3.3 Type of inconsistency 29
3.4 Word categories. 32
3.5 Summary of random effects of Study 1 (Number of observations: 22351) . . 34
3.6 Results of fixed effects 34
3.7 Median, mean and SD of syllable duration (ms). 36
3.8 Syllable duration by sentence position (ms) 40
3.9 Number of word by category 42
3.10 Number of syllables of each syllable structure type 46

4.1 Summary of random effects of Study 2 (Number of observations: 9180) . . . 59
4.2 Summary of fixed effects of Study 2 60
4.3 Syllable duration by positions 61
4.4 Syllable duration of trisyllabic word by syllable structure 67
4.5 Eight pairs of trisyllabic words (right-branching and left-branching paired trisyllabic words without *de* attachments) 68

5.1 Length information of Chinese idioms 76
5.2 Summary of random effects of Study 3(Number of observations: 55440) . . . 84
5.3 Summary of fixed effects of Study 3 84
5.4 Overall syllable duration of Study 3 85
5.5 12 identical characters used on 4 syllable positions in 48 quadrisyllabic idioms. 87
5.6 Syllable duration of identical character (i represents the identities of the 12 identical characters in Table 5.5). 88
5.7 Repeated identical characters on different positions in 60 quadrisyllabic idioms. 92
5.8 Mean, median, and SD of content characters and function characters in idioms 96
5.9 Syllable duration by syllable structure 97

List of Figures

2.1 Map of three Mandarin speaking regions investigated in this dissertation (Beijing, Taiwan and Malaysia) . . . 7
2.2 Mandarin syllable structure . . . 11
2.3 Syllable length classified by syllable type (Shih and Ao, 1997) . . . 13

3.1 Original text to transcription data . . . 29
3.2 The distribution of syllable duration. . . . 35
3.3 Syllable duration by word positions. . . . 38
3.4 Syllable position by sentence position. . . . 39
3.5 Syllable duration by word positions and sentence positions . . . 41
3.6 Word Duration by Category . . . 43
3.7 Correlation between word frequency and syllable duration . . . 44
3.8 Duration of high frequency monosyllabic particles . . . 45
3.9 Syllable duration by syllable structures . . . 47
3.10 Interaction between region and syllable position on syllable duration of disyllabic word . . . 48
3.11 Regional differences on syllable duration of disyllabic word . . . 49

4.1 Morphological structure of trisyllabic words . . . 52
4.2 Mean of syllable duration by syllable position and region . . . 61
4.3 Interaction of branching structure and syllable position . . . 63
4.4 Interaction between particle *de* and syllable position . . . 64
4.5 Particle *de* attachment effect . . . 65
4.6 Syllable duration by structures . . . 66
4.7 Syllable duration of the second syllables at paired left-branching and right-branching words (5 pairs of trisyllabic words with frequency differences) . . . 69
4.8 Syllable duration of the second syllables at paired left-branching and right-branching words . . . 70
4.9 Interaction between particle *de* and syllable position for each region . . . 72

5.1 Structure of Chinese idioms . . . 77
5.2 Seven types of character duplications of quadrisyllabic idioms . . . 78
5.3 Syllable duration by positions among three regions . . . 86
5.4 Syllable duration of identical character by position and idiom (i represents the identities of the 12 identical characters.) . . . 88

5.5 Duration patterns of four structures of quadrisyllabic idioms. 90
5.6 Distribution of syllable duration of reduplicated syllable 91
5.7 Duration of reduplicated syllables by syllable positions 94
5.8 Effect of *de* attachment on syllable duration 95
5.9 Syllable duration of content and function characters by syllable positions . . 97
5.10 Syllable duration by syllable structure 98
5.11 Correlation of syllable duration 99
5.12 Metrical pattern of 4-syllable words in Kochanski et al. (2003) 100

6.1 Disyllabic long-short pattern by multi-syllabic word types 103
6.2 Correlation between word frequency and syllable duration 104
6.3 Individual difference on distribution of syllable duration among three Mandarin region 107
6.4 Gender difference on syllable duration (pooled all three studies) among three Mandarin regions 108

List of Abbreviations

bj	Beijing. For example, 'bj01' represents the subject number of first Beijing subjects.
ms	Malaysia
tw	Taiwan
st	Sentence in a story (Study 1). For example, 'st05' represents the fifth short sentence in a story.
fst	Initial short sentence in a story (Study 1)
mst	Middle short sentence in a story (Study 1)
lst	Last short sentence in a story (Study 1)
wd	Word in a short sentence (Study 1). For example, 'wd03' represents the third word in a sentence.
fwd	Initial word in a sentence (Study 1)
mwd	Middle word in a sentence
lwd	Last word in a sentence
de	Particle *de*

Chapter 1

Introduction

The purpose of this dissertation is to explore the duration patterns of Mandarin words among different Mandarin dialectal regions. Duration contrasts can convey many types of information, including language background, word structure, word frequency, speech genre, intention, emotions, etc., depending on how they are produced and how they are shaped by linguistic, social, and psychological factors in both speech production and perception.

Among different Mandarin areas, there is an impression that speakers of Beijing Mandarin speak faster than speakers of Taiwan Mandarin or Mandarin speakers in Southeast Asia (e.g. Singapore, Malaysia, and Indonesia). What are the phonetic features leading to such an impression? In speech communication, listeners perceptually integrate all the linguistic cues to process the received information. The impression of variability may arise from listeners' native language background or dialectal phonological systems. Listeners can observe the patterns of fine-grained phonemic contrasts in the phonological system, and simultaneously be sensitive to the acoustic cues such as pitch and duration in speech communication. Duration is one of the phonetic cues listeners use to form an impression about the speaker.

Most listeners' perception of speech is holistic. They typically do not intuitively differentiate one cue from the others. Though there are talented comedians and impersonators who may capture a few features of a person or a region and exaggerate them for theatrical effect. In such acts, duration is often one of the features being imitated. In everyday speech, duration interacts with other linguistic cues and multiple social factors in a natural environment. Factors affecting duration do not occur in isolation. The combination of such

factors and their interactions creates endless possibilities of a richly intricate speech system.

We ask the following questions to investigate duration patterns in speech. What are these factors and how do they interact with each other? How to create a dynamic speech context to cover and examine factors as natural as possible in linguistic research? How to use statistical model to uncover duration patterns and to describe them quantitatively?

On the one hand, we can investigate duration factors by using controlled experiments. This method works well on verifying known factors leading to duration variations, but cannot cover all possible factors and their interactions outside the controlled environment. It also does not provide an opportunity for exploratory studies to observe unknown factors and interactions. On the other hand, we can use large speech corpus including varied speech genres which may contain potential factors and their interactions in a natural speech environment. But building speech corpus is a high-cost and time-consuming project due to the work of data collection, transcription, and annotation. Furthermore, corpus may not contain strictly controlled environment that can unambiguously verify the effect of a known factor. Therefore, this dissertation will combine the advantages of both research methods using controlled experiment in natural sentence context and supplement with corpus coding. This method allows us to investigate key factors in a style similar to traditional controlled experiments, while we can also investigate factors such as word frequency and suffixation that exist in the data using corpus style coding.

The experiments are designed to have a comprehensive coverage of Mandarin words with different length and different internal structures. Speakers from three Mandarin speaking regions are recruited. The northern speakers have stronger long-short durational contrast, which reveals the underlying rhythmic pattern. The dynamics of suffixation in the northern speech and southern speech are different where northern speakers incorporate the suffix into the words causing internal changes of the lexical structures and the effect is obvious from duration.

We observe consistent frequency effect, as in all languages, where higher frequency words

lead to shorter duration. This is a robust pattern consistent with the prediction of information theory. Interestingly, the frequency effect is not observed in the reading of quadrisyllabic idioms which is a relic of Classical Chinese commonly used in Modern Chinese, indicating literary and education status. When Mandarin speakers talk it seems that they code switch between Modern Chinese and Classical Chinese. When they speak Modern Chinese, the duration pattern is affected by lexical structures and frequency, among others. When they speak classical Chinese, the predominant effect comes from the internal structures of quadrisyllabic idioms, frequency no longer has an effect.

Chapter 2 will review the literature on durational studies. Following the literature review chapter, Chapter 3 will present Study 1 on monosyllabic and disyllabic words. Chapter 4 will present Study 2 on trisyllabic words. Chapter 5 will present Study 3 on quadrisyllabic words. Chapter 6 will synthesize the general findings on word duration patterns in these three studies, and discusses further research direction. Chapter 7 will conclude the dissertation.

Chapter 2

Literature Review

Duration is an important aspect of speech and plays an essential role in both phonetic and phonological studies. An understanding of duration lays the foundation for many aspects of speech technology. For example, it is used in speech fluency testing. Duration can differentiate the fluency levels of second language learners in the first few years of second language acquisition. Another example is text-to-speech synthesis, which is popular in many apps of mobile phones. The synthesis systems must predict duration accurately to generate natural sounds in speech. One of the current demands of these synthesizers is that they require variations of duration within words from regional dialects. This is one of the focuses of this dissertation.

In duration research, many linguistic factors need to be taken into consideration: phoneme identity, phoneme context (surrounding phonemes), morphosyntactic structure, stress, etc. Moreover, many social factors are also related to duration—for example, dialect, gender, age, race, and education background—these are also potential areas for improving the accuracy of duration measurement and modeling.

2.1 Regional Variation of Mandarin

This dissertation investigates Mandarin speech. However, Mandarin spoken in different regions may be affected by local dialects. Hence, this section first reviews Chinese dialects and their distributions.

Chinese dialects can be divided into Mandarin and non-Mandarin groups. The non-

Mandarin group generally refers to the southern dialects, including Wu, Xiang, Min, Yue (Cantonese), Gan, and Kejia (Hakka) (Ramsey, 1987; Norman, 1988; Sun, 2006). The Mandarin group covers most of northern China and some parts of southwest China. Compared to the Mandarin group, southern dialects have different phonological systems and more complicated tonal systems in general. However, the Mandarin group has a cross-regional consistency of the phonological system. The Mandarin group also can be divided into different subcategories, but in this dissertation, the concept of Mandarin only refers to the Standardized Mandarin (Lee and Zee, 2003).

Mandarin is also known as Standard Chinese, *Putonghua*, *Guoyu*, and *Huayu*, and it is the official language of the People's Republic of China (Mainland China), the Republic of China (Taiwan), and is one of the four official languages of Singapore. It is also a major language in some Southeast Asian countries that have large Chinese populations, such as Malaysia, Indonesia, and Thailand.

In Mainland China, the movement of *Putonghua* promotion by the government has made *Putonghua* become dominant and dialects become less frequently used since 1950s. In Taiwan, before the 1940s, the Min dialect was the predominant language. After the Second World War, the government vigorously promoted *Guoyu* in Taiwan. In recent years, there is a revival of Min or Taiwanese though *Guoyu* is still the dominant language.

In Southeast Asian nations, Chinese immigrants were mostly from the coastal provinces of China, such as Guangdong, Guangxi, Fujian, and Hainan. The immigration situation led to a complicated dialect situation inside the Chinese communities in these countries. Mandarin is one of the official languages or major languages used in many Southeast Asian countries.

The Chinese ethnic group comprised around 25% of the population in Malaysia in 2010 (Lin, 2016). In the education system of Malaysia, Chinese communities have a distinctive Chinese school system—from elementary school to high school (Brown, 2007; Lin, 2016), where Mandarin is the dominant instructional language.

In Singapore, the Chinese ethnic group comprised 74.3% proportion of the population (Lin, 2016), and the government has begun to carry out its 'Speak Mandarin Campaign' to encourage the Chinese population to speak Mandarin instead of Chinese dialects since 1970s. (Newman, 1988; Chang, 2008).

Li (2014) proposed the concept of Supra-Chinese, the Chinese common language based on *Putonghua* or *Guoyu* among Chinese communities, which shares the linguistic core and covers the variation of Mandarin among different communities in Southeast Asia or worldwide. The Supra-Chinese is a conceptual language description, which requires more research and evidence on its linguistic properties and status. There are two possible directions for the future of Mandarin regional variation: greater convergence or greater divergence. On the one hand, it is possible for all the variants to reduce the distinctive features and merges into one worldwide unified speech community. On the other hand, it is also possible that each variety evolves into more distinctive linguistic properties like the situation of English around the world (Li, 2014) .

Therefore, studies on cross-region duration patterns will help to describe the current situation of language usage and to enumerate factors contributing to durational differences. In my dissertation, three Mandarin regions—Beijing, Taiwan and Malaysia (Figure 2.1) are selected to investigate the word duration patterns within- and cross-region.

2.2 Mandarin Words

The definition of word is an important issue in the linguistic analysis of Mandarin. The segmentation of word is not as straightforward as in English because of its orthographic conventions and morphology. The orthographic system of Chinese does not require a space between characters, and there are no systematic morphological markers, such as n tense, number, and case in Mandarin.

Chinese scholars used the concept of 字 (*zi*, Chinese character) to refer the lexical unit.

Figure 2.1: Map of three Mandarin speaking regions investigated in this dissertation (Beijing, Taiwan and Malaysia)

Chao (1968) called *zi* a 'sociological word', which is the unit that 'the general, nonlinguistic public is conscious of, talk about, has an everyday term for' (Packard, 2000, p.14). A Chinese character usually is a monosyllable and may correspond to different levels of lexical representations. For example, the character '老' (*lao*, old) in the word '老人' (*lao ren*, old people) , '老' (*lao*) is considered to be an adjective with the semantic meaning of 'old', but in the word '老师' (*lao shi*, teacher), it is a prefix to form a disyllabic word. When *lao* is a prefix (bounded morpheme), it does not have the semantic meaning of 'old'.

Packard (2000) clarified that "the status of *zi* as the sociological word in Chinese is just as salient as the status of the word as the sociological word in English", and 词 *ci* can be considered to be 'syntactic word', "which is used mostly as a technical term by specialists in language and linguistics". Based on the criteria of free-bound and content-function, a morpheme can be classified into one of four different categories (Packard, 2000, pp.67-69):

(a) [+free, +function] → Function word;

(b) [+free, -function] → Root word;

(c) [-free, -function] → Bound root;

(d) [-free, +function] → Affix;

Zhang (1984) and Packard (2000) used the concept of 'compositionality' to differentiate *ci* and *non-ci* in Chinese. For example,

黑鸟 *heiniao* black-bird 'black bird'

黑道 *heidao* black-road 'underworld'

Heiniao consists of two characters and two morphemes, and it is a phrase made up of two *ci* based on the productivity of the two constituents; *heidao* consists of two characters and two morphemes, but it is one *ci* because of its lack of compositionality. Therefore, *heiniao* is a phrase rather than a word in Chinese. But *heidao* is a word, and it will lose the idiomatic meaning if the morpheme is replaced.

Besides compositionality, word frequency is another factor affecting the perception of *ci* in Mandarin. For example,

老人 *laoren* old-people 'the old' (Frequency：33,483 in PKU-CCL Corpus[1])

老楼 *laolou* old-building 'old building' (Frequency：50 in PKU-CCL Corpus)

The word *laoren* is a high frequency word compared to the word *laolou*. Native speakers are more likely to consider *laoren* as a disyllabic word, despite of the same syntactic structure in these two expressions.

2.3 Word Length

The length or size of the linguistic unit is vital for duration research. This dissertation will address the words with different length from one syllable to four syllables. To get the overall impression of word length (how many syllables/characters in one word) in Chinese, the word list of the Contemporary Chinese Corpus is used to calculated the word length information in Table 2.1.

The size of the Contemporary Chinese Corpus is 100 million Chinese characters. In this corpus, 70% of documents are from the period 1919 to 1992, and 30% documents are in the period from 1993 to 2002. The documents are sampled from different genres and disciplines, including 50% social sciences and humanity, 30% natural sciences, and 20% various business and official documents (Zhan et al., 2006). Among all the materials in this corpus, one fifth materials (20 million Chinese characters) is annotated using the *Standard of the POS (part of speech) Tag of Contemporary Chinese for Chinese Information Processing* (GBT20532-2006). The word list used in Table 2.1 is based on the annotated data, and only includes words with frequency of at least 50.

Table 2.1 displays the word length information of Chinese. Monosyllabic words constitute

[1]Peking University Center for Chinese Linguistics corpus: CCL corpus includes 783,463,175 characters text data from Chinese text materials.

around 16% of words (types) in Chinese. Disyllabic words are about 72%, and this means that disyllabic words are the most common word type in Chinese. The trisyllabic words are about 8% of the list. The quadrisyllabic words are around 3.5% in this list. Thus, in total, these four word forms cover 99.5% of words in Chinese. And this dissertation will address the duration patterns of monosyllabic and disyllabic forms in Study 1, trisyllabic forms in Study 2, and quadrisyllabic forms in Study 3, respectively.

Syllable/Word	**Num. of words**	**Percentage in the list**
1	2391	16.30%
2	10476	71.60%
3	1181	8.10%
4	506	3.50%
5	43	0.29%
6	22	0.15%
7	8	0.06%
8	1	0.01%
10	1	0.01%
Total	14629	100%

Table 2.1: Chinese word length

2.4 Durational Variation

This section discusses the major linguistic factors related to the variability of duration observed in speech production, including syllable structure, articulation, boundary and position, word frequency, prosodic condition, regional variation.

2.4.1 Syllable and Syllable Structure

In Mandarin, syllable duration is sensitive to syllable structure. Contrary to the prediction of the isochrony theory (Lehiste, 1977; Tuller and Fowler, 1980, 1981; Roach, 1982; Dauer, 1983; Wagner and Dellwo, 2004), where Chinese is classified as syllable-timed language

and different syllables are expected to have similar duration in the same prosodic context, quantitative studies show that syllable duration reflects the composition of the syllable. If there are more sounds in a syllable the duration is longer.

Mandarin has around 400 syllables (excluding tones). A syllable may contain a consonantal onset (C), a pre-nuclear on-glide (G), a nucleus (V), and a nasal coda (alveolar nasal or velar nasal) or an off-glide (X). The C is traditionally referred to as the initial, and the GVX is referred to as the final (see Figure 2.2 for the illustration of syllable structure of Mandarin). X is used to represent either nasal coda or off-glide in the final. In Chinese syllable structure, the V is obligatory, while C, G and X are optional, so that the maximal syllable structure is CGVX while the simplest one is V (Chao, 1968; Chen, 2000; Duanmu, 2007, 2009). The maximal syllable structure is shown in Figure 2.2.

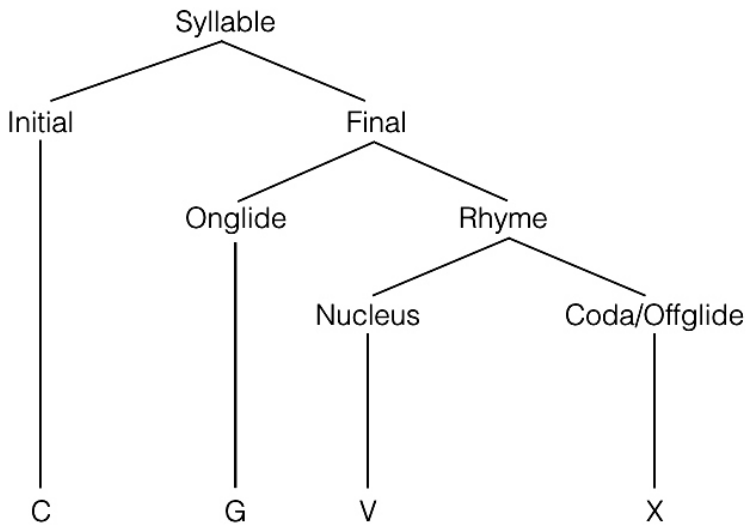


Figure 2.2: Mandarin syllable structure

Based on the frame of syllable structure, the 400 Chinese syllables can be divided into 12 subcategories regard to their syllable structures: V, VG, GV, CV, VN, CGV, GVG, CVG, CVN, GVN, CGVG and CGVN. They can fully cover all the possible Chinese characters. Although there are around 3000 to 5000 frequently used Chinese characters (around 1200

syllables), without tonal differences, there are approximately 400 Chinese syllables.

Syllable duration is related to the number of phonemes in the syllable and the way the phonemes are articulated or co-articulated in the speech. Duration can serve as an important acoustic cue in phoneme contrast cross-language. Some languages have length contrast in vowels or consonants. Even in language without phonemic length contrast, different phonemes may have different duration values that play a role in phoneme perception. For example, the significance of voice-onset time (VOT) for distinguishing stops is well established in a variety of languages or dialects such as American English (Lisker and Abramson, 1964, 1967; Klatt, 1975a; Keating, 1984), British English (Docherty, 1992), Korean (Kim, 1965; Han and Weitzman, 1970; Cho et al., 2000), Japanese (Riney et al., 2007) and Mandarin Chinese (Rochet and Yanmei, 1991).

In Mandarin, many studies have been conducted on duration measurement. Qi and Zhang (1982) measured the duration of 22 consonants in disyllabic words in Standard Mandarin, and found out that articulation manner, aspiration, and following vowel are related to consonant duration, but tonal condition and syllable length are not related to consonant duration. Feng (1985) investigated the initial, rhyme and tone duration in Beijing Mandarin, and reported that consonant duration is related to articulation manner and rhyme duration is related to tongue shape and articulation place. Ren (1985) measured the phoneme and syllable duration in monosyllabic, disyllabic, trisyllabic, and quadrisyllabic artificial words, and further built duration rules to predict duration value under different linguistic conditions. Duanmu (1994) reported the average syllable duration in Mandarin and discussed the relationship between syllabic weight and syllabic duration. Shih and Ao (1997) reported the duration patterns of Mandarin phonemes and syllables under various phonetic conditions in a speech corpus of Beijing Mandarin. van Santen and Shih (2000) reported that syllable duration is related to the number and type of the segments contained in the syllable.

Moreover, syllables with different number of segments also exhibit durational differences. Figure 2.3 from Shih and Ao (1997) shows that syllable duration of the most complicated

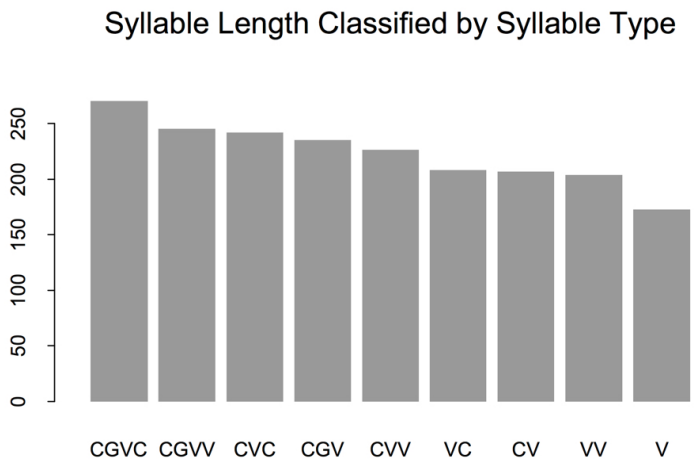


Figure 2.3: Syllable length classified by syllable type (Shih and Ao, 1997)

syllable structure CGVC (which is corresponding to CGVG and CGVN in this dissertation) is 1.5 times of the simplest syllable structure V with the shortest duration. And the results in their research show the duration of syllable is related to the number of segments: syllable with 4 phonemes > syllables with 3 phonemes > syllable with 2 phonemes > syllable with 1 phoneme. However, syllable duration is not simply computed by adding segments duration together in a syllable. The duration is typically shorter than the expected sum of the components, possibly due to the coarticulation in speech.

2.4.2 Articulation

Articulation is another major effect leading to durational variation. Speech articulation is a physical movement process, which takes time for speaker to finish the movement of human articulatory apparatus. Sound combination that requires more articulatory movement takes longer durations, such as low vowel in the context of consonants. Articulatory place assimilation effectively reduces duration, because such assimilation reduces the articulatory distance and movement. Nelson (1983) used the physical principles to explain the motor

control strategies in articulation. Speech articulation is highly related to movement time, movement distance and velocity, energy, peak acceleration, and rate of change of acceleration. During the articulation process, minimal duration of movement is the necessary time window for the speaker to finish the articulation of the segment and to maintain the accuracy in language communication. Klatt (1976) used the concept of 'Dmin' (minimal duration) to indicate that the absolute minimal duration that is required to execute a satisfactory articulatory gesture.

Articulatory phonology (Browman and Goldstein, 1986, 1992, 1995) proposed that gestures are the basic units in phonological system, and speech is a gestural system. Speech is a physical process of coordination of gestures during a certain time. Many studies (Byrd, 1994a, 1996; Lucero et al., 1997; Byrd, 2000; Lee et al., 2006) reported variation of articulatory timing on segment, syllable and word levels.

2.4.3 Boundary and Position

In natural speech production, there is a durational adjustment of segment and syllable duration based on the morphological structure, syntactic structure and prosodic structure. The boundaries of syllable, word, phrase, sentence, and utterance may change the articulation timing and the duration pattern. Many early studies reported the lengthening and shortening effects related to the boundaries, such as clause-final lengthening effect (Gaitenby, 1965; Lindblom and Rapp, 1973; Oller, 1973; Klatt, 1975b), non-phrase-final shortening and non-word-final shortening effects (Lindblom and Rapp, 1973; Klatt, 1975a), non-initial-consonant shortening effect (Klatt, 1974; Umeda, 1977) and polysyllabic shortening effect (Lindblom and Rapp, 1973; Lehiste, 1975).

In order to explore the duration patterns on words, it is important to consider the morphological structure, the syllable position in the word, the word position in the phrase or sentence, and the sentence position in the utterance. Wightman et al. (1992) examined the segmental lengthening before prosodic boundaries and reported that lengthening can

appear at multiple boundary levels. Based on the hierarchy, higher level boundaries have greater lengthening effect than lower level boundaries. It means the lengthening effect is associated with the level of the boundary. van Santen (1992) measured vowel duration under the effect associated with syntactic boundary, and the results confirm the lengthening effect on boundaries.

Syllable before boundaries usually tend to have longer duration (Gaitenby, 1965; Lindblom and Rapp, 1973; Oller, 1973; Klatt, 1975b; van Santen and Olive, 1990) in English. In Mandarin Chinese, many studies agree that middle syllables in a multi-syllabic word are the weak syllables with less degree of stress and shorter duration (Yip, 1980; Zhang, 1988; Shih, 2005).

Syllables in different positions of a word also show durational variation. In Mandarin, Kochanski et al. (2003) reported that there is a strong-weak pattern of prosodic strength on disyllabic, trisyllabic and quadrisyllabic words, and they also reported that prosodic strength is correlated with syllable duration. Chen (2006) reported the experimental result that the second syllable on the weak position of a trisyllabic word has the shortest duration.Xu and Wang (2009) found that syllables at boundary positions, both initial and final, in three-character and four-character phrases in Mandarin display longer duration than syllables on other positions.

Boundary and position not only affect syllable duration but also affect phoneme duration in Chinese. Ho (1977) investigated the duration pattern in different sentence types, and the result shows that duration of syllable nuclei varies according to word position, and the duration of syllable nuclei is more distinctive in the word with longest duration in the sentence final position. In Shih and Ao (1997), they reported that the consonant lengthening effect is found at the initial positions of word and phrase, and the vowel lengthening effect is found at the phrase final position.

2.4.4 Word Frequency

Word frequency plays an important role in predicting duration variation. Following Zipf's law (Zipf, 1932; Mandelbrot, 1953; Booth, 1967; Li, 1992; Baayen, 2001; Ha et al., 2002), which states that, in human languages, there is a linear relationship between the log-transformed word frequency and the log-transformed word frequency rank. We expect that the log-transformed word frequency information is an important factor in explaining duration, and should be included in the statistical model predicting duration.

Word frequency affects the speed of word recognition. Many psycholinguistic studies reported that low frequency words require longer duration in word recognition and in reading (Howes and Solomon, 1951; Solomon and Howes, 1951; Wright, 1979; Rayner and Duffy, 1986; Monsell et al., 1989; Allen et al., 1995).

The frequency-related word recognition speed is reflected in the duration measurement of speech production. Just as in speech processing, word frequency is also inversely related to duration in speech production. Fosler-Lussier and Morgan (1999) reported that the variation of pronunciation is related to word frequency. For example, syllables in high frequency words tend to be shortened the most. Jurafsky et al. (2001) had reported frequency can be used to predict linguistic reduction. High frequency words usually have shorter duration than low frequency words within the same context in speech (Pluymaekers et al., 2005), and speakers produce high frequency words more quickly than low frequency words (Oldfield and Wingfield, 1965; Forster and Chambers, 1973; Balota and Chumbley, 1985; Fosler-Lussier and Morgan, 1999). In the analysis of duration patterns in this dissertation, word frequency is one of the factors being considered.

2.4.5 Prosodic Condition

Duration is also related to prosodic features, such as intonation, stress, focus, and prominence. The same phonetic segment can be produced with different duration in different

prosodic environment. For instance, segments are lengthened via stress. In English, vowel and consonant are lengthened within a stressed syllable (Umeda, 1975, 1977; Lehiste, 1975; van Santen, 1992). In Arabic, de Jong and Zawaydeh (2002) and de Jong (2004) reported that Arabic vowel duration shows similar lengthening pattern as in English that vowel duration correlates with stress and focus in different phonetic environments.

In Mandarin, Shen (1993) reported that vowels display an approximate ratio of 3:2 on duration between stressed and unstressed conditions. Stress and focus have effects on syllable duration. Jin (1996) examined the relationship between sentence stress and three acoustic features—f0, syllable duration and intensity in Mandarin, and reported that a stressed syllable leads to extension of pitch range and longer syllable duration. Chen and Xu (2006) reported that syllable in Mandarin with neural tone, which does not bear lexical stress, have shorter duration than a syllable with full tone. Chen (2006) found that there is a gradual increase of syllable duration on syllable with different degree of emphasis.

2.4.6 Regional Variation

Based on the durational variation, rhythm measures have been widely used to discriminate languages or language varieties (White and Mattys, 2007; Jacewicz et al., 2007; Loukina and Kochanski, 2010; Jacewicz et al., 2009; Clopper and Smiljanic, 2015). Duration can be used as a parameter to capture rhythm variations and classify languages. In British English, Loukina and Kochanski (2010) reported that duration differences among dialects of British English can be used as a rhythm feature to capture the dialectal differences. In American English, many studies reported there are characteristic differences in vowel duration, consonant duration and articulation rate that distinguish American English dialects (Jacewicz et al., 2007, 2009; Clopper and Smiljanic, 2015). For example, the southern dialect has longer pauses and highly variable vowel duration; conversely, the New England dialect has shorter pauses and larger variance of consonant duration.

In Chinese, Duanmu (1994) examined the syllable duration in sentence frame in Beijing

Mandarin and Shanghai dialect, which belongs to the Wu dialect, and found that the average syllable duration in Mandarin is 215 ms and average syllable duration in Shanghai is 162 ms. Deng et al. (2006) investigated the duration of monosyllable in both Beijing Mandarin and Taiwan Mandarin, and reported that there are durational differences of monosyllables between Beijing and Taiwan. Chang and Shih (2015) investigated contrastive focus realization of the alveolar-retroflex sibilants contrast in Taiwan and Beijing Mandarin and reported that focal prominence increase frication duration and amplitude in Mandarin. Compared with Beijing Mandarin, Taiwan Mandarin has longer syllable duration and frication duration in both focused and unfocused conditions. Wu and Shih (2015) investigated the duration pattern of the diminutive suffix *'er'* with its host syllable among three Mandarin regions, and found that Beijing Mandarin speech has the shortest duration; Taiwan Mandarin speech has the longest duration; Mandarin of central China is in between. Studies on cross-regional duration patterns help to capture characteristic features of regional dialects, which may reflect the historical development over time.

2.5 Research Approach

In many previous studies on Mandarin duration, speech data was collected mostly from laboratory speech, where a series of phonetic, morphologic, syntactic and prosodic factors were manipulated to keep the target stimuli under identically controlled language contexts (Qi and Zhang, 1982; Feng, 1985; Ren, 1985). These studies usually used the same carrier sentence or limited contextual conditions to maintain tightly controlled linguistic conditions in which the stimuli occur. The experimental approach is an effective way to collect balanced data to examine selected factors. It is also a useful approach to evaluate whether a particular factor leads to significant duration difference.

However, there are several theoretical and practical concerns of the experimental approach. Firstly, it is hard for the controlled experiment in the laboratory speech to cover

all the factors in natural speech, especially their interactions. Secondly, using the highly controlled stimuli may lose a certain degree in naturalness of the speech data. When speakers repeatedly produce the target elements in a fixed context, it is possible for them to be aware the experiment design and inadvertently affect the experiment results. Speakers may unconsciously emphasize the keywords, or speak in monotone after many repetitions. Thirdly, although the experimental result may explain the experimental data very well under the combination of predefined factors, it may have less predictive power or have problem generalize to new data due to over-fitting, having context outside the experimental design, or having interaction effect not included in the experiment. The latter point explains why it is hard to integrate factors from different experimental studies and predict the interaction in a coherent system. This approach was attempted by Allen et al. (1987) to build the most successful formant synthesis system still in use today, the MITalk system. Allen et al. (1987) collected all available experimental result to build duration rules with different factors and conditions to predict the duration of MITalk. They used 11 durational rules based on more than 20 studies (see Table 2.2) in their duration module. These 20 studies are from different scholars in different time periods, and focus on different phonetic and prosodic factors leading to durational difference.

If the interaction effect of two factors is not captured, combining the result of 2 studies can lead to big errors. For example, clause-final lengthening is a robust effect observed consistently across languages. Lengthening for emphasis is another strong and robust effect. Putting them together produces unacceptably long duration for clause-final emphasized word (van Santen, 1994), because individual studies did not include environment which includes data to inform about the interaction effects of final lengthening and emphasis. If the coordination and unification of these rules are not systematic it may cause problems in the application of the system. Allen et al. (1987) spend decades training the MITalk duration system adjusting the weights or coefficients combining features into rules. The most difficult task is to manage the rules to capture the interactions of factors in an integrated system.

Rule	**Reference**
Pause insertion	Eisler (1968) Cooper et al. (1978)
Clause-final lengthening	Gaitenby (1965) Lindblom and Rapp (1973) Oller (1973) Klatt (1975b)
Non-phrase-final shortening	Lindblom and Rapp (1973) Klatt (1975b)
Non-word-final shortening	Lindblom and Rapp (1973) Oller (1973)
Polysyllabic shortening	Lindblom and Rapp (1973) Lehiste (1975)
Non-initial-consonant shortening	Klatt (1974) Umeda (1977)
Unstressed shortening	Umeda (1975) Umeda (1977) Lehiste (1975)
Lengthening for emphasis	Bolinger (1972) Carlson and Granström (1973) Umeda (1975)
Postvocalic context of vowel	House and Fairbanks (1953) Peterson and Lehiste (1960) Klatt (1975b)
Shortening in clusters	Klatt (1973)
Lengthening due to plosive aspiration	Peterson and Lehiste (1960)

Table 2.2: Duration rules of MITalk system (based on the Chapter 9 in *From text to speech: The MITalk system* (Allen et al., 1987)).

The same situation occurs in Mandarin. Ren (1985) reported the duration rules regarding different linguistic conditions in Mandarin. When these rules are applied to duration modeling, the lack of interactions among different linguistic conditions leads to the same problems as in the MITalk system.

Speech corpora can provide better solutions to solve the above problems. Many studies focusing on duration modeling in Mandarin speech synthesis system use corpus approach to investigate the duration pattern (Shih and Ao, 1997; Lai and Chen, 2001; Chen et al., 2003).Shih and Ao (1997) used a greedy algorithm to build a Mandarin database with 427 sentences, which cover 14 factors and their three-way interactions with more than 8000 interaction scenarios covered and coded in the database. Lai and Chen (2001) included 455 sentences and 200 long paragraphs. Chen et al. (2003) used a speech database including a phonetic-balanced sentence set with 455 sentences and a paragraph set with 300 paragraphs in duration modeling. Corpus-based duration studies are effective in estimating duration for all sounds or syllables in a language within a variety of contexts. Nonetheless, the primary practical concern of speech corpora maybe the size of the database. The construction of large size database is a high-cost and time-consuming work. It is time-consuming to do text transcription, audio segmentation, data annotation and labeling. Most importantly, the corpus-based approach must address and resolve ambiguity in the annotation of the data, which requires multiple transcribers' labels to come up with a standardized transcription and labeling guidelines. These guidelines need to be tested and revised iteratively before the finalized guideline is applied to corpus transcription and labeling.

When large speech database is hard to be constructed in a low-cost way, there is a tradeoff between corpus approach and experiment approach. This dissertation combines the experiment and corpus approaches together to explore the duration patterns of Chinese words with different length and structures. To ensure the speech data is as natural as possible, Study 1 (in Chapter 3) uses stories to examine monosyllabic and disyllabic words, and Study 2 (in Chapter 4) and Study 3 (in Chapter 5) use different carrier sentences to investigate

trisyllabic and quadrisyllabic words in sentence reading tasks. In all studies, speakers' region, gender, word length and internal structure of word are controlled. Each stimulus occurs in a natural sentence rather than in a carrier sentence. In this way, participants do not have information to guess the purpose of the experiment, nor can they identify the stimuli being tested in the sentences. Potential factors from the sentence that may affect the duration of the stimuli are coded following the procedures commonly used in corpus approach.

Given current development in computing power, data storage, and statistical modeling, managing integrated system like MITalk system is no longer a problem. For example, Riley (1991) implemented the CART (Breiman et al., 1984) algorithm to train duration model for text-to-speech synthesis, which is easy to maintain. van Santen (1994) used regression models in a tree-structure to predict duration. The system is also easy to maintain. Later corpus-based duration models have all advanced greatly from the foundation laid down by Allen et al. (1987). These models predict duration effects above word level with great success, including lengthening effects on all levels of prosodic boundaries, word stress, emphasis and their interactions, the effect of syntactic structure on duration, function words and content words, questions and statements among other. What is missing in the literature from both corpus research and controlled experiment on Mandarin duration is the effect of word-internal structure on syllable duration. So this dissertation attempts to fill the gap by studying the duration patterns associated different word-internal structures.

Mandarin trisyllabic words have two major branching structures—left-branching ([[AB]C]) and right-branching ([A[BC]]). Mandarin quadrisyllabic words have four possible structures—[[A][B][C][D]], [[AB][CD]], [[ABC]D] and [A[BCD]]. My dissertation includes these word-internal structures. Furthermore, in quadrisyllabic words, reduplication has a profound influence on duration, so I study all possible reduplication patterns in quadrisyllabic words.

2.6 Research Questions

This dissertation investigates the cross-regional word duration patterns of Mandarin words among speakers from Beijing, Taiwan, and Malaysia. The general questions addressed in this dissertation are:

(1) What are the duration patterns of Mandarin words with different length and structures?

(2) Are there any universal duration patterns in Mandarin words?

(3) What are the functions of linguistic factors, social factors, and their interactions?

Three speech production studies are conducted to focus on exploring the duration patterns and answering the questions.

2.7 Summary

This chapter introduced the regional variation of Chinese and the distribution of words by different word length, and also reviewed the phonetic, morphological, prosodic and sociolinguistic factors leading to duration variation. In addition, this chapter also discussed the major research approaches used in previous works focusing on duration measurement and modeling, and then proposed the research methodology used in this dissertation. Chapter 3 will present Study 1 on monosyllabic and disyllabic words.

Chapter 3

Study 1: Duration Patterns of Monosyllabic and Disyllabic Words

This dissertation examine duration patterns of Chinese words with various length. Chapter 3 will report and discuss the duration patterns on Chinese monosyllabic and disyllabic words in story reading task.

3.1 Introduction

To investigate duration patterns of words with different length, it is necessary for us to have an overall picture of Chinese word length. In Mandarin Chinese, the cumulative frequency of monosyllabic and disyllabic words is around 88% (Table 2.1) in the word list compiled from a 20 million annotated data set from the Contemporary Chinese Corpus, including words occurring 50 or more times in the corpus. It means that monosyllabic and disyllabic words cover 88% Chinese language use in daily life. Therefore, investigation of the duration of monosyllabic and disyllabic words is one of the major tasks of Mandarin word duration research.

Study 1 combines the approach of controlled experiment and the corpus construction approach together to build a speech corpus of monosyllabic and disyllabic words used within a context of story. This chapter will first introduce the research goal and the detailed research questions/ hypotheses of Study 1 in Section 3.2. The methodology section (Section 3.3) will address how the factors related to durational variation are designed in the materials and how these factors are coded in the corpus. Next, this chapter will present the statistical analysis and report the major findings of Study 1 (Section 3.4). Finally, the chapter will discuss the

durational patterns based on these factors (Section 3.5) and draw a conclusion (Section 3.6).

3.2 Research Questions

Chapter 2 reviewed the major linguistic factors leading to duration variation. The goal of this study is to investigate the effects of factors and their interactions on duration variation of monosyllabic and disyllabic words. The research questions addressed in Study 1 are:

(i) Do sociolinguistic factors (region and gender) affect duration?

(ii) On the syllable level, do syllable structure and syllable position affect duration?

(iii) On the word level, do word category, word frequency, word length, and word position in a sentence affect duration?

(iv) On the sentence level, do sentence length (number of syllables), sentence position in a story affect duration?

3.3 Methodology

3.3.1 Participants

Thirty native speakers of Mandarin from three regions–Beijing, Taiwan, and Malaysia, participated in this study. Their gender and age information are in Table 3.1. None of the participants reported speech or hearing disorders. Participants were paid 15 dollars or 100 RMB for their participations.

	Number of Subjects (Gender)	**Age Mean (SD)**
Group-Beijing	10 (5M and 5F)	23.7 (3.65)
Group-Taiwan	10 (5M and 5F)	28.6 (3.13)
Group-Malaysia	10 (5M and 5F)	27.1 (5.10)

Table 3.1: Subject information

Ten native speakers of Beijing Mandarin have never had long-term (more than four months) oversea study experience before age 20. Furthermore, their parents are all native speakers of Beijing Mandarin, who were born and raised in Beijing.

Ten native speakers of Taiwan Mandarin were recruited. None of them had experienced living outside of Taiwan for a period longer than six months before the age of 20. All of the participants reported Taiwan Mandarin as their native and first language. Five of them reported a dialect background—Min or Hakka, but only two reported a native-level proficiency in their dialect language.

Ten native speakers of Malaysia Mandarin were recruited. They were born and raised in Malaysia and their parents are at least second generation Chinese-Malaysian. They all went to Chinese elementary school, middle school, and high school. These schools are a part of a separate school system in Malaysia, where the language of instruction is standard Chinese (see Section 2.1 Regional variation). None of them have long-term oversea experience before college.

The same thirty participants also took part in the sentence reading task in Study 2 and 3; the methodology sections of Study 2 and 3 will not list the participants' information again.

3.3.2 Materials

Six stories from *Aesop's Fables* (Appendix A) are used in the story reading task in this study. The story line comes from *Aesop's Fables*, but the stories are translated into Chinese and modified to fit specific word and contextual conditions for the research purpose of this study. All six stories were examined by four native speakers of Mandarin—two native speakers of Beijing Mandarin and two native speakers of Taiwan Mandarin. Each story was included in the reading task only after the four examiners reached an agreement on lexical use and grammaticality.

Table 3.2 displays the detailed information of each story. The length of each story is from 108 to 227 Chinese characters (syllables), with an average length of 148 syllables per story,

with a total of 885 syllables in all six stories. Disregarding tonal differences and repetitions, 210 unique syllables are used in these stories. Syllables used in these stories cover around 53% syllables in Standard Mandarin with approximately 400 syllables.

	Number of syllable	Monosyllabic word (type)	Monosyllabic word (token)	Disyllabic word (type)	Disyllabic word (token)	**Word TTR**
Story1	108	27	40	24	34	61%
Story2	113	29	51	23	31	63%
Story3	103	22	43	26	30	66%
Story4	184	42	70	38	57	63%
Story5	227	42	83	52	72	61%
Story6	150	31	56	32	47	61%
Total	885	197	345	208	271	66%

Table 3.2: Story material word counts

Many previous works provided the algorithm and analysis of Chinese text segmentation (Wu and Tseng, 1993; Lin et al., 1993; Sproat et al., 1996; Chien, 1997; Cheng et al., 1999; Dai et al., 1999; Sproat and Shih, 2002). In this study, the text script of all six stories are segmented into Chinese word by the toolkit NLPIR (Zhang, 2014) with manual check. The distribution of word types in Table 3.2 is based on the result of segmentation. Disregarding duplicate words, each story has 51 to 94 words, including both monosyllabic words and disyllabic words. The TTR (type/token ratio is used to indicate the richness of the text) of the materials is around 66%.

3.3.3 Procedure

After participants had completed the language background forms, they proceeded with the story reading task. Given the different orthographic systems used among these three Mandarin regions, all the materials come in two versions—simplified and traditional Chinese character versions. The different versions are distributed to the participants based on their self-reported place of origin and preferred orthographic system.

Each story is presented in Chinese orthography on a computer screen. The computer screen is placed in front of the participants. Participants were asked to read the stories at their normal speech rate. The audio recordings were conducted in a sound-attenuated booth. An AKG-C520 head-worn condenser microphone was used to record acoustic signals onto a Marantz PMD-570. The sample rate of the recording is 48000Hz.

Data from the reading task was manually checked against the original text of the stories. If there have inconsistencies between the original text and the individual participant's speech production, the differences are precisely transcribed based on the production data. Audio files with their corresponding text transcriptions were submitted for automatic segmentation by using the Penn Phonetics Lab Forced Aligner (Yuan and Liberman, 2008, 2009) with manual check. Syllable duration is measured in this study.

3.3.4 Data Transcription

This section will introduce the corpus construction and data coding.

Transcription of Speech data

Participants were asked to read the stories as accurately as they can in the reading task, but each participant produced various speech data, which may be inconsistent with the original story text to a certain extent. Figure 3.1 demonstrates the process from original text to transcription of participants' speech data.

For the transcription of each participant, the inconsistencies or mistakes are caused by different reasons during the speech production. For example, the sentence

他们	正好	看到	有	个	人	走	过
tamen	*zhenghao*	*kandao*	*you*	*ge*	*ren*	*zou*	*guo*
they	just	see	have	Classifier	person	walk	Aspect

They just saw a person passed by.

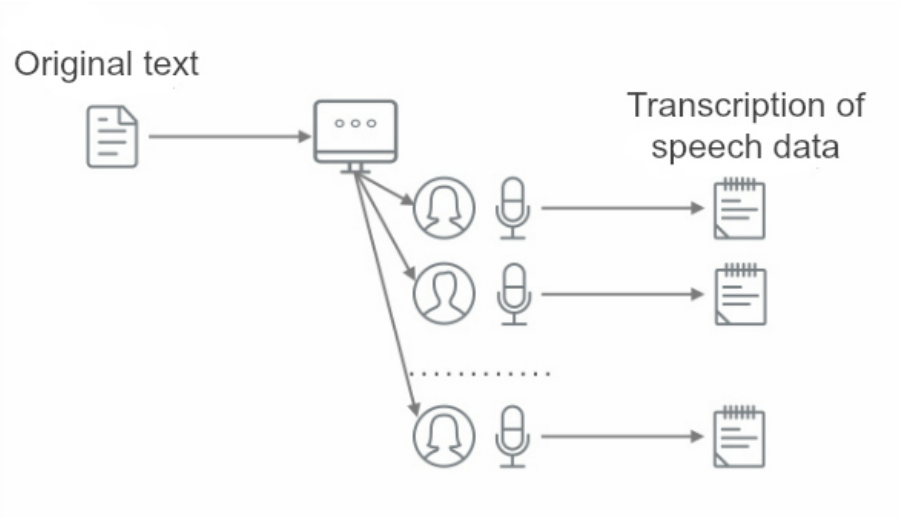


Figure 3.1: Original text to transcription data

can be produced in several different ways by participants. The different types of mistakes are listed in Table 3.3.

Type of Mistake	**Inconsistent Production**
Insert	*tamen zhenghao kandao you **yi**ge ren zou guo* (Insert 'yi')
Delete	*tamen zhenghao kandao you ~~**ge**~~ ren zou guo* (Delete 'ge')
Replace	*tamen **ganghao** kandao you ge ren zou guo* (Replace 'zhenghao' with its synonym 'ganghao')
Break	*tamen zhenghao**,** kandao you ge ren zou guo* (Break the sentence with a detectable pause)
Repeat or Repair	*tamen zhenghao **kan kandao** you ge ren zou guo* (Repeat the syllable 'kan')
Mix	*tamen **ganghao ,** kandao you **yi** ~~**ge**~~ ren zou guo* (Mix any two or more mistakes together)

Table 3.3: Type of inconsistency

Table 3.3 displays six possible mistakes produced by participants.

(1) Insertion: participant inserted a syllable or syllables in a sentence.

(2) Deletion: participant deleted a syllable or syllables in a sentence.

(3) Replacement: participant replaced a word with its synonym having the identical number of syllables in the sentence.

(4) Break: participant broke the sentence into different parts with a pause or filler words (*a* 啊, *e* 哦, *um* 嗯, *o* 噢, *zhege* 这个 *and nage* 那个).

(5) Repeat/repair: participant repeated or repaired certain parts of the speech.

(6) Mix: mistake combined multiple mistakes among insertion, deletion, replacement, break and repeat.

All the sentences with mistakes are coded in the corpus. When participant made the first three kinds of mistakes, they may not be consciously aware. These three kinds of mistakes may not cause disfluency in speech production. Wu (2011) discussed that language disfluency is related to repairs, hesitation, repetitions, filler word insertion, and pause etc. They may have effect on duration. In this study, 'break', 'repeat', 'and repair' can affect the fluency and grammaticality of the sentence. For example, *kan* "look at"or *kandao* "see" are both grammatical expressions, but '*kan kandao*' in mistake (4) is not grammatical and it may further affect the syllable duration.

3.3.5 Corpus

In total, there are 27,251 syllables (Beijing 9,026 syllables, Malaysia 9,281 syllables, and Taiwan 8,944 syllables) from 30 speakers in the transcription. These 27,251 syllables are from 18,953 words (10,655 monosyllabic words and 8,298 disyllabic words) in 3,391 sentences.

Among 3,391 sentences, there are 684 sentences labeled as 'wrong' sentences, which are not consistent with the original text. Shriberg (2001) reported that in natural conversation speech produced by native speakers, around 10% of the word and over 1/3 of the utterances are affected by disfluency. In Study 1, 684 'wrong' sentence is around 20% of the total data.

It means the probability of 'wrong' sentence produced by participants is not very high in this reading task.

In order to examine the accurate patterns of monosyllabic and disyllabic words in Mandarin, sentences with mistakes should be ruled out from the database. Therefore, 2,707 sentences without mistake produced by 30 participants are used in the further analysis. In term of word, there are 8625 monosyllabic words and 6,863 disyllabic words in these 2,707 sentences, including 22,351syllables (Beijing 7,249 syllables, Malaysia 7,485 syllables, and Taiwan 7,617 syllables) in total. Each syllable is coded via the following eleven factors.

(1) Region: with three levels—Beijing, Taiwan and Malaysia;

(2) Speaker: 30 speakers;

(3) Gender: male and female;

(4) Syllable positions: with three levels–1s-1(syllable in the monosyllabic word), 2s-1(the first syllable in a disyllabic word) and 2s-2 (the second syllable in a disyllabic word);

(5) Syllable structure: with twelve levels–V, VG, GV, CV, VN, GVG, CVG, CGV, GVN, CVN, CGVG and CGVN covering all the possible syllable structures in Mandarin;

(6) Word category: for the word where the syllable is, the word category is coded based on the result of word segmentation from NLPIR (see Section 3.3.2). There are eleven Chinese word categories used in the corpus (Table 3.4).

(7) Normalized word frequency: the data of word frequency is based on the word list of *The Contemporary Chinese Corpus*. The normalized frequency is calculated via log algorithm.

(8) Word position in the sentence: the word position is labeled with three levels–fwd (the first word in a sentence), mwd (the middle word in a sentence) and lwd (the last word in a sentence);

Chinese	Word Category	Abbreviation in Coding
ming ci	noun	n
dong ci	verb	v
xingrong ci	adjective	a
fu ci	adverb	d
dai ci	pronoun	r
lian ci	conjunction	c
shuliang ci	numeral; classifier; num-cl. compound	nq
jie ci	preposition	p
fangwei ci	noun of locality	f
tan ci	interjection; exclamation	y
zhu ci	particles; aspects	u

Table 3.4: Word categories.

(9) Sentence position in the story: the sentence position is labeled with three levels–fst (the first sentence in a story), mst (the middle sentence in a story) and lst (the last sentence in a story);

(10) Syllable count in each sentence;

(11) Word count in each sentence;

3.4 Data Analysis and Result

Study 1 used corpus method to cover the various factors to approximate the natural speech environment with unbalanced data of each factor. The traditional linear regression model with the requirement of balanced data is not the optimal way to build the duration model. However, because of the advantage in dealing with missing values, the mixed effects model is often preferred over the traditional approaches. Therefore mixed effects regression model not requiring balanced data is adopted in the data analysis.

3.4.1 Statistical Analysis

According to the properties of the factors, 9 fixed effects and one random effect are used to predict the dependent variable—syllable duration in monosyllabic words and disyllabic words. The fixed effects are region, gender, syllable positions, syllable structure, word category, word frequency, word position, sentence position, syllable count of a sentence. The random effect is the speaker effect. Because each speaker of this study occurs within one and only one level of region factor, there are two strength measurements on each speaker within each of the regions and combine with the average strength for speaker. Therefore, speaker is nested within region in grouping factors, and the nesting of factors needs to be considered in the model. Moreover, in order to examine the cross-regional differences, it is necessary to take the interaction of two variables—region and syllable position in consideration.

The mixed effects regression model used to analyze the syllable duration via 'lmer' function in R package lme4 (package version 1.1-9, `https://CRAN.R-project.org/package=lme4`) (Bates et al., 2014) is in the following formula. '1|region:speaker' indicates that speaker is nested with region, and 'word category $*$ log(word-frequency)' represent the interaction between word frequency and word category.

duration $\sim$ region $*$ syllable position+ gender + syllable count + sentence position
+ word position + word category + log(word-frequency)
+ syllable structure + (1|region:speaker)

Table 3.5 summarized the result of random effect in this model. The standard deviation is the measure of how much variability in the random effect that we added into the model. In the model, speaker is nested within region, so the result show variability of speaker among three regions respectively. The standard deviations show that the noise from speakers of Beijing, Malaysia and Taiwan are different. Taiwan speakers have largest variability (19.27 ms), and Malaysia speakers have the smallest one (13.21 ms). The standard error of residual

(51.28 ms) stands for the variability that is not due to the random effect.

Random effects:	**Groups**	**Names**	**Std.Dev.**
	Speaker	(Intercept)	17.94
		Malaysia	13.21
		Taiwan	19.27
	Residual		51.28

Table 3.5: Summary of random effects of Study 1 (Number of observations: 22351)

Mixed effects regression models do not automatically report p-value of each effect. In this study, I use 'mixed' function in R package 'afex' (R package version 0.15-2. `http://CRAN.R-project.org/package=afex`) (Singmann et al., 2015) to calculate the p-values for all the fixed effects of the model.

Effect	**Stat (Chisq)**	**p value**
Region	1.29	0.56
Gender	2.52	0.14
Syllable count	0.02	0.89
Sentence position	120.21	0.001 ***
Syllable position	515.1	0.001 ***
Syllable structure	4331.87	0.001 ***
Word position	3204.85	0.001 ***
Word category	1909.65	0.001 ***
Word frequency	1504.29	0.001 ***
Region $*$ Syllable position	713.19	0.001 ***

Table 3.6: Results of fixed effects

Table 3.6 summarizes the results of fixed effects in the model. Sentence position (p=0.001), syllable position (p=0.001), syllable structure (p=0.001), word position (p=0.001), word category (p=0.001), word frequency (p=0.001), and the interaction between region and syllable position (p=0.001) are significant in the model; region (p=0.56), gender (p=0.14), syllable count (p=0.89) are not significant.

3.4.2 Syllable Duration

In Figure 3.2, the density plots represent different syllable positions: pink represents the syllables in the monosyllabic words; green represents the first syllables in the disyllabic words; and blue represents the second syllables in the disyllabic words. The dash lines represent the mean values of the corresponding syllable positions. Among all three conditions, there is a large overlap area. The mean values of these three conditions are really close to each other (Table 3.7), but the distribution of syllables in monosyllabic words are shorter than syllables in disyllabic words.

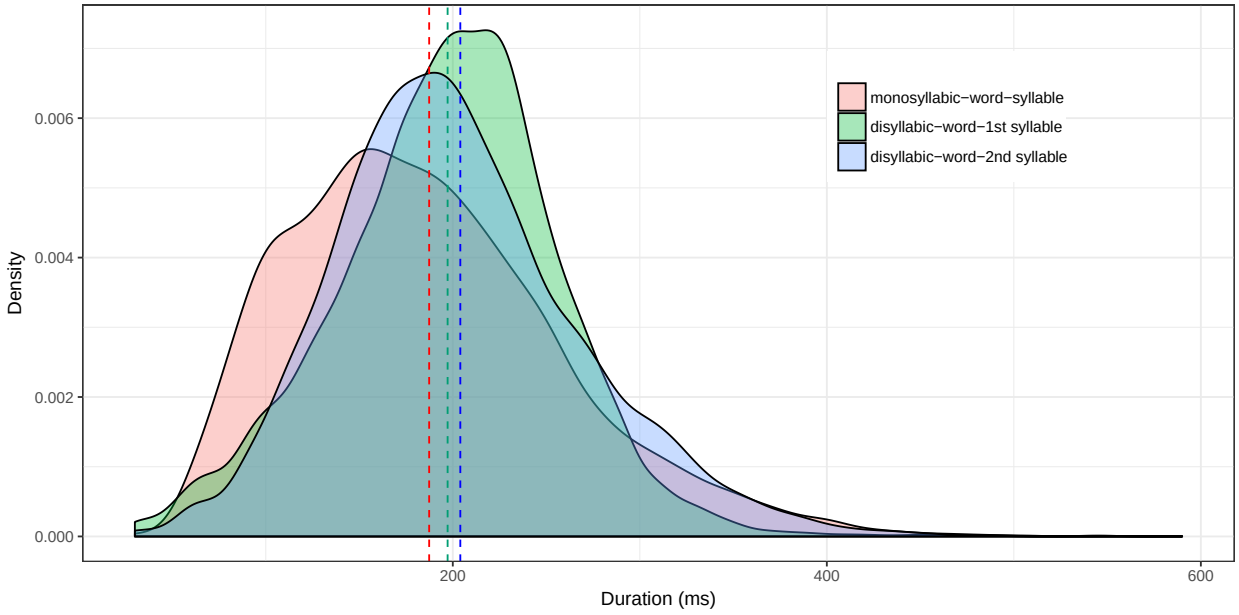


Figure 3.2: The distribution of syllable duration.

Furthermore, Table 3.7 displays the median, mean, and standard deviation for each syllable status. Syllables in monosyllabic words have smaller mean and median than the syllables in disyllabic words, but with larger standard deviation caused by the high frequency monosyllabic function words (discussed in Section 3.4.5).

Syllable Position	Median (ms)	Mean (ms)	SD (ms)
Syllable of Monosyllabic Word	180	185.1	72.9
1st Syllable of Disyllabic Word	200	197.3	58.9
2nd Syllable of Disyllabic Word	200	202.4	66.1

Table 3.7: Median, mean and SD of syllable duration (ms).

3.4.3 Word Position

Word position has significant effect (p=0.001) on syllable duration in both monosyllabic and disyllabic words. In Figure 3.3, the upper panel is the syllable duration of monosyllabic words and the lower panel is the syllable duration of disyllabic words by word positions. There are three subplots in each panel: fwd, lwd, and mwd represent the first word, last word, and middle word in sentence respectively. 2s-1 and 2s-2 represent the first and second syllable positions of a disyllabic word. In monosyllabic word, Figure 3.3 (a) shows that last word > first word > middle word. In disyllabic word, Figure 3.3 (b) displays that the two syllable positions in last word of a sentence are all longer than two syllable positions in in first word and middle word. For the last word in a sentence, the duration pattern of the disyllabic word shows short-long pattern. Instead, the middle word and first word all display the long-short pattern.

The results prove that the sentence final position has lengthening effect on both monosyllabic word and disyllabic word; moreover, this sentence final lengthening effect prolongs both syllable positions of disyllabic word, and has stronger effect on the last syllable than the first syllable of the disyllabic word. These results from Study 1 are high consistent with the sentence final effect in the previous studies (Gaitenby, 1965; Lindblom and Rapp, 1973; Oller, 1973).

It's also worth noting that the duration pattern of disyllabic words on the non-sentence-final positions is long-short pattern. In term of fwd in the Figure 3.3 (b), the first syllables are marginally longer than the second syllables. In term of mwd in Figure 3.3 (b), the first syllables are obviously longer than the second syllables in the disyllabic words. For non-

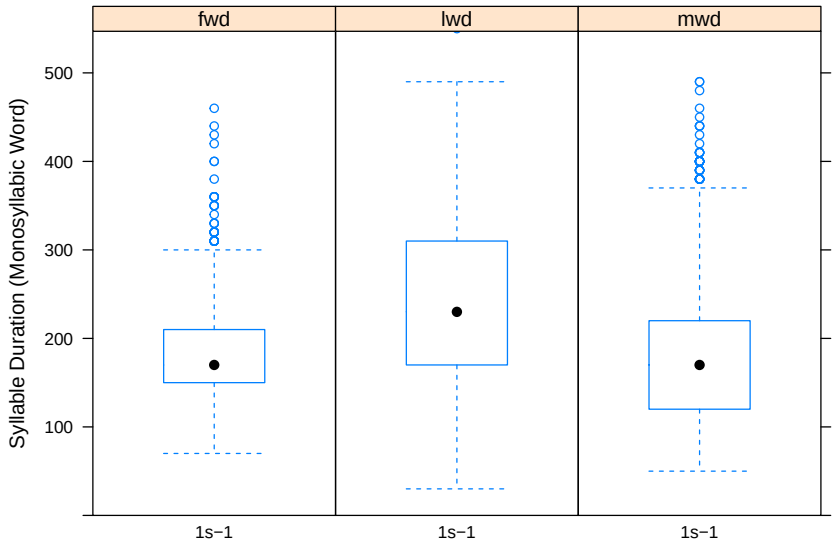
sentence-final words (fwd and mwd), the first syllables (mean=191ms) are 5ms longer than the second syllables (mean=186ms) on average. Therefore, syllable duration of disyllabic word on the non-sentence-final position shows long-short pattern, which is the opposite of Xu and Wang (2009).

Without sentence final lengthening effect, Mandarin disyllabic words show long-short pattern on these two syllable positions. Feng (1998, 2005) proposed that Chinese words are parsed from left to right into trochaic disyllabic foot. Duanmu Duanmu (2000) and Sui (2013) also proposed that polysyllabic words in Standard Chinese have the trochaic foot from left to right. The long-short pattern of disyllabic word could be taken as the phonetic evidence to support the trochaic form of Chinese words. Furthermore, Kochanski et al. (2003) reported the strong-weak pattern of prosodic strength on polysyllabic Chinese words. Having the longer duration on the first syllable and the shorter duration on the second syllable could be seen as the durational performance of the strong-weak prosodic strength pattern in disyllabic words.

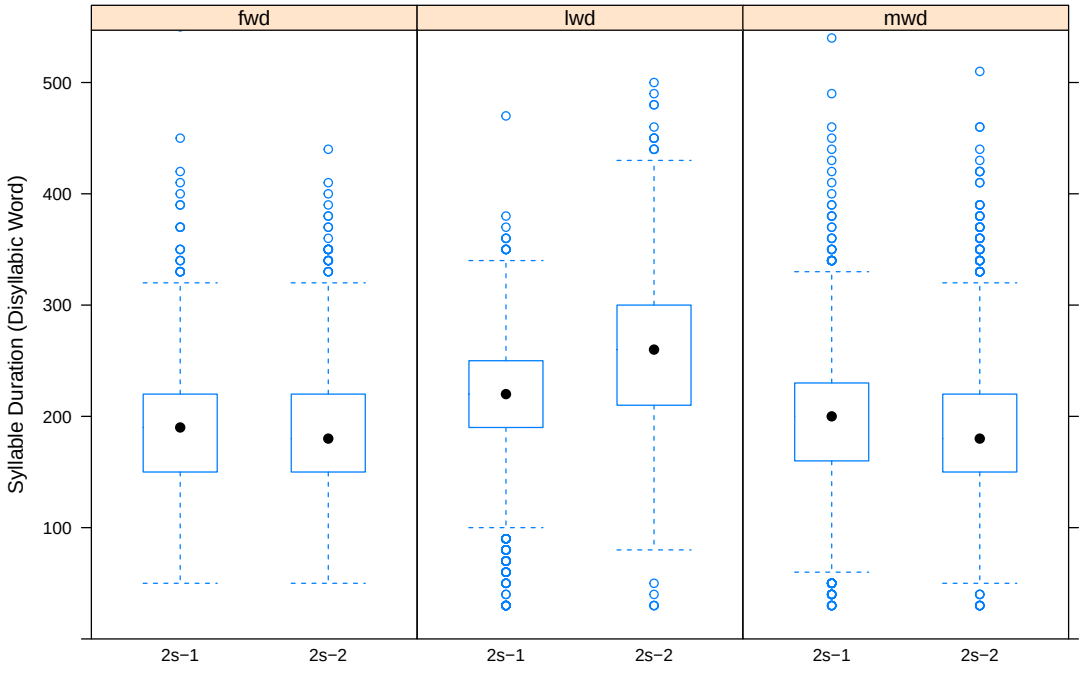
3.4.4 Sentence Position

In the mixed effects regression model, sentence position has significant effect (p=0.001) on syllable duration in both monosyllabic and disyllabic words. In Figure 3.4, the upper panel is the syllable duration of monosyllabic words and the lower panel is the syllable duration of disyllabic words. There are three subplot in each panel: fst, lst, and mst represent the first sentence, last sentence, and middle sentence in story respectively. 2s-1 and 2s-2 in the lower panel represent the first and second syllable positions of the disyllabic word. In Figure 3.4 (a), monosyllabic words in the first sentence of the stories are longer than monosyllabic words in the middle sentences and last sentences. In Figure 3.4 (b), both the first syllables and the second syllables in the first sentences are longer than the corresponding syllables of disyllabic words in the middle sentences and last sentences.

Regardless of the difference between monosyllabic words and disyllabic words, Table

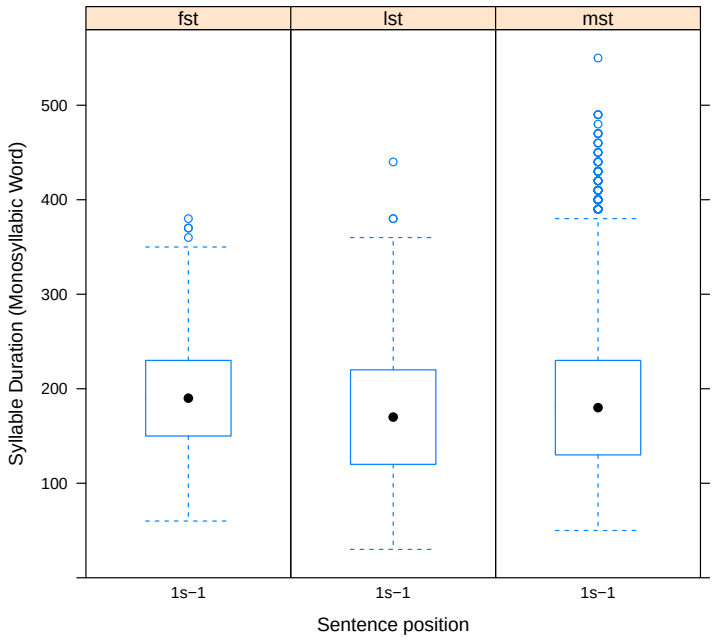


(a) Syllable of Monosyllabic Word

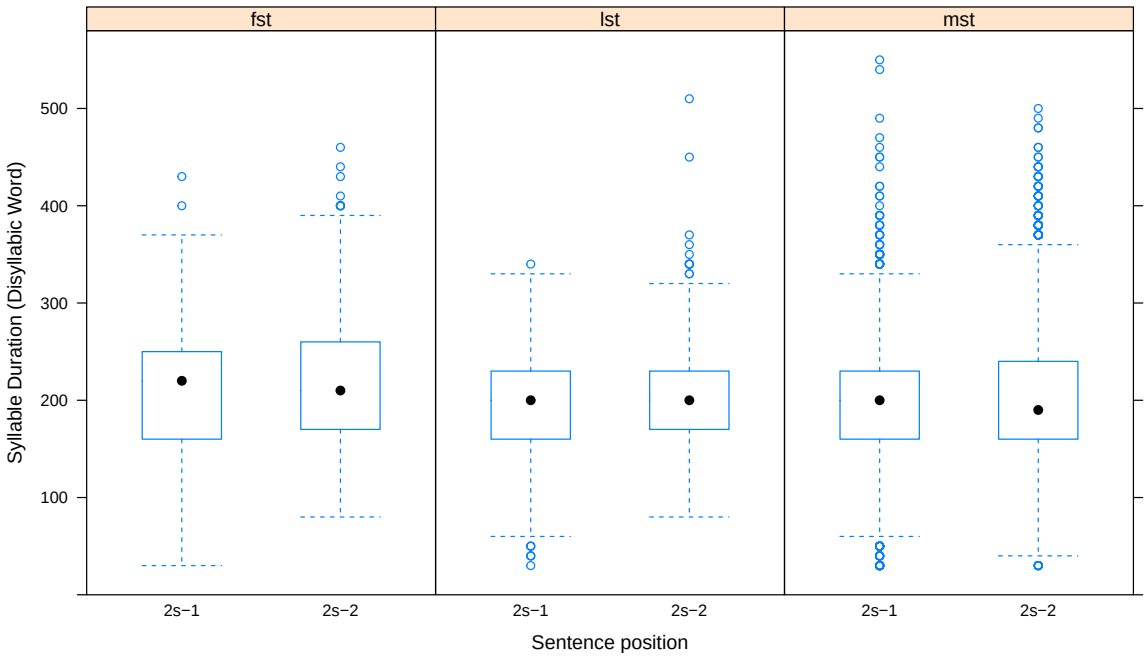


(b) Syllable of Disyllabic Word

Figure 3.3: Syllable duration by word positions.



(a) Syllable of Monosyllabic Word



(b) Syllable of Disyllabic Word

Figure 3.4: Syllable position by sentence position.

3.8 shows that syllables in the middle sentences and final sentences have identical medians (190ms) and similar means (193.4ms and 190.9ms). But syllables in the first sentences have larger median (200ms) and mean (207.6ms). These differences suggest that the first sentence in a paragraph or story leads to longer syllable duration than other sentence positions.

Sentence position	**Median (ms)**	**Mean (ms)**	**SD (ms)**
First sentence	200	207.6	69.7
Middle sentence	190	193.4	67.3
Last sentence	190	190.9	61.9

Table 3.8: Syllable duration by sentence position (ms)

There are two possible reasons to explain the first sentence lengthening effect. First, the information in the first sentence of a paragraph is all new information for speakers. Second, speech production is an activity that requires planning, even in a reading task. Speech is planned at different levels (syntax and phonology) and different units (syllable and word) (Fromkin, 1973; Garrett, 1975; Shattuck-Hufnagel, 1979; Meyer and Gordon, 1985; Levelt, 1993). On articulation of speech planning, speakers control the volume of inhaled air (McFarland and Smith, 1992; Winkworth et al., 1995). When speaker meet a new paragraph, especially the new information in the first sentence, they need to have longer time to finish linguistic planning and motor programming.

Based on the discussion of word position and sentence position in Section 3.4.3 and Section 3.4.4, it suggests that the first sentence position and last word position both have lengthening effects on syllable duration. Therefore, it is reasonable to predict that words under these two conditions, namely the last word in the first sentence, should have long syllable duration. In Figure 3.5, each subplot represents the a type of word position, and within each subplot, the three box plots represent the three different sentence positions. In the second subplot, it is clear that last words in three different types of sentences have longer syllable duration than the other two subplots, and the last word (lwd) in the first sentence (fst) has the longest syllable duration.

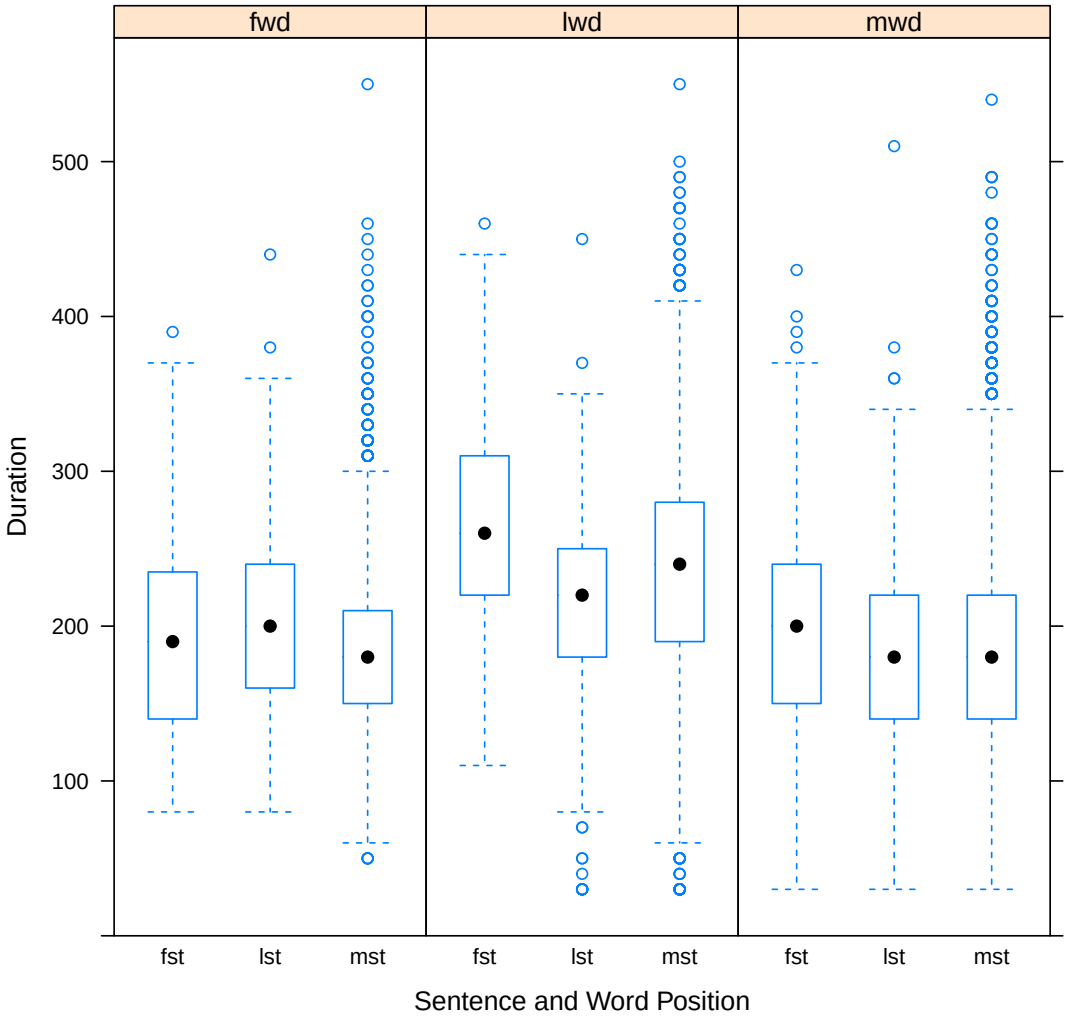


Figure 3.5: Syllable duration by word positions and sentence positions

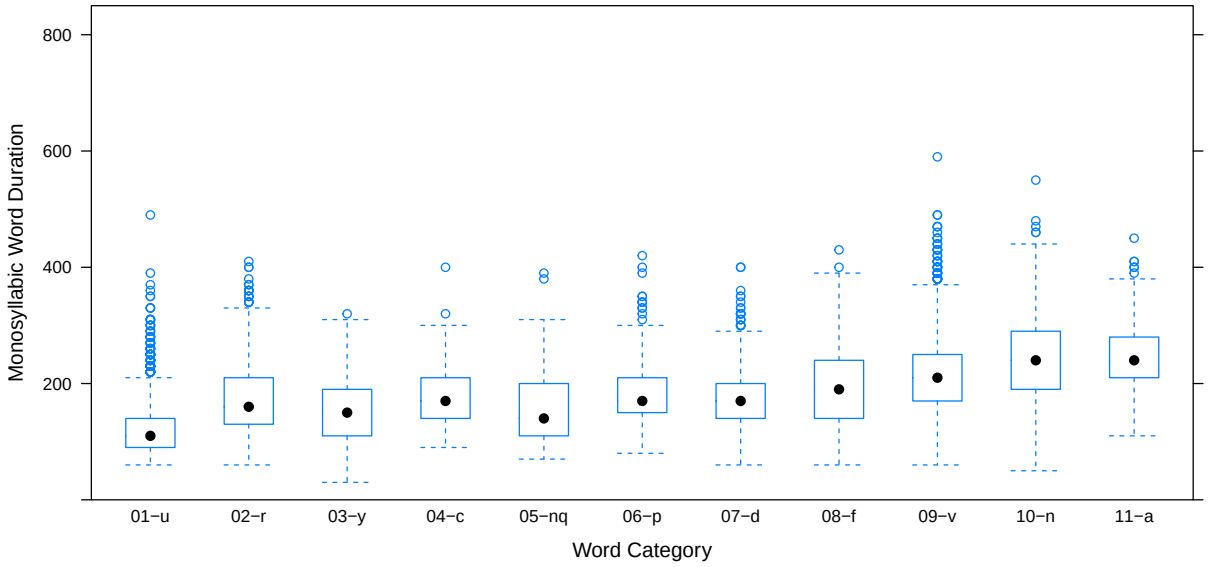
3.4.5 Word Category

In this section, word duration by categories are examined. Monosyllabic word and disyllabic words are classified into eleven word categories (Table 3.4). Table 3.9 lists the number of word of each word category. In order to demonstrate an overall impression about the category effect on word duration, the duration of disyllabic word is discussed in this section. The duration of the disyllabic word is the sum of the duration of the first syllable and the duration of the second syllable.

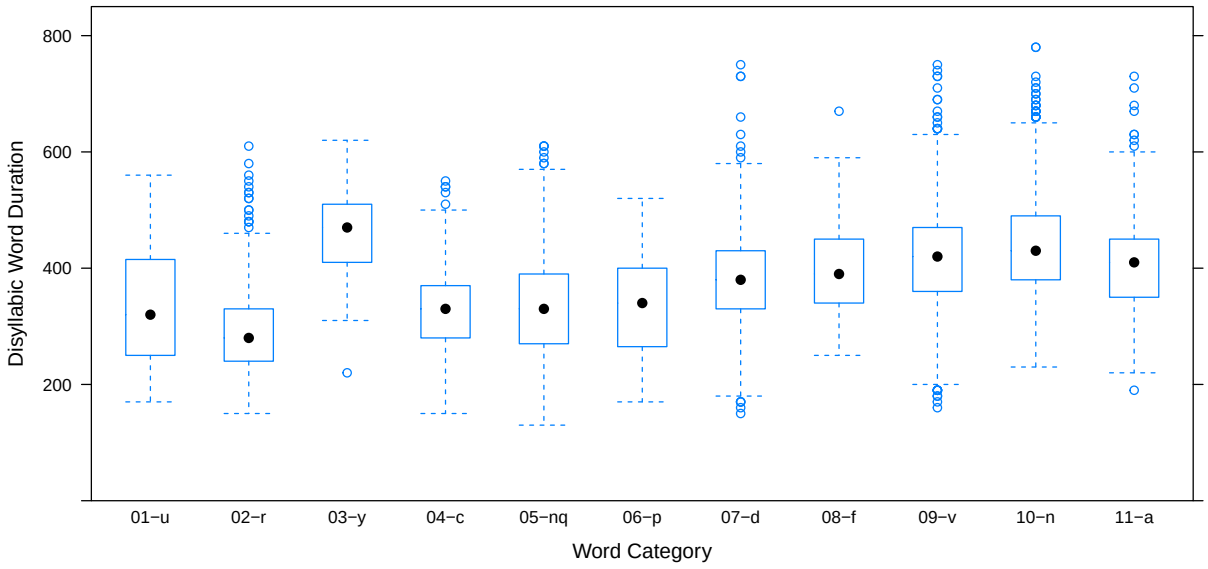
Word Category	**Monosyllabic Word**	**Disyllabic Word**
noun(n)	786	2604
verb(v)	2223	1597
adjective(a)	526	321
adverb(d)	1084	824
pronoun(r)	796	469
conjunction(c)	86	208
numeral or classifier (nq)	129	583
position (f)	317	117
proposition (p)	504	83
interjection or exclamation (y)	304	30
particle (u)	1870	27
Total:	8625	6863

Table 3.9: Number of word by category

Figure 3.6 displays the word duration differences among these 11 word categories. The result of monosyllabic word duration in the upper panel demonstrates that content words such as adjective, noun and verb have longer word duration, and function words such as particle, interjection and conjunction have shorter duration. Function words usually are high frequency words, and word frequency also have effect on word duration (see more discussion on word frequency in Section 3.4.6). The lower panel of Figure 3.6 displays the word duration of disyllabic word. Except the two categories of particle and interjection, the other word categories show the similar pattern in monosyllabic words. Both disyllabic particle (30 samples) and interjection (30 samples) have extremely small sample size in the corpus, and this may be the reason leading to large variance and inconsistency.



(a) Monosyllabic Word



(b) Disyllabic Word

Figure 3.6: Word Duration by Category

3.4.6 Word Frequency

Word frequency is another significant fixed effect (p=.001) on syllable duration. High frequency words generally have better visual recognition and shorter duration than low frequency words in reading and speaking (Section 2.4.4). Therefore, high frequency words are generally produced faster than low frequency words. It means there should have a negative correlation between word frequency and syllable duration.

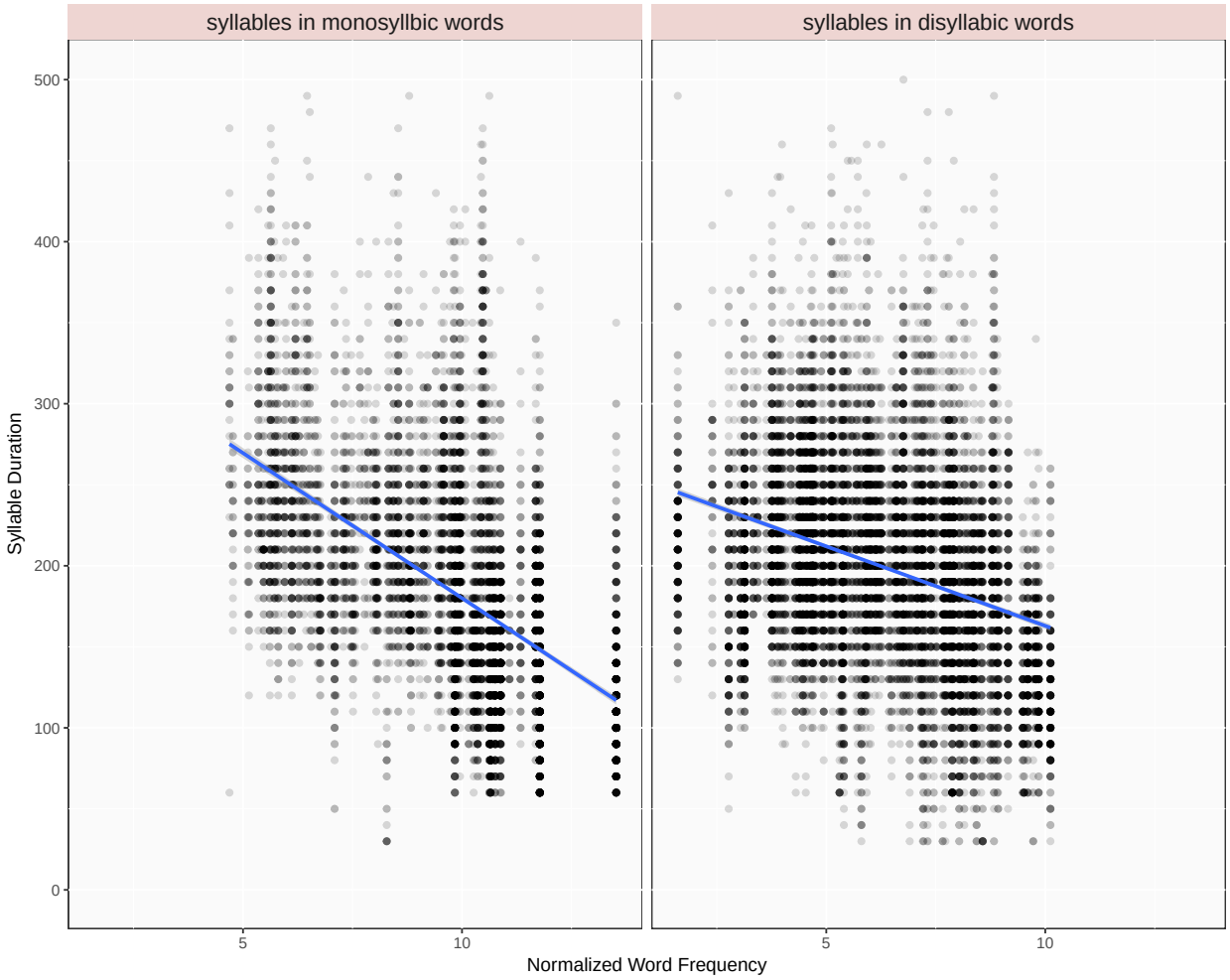


Figure 3.7: Correlation between word frequency and syllable duration

Figure 3.7 shows the relationship between syllable duration and word frequency. The left panel represents the duration data of syllables in the monosyllabic words, and the right panel is the duration data of syllables (both the first and second syllables) in the disyllabic words.

Spearmen correlation tests are performed on both data sets respectively. For syllables in the monosyllabic words, the result of Spearmen correlation test indicates that there is a true correlation between syllable duration and word frequency (S = 1.7081e+11, p < 2.2e-16), and the Spearmen correlation coefficient (r=-0.5973) indicates there is a relatively strong negative correlation between duration and frequency. For syllables of disyllabic words, the result of Spearmen correlation test indicates that there is a true correlation between syllable duration and word frequency (S = 5.7395+12, p < 2.2e-16), and the Spearmen correlation coefficient (r=–0.3317) indicates there is a moderate negative correlation between these two variables.

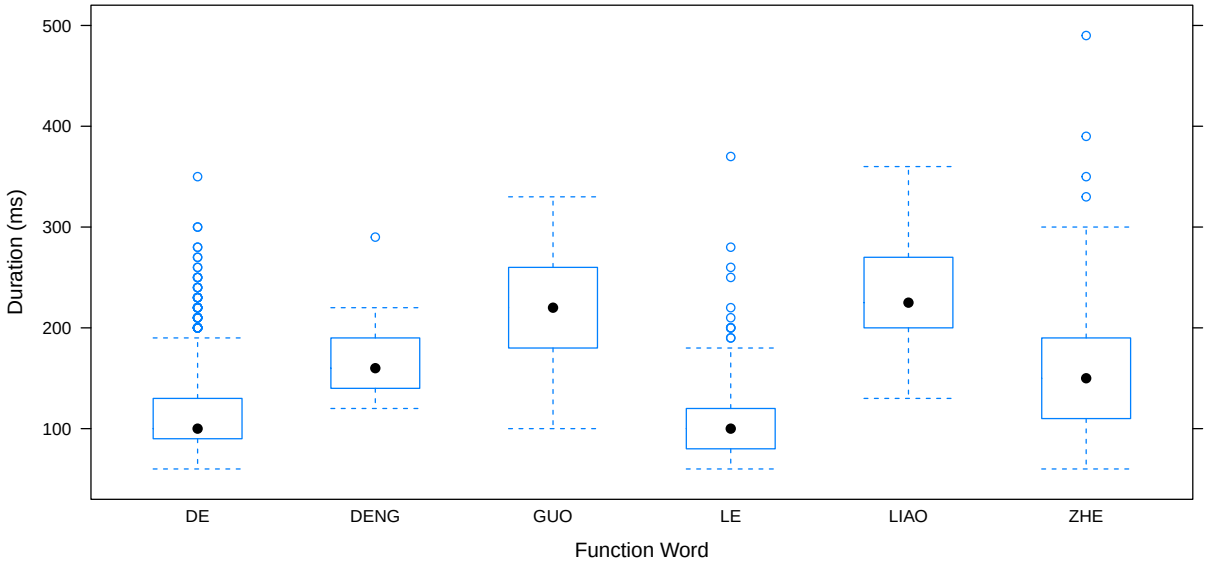


Figure 3.8: Duration of high frequency monosyllabic particles

High frequency function words often have reduction in speech. Take the monosyllabic particles in Chinese as examples, Figure 3.8 display duration of six particles, and the two shortest particles are *de* 的 and *le* 了 with duration around 100 ms. *De* (1361 samples) and *le* (773 samples) are the most frequent words used in the corpus of Study 1. It is consistent with the result based on large scale corpus of Chinese. For example, *de* and *le* are the top two words in the word list of the Contemporary Chinese Corpus data set (20 million Chinese

characters): *De* is the most frequent word with the percentage of 7.795% and *Le* is the second most frequent word (1.362%) in the Contemporary Chinese Corpus. High frequency monosyllabic function words exhibit the shortest syllable duration among all the categories.

3.4.7 Syllable Structure

Syllable structure is highly related to syllable duration (Section 2.4.1). Syllables with more phonemes tend to have longer syllable duration. There are 12 possible syllable structures in Chinese, but Chinese syllables do not show even distribution on these 12 categories.

	1 Phoneme	**2 phonemes**	**3 phonemes**	**4 Phonemes**
Syllable structure	V: 1279	VG: 42 GV: 275 CV: 8735 VN: 130	GVG: 594 CVG: 2714 CGV: 1074 CVN: 3566 GVN: 505	CGVG: 1224 CGVN: 2213

Table 3.10: Number of syllables of each syllable structure type

In the corpus of Study 1, there are 22,351syllable in total, and Table 3.10 lists the number of syllables of each syllable structure type. In Mandarin Chinese, the most common syllable structure is CV (38.3%), and the least common one is VG (0.23%).

Figure 3.9 demonstrates the distribution of syllable duration among all the syllable structures. The results agree with the findings in Shih and Ao (1997). Syllable structures with more phonemes tend to have longer duration and syllables with less phonemes tend to have shorter duration. Moreover, the maximal structure CGVN has the longest duration, and the simplest structure V has the shortest duration.

3.5 Discussion

The statistical results in Section 3.4.1 indicate that the interaction between region and syllable position is significant (p=0.001). It means there is regional differences among these

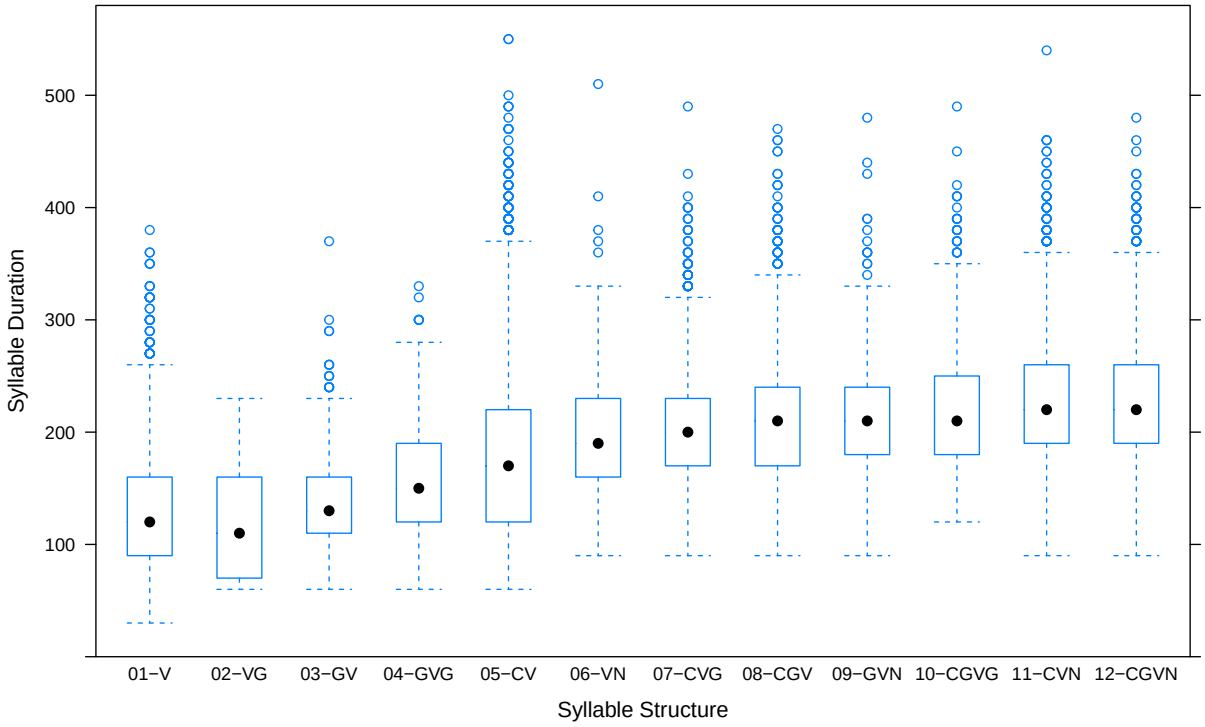


Figure 3.9: Syllable duration by syllable structures

three regions. Duration of syllables in disyllabic words are different on account of the syllable position within the words. The result in Table 3.7 shows that the second syllables have slightly longer mean than the first syllables.

It seems that there is a short-long duration pattern of the disyllabic words. Xu and Wang (2009) reported that in a short phrase of 1–4 syllables, duration is longest in the final position. It seems that the short-long syllable duration pattern of the disyllabic word in Study 1 is consistent with Xu and Wang (2009). However, Xu and Wang (2009) used the controlled experimental design, and all the target stimuli are at either the sentence final or the sentence initial positions of four different carrier sentences. Many early works had reported the lengthening effect of sentence final (Gaitenby, 1965; Lindblom and Rapp, 1973; Oller, 1973) and the non-word-final shortening effect (Lindblom and Rapp, 1973; Oller, 1973; Klatt, 1975b). And these two effects may cause the duration pattern of short-long in disyllabic words in Xu and Wang (2009).

Moreover, in order to guarantee the minimal dialectical variability, Xu and Wang (2009)'s study only recruited Mandarin speakers born and raised in Beijing as participants. Although the regional effect is not a significant effect on duration of syllables in the mixed effects model of Study 1, the interaction between syllable position and region is significant.

Figure 3.10 does display that for syllable positions of disyllabic words, there are regional distinctions. Taiwan speakers have longer duration of the second syllables in disyllabic words. It may indicate that Taiwan speakers show stronger lengthening effect on disyllabic words at sentence final positions. In Section 3.4.3, the duration pattern of disyllabic words are displayed under different conditions of word position, and the result show there has lengthening effect of the final position.

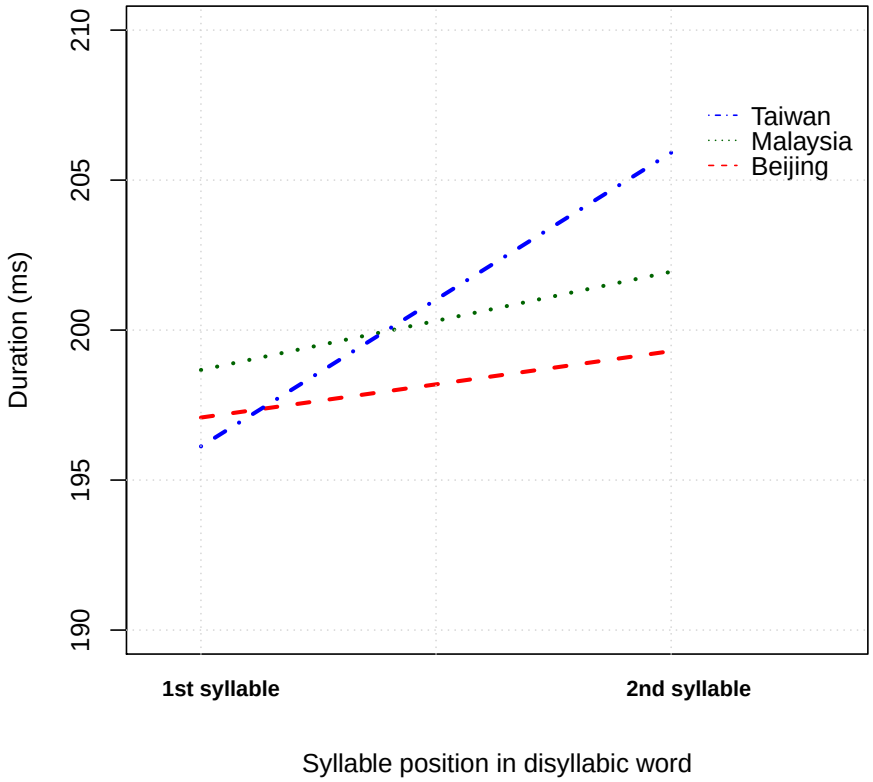


Figure 3.10: Interaction between region and syllable position on syllable duration of disyllabic word

Beijing speakers do not show difference between the first syllable and the second syllable, but speakers form Taiwan and Malaysia do have differences between these two syllable positions. According to the box plots in Figure 3.11, Taiwan and Malaysia show that there are durational differences between first syllable and second syllable. In addition, Beijing shows the long-short duration pattern, which is the opposite of Xu and Wang (2009). The result of regional difference on duration pattern will help to describe the current situation of language among different Mandarin regions.

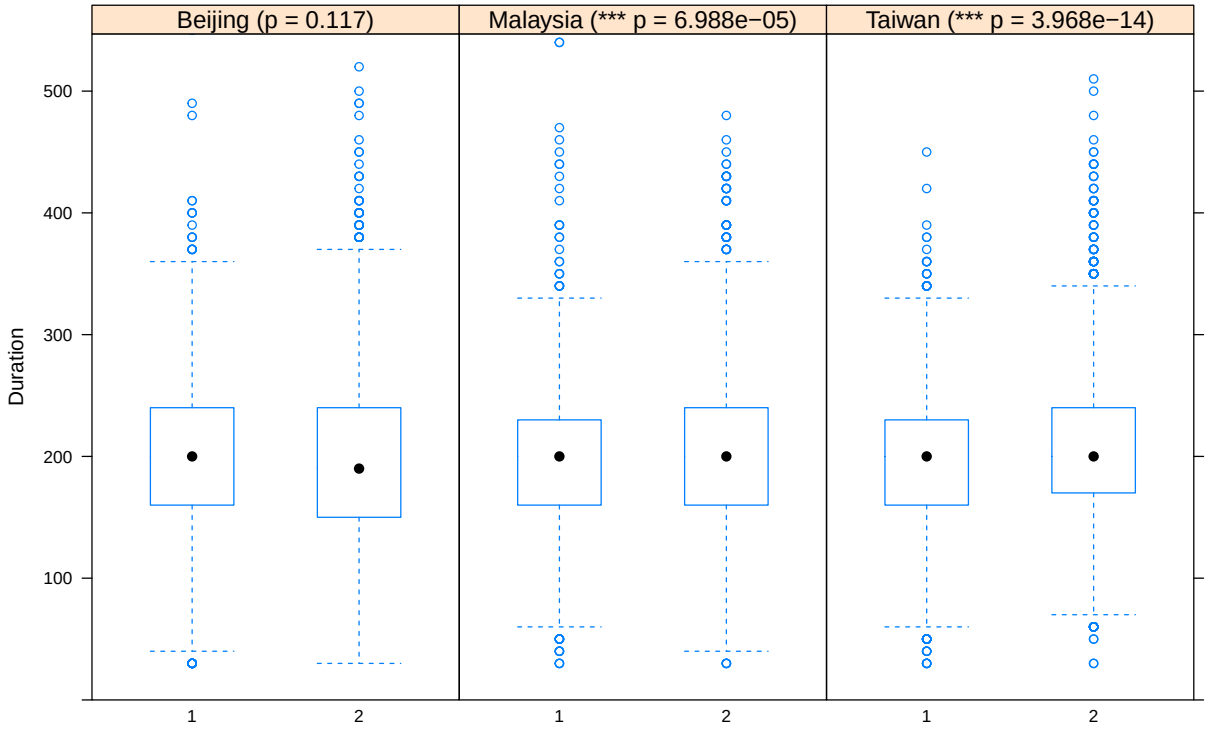


Figure 3.11: Regional differences on syllable duration of disyllabic word

3.6 Summary of Study 1

This study investigates the syllable duration of monosyllabic and disyllabic words among three different Mandarin regions. Both experimental and corpus research approaches are

adopted in this study. On the experimental aspect, six stories are used as experiment materials with the constraints on story length and word length in the reading task, and participants region, gender, age are highly controlled. On the corpus aspect, each speaker's individual production was transcribed and coded in the speech corpus. Based on the research approach and the characteristics of the data, the mixed effects model is used in the statistical analysis, and the results show that sentence position, word position, word category, word frequency, syllable status, syllable structure, and the interaction between region and syllable position are significant effects. The further examination first demonstrated that syllables in disyllabic words show a long-short pattern in Mandarin. Second, sentence at the beginning of a story and word at the sentence-final position both show lengthening effects on syllable duration. Third, there is a negative correlation between word frequency and syllable duration and function words tend to have shorter syllable duration than content words. Finally, syllable structures with fewer components tend to have shorter duration.

Chapter 4

Study 2: Duration Patterns of Trisyllabic Words

Study 2 aims to describe and analyze the duration patterns of trisyllabic words in Mandarin Chinese. Based on the result of large-scale corpus, trisyllabic words account for around 8% of Chinese words (Table 2.1). More importantly, Study 2 will examine the word-internal branching effect on syllable duration of trisyllabic words.

This chapter will first introduce the branching structures of Chinese trisyllabic words and how we use controlled materials and corpus method to examine the structure effect among other factors in Section 4.1 and 4.3. Section 4.4 will report the statistical results of the examination and Section 4.5 will further discuss the interactions of factors leading to variations of syllable duration in trisyllabic words.

4.1 Branching Structures of Trisyllabic Words

There are two major syntactic structures of trisyllabic words—left-branching structure and right-branching structure. Figure 4.1 displays these two structures, where s1, s2, and s3 refer to the three different syllable positions correspondingly. The left panel displays the example 生产线 *sheng chan xian* (production line) with left-branching morphological structure, and the right panel demonstrates the word 年产量 *nian chan liang* with the right-branching morphological structure. In the left-branching trisyllabic word *sheng chan xian*, the monosyllabic noun *xian* (line) is the head of the NP, and the disyllabic word *sheng chan* (produce) is the modifier. In the right-branching trisyllabic word *nian chan liang*, the disyllabic noun *chan liang* (output; production) is the head and the monosyllabic adjective *nian* (annual) is

the modifier. In order to investigate the branching effect on duration pattern of trisyllabic word, this study will control the syllable used at the second syllable position. For example, in Figure 4.1, the identical character 产 *chan* is used in both branching structures. Section 4.3.1 will explain the experimental design on the controlled syllable.

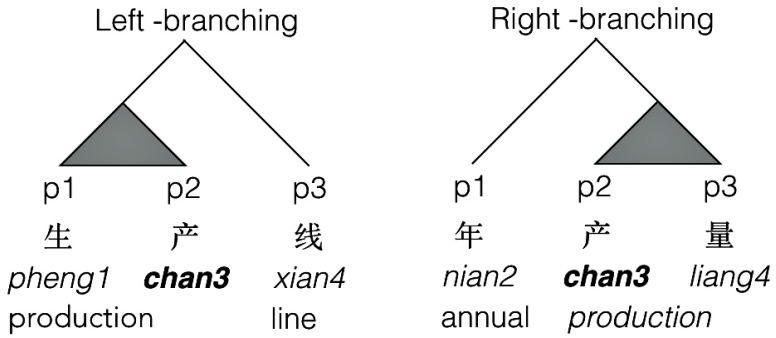


Figure 4.1: Morphological structure of trisyllabic words

There are many syntactic and phonetic studies on Chinese trisyllabic words. Lu and Duanmu (1991) reported that for trisyllabic NN (noun-noun) phrases in Mandarin, the right-branching (monosyllabic modifier + disyllabic noun) is the disfavored pattern, and the left-branching (disyllabic modifier + monosyllabic noun) is the favored pattern, because the trochaic foot is the default form in Chinese. Feng (1998, 2005) proposed that Chinese words are parsed from left to right into trochaic disyllabic foot, which leads to the left-branching structure with a trochaic foot at the word beginning being the favored pattern in trisyllabic words of Mandarin. Duanmu (2000) also reported that polysyllabic word in Standard Chinese has the trochaic foot from left to right. If there is a disyllabic trochaic pattern in Chinese trisyllabic words, it will lead to the prosodic difference of the two syllables in the trochaic metrical foot. Namely, there should have a durational contrast between syllables.

Phonetically, syllables and phonemes on boundary positions usually possess longer duration (Gaitenby, 1965; Lindblom and Rapp, 1973; Klatt, 1973; van Santen and Olive, 1990) and non-boundary ones are in prosodically weak positions with shorter duration. In Man-

darin Chinese, many studies agree that the non-boundary middle syllable in trisyllabic word is the weak syllable with less degree of stress, prosodic strength and shorter duration (Yip, 1980; Zhang, 1988; Kochanski et al., 2003; Shih, 2005; Chen, 2006; Xu and Wang, 2009).

One of the main research questions is to examine the duration pattern of the trisyllabic words with both left-branching and right-branching structures and explore the relation between duration pattern and structures.

4.2 Research Question

Study 2 aims to address the following questions:

(i) What is the duration pattern for trisyllabic Chinese words? More specifically, is durational pattern related to word structure?

(ii) Are there any linguistic and social factors having effects on duration pattern of trisyllabic words?

For research question (i) I hypothesize that syllables show within-word duration variation on different positions, and I expect that there is a durational difference on two different morphological structures. For research question (ii) I hypothesize that there are factors and their interactions affecting duration pattern. This study will use the sentence-reading task to collect speech production data from three Mandarin speaking regions—Beijing, Taiwan, and Malaysia.

4.3 Methodology

Study 2 will adopt the same research approach—combination of experimental control and corpus approaches together, which is used in Study 1. Firstly, Study 2 will manipulate morphological structures of trisyllabic words in experiment materials to get balanced data on

both left-branching and right-branching words (see Section 4.3.1 for experiment materials). Secondly, Study 2 will control speakers' regions and genders to have balanced data on these two social factors and their interaction. Thirdly, in order to ensure the speech data is as natural as possible, Study 2 will use different carrier sentences to carry different trisyllabic words in a random order in the sentence reading task. Finally, trisyllabic words in different carrier sentences are coded via corpus approach.

4.3.1 Materials

Thirty-four trisyllabic words are tested in a sentence reading task (see Appendix B for target words and their carrier sentences). These trisyllabic words belong to two different morphological structures—left-branching and right-branching structures. These 34 words are all NPs with the structure of modifier + noun. An identical Chinese character is used in the second syllable positions of a pair of trisyllabic words with left-branching and right-branching structures respectively. For example, the left-branching word *sheng chan xian* and the right-branching word *nian chan liang* form a pair of trisyllabic words with the identical character 产 *chan* on the second syllable positions.

There are two criteria on the selection of the target words in this study.

(1) First, on the second syllable position of both left-branching and right-branching words, the study uses the identical Chinese character. Therefore, the 34 target trisyllabic words are grouped into 17 pairs based on the identical second syllable. For example, according to the identical character 产 *chan*, words—*sheng chan xian* and *nian chan liang* (in Figure 4.1) are paired trisyllabic words. The purpose of the same second syllable in the paired left-branching and right-branching words is to test whether there are durational difference of syllables between the two morphological structures. If there is no branching effect, the identical second syllables in both structures should have similar syllable duration.

(2) The second criterion is that tone sandhi does not appear on any two adjacent characters among these three characters in each word. For example, the words 养老院 *yang3 lao3 yuan4* (nursing home) and 母老虎 *mu3-lao3-hu3* (tigress) share the identical second character 老 *lao*, so they meet the first criterion. But the first two characters 养老 *yang3 lao3* in *yang3 lao3 yuan4* and the second two characters 老虎 *lao3 hu3* in *mu3 lao3 hu3* have tone sandhi and switch into *yang2 lao3* and *lao2 hu3*, resulting in tonal difference of the syllable, so they do not meet the second criterion. Therefore, these examples are eliminated from the experiment materials.

Under these two criteria, 17 pairs of trisyllabic words are used in this study, and they cover most such cases in Mandarin. Target words were put into sentence middle position to avoid sentence final lengthening effect. All thirty-four carrier sentences are with correct grammar forms and acceptable semantic meanings. The average sentence length is 14 syllables/sentence with the range from 10 to 15 syllables/sentence. There are 34 carrier sentences in total to prevent subjects from knowing which words in the sentences are being tested.

Two graduate students majoring in Chinese Linguistics annotate the structures of the 34 target words via the forced choice task. The annotators are asked to identify the 34 words to their corresponding branching categories between left-branching and right-branching. And the inter-rater reliability (Cohen's kappa=1) shows that the strength of agreement is perfect.

4.3.2 Procedure

Thirty-four sentences are randomized and presented on a computer screen in front of each participant. All participants from Taiwan choose to use the traditional character version and all participants from Beijing and Malaysia choose to use the simplified character version. 34 stimuli sentences are repeated three times, yielding a total of 3060 recorded sentences (34 sentences $\times$ 3 repetitions $\times$ 30 participants = 3060). All the acoustic data is segmented and labeled by using the Penn Phonetics Lab Forced Aligner (Yuan and Liberman, 2008,

2009) with manual inspection. Syllable duration of each syllable in each trisyllabic word is measured. In total, there are 9180 syllables from the trisyllabic words in 3060 recordings. Syllable duration data is used in the statistical analyses in Section 4.4.

4.3.3 Corpus

According to the experimental materials and speakers' region and gender information, morphological structure, syllable position, repetition of sentence, speaker gender and region of Mandarin are factors, which have balanced data in the speech database. Section 4.3.1 and Section 4.3.2 explained these experimental controlled factors in Study 2, and this section will focus on how to apply the corpus method to build the speech corpus of trisyllabic words in Chinese.

Corpus coding factors

Word frequency, syllable structure and prosodic condition of the word are labeled via corpus approach.

In Study 2, word frequencies information is from Peking University Center for Chinese Linguistics corpus[1]. CCL corpus includes text data of 783,463,175 characters. If the word frequencies between the left-branching and right-branching stimuli are different, there might be a frequency effect involved in duration variation.

For syllable structure, there are 102 syllables in total in the thirty-four target trisyllabic words. Without tonal difference, these 102 syllables share 65 different syllable identities. To avoid the overfitting in statistical analysis, 65 syllable identities are classified into eight categories based on their syllable structures. In the corpus, 9,180 syllables from trisyllabic words are categorized into eight syllable structure types: CGVG (450 samples), CGVN (1,710 samples), CVG (450 samples), CVN (3,150 samples), GVN (450 samples), CGV (720 samples), CV (2,160 samples), and V (90 samples). These 8 syllable structure types maintain

[1]CCL corpus `http://ccl.pku.edu.cn:8080/ccl_corpus/CCLCorpus_Readme.html`

the categorical information of the segment inside the syllable and would be helpful to make the model more effective.

Prosodic context of the trisyllabic word is also a consideration in this research. Among these thirty-four target trisyllabic words in this study, nine of them are followed by a particle '*de*'. As mentioned in Study 1, *de* is the most commonly used morphological particle in Mandarin, and the function of *de* is to connect the modifier to its head. Based on the analysis in Shi and Li (2002), there are five types of constructions of *de*: relative clauses, genitive, adjectival, associate and adverbial phrases. The nine target words with particle *de* attachment among 34 stimuli are genitive and associated construction in this study. After the particle attachment, the trisyllabic word and the particle form a new speech unit, which is a modifier of the following noun. In the speech database, trisyllabic words are labeled via the conditions with or without particle *de* attachment.

Individual speaker's speech rate of the sentence may also have effect on syllable duration. If a participant reads a sentence really fast, the syllables of the trisyllabic word may have shorter syllable duration. Therefore, for each recorded sentence in Study 2, the speech rate is calculated by the following equation and further used in the data analysis:

$$\textbf{Sentence Speech Rate} = \frac{Ending\,time_{\,the\,last\,syllable} - Starting\,time_{\,the\,first\,syllable}}{Number\,of\,syllables\,of\,the\,sentence} \quad (4.1)$$

4.4 Data Analysis and Result

The five experimental controlled factors all have balanced data (Section 4.3.1 and Section 4.3.2), and the three corpus-method coding factors (Section 4.3.3) all have unbalanced data in this study. Therefore the mixed effects model without requirement of balanced data on factors is performed on the syllable duration corpus.

4.4.1 Statistical Analysis

In the mixed effects regression model, the dependent variable is the syllable duration, and each spoken syllable is coded via the following factors:

(1) Region: with three levels—Beijing, Taiwan, and Malaysia;

(2) Speaker: 30 speakers;

(3) Gender: with two levels—female and male;

(4) Syllable position: with three levels—p1, p2, and p3 represent the first, second, and third syllable positions respectively;

(5) Morphological structure: with two levels—left-branching and right-branching;

(6) Syllable structure: with eight levels—V, CV, CGV, GVN, CVN, CVG, CGVN, and CGVG;

(7) Particle *de* attachment: with two levels—*de* and no-*de*;

(8) Normalized word frequency: is calculated via logarithm;

(9) Speech rate of sentence.

Eight fixed effects with interactions and two random effects are used to predict syllable duration. The eight fixed effects are region, gender, branching, syllable position, syllable structure, particle *de* attachment, word frequency, and speech rate. The interactions among region, syllable position, branching and particle *de* attachment are used in the mixed effects model. The two random effects are repetitions and speakers. Since each speaker occurs within one and only one level of region, so there are two strengths measurements on each speaker within each of the regions and combine with the average strength for speaker. Therefore, speaker is nested within region.

The formula of mixed effects regression model used to analyze the syllable duration via 'lmer' function in R package lme4 is in the following equation. The form of "syllable position ∗ branching ∗ particle attachment" represents the three-way interaction among these factors and the form of "1|region:speaker" indicates that speaker is nested with region:

syllable duration ∼ gender + region∗syllable position∗branching∗particle attachment
+ log(word frequency) + syllable structure + speech rate
+ (1|repetition) + (1|region:speaker)

4.4.2 Result

Table 4.1 summarizes the results of the random effects. The standard deviation is the measure of how much variability in the random effect being added into the model. In the results, repetition has much less variability (sd=6.137 ms) in the data. But speaker has much more variability than repetition. Because the speaker is nested within region, the results show that the variability of speaker among three regions respectively. The standard deviations show that the noise from speakers of Beijing, Malaysia and Taiwan are around 10, 8 and 1 ms respectively, though Taiwan is the one with smaller variability. The standard error of residual (32.147) stands for the variability that is not due to either speaker or repetition.

Random effects:	**Groups**	**Names**	**Std.Dev.**
	Speaker	(Intercept)	9.776
		Malaysia	8.414
		Taiwan	1.021
	Repetition	(Intercept)	6.137
	Residual		32.147

Table 4.1: Summary of random effects of Study 2 (Number of observations: 9180)

Table 4.2 summarizes the results of seven fixed effects in the model. Eight fixed effects

are examined in the mixed effects model individually. Syllable position (p=0.001), syllable structure (p=0.001), particle attachment (p=0.03), word frequency (p=0.001), speech rate (p=0.001), and branching (p=0.03) are significant.

The interactions among region, syllable position, branching and particle attachment are also examined in the mixed effects model. For the two-way interaction, the interaction between region and syllable position (p=0.001), the interaction between syllable position and branching (p=0.001), the interaction between syllable position and particle attachment (p=0.001), and the interaction between branching and particle attachment (p=0.03) are significant. Among these four factors, four three-way interactions are all significant. More importantly, the four-way interaction of these four factors is significant (p=0.001). Section 4.5 will discuss the interaction effect on syllable duration.

Effects	**Stat**	**p.value**
Region	2.00	0.37
Gender	1.02	0.33
Branching	4.11	0.03 *
Syllable position	106.3	0.001 ***
Syllable structure	967.08	0.001 ***
Particle attachment	4.69	0.03 *
Word frequency	199.12	0.001 ***
Speech rate	560.29	0.001 ***
Region: Particle attachment	2.86	0.26
Region: Syllable position	34.47	0.001 ***
Region: Branching	5.12	0.07 +
Syllable position: Branching	19.06	0.001 ***
Syllable position:Particle attachment	116.94	0.001 ***
Branching:Particle attachment	4.19	0.03 *
Region: Syllable position: Branching	16.05	0.004 **
Region: Syllable position: Particle attachment	43.77	0.001 ***
Region: Branching: Particle attachment	7.30	0.03 *
Syllable position: Branching: Particle attachment	23.1	0.001 ***
Region: Syllable position: Branching: Particle attachment	20.51	0.001 ***

Table 4.2: Summary of fixed effects of Study 2

4.4.3 Syllable Duration

Syllable position is a significant effect (p=0.001) on syllable duration. Table 4.3 lists the median, mean and SD of syllable duration of each syllable position, and p1, p2 and p3 represent the three syllable positions. The data shows that syllable durations are different among positions. In general, the syllable duration values on different syllable positions of trisyllabic words are different. The first syllables have the longest syllable duration. The second syllables have the shortest syllable duration. And the last syllable position's duration is shorter than the first syllable but longer than the second syllable.

	p1	**p2**	**p3**
Median (ms)	220	190	200
Mean (ms)	215.7	194.8	203.3
SD (ms)	39.7	40.5	47.8

Table 4.3: Syllable duration by positions

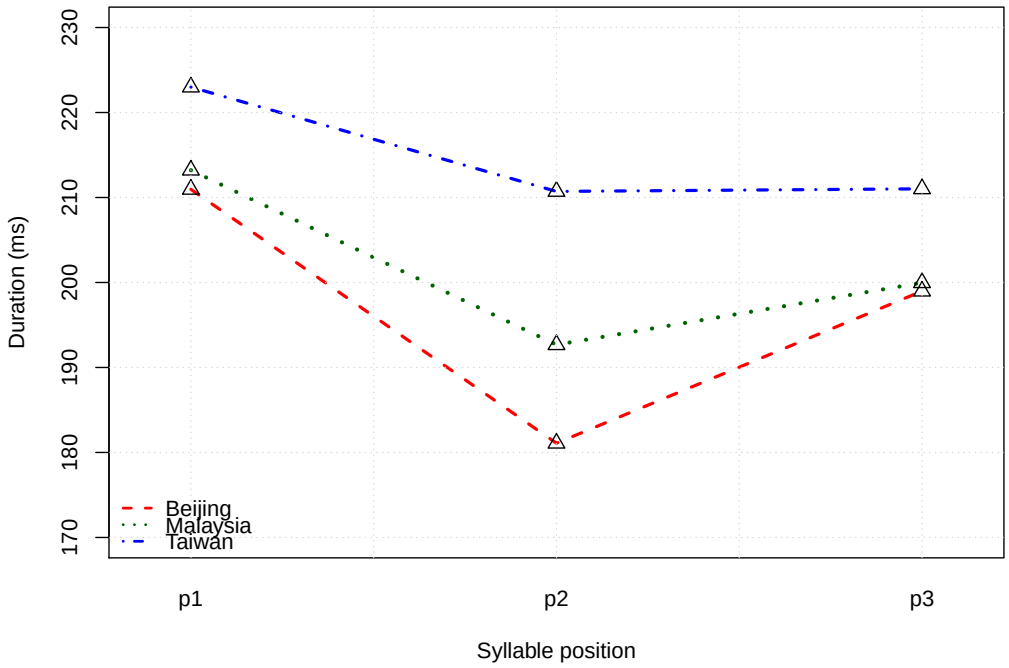


Figure 4.2: Mean of syllable duration by syllable position and region

Figure 4.2 demonstrate the mean value of syllable duration by syllable position and region. Although region is not a significant effect, the interaction (p=0.001) between region and syllable position is significant. In Figure 4.2, the red, blue and green lines correspondingly represent Beijing, Taiwan and Malaysia, and the triangles represent the mean values of three syllable positions. The data shows that Beijing and Malaysia have the similar duration pattern on all three syllable positions that the duration rank from the longest one to the shortest one is: p1 > p3 > p2. Taiwan has a slightly different order, where the first syllable is the longest syllable, but the second and the third syllables have similar duration. According to the order of syllable duration among positions, we can conclude that Beijing Mandarin has the strongest cross-positional rhythmic contrast, and Taiwan Mandarin does not show durational contrast on the last two positions.

4.4.4 Branching Structure

This section focuses on the branching effect (p=0.03) on syllable duration, especially the controlled syllable at the second syllable position. Syllables on the second syllable positions are the controlled syllables in these 17 pairs of trisyllabic words with identical syllable on left-branching and right-branching structure, so these identical syllables on the second syllable position provide the possibility to examine the effect of internal structure on trisyllabic words.

Figure 4.3 demonstrates the branching effect on syllable duration among three syllable positions respectively. The solid line represents right-branching trisyllabic words and the dashed line represents the left-branching trisyllabic words. Among these three syllable positions, only the second syllable position has the controlled identical characters. The second syllables of left-branching words have longer syllable duration than second syllables of right-branching words. It can be explained by the lengthening effect of the branching boundary inside the trisyllabic word. In a left branching word, the first two syllables form the disyllabic modifier in the word, and the second syllable of the word is also the last syllable of

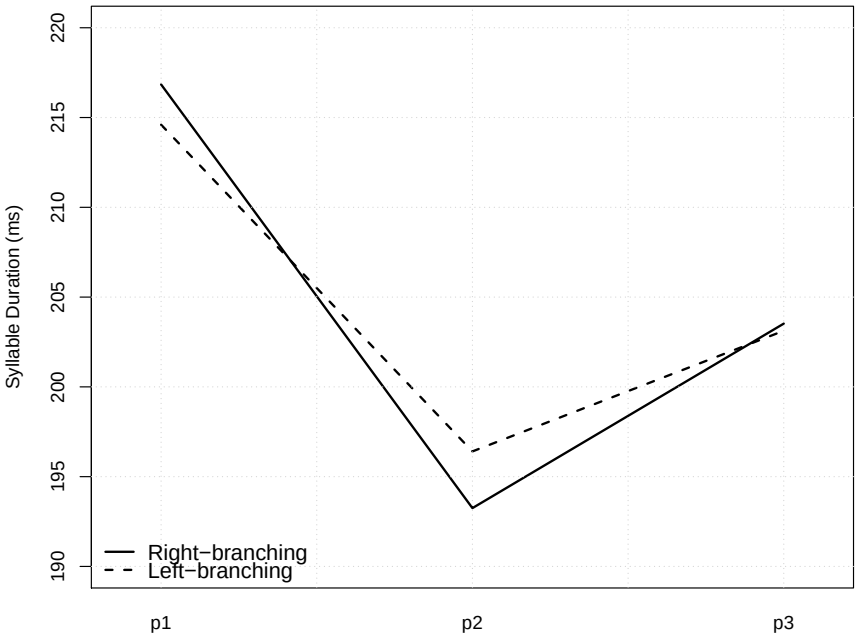


Figure 4.3: Interaction of branching structure and syllable position

the disyllabic modifier. The lengthening effect of the branching boundary applies on the final syllable of the modifier, namely on the second syllable of the left-branching word. In a right-branching word, the first syllable is the modifier, and the second syllable is at the beginning position of the disyllabic noun without any lengthening effect. Therefore, the second syllables in the left-branching words have longer duration than the second syllables in the right-branching words.

4.4.5 *De* Attachment

Particle *de* attachment is a significant factor (p=0.03) on syllable duration modeling of trisyllabic words. This section will discuss how the particle *de*'s attachment change the prosodic unit and how Mandarin speakers from different regions react to the change of prosodic unit.

The particle *de* is the most common particle with the highest word frequency in Chinese. In spontaneous speech, speakers tend to use a schwa [ə] with neutral tone to replace the

original form [tə]. Syntactically, when the particle *de* attaches to the host trisyllabic word, they are combined to form a quadrisyllabic form. For example, the word *sheng chan xian* is a left-branching trisyllabic word. When *de* attaches to *sheng chan xian*, the quadrisyllabic form–*sheng chan xian de* become the modifier of the flowing noun *gong ren* in the carrier sentence.

手机	生产线	的	工人	没有	假期
Shou ji	*sheng chan xian*	*de*	*gong ren*	*mei you*	*jia qi*
Cellphone	production line	DE	workers	have no	vocation

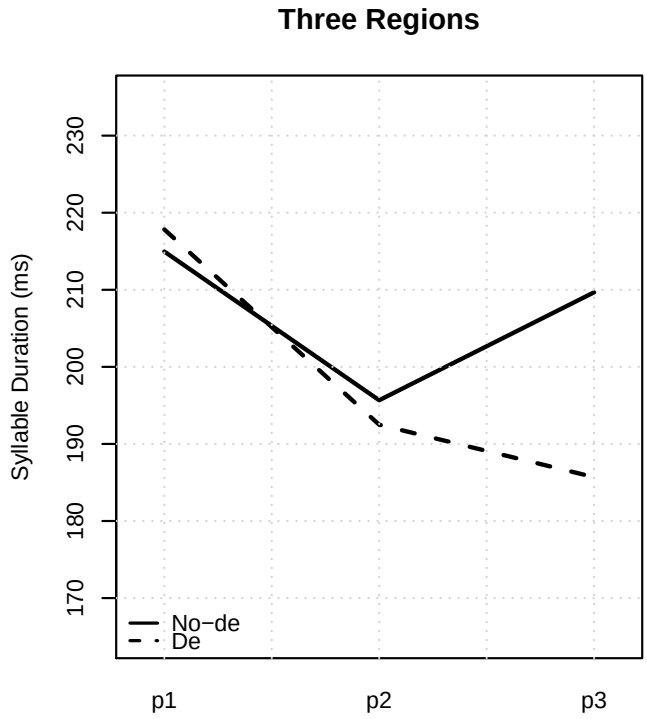


Figure 4.4: Interaction between particle *de* and syllable position

Figure 4.4 and Figure 4.5 demonstrates the *de* effect on trisyllabic words. The solid line represents syllable duration in trisyllabic word without particle attachment (no-*de*), and dash line represents the syllable duration in trisyllabic word with particle attachment (*de*). On the first and the second syllable positions, syllable duration of words with particle and words

without particle are very close (±5ms) to each other, but on the third syllable position, the syllable duration with particle is 25ms shorter than the syllable without particle. The results indicate that particle *de* attachment causes the shortening on the third syllable. Moreover, the statistical results also show that particle attachment interacts with other factors, for instance, region and branching. Section 4.5 will discuss these interaction effects.

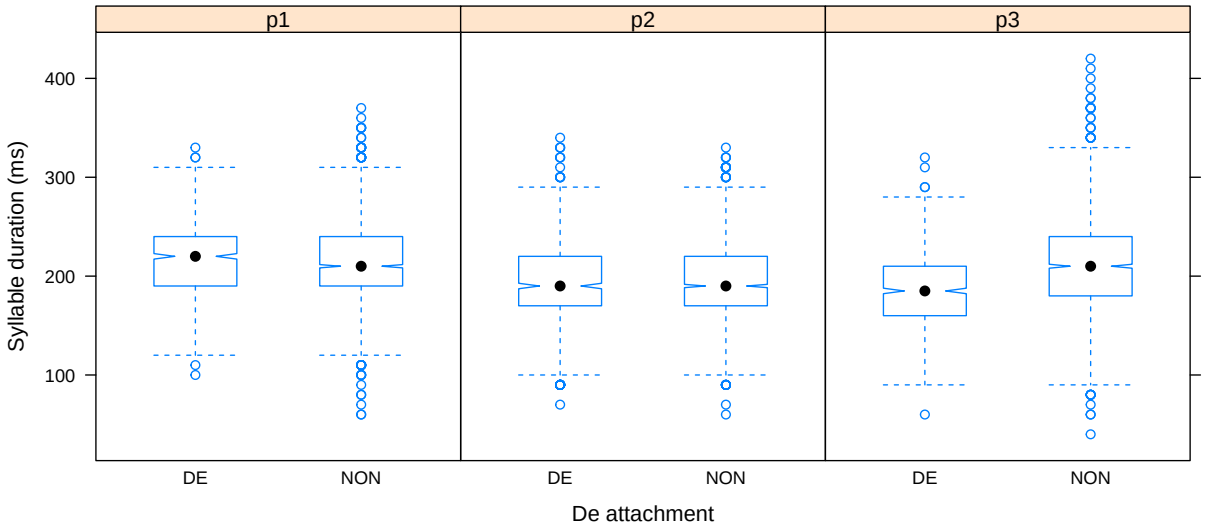


Figure 4.5: Particle *de* attachment effect

4.4.6 Word Frequency

Word frequency is also a significant fixed effect (p=.001) on syllable duration of trisyllabic words. A number of studies show that phoneme, syllable, and word identifications are related to speech rate (Miller, 1981; Miller et al., 1984; Vitevitch et al., 1997). The high frequency words generally lead to the quick identification and shorter duration. For syllables of trisyllabic words, the result of Spearmen correlation test indicates that there is a true correlation between syllable duration and word frequency(S = 1.4785e+11, p-value < 2.2e-16), although the correlation coefficient r=-0.1467038 indicates the negative correlation is weak.

Interestingly, the negative correlations between frequency and syllable duration on three different positions are different. The correlation coefficient between word frequency and the syllable duration on the first syllable position is r= -0.127; the one between word frequency and the second syllable position is r=-0.159; and the one between word frequency and the third syllable position is r=-0.165. It indicates the frequency effect is not applied to all the syllable positions evenly.

4.4.7 Syllable Structure

Syllable structure is a significant fixed effect (p=0.001) in the model. Figure 4.6 exhibits the durational differences among 8 syllable structures in trisyllabic words. The rank of syllable structure is based on the ascending order of medians and mean in Table 4.4.

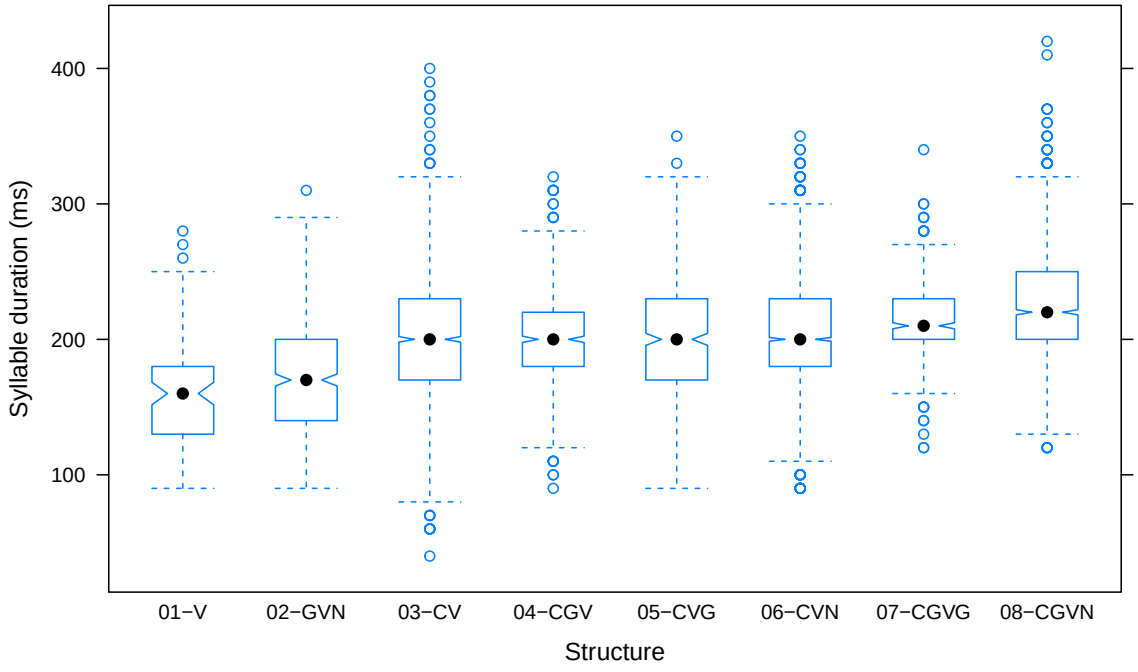


Figure 4.6: Syllable duration by structures

This pattern is similar to the syllable duration result of monosyllabic and disyllabic words in Study 1 (Figure 3.9 in Section 3.4.7). First, the simple structure types–V, GVN without

Structure	Number of syllables	Median (ms)	Mean (ms)	SD (ms)
01-V	90	160	164	43
02-GVN	450	170	170	45
03-CV	2160	200	198	47
04-CGV	720	200	200	36
05-CVG	450	200	202	38
06-CVN	3150	200	204	39
07-CGVG	450	210	215	30
08-CGVN	1710	220	226	42

Table 4.4: Syllable duration of trisyllabic word by syllable structure

consonantal onsets have the shorter duration. Second, two maximal structures–CGVG and CGVN have longer duration than other structures, and one of the maximal structures–CGVN with both consonantal onset and nasal final has the longest syllable duration. Third, structures–CV, CGV, CVG and CVN have similar mean and median values around 200 ms, but the distribution of CV is more dispersed than CVG, CGV and CVN.

4.5 Discussion

This section mainly focuses on the discussion of interactions among different fixed effects in the statistical model.

4.5.1 Duration of Identical Character

The results of the mixed effect models show that branching (p=0.03), particle attachment (p=0.03), and their interaction (p=0.03) are all significant on syllable duration. There are 8 pairs of trisyllabic words without particle *de* attachment on both branching structures (Table 4.5), therefore we don't expect shortening of duration caused by *de*.

In Figure 4.7, the x-axis of each individual plot represents the duration value of the second syllable in the left-branching trisyllabic word, and the y-axis represents the duration value of the second syllable in the right-branching trisyllabic word. The red, blue and green

	Character	Right-branching Word	Left-branching Word
1	行 *xing*	小行星 *xiao xing xing* asteroid	旅行家 *lü xing jia* traveller
2	冰 *bing*	电冰箱 *dian bing xiang* refrigerator	破冰船 *po bing chuan* icebreaker
3	作 *zuo*	农作物 *nong zuo wu* crops	工作日 *gong zuo ri* workday
4	皇 *huang*	土皇帝 *tu huang di* local despot	保皇党 *bao huang dang* royalists
5	水 *shui*	白水晶 *bai shui jing* rock Crystal	山水画 *shan shui hua* landscape painting
6	豆 *dou*	咸豆浆 *xian dou jiang* salty soybean milk	红豆冰 *hong dou bing* red bean with milk ice
7	镜 *jing*	慢镜头 *man jing tou* slow-motion	眼镜蛇 *yan jing she* cobra
8	功 *gong*	肝功能 *gan gong neng* liver function	成功率 *cheng gong lü* success rate

Table 4.5: Eight pairs of trisyllabic words (right-branching and left-branching paired trisyllabic words without *de* attachments)

circles indicate the sample points from Beijing Mandarin, Taiwan Mandarin and Malaysia Mandarin speakers respectively (30 samples/region). The text on the top of each subplot is the syllable identity of the second syllable in the paired trisyllabic words. The texts in pink at the upper-left corner indicate the particle attachment situation—'no-de' and the normalized word frequency of the right-branching trisyllabic word; and the texts in purple at the bottom-right corner indicate the particle attachment situation and the normalized word frequency of the left-branching trisyllabic word). There are five separate subplots, and each subplot is an individual controlled second syllable in the paired trisyllabic words with notable frequency difference. The diagonal line is the y=x line.

Since the second syllables in the trisyllabic word pairs are controlled identical characters in this study, so hypothetically, without frequency effect and branching effect each second syllable in these two branching structures should have similar syllable duration. In other words, if we plot the syllable duration in the coordinate system, ideally, all the points should fall on the line: y=x. All the five syllables in Figure 4.7 are in the word pairs

without particle *de* on both structures. Under this circumstance, high frequency words' second syllables should have shorter duration than low frequency words' second syllables, and paired words with similar normalized frequencies should have similar second syllable duration values distributed along the line of y=x.

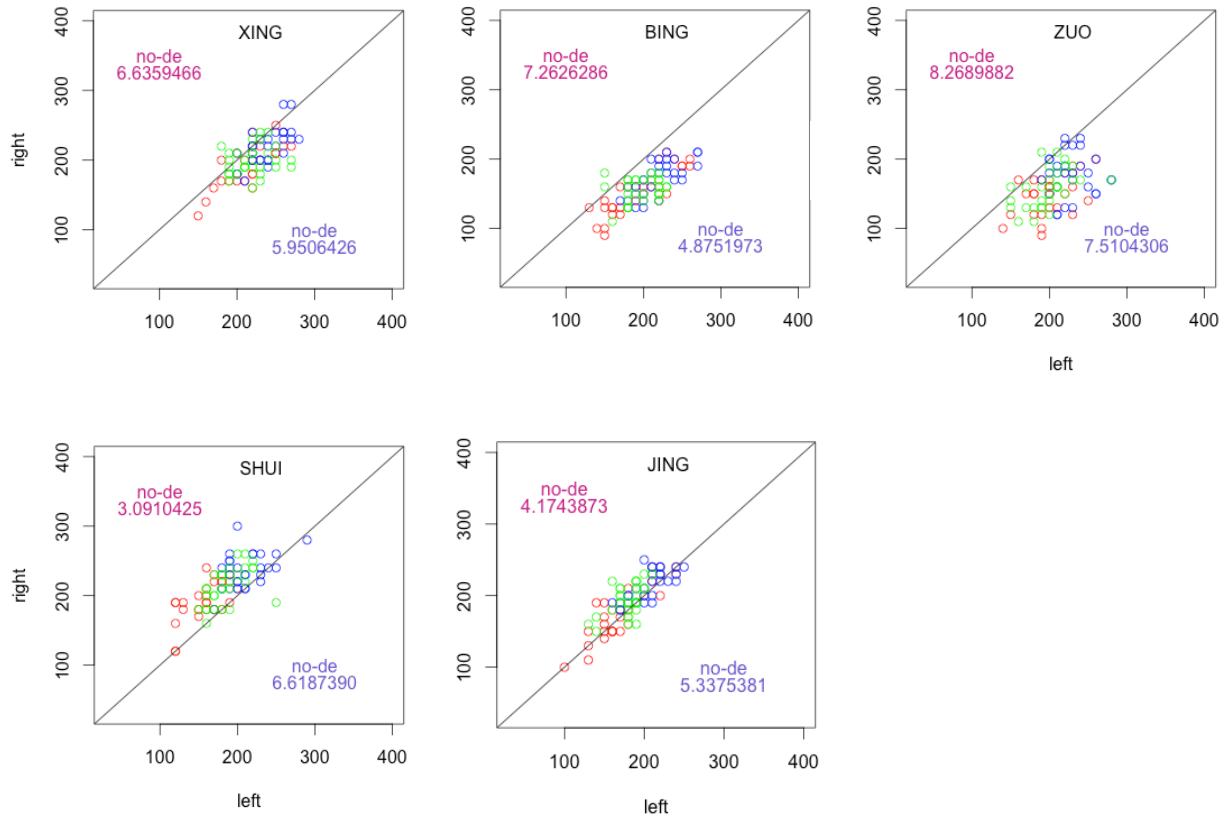


Figure 4.7: Syllable duration of the second syllables at paired left-branching and right-branching words (5 pairs of trisyllabic words with frequency differences)

In the three upper subplots of Figure 4.7, all the right-branching words have higher frequencies than their counterparts and the second syllables in right-branching words have shorter duration. Furthermore, almost all the samples are below the diagonal line. In the two lower panel subplots of Figure 4.7, the left-branching words have higher word frequencies than the right-branching words, and almost all the sample points are above the diagonal line.

When the paired trisyllabic words have word frequency that are very different, there is a

noticeable durational difference between the second syllables within different word structures. When there is a high frequency word on particular branching structure, almost all the thirty speakers from three different geographic regions shorten the syllable duration of this word. This agreement of speakers leads to the seemingly strong branching distinction on duration of the second syllable between the left-branching structure and the right-branching structure, but this robust effect is not due to the branching structures.

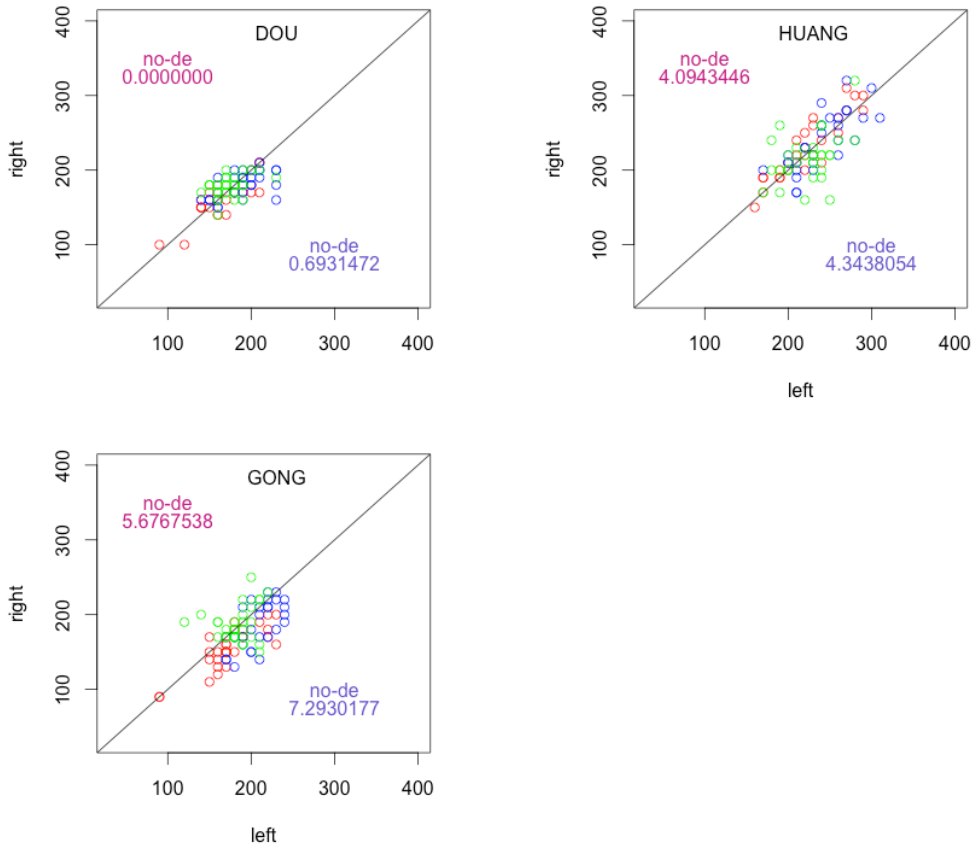


Figure 4.8: Syllable duration of the second syllables at paired left-branching and right-branching words

For the paired words with similar word frequencies, in Figure 4.8, there are two subplots at the upper panel, and sample points in each subplot almost symmetrically distribute along the diagonal line. In the first subplot of syllable DOU, when left-branching word and right-branching word both have extremely low word frequencies (0 and 0.7), speakers

also produced the similar duration of the second syllables of both branching structures. In the second subplot of syllable HUANG, when left-branching word and right-branching word have similar frequencies (4.34 and 4.09), speakers produced the second syllable with similar syllable duration.

Among these 8 word pairs, there is only one exception of the syllable duration. The lower panel of Figure 4.8 displays the exceptional case of syllable–GONG. The right-branching word '*gan gong neng*' has the lower normalized word frequency than the left-branching word '*cheng gong lü*'. Thus, based on the prediction, syllable *GONG* in the right-branching should have longer syllable duration than the corresponding one in the left-branching, and most of the sample points should be above the line of y=x. But the subplot of GONG exhibits that most of the samples are below the line. There may be other factor/factors leading to the exception of GONG, such as articulatory effect, which needs to be further studied in articulatory research on Chinese words duration.

In this section, the syllable duration of the identical character is discussed based on data of the paired words without particle attachment, and the results support the prediction that the high word frequency leads to the shortening effect on all the syllables within the word.

4.5.2 Regional Difference

In the general result of particle attachment (Section 4.4.5), Figure 4.4 displays that particle *de* attachment causes the shortening of the third syllable. This section will discuss the interaction (p=0.001) among region, syllable position, and particle attachment.

The left panel of Figure 4.9 displays the durational difference under two particle conditions in Beijing Mandarin. There is almost no durational difference on the first syllable between the words with and without particle. The second syllable of the word with particle is 7.5ms shorter than the second syllable of the word without particle. The third syllable of the words with particle is around 40ms shorter than the third syllable of the words without particle.

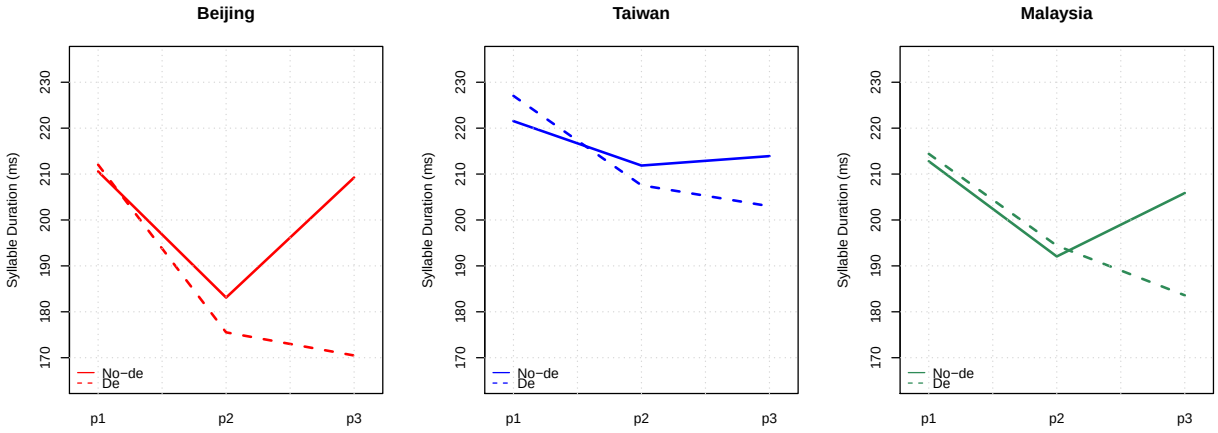


Figure 4.9: Interaction between particle *de* and syllable position for each region

The middle panel of Figure 4.9 shows the durational difference in Taiwan Mandarin. Words with particle have longer syllable duration on the first syllable, but shorter syllable durations on the second and third syllable. It is important to note that the durational difference on the third syllables between words with particle and words without particle is smaller than Beijing and Malaysia.

The right panel of Figure 4.9 displays the situation in Malaysia Mandarin. For the first two syllables, there is almost no durational difference between words with and without particle, but words with *de* do have shorter duration on the third syllable.

In sum, Figure 4.9 illustrates that the particle *de* shortens the duration of the third syllable in all three Mandarin regions. Different regions show different degrees of shortening on the third syllable. Speakers of Beijing Mandarin have a higher degree of shortening from the particle attachment, and it may indicate that Beijing speakers are more sensitive to the prosodic reorganization with regard to the length change of a speech unit. Speakers of Taiwan Mandarin may handle the particle attachment in a different way with less degree of shortening. Therefore, Beijing Mandarin speakers have the strongest rhythmic contrast and Taiwan Mandarin speakers show less rhythmic contrast.

4.5.3 Duration Pattern

This long-short pattern on the first two syllables in this study supports the theoretical analysis proposed and discussed in many previous works (Lu and Duanmu, 1991; Feng, 1998, 2005; Duanmu, 2000; Sui, 2013). Their theoretical analysis predicted that the trochaic disyllabic foot is the predominantly prosodic form of Chinese words. Moreover, having the shortest duration on the second syllable also supports the view that non-boundary syllable in a trisyllabic word is the weak syllable with shorter duration, unstable tone, less degree of stress and prosodic strength (Yip, 1980; Zhang, 1988; Kochanski et al., 2003; Shih, 2005). For example, Kochanski et al. (2003) reported the strong-weak pattern of prosodic strength on trisyllabic word in Mandarin. The second syllable in the trisyllabic form is the weak syllable. Shih (2005) reported that tone2 reduction in Mandarin trisyllabic word is on the prosodic weak position, which is the second syllable position. Speakers usually do not put enough effort in their articulations to produce the syllable on the prosodic weak position. Hence it is hard to maintain the prosodic contrast on a weak syllable and then leads to shorter syllable duration in the middle position of trisyllabic word.

Both left-branching and right-branching exhibit the long-short duration pattern on the first two syllable positions regardless of the distinction between these two structures. It suggests that the duration pattern may be independent of syntactic structure. The relationship between duration pattern and morphosyntactic structure of multi-syllabic words will be further discussed in Chapter 5 and Chapter 6.

4.6 Summary of Study 2

Mandarin speakers produce trisyllabic words under different conditions such as word internal structure, syllable structure, syllable position, particle attachment, word frequency, and speech rate, and these factors lead to the variance of syllable duration of trisyllabic words. Study 2 focused on the construction of the speech database of trisyllabic words and the

analysis of the duration pattern under multiple factors and their interactions. Based on the research method and the properties of the speech data in this study, the mixed effects model was conducted on syllable duration data of trisyllabic words.

Syllable duration of trisyllabic word displays the pattern that the initial syllable has the longest duration and the second syllable has the shortest duration. The pattern of syllable duration is consistent with the hypothesis that syllables duration show within-word variation on different positions. This pattern is also consistent with previous research on prosodic strength of polysyllabic Mandarin words and tonal reduction on the weak syllable in trisyllabic words (Kochanski et al., 2003; Shih, 2005). Moreover, this pattern also supports the theoretical analysis on trochaic foot in Chinese (Lu and Duanmu, 1991; Feng, 1998, 2005; Duanmu, 1999, 2000; Sui, 2013).

To address the relationship between syllable duration and morphological structure, Study 2 investigated the duration difference of the controlled syllables in word pairs with two branching structures, and the results show the second syllables of left-branching words have shorter syllable duration than the second syllables of right-branching words. Moreover, particle *de* attachment and word frequency both have effects on the duration of the second syllables.

The result of the study also indicates that Beijing Mandarin speakers show stronger sensitivity to rhythmic contrasts than Taiwan Mandarin speakers, and this rhythmic difference may be related to the impression that speakers of Beijing Mandarin speak faster than other Mandarin regions.

Chapter 5

Study 3: Duration Patterns of Quadrisyllabic idioms

In Chinese, there is a multitude of idiomatic expressions or proverbs, which are most in the quadrisyllabic form, named as 成语 (*chengyu*). They are widely used in both spoken language and written language in Mandarin with strong rhythmic or prosodic patterns.

Chinese idioms have unique linguistic properties. Syntactically, grammar of Modern Chinese cannot be used to annotate Chinese idioms because *chengyu* involves considerable Classical Chinese syntactic structures. Semantically, many *chengyu* cannot be understood from the literal meanings of its components, because they are rich in metaphor, simile, and analogy. Phonetically, *chengyu* usually have a distinctive tempo or prosodic pattern, and they are typically in quadrisyllabic form. To understand Chinese idioms, language learners need to be familiar with the allusions, which are related to historical figures, history stories, or special events in China.

5.1 Introduction

5.1.1 Idiom Length

The length or size of a linguistic unit is vital for the prosodic pattern. To get an impression of the length (number of characters) of *chengyu*, a *chengyu* dictionary is used to calculate the number of characters in each idiom. The *chengyu* list is from a public online resource based on the corpus of *Ziguang Pinyin Method*[1], a popular Chinese character input method developed by Thunisoft Inc. There are 47,775 idioms in this list, which is the most complete

[1] `http://www.datatang.com/data/13498`

compilation of idiomatic expressions in Mandarin. Table 5.1 displays the length information of *chengyu*. Quadrisyllabic form takes up around 93% of the idiomatic expressions of Chinese.

Idiom Length (syllable)	**Number of Idiom**	**Percentage**
3	386	0.81%
4	**44511**	**93.17%**
5	689	1.44%
6	525	1.10%
7	408	0.85%
8	1012	2.12%
9	41	0.09%
10	153	0.32%
11	11	0.02%
12	20	0.04%
13	0	0%
14	19	0.04%
Total	47775	100.00%

Table 5.1: Length information of Chinese idioms

5.1.2 Idiom Structure

The internal structures of quadrisyllabic idioms may also have effects on duration pattern. There are four possible structures of Chinese idioms, and Figure 5.1 demonstrates them with examples.

In Structure 1, all the four characters are in parallel semantic status in the expression. For example, 青红皂白 *qing hong zao bai* originally refers four different colors–green/cyan, red, black and white. Therefore these four characters are parallel with each other in the syntactical structure. The word *qing hong zao bai* in Modern Mandarin is used in the context of someone not being able to distinguish right from wrong.

In Structure 2, the four characters are divided into two constituents—the first group (the first character and the second character) and the second group (the third character and the fourth character) are in parallel status. For example, in 家喻户晓 *jia yu hu xiao* (be known to every family and household), the first group *jia yu* (be known to every family) and the

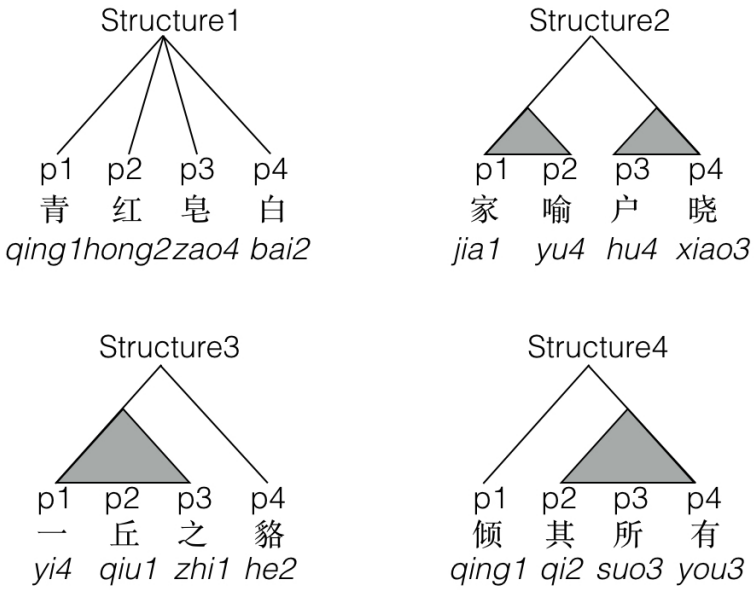


Figure 5.1: Structure of Chinese idioms

second group *hu xiao* (be known to every household) are parallel with each other. In *jia yu hu xiao*, the first group and the second group are all VPs. It is also possible for two NPs, or other structures of phrases to form an idiom. For example, in the idioms 血雨腥风 (*xue yu xing feng*), which is used to indicates the tyrannical rule or brutal war, both *xue yu* (bloody rain) and *xing feng* (sanguinary wind) are NPs.

In Structure 3, the first three characters are grouped to describe the fourth character, which is usually a noun. For example, in 一丘之貉 *yi qiu zhi he*, the first three character *yi qiu zhi* (the same lair) is the modifier to modify the noun *he* (jackals). The literal meaning of this word is 'being jackals of the same lair', with the metaphor meaning—birds of a feather.

Structure 4 usually appears in subject-predicate constructions or verb-object constructions. For example, in 倾其所有 *qing qi suo you*, the first character *qing*(give) is the main verb, and the rest of the idiom—*qi suo you* (everything he has) is the object.

The possibility and the number of words of each structure type in language use are not equal. The first two structures are more common than the last two structures in Chinese idioms.

5.1.3 Duplication of Character

In Mandarin quadrisyllabic idioms, it is common for an identical character repeatedly appearing at two different syllable positions of one idiom. Between these two identical characters, the duplicated character may have shorter syllable duration than the first one due to the repetition effect. Therefore, durational research needs to pay attention to the shortening effect from duplication of syllable in words. Figure 5.2 displays all the possible situations for the identical character appears twice on two different syllable positions within one quadrisyllabic word[1]:

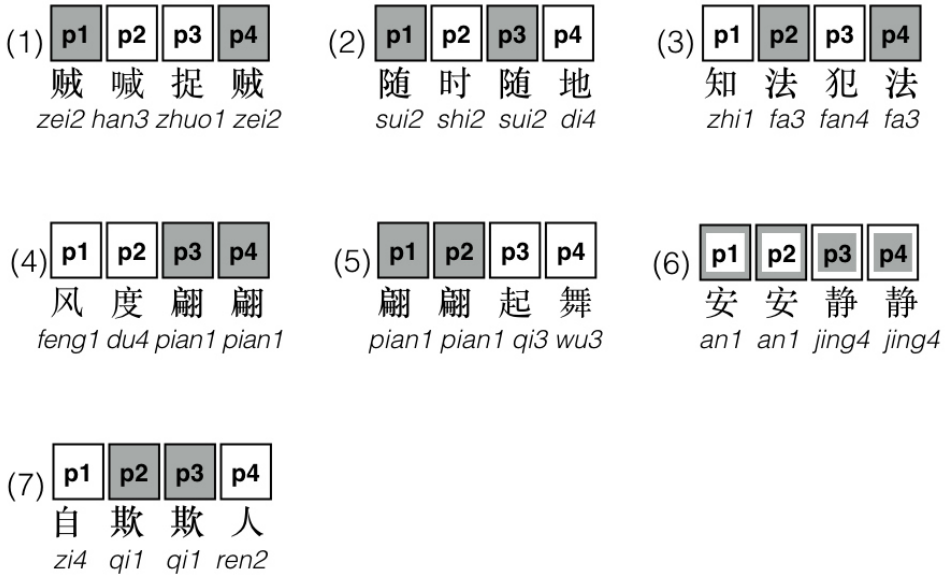


Figure 5.2: Seven types of character duplications of quadrisyllabic idioms

In Figure 5.2, each square represents one syllable position, and p1, p2, p3 and p4 represent four positions respectively. The grey squares are the positions having the identical characters. In Type(1), the characters appear at the first and the fourth positions; in Type (2), the

[1]Examples used in Figure 5.2:
1) *zei han zhuo zei*: play the trick of a thief crying "Stop thief!"
2) *sui shi sui di*: whenever and wherever possible;
3) *zhi fa fan fa*: know the law but break it;
4) *feng du pian pian*: be dapper in appearance;
5) *pian pian qi wu*: rise and dance in a happy mood;
6) *an an jing jing*: peaceful and serene;
7) *zi qi qi ren*: fool oneself as well as others.

identical characters are at the first and the third positions; and in Type (3) the repeated characters are on the second and the fourth positions. Among these three types, the repeated characters are not next to each other. In Type (4) and (5), the repeated characters are next to each other and either at the first two positions or the last two positions. Type (6) is a combination of Type (4) and (5). Two pairs of duplicated characters appear in one idiom—the first pair takes the first two positions, and the second pair takes other two positions. These six types commonly exist in Mandarin idioms. In Type (7), the second and the third positions have the identical characters. Idioms with two identical characters at the two middle positions are extremely rare in Chinese idioms. Study 3 will address the issue of duplicated characters in the duration pattern of quadrisyllabic idioms.

5.1.4 Content Character and Function Character

As mentioned at the beginning of this chapter, Chinese idioms involve considerable Classical Chinese grammar, including Classical Chinese word categories. Study 1 exhibits that word category is a significant factor leading to duration variance. In order to investigate the word category effect on syllable duration in idioms, it is necessary to annotate the idioms via Classical Chinese word categories. In Classical Chinese grammar (Norman, 1988; Pulleyblank, 2010), the minimal lexical unit is the monosyllabic *zi* (see discussion in Section 2.2), and the concepts of 实字 *shi zi* (full word; content word) and 虚字 (empty word; function word) *xu zi* are used to classify the *zi* categories. "The former, also called content words, correspond to nouns, verbs, and adjectives, and the latter to particles whose main function is to show grammatical relationships" (Pulleyblank, 2010). For example, in the idiom '一丘之貉 *yi qiu zhi he*' used in Figure 5.1, the third character '之 *zhi*' is a function word and the other three characters are content words in Classical Chinese. Since function character and content character may lead to durational differences, it is necessary to annotate the categories of the characters used in quadrisyllabic idioms.

5.2 Research Questions

Native speakers of Chinese use numerous quadrisyllabic idioms in both written and spoken form. Using idioms is considered as a linguistic parameter related to the education level of the speaker. An understanding of the duration pattern of quadrisyllabic words is important to duration modeling. In this study, duration of each syllable of a quadrisyllabic idiom is measured and fed into the statistical model for further analysis. Study 3 uses the sentence reading task to investigate the following research questions:

(i) Do social (e.g., gender and region) and linguistic factors (e.g., word internal structure, word frequency, word category and syllable structure) affect the duration pattern of quadrisyllabic idioms?

(ii) Do internal structures of Chinese idioms affect the duration pattern of idioms? If so, why?

In Study 3, a research approach which combines the experimental design approach and corpus approach is used. Section 5.3 describes the method with detailed information.

5.3 Methodology

5.3.1 Materials

154 quadrisyllabic idioms are selected based on the research purpose, and these idioms are put into the sentence middle positions of 154 unique carrier sentences (Appendix C). The average sentence length is 14.48 syllables with the range of 12∼18 syllables. On average, the first syllable of the quadrisyllabic idiom appears at the 6.5 syllable position of the sentence.

One native speaker of Beijing Mandarin and two native speakers of Taiwan Mandarin examined the 154 quadrisyllabic *chengyu* and their corresponding carrier sentences. After

these three examiners reach the agreement on the grammaticality of all the sentences, the sentences are used in the sentence reading task.

5.3.2 Procedure

154 sentences are randomized and presented one by one on a computer screen in front of the participants. The participants were asked to read each sentence at their normal speech rate without pause in the sentence.

This procedure was repeated three times, yielding a total of 13860 recorded sentences (154 sentences $\times$ 3 repetitions $\times$ 3 region groups $\times$ 10 subjects = 13860). The audio recording environment is identical to Study 1.

All the acoustic data was segmented and labeled by using the Penn Phonetics Lab Forced Aligner (Yuan and Liberman, 2008, 2009) with manual inspection. Data of syllable duration in these quadrisyllabic idioms is used in the corpus construction and data analysis. There are 55440 syllables (4 syllables/idiom $\times$ 13860 recorded idioms) in total.

For the internal structure of idiom, five native speakers of Mandarin with Master degree in Chinese Linguistics annotated the structures of 154 target idioms via the forced choice task separately. The annotators were asked to identify the 154 idioms to their corresponding structure type among the four structures discussed in Section 5.1.2. The agreement rate among these five annotators is 91.0%.

For content or function categories of each *zi*/(syllable/character) of a quadrisyllabic idiom, three annotators majoring in Chinese Linguistics labeled the category respectively, and the agreement rate among three annotators is around 98%.

For the 'duplicated status' of a syllable, the syllables without duplication within an idiom are labeled as 'nr' (non-repeated) syllable; the paired repeated syllables are labeled as 'r1' and 'r2' respectively based on their orders in the word. For instance, the example *zei han zhuo zei* used in Figure 5.2 Type (1), the first syllable *zei* is coded as 'r1'; the repeated *zei* is coded as 'r2'; the two middles non-repeated syllables are both coded as 'nr'.

For speech rate, individual speaker's speech rate of each sentence is calculated in Study 3 by the same equation used in Study 2 (Section 4.3.3).

The annotation results (structure, category, and duplicated status) and the speech rate are further used in the coding with other factors in the corpus of Study 3 (Section 5.3.3).

5.3.3 Corpus

In the corpus of Study 3, each syllable is labeled with the following factors:

(1) Speaker identity (30 speakers);

(2) Mandarin regions—Beijing , Taiwan, Malaysia;

(3) Repetition;

(4) Speech rate of the sentence;

(5) Word frequency (based on the frequency information from PKU CCL corpus);

(6) Particle *de* attachment;

(7) Structure type of the idiom—str40, str22, str31, and str13 represent the four structures demonstrated in Figure 5.1 correspondingly;

(8) Syllable structure—V, VG, GV, CV, VN, GVN, GVG, CVG, CGV, CVN, CGVG, and CGVN;

(9) Syllable position in the quadrisyllabic idiom—p1, p2, p3, and p4;

(10) Classical Chinese word category—cont (abbreviation of content character) and func (abbreviation of function character);

(11) Duplicated status of the character—r1, r2, and nr.

All the syllable duration data is used in the data analysis in Section 5.4 to explore the duration patterns of Mandarin quadrisyllabic idioms.

5.4 Data Analysis and Result

According to the design of experimental materials and the characteristics of the corpus, the mixed effects regression model is used in Study 3 to analyze data.

5.4.1 Statistical Analysis

The fixed effects used in the model are region, gender, syllable position, syllable structure, idiom structure, idiom frequency, particle attachment, Classical Chinese word category, duplicated status, speech rate, and interactions among syllable position, idiom structure, and region. The random effects used in this model are repetition and speaker. As with the circumstance in Study 2, because speaker is also nested within region, the nesting of factors need to be handled in the mixed model. The formula of the mixed effects regression model used in this study is:

$$\begin{aligned}\textbf{syllable duration} \sim\ & \text{gender} + \text{syllable structure} + \log(\text{idiom frequency}) \\ & + \text{idiom structure} * \text{syllable position} * \text{region} + \text{particle attachment} \\ & + \text{Classical Chinese word category} + \text{duplicated status of syllable} \\ & + \text{speech rate} + (1|\text{repetition}) + (1|\text{region:speaker})\end{aligned}$$

5.4.2 Result

Table 5.2 summarizes the results of the random effects in the mixed effects model. In the results, repetition has much less variability (SD=1.607) in the data. It means the repetition does not lead to large variance. However, speaker has much more variability than repetition. Because the speaker factor is nested within region factor, the results show the variability of speaker among three regions respectively. The standard deviations show that the noise from speakers of Beijing, Malaysia are closed to each other, but speakers from Taiwan show larger variance (around 9). The standard error of residual (37.6) stands for the variability

that is not due to either speaker or repetition.

Random effects:	**Groups**	**Names**	**Std.Dev.**
	Speaker	(Intercept)	6.53
		Malaysia	0.35
		Taiwan	9.07
	Repetition	(Intercept)	1.607
	Residual		37.62

Table 5.2: Summary of random effects of Study 3(Number of observations: 55440)

Effects	**Stat**	**p.value**
Region	4.73	0.12
Gender	0.11	0.80
Idiom frequency	1.04	0.30
Syllable structure	10810.37	0.001 ***
Idiom structure	35.95	0.001 ***
Syllable position	297.90	0.001 ***
Particle attach	4.72	0.02 **
Duplicated status	540.02	0.001 ***
Classical Chinese word category	119.47	0.001 ***
Speech rate of sentence	4015.79	0.001***
Region: Idiom structure	36.13	0.001***
Idiom structure: Syllable position	190.70	0.001***
Region: Syllable position	71.60	0.001***
Idiom structure: Region: Syllable position	94.90	0.001***

Table 5.3: Summary of fixed effects of Study 3

Table 5.3 summarizes the results of fixed effects and interactions used in the model. Severn fixed effects idiom structure (p=0.001), particle *de* attachment (p=0.02), speech rate (p=0.001), syllable structure (p=0.001), syllable position (p=0.001), word category (p=0.001), and duplicated status (p=0.001) are significant in the syllable duration model. The interactions among region, idiom structure, and syllable position are also significant (p=0.001).

5.4.3 Syllable duration

Syllables on different positions of quadrisyllabic idioms show durational differences (Table 5.4). The first syllable has the longest average duration (215 ms), and the third one has the second longest one (202 ms). Syllable on the second and the fourth positions have similar average duration, and both of them are shorter than the other two. The general syllable duration pattern of Chinese quadrisyllabic idioms according to the order of syllable position is long, short, peripheral long, and short.

	P1	**P2**	**P3**	**P4**
Mean (ms)	215	193	202	193
Median (ms)	220	190	200	190
SD (ms)	52	48	46	47

Table 5.4: Overall syllable duration of Study 3

Figure 5.3 further display syllable duration of four syllable positions among three regions, and Beijing speakers show stronger durational contrast than speakers from the other two regions on the first two syllables. The odd syllable positions have longer syllable duration. The even positions have shorter syllable duration. Therefore, the whole pattern of quadrisyllabic word could be considered as the combination of two long-short sub-patterns (See discussion in Section 5.5). Furthermore, the 'long-short' durational pattern is the pattern of non-sentence final disyllabic word reported in Study 1 and the pattern of the first two syllables of trisyllabic word reported in Study 2.

5.4.4 Syllable Position

Syllable position is a significant effect (p=0.001) on syllable duration in the mixed effect model. Section 5.4.3 discussed the durational differences among four syllable positions and reported that the long-short pattern on each metrical foot. To examine the syllable position further, there are 48 quadrisyllabic idioms in materials, and they are listed in Table 5.5 (see Appendix D for the meaning of each idiom used in this table). Among these 48 idioms,

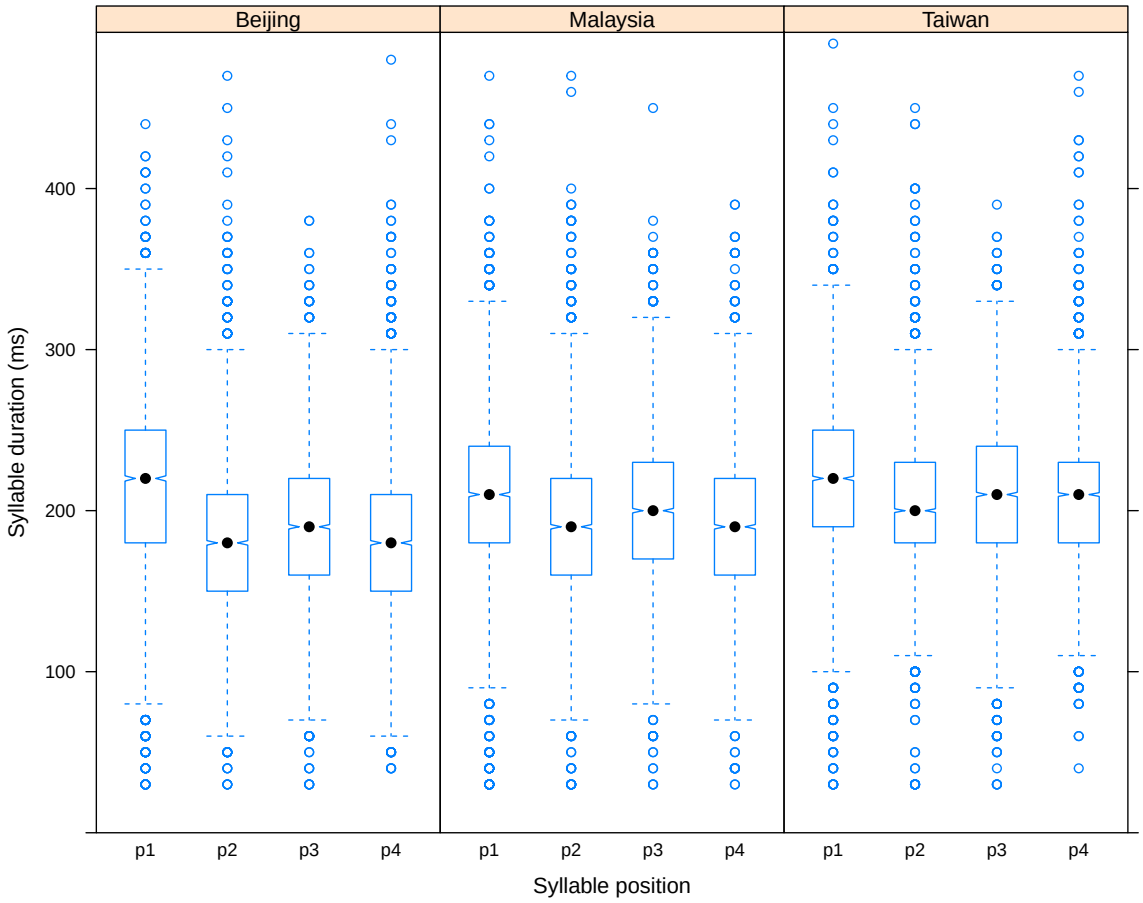


Figure 5.3: Syllable duration by positions among three regions

every four of them are categorized into a group, and there are 12 groups. Within each group, a Chinese character is assigned to four different syllable positions of four different idioms respectively. For example, the character 国 *guo* (country) in Group 6 is placed in four syllable positions in four different quadrisyllabic idioms separately:

1) 国破家亡 *guo po jia wang*: *guo* is in the first syllable position (p1);

2) 卖国求荣 *mai guo qiu rong*: *guo* is in the second syllable position (p2);

3) 皇亲国戚 huang qin guo qi: *guo* is in the third syllable position (p3);

4) 闭关锁国 *bi guan suo guo*: *guo* is in the fourth syllable position (p4).

1 当 ***dang***	2 上 ***shang***	3 尔 ***er***
当家做主 *dang jia zuo zhu* 门当户对 *men dang hu dui* 对酒当歌 *dui jiu dang ge* 旗鼓相当 *qi gu xiang dang*	上行下效 *shang xing xia xiao* 天上人间 *tian shang ren jian* 成千上万 *cheng qian shang wan* 箭在弦上 *jian zai xian shang*	尔虞我诈 empher yu wo zha 莞尔一笑 *wan er yi xiao* 温文尔雅 *wen wen er ya* 新婚燕尔 *xin hun yan er*
4 来 *lai*	**5 地 *di***	**6 国 *guo***
来龙去脉 *lai long qu mai* 南来北往 *nan lai bei wang* 心血来潮 *xin xue lai chao* 古往来来 *gu wang jin lai*	地广人稀 *di guang ren xi* 落地生根 *luo di sheng gen* 天经地义 *tian jing di yi* 死心塌地 *si xin ta di*	国破家亡 *guo po jia wang* 卖国求荣 *mai guo qiu rong* 皇亲国戚 *huang qin guo qi* 闭关锁国 *bi guan suo guo*
7 家 *jia*	**8 事 *shi***	**9 和 *he***
家喻户晓 *jia yu hu xiao* 万家灯火 *wan jia deng huo* 万贯家财 *wan guan jia cai* 小户人家 *xiao hu ren jia*	事半功倍 *shi ban gong bei* 无事生非 *wu shi sheng fei* 置身事外 *zhi shen shi wai* 见机行事 *jian ji xing shi*	和睦相处 *he mu xiang chu* 风和日丽 *feng he ri li* 一团和气 *yi tuan he qi* 心平气和 *xin ping qi he*
10 知 *zhi*	**11 余 yu_1**	**12 雨 yu_2**
知书达理 *zhi shu da li* 明知故问 *ming zhi gu wen* 迷途知返 *mi tu zhi fan* 未卜先知 *wei bu xian zhi*	余音绕梁 *yu yin rao liang* 茶余饭后 *cha yu fan hou* 心有余悸 *xin you yu ji* 游刃有余 *you ren you yu*	雨过天晴 *yu guo tian qing* 血雨腥风 *xue yu xing feng* 风调雨顺 *feng tiao yu shun* 倾盆大雨 *qing pen da yu*

Table 5.5: 12 identical characters used on 4 syllable positions in 48 quadrisyllabic idioms.

If there is no syllable position effect, the syllable duration of the identical characters inside different idioms should have equal duration. However, in Table 5.6, the mean, median and SD vary in duration depending on the position. For these 12 identical characters, the first

syllable position leads to longest average syllable duration, and the third syllable position has the second longest duration. The second and the fourth positions have similar average syllable duration.

	Idiom.*i1*-p1	**Idiom.*i2*-p2**	**Idiom.*i3*-p3**	**Idiom.*i4*-p14**
Mean (ms)	217.6	184.4	195.2	182.8
Median (ms)	220	180	190	180
SD (ms)	52.6	46.1	50.3	46.5

Table 5.6: Syllable duration of identical character (i represents the identities of the 12 identical characters in Table 5.5).

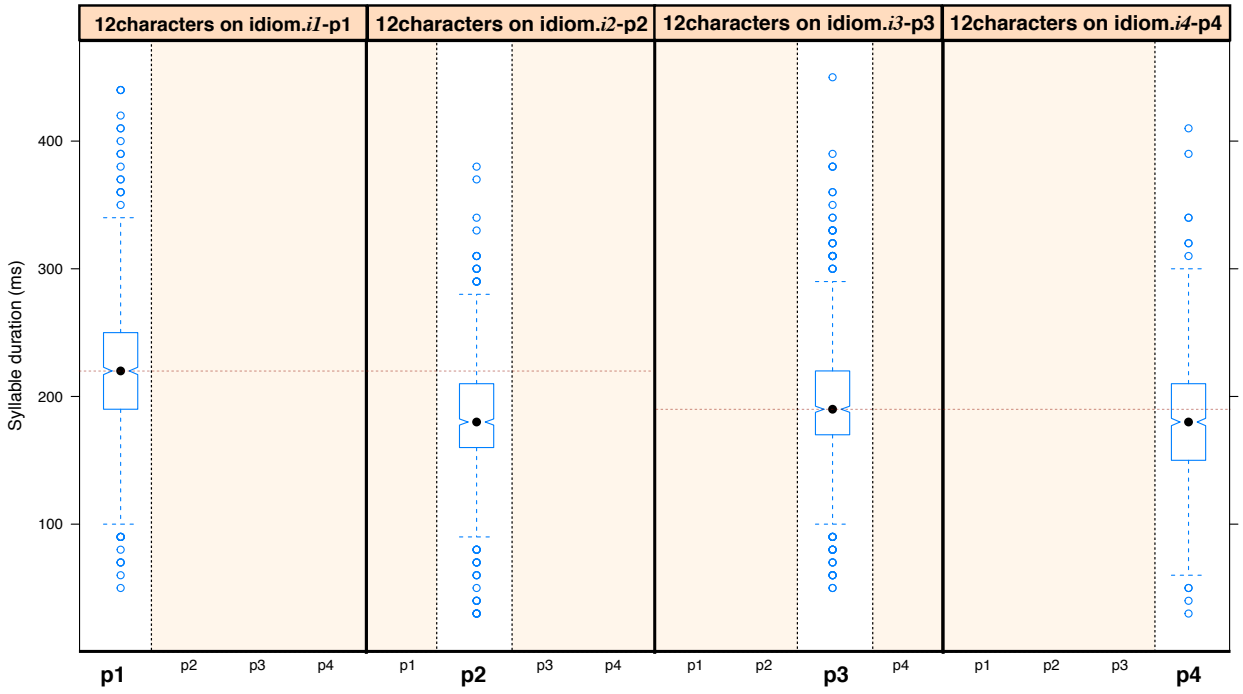


Figure 5.4: Syllable duration of identical character by position and idiom (i represents the identities of the 12 identical characters.)

Figure 5.4 further displays the position effect on the syllable duration. The four panels refer to four positions of the 48 quadrisyllabic idioms. For instance, the first panel with the label 'idiom.*i1*-p1' represents 12 idioms having the 12 target characters on their first syllable positions. In order to illustrate the distinction, only the target position is plotted

inside each panel (the white area), and the other three positions inside the same panel are not plotted. The two horizontal lines mark the median values of positions p1 and p3. There are two long-short contrasts appear on the first two positions and the last two positions repeatedly. Namely, these cross-idiom durational contrasts of identical characters are related to the metrical pattern–trochee discussed in Section 5.4.3. Disyllabic trochaic foot has longer syllable duration of the 'strong' syllable and shorter syllable duration of the 'weak' syllable.

5.4.5 Idiom Structure

Section 5.1.2 explained the structures of quadrisyllabic idioms. Structure is an significant effect (p=0.001) and interacts with syllable position in duration modeling. The four structures in Figure 5.5 are used in the data coding and analysis, and this section discusses the duration pattern under the conditions of different syntactic structures. The four labels ''str40', 'str22', 'str31' and 'str13' are used to refer to Structure 1 to 4 in Figure 5.1.

According to the annotation results from 5 annotators, among 154 idioms used in Study 3, there are 118 idioms with 'str22' structure, 16 idioms with 'str13' structure, 16 idioms with 'str40' structure, and 4 idioms with 'str31' structure. Figure 5.5 displays the durational pattern of each structure. For 'str13', 'str40' and 'str22', they all show long-short patterns on the first two syllables and the last two syllables regardless of the structural distinctions. Among these three structures, only 'str22' has the syntactic disyllabic groups, and the other two structures—'str31' and 'str40' do not form any disyllabic group, but all these three structures share the similar durational pattern. It means there is a disyllabic prosodic pattern applying on all these three structure types, and this disyllabic prosodic pattern is independent of the internal syntactic structure of quadrisyllabic idioms.

This result supports the view that in Chinese, disyllable is the basic or minimal prosodic foot discussed in many previous works (Shih, 1986; Feng, 1998, 2005; Duanmu, 1999, 2000). And more importantly, the cross-structure pattern that different syntactic structures share same durational pattern also suggests that metrical and prosodic pattern can be independent

of syntactic structures.

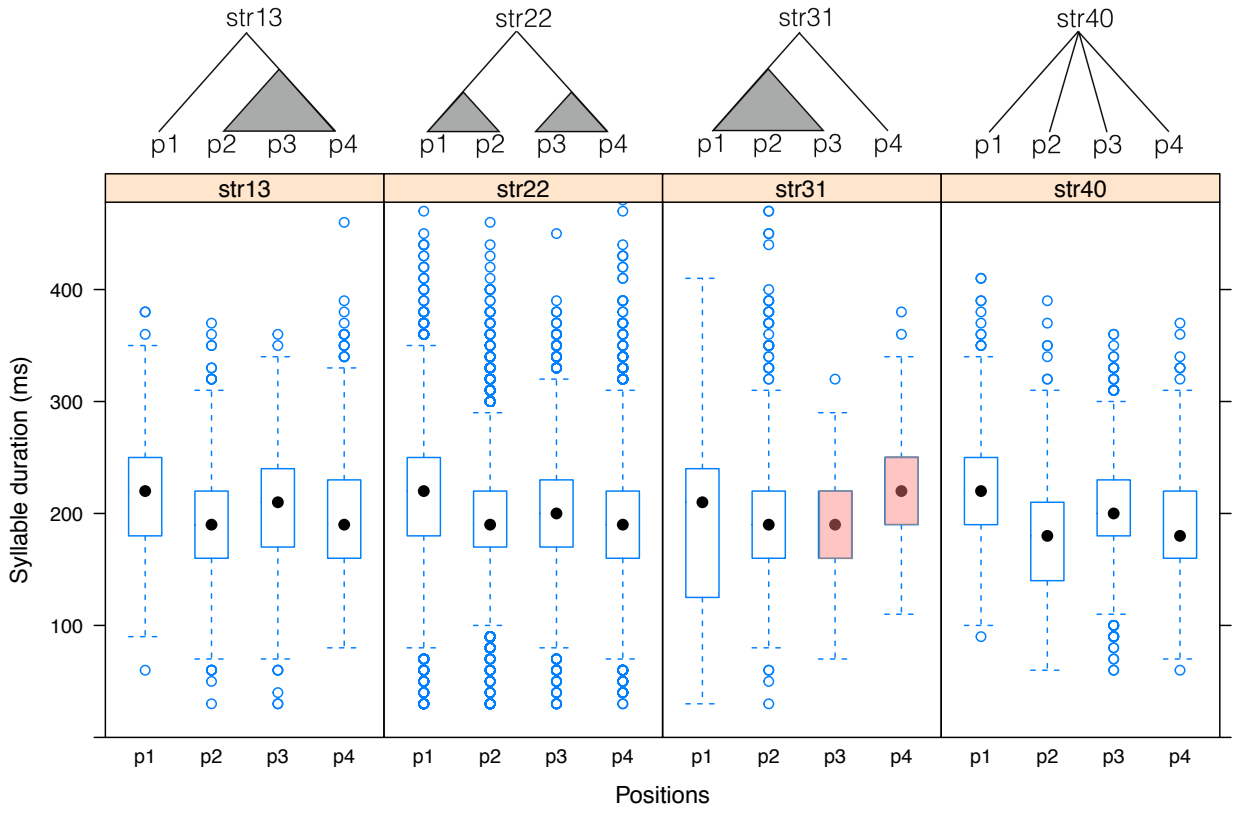


Figure 5.5: Duration patterns of four structures of quadrisyllabic idioms.

The structure 'str31' differentiates itself from other three structures. The first two syllables keep the long-short pattern, but the last two syllables show the short-long pattern. Moreover, the last syllable is the longest syllable in the quadrisyllabic idioms. There are only four idioms[1] with the 'str31' structure, and three of them have the function characters on the third syllable positions. The duration of function character is shorter than content character, and this may be the reason leading to the short duration on the third position. The last syllables in these four idioms with 'str31' structure are the syntactic heads. Though participants were asked to read the sentences without stress and emphasis, the syntactic head,

[1]There are four idioms with structure 'str31' in Study 3: 从一而终 *cong yi er zhong*, 而立之年 *er li zhi nian*, 后起之秀 *hou qi zhi xiu* and 一衣带水 *yi yi dai shui*

especially these following the content word, may attract the stress and emphasis. Speakers treat the last syllables of 'str31' idioms with stress or emphasis, so the last syllables have the longest duration.

In sum, the results suggest that the default and basic metrical unit in Chinese is disyllabic form. For quadrisyllabic idioms with internal structures—'str40', 'str22' and 'str13', the prosodic structure is independent of syntactic structure. For quadrisyllabic idioms with 'str31' structure, speakers may apply stress or emphasis on the last syllable. These findings can be used to predict the duration pattern of idioms with more than four syllables.

5.4.6 Character Reduplication

In quadrisyllabic words, there are many cases that the identical characters appears on different positions (discussed in Section 5.1.3), and the duplicated status of syllable is a significant effect (p=0.001) in the model. Figure 5.2 (in Section 5.1.3) exhibits all the seven possible types of duplicated character in quadrisyllabic idioms. Except for the uncommon Type (7), there are ten idioms selected for each type in the experiment materials. In total, there are sixty quadrisyllabic idioms (Table 5.7).

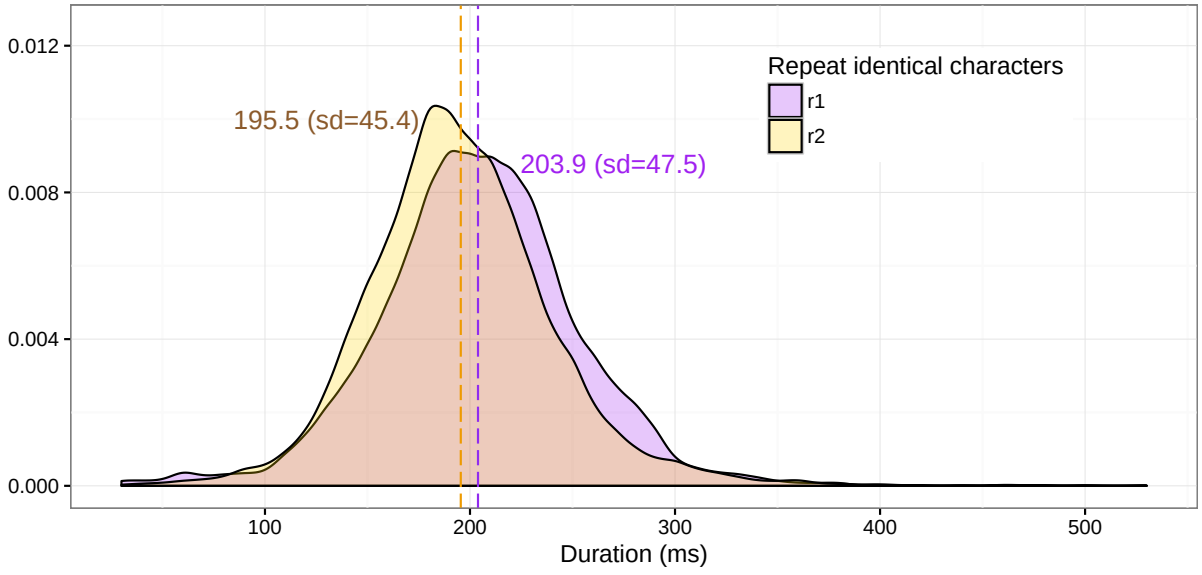


Figure 5.6: Distribution of syllable duration of reduplicated syllable

Group 1	Group 2	Group 3
错上加错*cuo shang jia cuo*	百依百顺 *bai yi bai shun*	爱理不理*ai li bu li*
话里有话*hua li you hua*	合情合理 *he qing he li*	心服口服*xin fu kou fu*
人无完人*ren wu wan ren*	绘声绘色 *hui sheng hui se*	知法犯法*zhi fa fan fa*
天外有天*tian wai you tian*	有勇有谋 *you yong you mou*	将计就计*jiang ji jiu ji*
微乎其微*wei hu qi wei*	随时随地 *sui shi sui di*	讨价还价*tao jia huan jia*
年复一年*nian fu yi nian*	无影无踪 *wu ying wu zong*	倚老卖老*yi lao mai lao*
痛定思痛*tong ding si tong*	任劳任怨 *ren lao ren yuan*	就事论事*jiu shi lun shi*
为所欲为*wei suo yu wei*	多才多艺 *duo cai duo yi*	自然而然*zi ran er ran*
贼喊捉贼*zei han zhuo zei*	没完没了 *mei wan mei liao*	在商言商*zai shang yan shang*
精益求精*jing yi qiu jing*	愣头愣脑 *leng tou leng nao*	不懂装懂*bu dong zhuang dong*
Group 4	**Group 5**	**Group 6**
天网恢恢*tian wang hui hui*	恢恢天网 *hui hui tian wang*	安安静静*an an jing jing*
含情脉脉*han qing mo mo*	脉脉含情 *han qing mo mo*	结结巴巴*jie jie ba ba*
大名鼎鼎*da ming ding ding*	鼎鼎大名 *ding ding da ming*	婆婆妈妈*po po ma ma*
文质彬彬*wen zhi bin bin*	彬彬有理 *bin bin you li*	大大方方*da da fang fang*
两手空空*liang shou kong kong*	空空如也 *kong kong ru ye*	兢兢业业*jing jing ye ye*
风度翩翩*feng du pian pian*	翩翩起舞 *pian pian qi wu*	匆匆忙忙*cong cong mang mang*
情意绵绵*qing yi mian mian*	绵绵不绝 *mian mian bu jue*	沸沸扬扬*fei fei yang yang*
无所事事*wu suo shi shi*	事事躬亲 *shi shi gong qin*	堂堂正正*tang tang zheng zheng*
得意洋洋*de yi yang yang*	洋洋得意 *yang yang de yi*	日日夜夜*ri ri ye ye*
长夜漫慢*chang ye man man*	漫漫长夜 *man man chang ye*	口口声声*kou kou sheng sheng*

Table 5.7: Repeated identical characters on different positions in 60 quadrisyllabic idioms.

Ten pairs of identical characters appear in p1 and p4 in Group 1; ten pairs of identical characters appear in p1 and p3 in Group 2; ten pairs identical characters appear on p2 and p4. In these three groups, identical characters are not next to each other. Group 4 and Group 5 have the same ten pairs of identical characters in the last two positions and the first two positions respectively. Group 6 have twenty pairs of identical characters in ten idioms–ten pairs on the first two positions and ten pairs on the last two positions. In these three groups, identical characters are next to each other. The meanings of these sixty idioms are listed in Appendix D.

Hypothetically, if there is a shortening effect of the reduplication, the first characters (r1) should have longer duration than the reduplicated one (r2). Figure 5.6 shows that there is a durational contrast between r1 and r2, and r1 indeed shows longer duration. However, Section 5.4.4 discussed the position effect that first and third positions have longer syllable duration and the second and the fourth positions have shorter duration. Moreover, the duration values of p2 and p4 are similar to each other (Table 5.6). Therefore, in order to rule out the position effect and further examine the repeated character effect, Figure 5.7 plot all the six groups separately. For each subplot in this figure, only the target paired identical characters' syllable positions are plotted.

In Figure 5.7's subplot 'group3-p2p4', the 10 pairs identical characters first appear in the p2 position, and repeatedly appear in the p4 position, and both p2 and p4 positions are 'weak' syllable positions in the two metrical feet of quadrisyllabic idioms. It is the possible paired positions to rule out the durational distinction caused by positions. The result in 'group3-p2p4' show that the paired identical characters indeed have durational difference, characters in p2 have longer duration than characters in p4.

Moreover, Group 5 and Group 4 used the same 10 paired identical characters on the two metrical feet–p1p2 and p3p4 respectively. The cross-group results show that: r1 at p3 of Group 4 and r1 at p1 of Group 5 are both at the 'strong' positions of the metrical feet, and their duration values have the similar distribution; r2 at p4 of Group 4 and r2 at p2

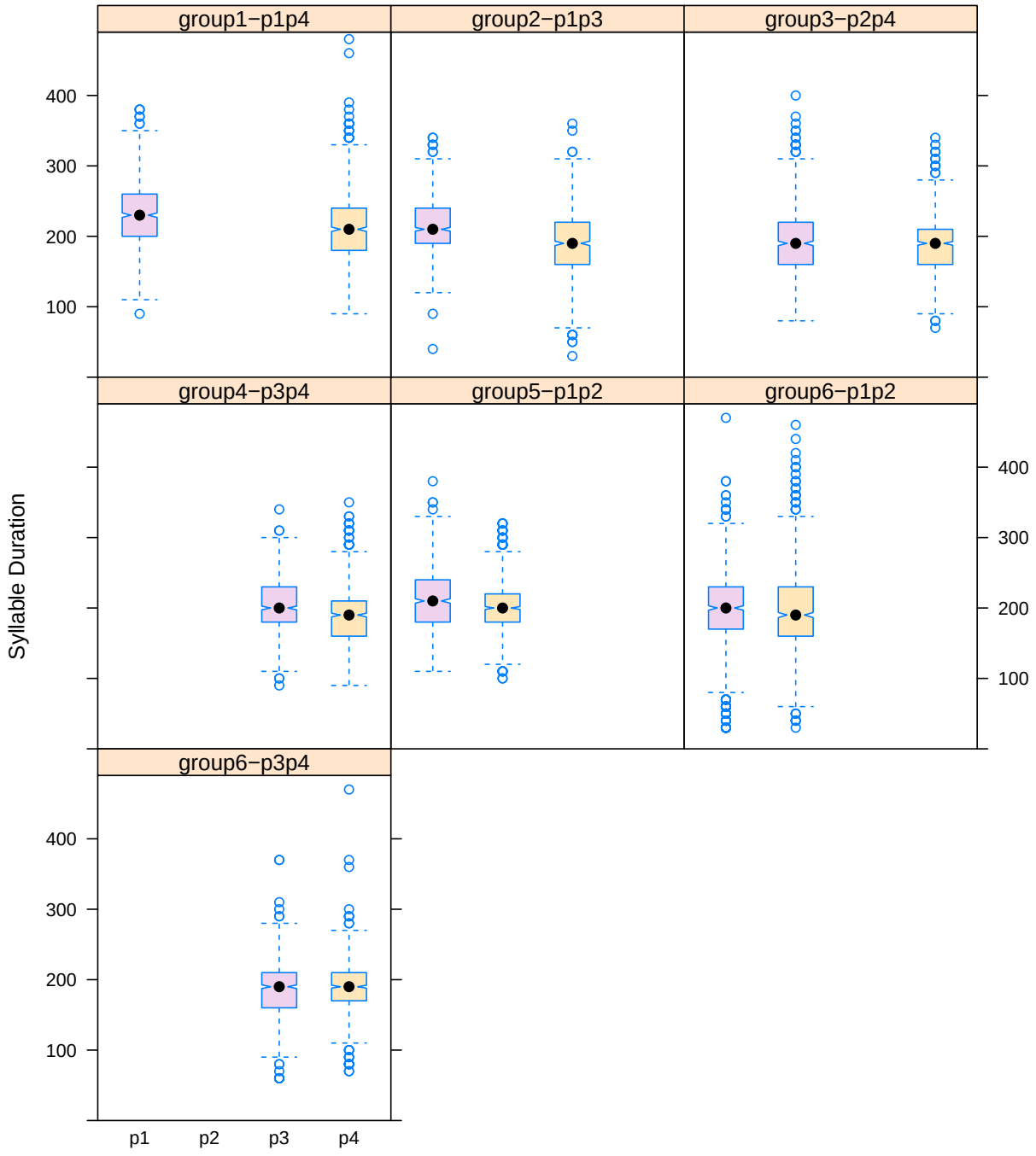


Figure 5.7: Duration of reduplicated syllables by syllable positions

of Group 5 are both at the 'weak' positions of the metrical feet, and they also have the similar distribution. It could be considered as the second evidence to prove that without the position effect caused by the metrical pattern, repeated characters have shorter duration than the first appeared ones.

For all subplots in Figure 5.7, only 'group6-p3p4' shows closed duration values, and all the other 5 cases demonstrate the clearly durational distinction between the first appeared characters and their corresponding repeated ones.

5.4.7 *De* Attachment

The particle *de* has a shortening effect on its host word. In Study 2, syllables of the trisyllabic words are affected by the particle *de* attachment. And the third syllables of the trisyllabic words are shortened most by the adjacent *de*. In quadrisyllabic idioms, the effect of *de* attachment is significant (p=0.02) on syllable duration.

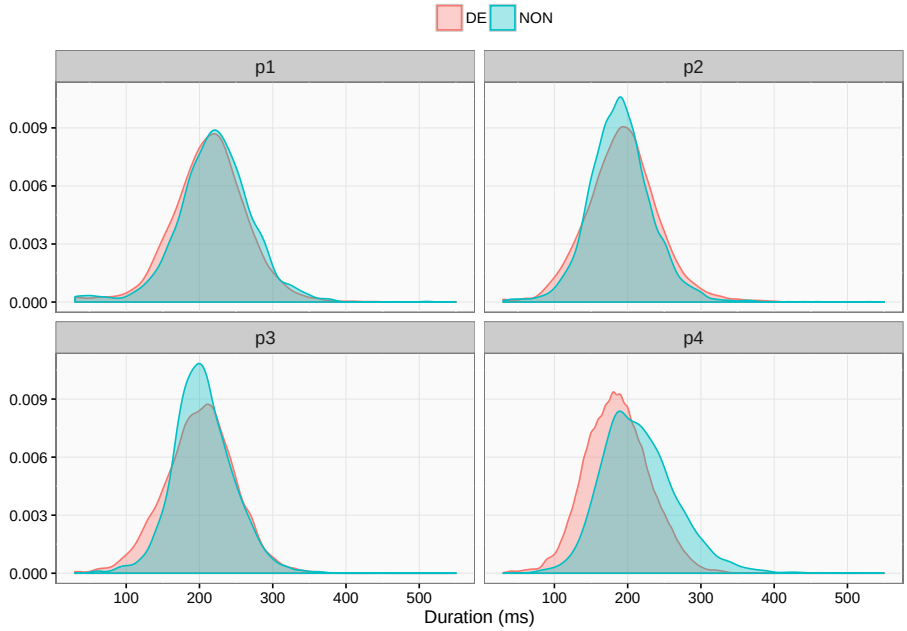


Figure 5.8: Effect of *de* attachment on syllable duration

Figure 5.8 exhibits the distribution of syllable duration under the two different conditions–

with and without *de* attachment on the four syllable positions of quadrisyllabic words correspondingly. From p1 to p4, the durational distinction between syllables with *de* attachment and without *de* attachment grow much larger and clearer. The syllable at the last position (mean=186.1, SD=44) with *de* attachment are shorter than syllable without *de* attachment on the same position (mean=215.7, SD=51) .

5.4.8 Content Character and Function Character

The result of word category effect in Study 1 shows that content words of Modern Chinese have longer duration than function words. As mentioned in Section 5.3.3, characters are divided into content and function categories in Classical Chinese. This section reportes the durational distinction between content character *shi zi* and function character *xu zi* in quadrisyllabic idioms. Table 5.8 shows that content characters are 20ms longer than function characters.

	Content	**Function**
Sample size	50040	5400
Mean (ms)	202.9	180.6
Median (ms)	200	180
SD (ms)	48.4	52.9

Table 5.8: Mean, median, and SD of content characters and function characters in idioms

In addition, Figure 5.9 displays that content characters also have longer duration than function characters on all the four syllable positions respectively, especially on the 'weak' positions of p2 and p4.

5.4.9 Syllable Structure

Syllable structure effect is significant in all three studies. Syllables in 154 quadrisyllabic idioms in Study 3 cover the 12 possible syllable structures in Chinese, and the sample size for each structure category is listed in Table 5.9. According to the rank of the medians

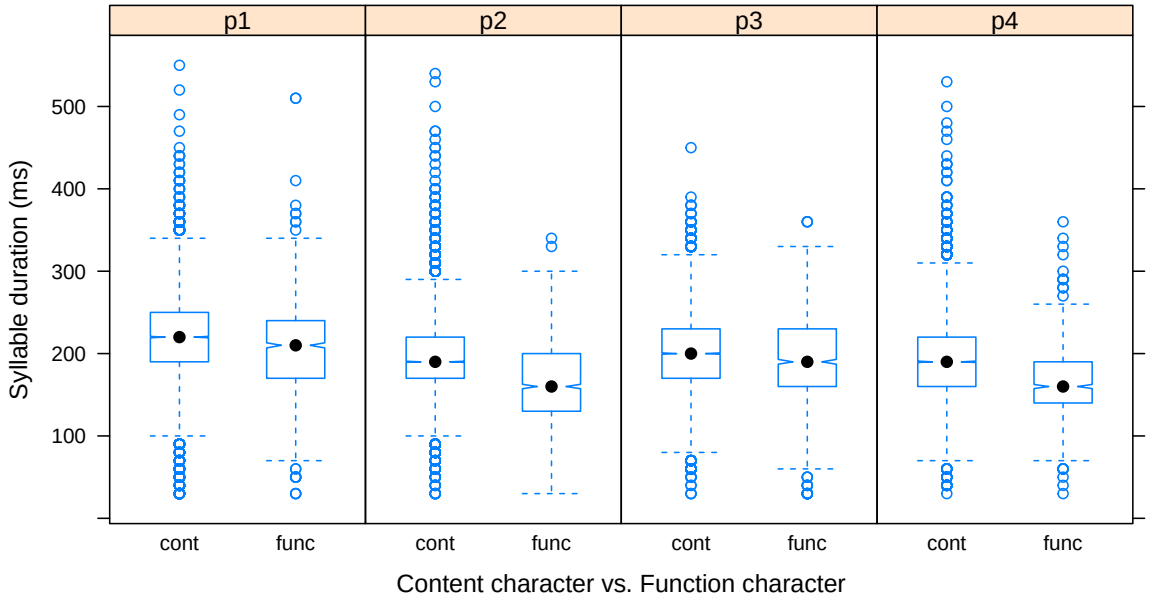


Figure 5.9: Syllable duration of content and function characters by syllable positions

Structure	Number of syllables	Median (ms)	Mean (ms)	SD (ms)
01-V	4050	150	159	56
02-VG	1260	160	161	57
03-GV	900	170	176	44
04-GVG	1620	180	180	41
05-VN	270	190	190	76
06-GVN	2610	190	188	45
07-CV	16200	190	192	46
08-CVG	4410	205	208	40
09-CGV	3600	210	213	40
10-CVN	13230	220	217	44
11-CGVG	2610	220	220	41
12-CGVN	4680	220	225	42

Table 5.9: Syllable duration by syllable structure

of these 12 structures, Figure 5.10 demonstrates the distribution of each type in ascending order, and the results show the pattern that syllables with simple structures have shorter duration and syllables with complicated structures have longer duration (Figure 5.10).

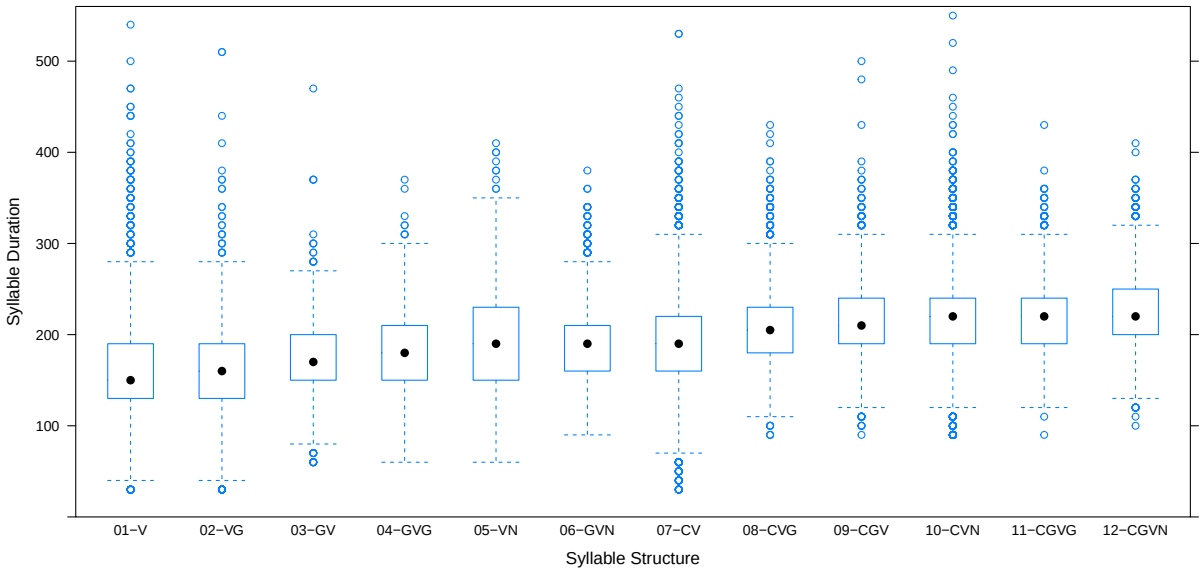


Figure 5.10: Syllable duration by syllable structure

5.5 Discussion

The result in Section 5.4.3 displays the pattern of quadrisyllabic word could be considered as the combination of two long-short sub-patterns. This disyllabic long-short pattern may relate to the dominant status of disyllabic word in Mandarin lexical system. The statistical data of word length based on large-scale corpus in Table 2.1 demonstrates that the majority form of Chinese word (71.6%) in Modern Chinese is the disyllabic word. This corpus result is also consistent with the calculation in many previous studies, for example, 69.8% in He and Li (1987) and 71% in Duanmu (1999). The dominant position of disyllabic word may lead to its durational pattern become the basic unit and primary form of the durational pattern in Chinese multi-syllabic words.

In addition, the pattern of quadrisyllabic words also supports the analysis that Chinese words are parsed from left to right into trochaic disyllabic metrical structure discussed in previous works (Feng, 1998, 2005; Duanmu, 2000; Sui, 2013).

The syllable duration pattern of quadrisyllabic idiom in Study 3 also displays two interesting facts (Figure 5.11). First, within each trochaic foot, there is a 'contrast' relationship between the two syllables. The correlation (r=0.085) between the duration of the first syllable and the duration of the second syllable and the correlation (r=0.008) between the duration of the third syllable and the duration of the fourth syllable are both extremely weak. Second, between the two trochaic feet, there is a 'coordinate' relationship between the first two syllables and the last two syllables. The correlation (r=0.303) between the sum duration of the first two syllables and the sum duration of the last two syllable is much stronger than the two within trochaic foot correlations.

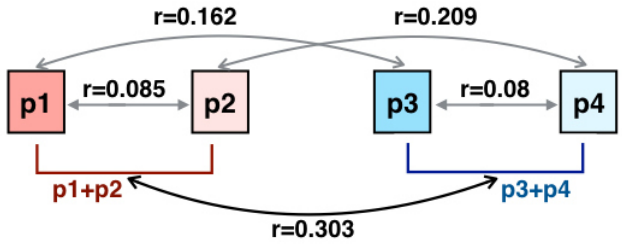


Figure 5.11: Correlation of syllable duration

In Figure 5.11, there are also two cross-position correlations—the correlation (r=0.162) of the first syllable and the third syllable; and the correlation (r=0.209) of the second syllable and the fourth syllable. These two cross-position correlations are stronger than the two within trochaic foot correlations.

The 'contrast' relationship within the trochaic foot indicates that duration is used to distinguish the two syllables. The 'coordinate' between two feet of one quadrisyllabic idiom indicates that there is a certain degree of consistency on durational patterns between two adjacent feet.

In metrical phonology, the abstract concepts of 'strong' and 'weak' are used to represent the speech strength (Liberman and Prince, 1977). Speech strength is related to the articulation effort (Browman and Goldstein, 1990). Kochanski et al. (2003) measured the prosodic strength and reported that strengths are correlated with syllable duration. Figure 5.12 displays the metrical patterns of 4-syllable words, and the two tree structures are based on the results from different annotators in Kochanski et al. (2003). 'The words are plotted as trees, and syllables are represented by the black arrowheads at the end of the line. The vertical position of the ith arrowhead is proportional to the metrical strength of the ith syllable'; 'the patterns for four syllable words have larger errors' and the double arrows display 'the range of the fitted solutions' (Kochanski et al., 2003). The tree diagrams show that the quadrisyllabic words could be broken up into pairs of disyllabic forms, and for each disyllabic form, there is a strong-weak pattern of prosodic strength.

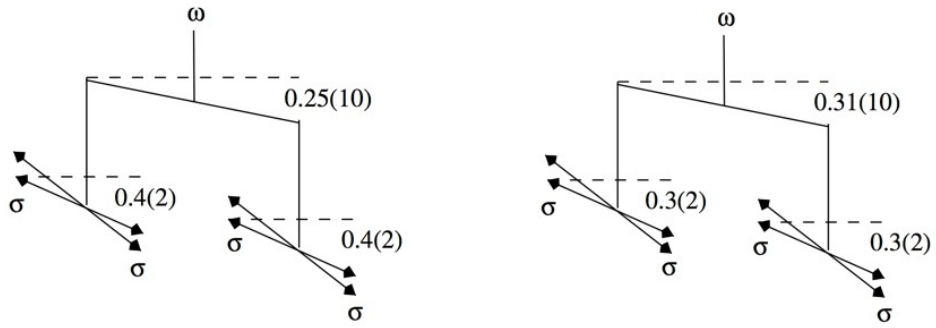


Fig. 8 and Fig. 9 in Kochanski et al. (2003)

Figure 5.12: Metrical pattern of 4-syllable words in Kochanski et al. (2003)

The duration pattern of quadrisyllabic idioms in Study 3 are coincident with the prosodic strength pattern measured by Kochanski et al. (2003). The combination of two long-short pattern on duration could be considered as the durational representation of prosodic strength pattern of quadrisyllabic words.

5.6 Summary of Study 3

Study 3 gives an overall examination on Chinese Quadrisyllabic idioms based on the linguistic factors, social factors and their interactions among three different Mandarin regions. In this study, 154 quadrisyllabic words were selected based on the various characteristics of Mandarin quadrisyllabic idioms, and each idiom was carried in a unique sentence in the sentence middle position.

A mixed effect model was performed on the duration data. The results indicate that syllable structure, syllable position, idiom structure, particle attachment, duplicated syllable, word category, and speech rate are all significant. The interactions among idiom structure, syllable positions and particle attachment are also significant. However, region, gender, and idiom frequency are not the significant effects. Section 5.5 discussed the syllable position effect, idiom structure, duplicated syllable, particle attachment, word category and syllable structure effects respectively. The syllable duration on different positions confirmed the statement that the trochaic disyllabic form is the basic metrical unit in Chinese. The results of syllable duration by different idiom structures reveals the relationship between duration pattern and structure types. The results demonstrate that not only the disyllabic form is the default metrical foot in Chinese, but also the prosodic structure can be independent of syntactic structures. For the duplicated syllables on the two different positions of an idiom, all the possible cases were examined in this study, and the result showed that the repeated characters have shorter duration. Particle *de* attachment displayed the shortening effect on syllable duration of quadrisyllabic words, especially at the last syllable position. According to the Classical Chinese word categories, both content characters and function characters were examined in the model, and the result showed that content characters have longer duration than function characters. Syllable duration on different syllable structures displayed the similar results in Study 1 and Study 2. The simple structures have shorter duration, and the complicated structures have longer duration.

Chapter 6

General Discussion

Chapter 3, 4 and 5 reported the duration patterns of monosyllabic, disyllabic, trisyllabic, and quadrisyllabic words in Mandarin. This Chapter will synthesize some findings from these studies and discuss the directions of further research. Section 6.1 will discuss the disyllabic rhythmic pattern in multi-syllabic words. Section 6.2 will address the relationship between rhythmic pattern and word frequency. Section 6.3 will explore the random effect of speaker and individual difference. Section 6.4 will briefly review the gender difference on duration patterns. At the end of this Chapter, Section 6.5 will address the directions of further research on Mandarin duration.

6.1 Rhythmic Pattern

The syllable duration results show that the disyllabic long-short rhythmic pattern is ubiquitous throughout all multi-syllabic words at non sentence-final positions, but the durational contrasts between the long syllable and the short syllable among words with different length are different. Figure 6.1 displays the distinction of durational contrast within each word type. The first two syllables of trisyllabic words and quadrisyllabic words show stronger contrasts than the two syllables of disyllabic word and last two syllables of the quadrisyllabic word. It indicates that for the first two syllables of Chinese multisyllabic words, the rhythmic contrast gets larger with the increasing of word length. In addition, for both trisyllabic and quadrisyllabic words, the long-short pattern is prevalent regardless of the syntactic branching structure of the word. Based on the long-short pattern and its independence, it is

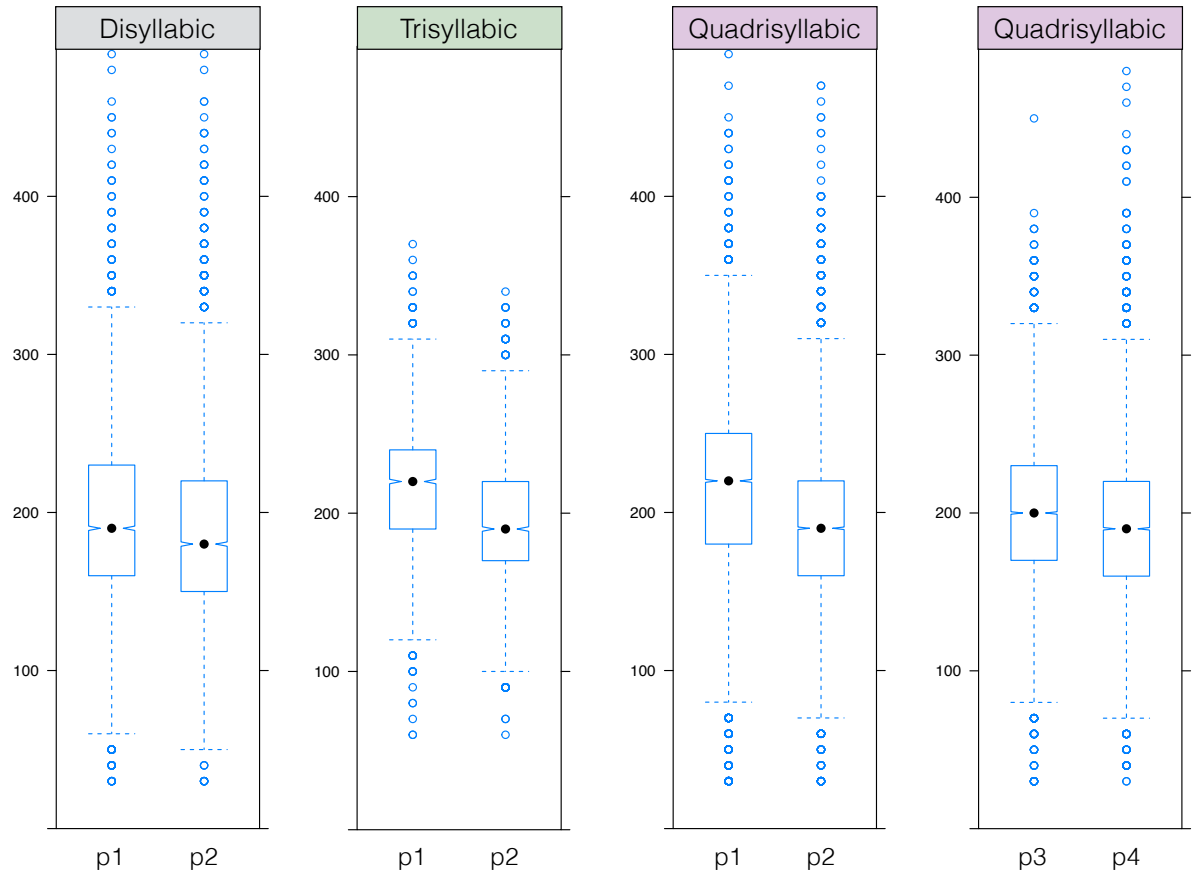


Figure 6.1: Disyllabic long-short pattern by multi-syllabic word types

possible to further suggest that words with more than four syllables can be processed from left to right into disyllabic feet with long-short durational pattern.

6.2 Word Frequency

Word frequency is one of the factors used in the duration models of Chinese words. The results of syllable duration of monosyllabic word, disyllabic word, and trisyllabic word all indicate that frequency effect is significant in duration modeling. However, the result of duration model of quadrisyllabic word (Section 5.4.2) shows that for syllable duration of quadrisyllabic word, word frequency is not a significant effect.

Figure 6.2 displays the relationship between word frequency and syllable duration. Four

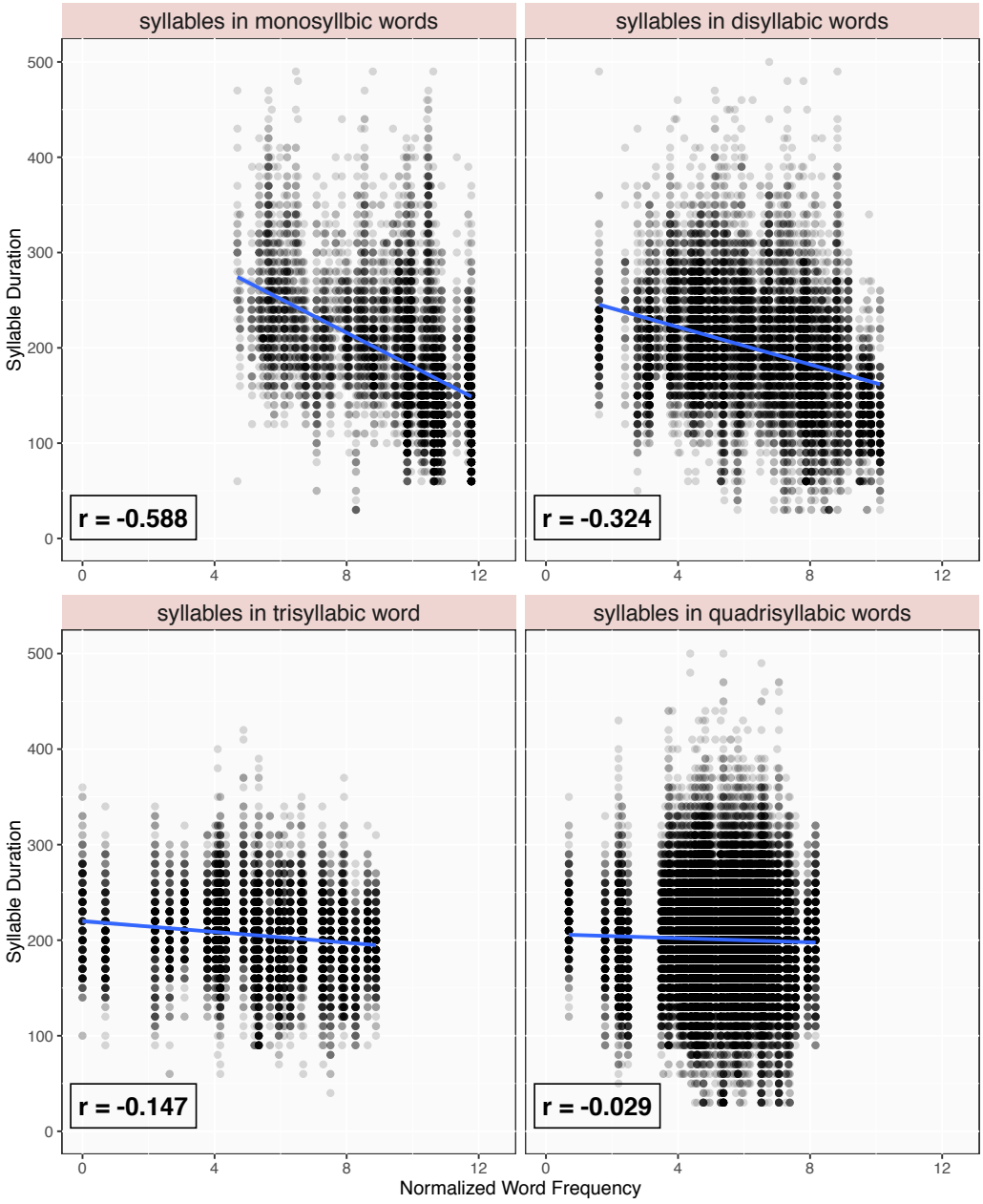


Figure 6.2: Correlation between word frequency and syllable duration

subplots in Figure 6.2 represent the monosyllabic, disyllabic, trisyllabic and quadrisyllabic words respectively. The negative correlations between word frequency and syllable duration become weaker with the increasing of word length. Syllable duration shows a strong negative correlation with word frequency in monosyllabic words, but there is a lack of correlation between word frequency and syllable duration in quadrisyllabic words.

This decrease of frequency effect is related to the rhythmic contrast in multisyllabic words. Section 6.1 discussed that along with the increase of word length, the durational contrast gets stronger. When contrast gets stronger, frequency effect becomes weaker. The lack of frequency effect on quadrisyllabic words may also be due to the code switching between Classical Chinese and Modern Chinese. Classical Chinese verse and poem have strong rhythmic patterns. Quadrisyllabic words inherit the characteristics of Classical Chinese, especially the rhythmic contrast in reading. When Mandarin speakers produce an idiom surrounded by Modern Chinese words, they may automatically switch between Classical Chinese and Modern Chinese and adjust the rhythmic contrast. Therefore, the frequency effect does word because of the code switching.

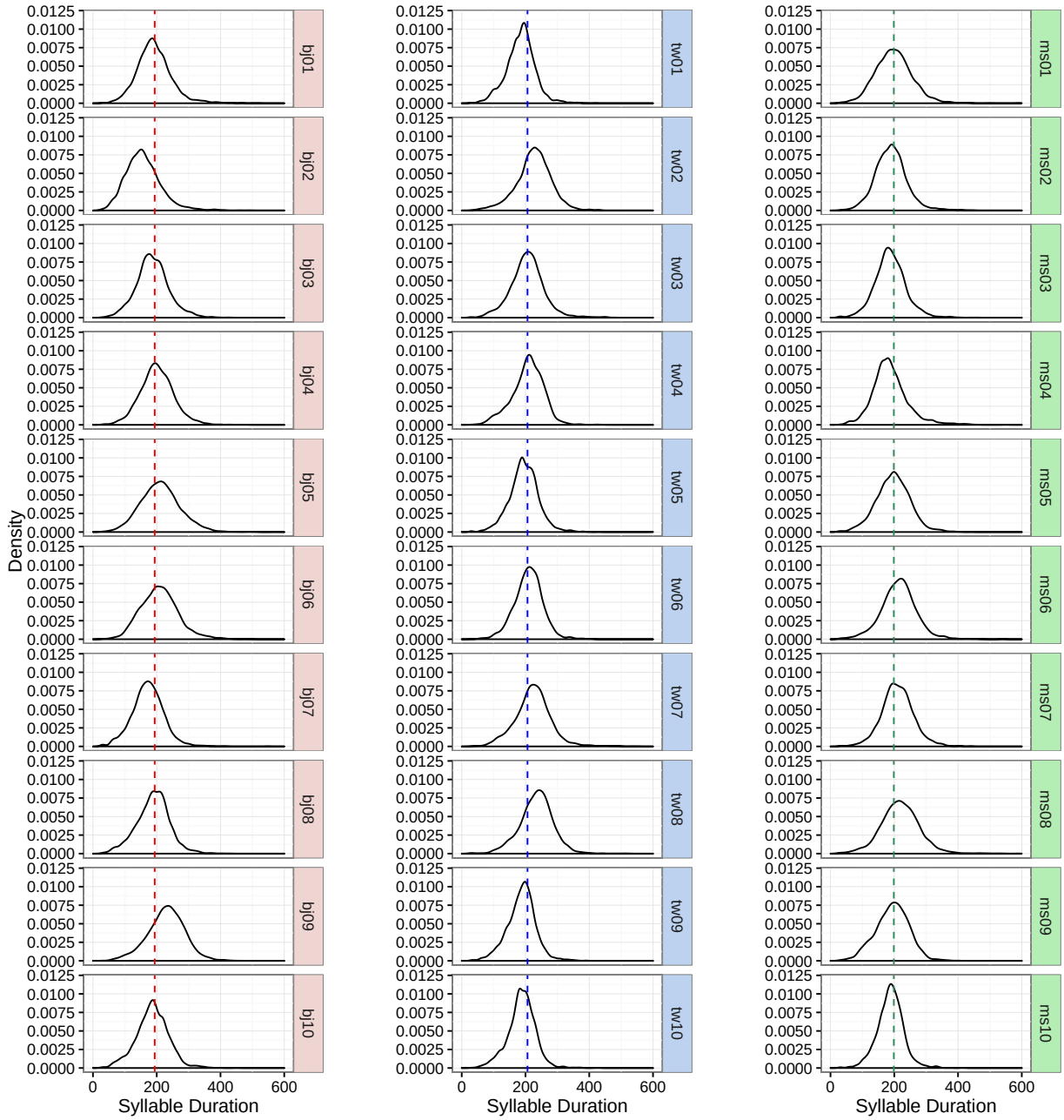
6.3 Individual Difference

This section concentrates on the individual differences among Mandarin speakers within each region. There is no homogeneous speech community, not only because of the gender and regional difference in speech production but also because of the individual differences from speakers. In linguistic research, there have been an increasing number of studies that have paid attention to individual differences on speakers' language ability and behavior (Ellermeier and Zimmer, 1997; Pennebaker and King, 1999; Ehrman et al., 2003), because individual difference could be the reason caused the within-region variation and may lead the direction of language development.

In this dissertation, speaker effect is a random effect and is nested within region in all

three studies. The results of models show that there are relatively small variances from the factor of speaker, but speaker effect is useful for construction of speaker-specific duration system. By using the mixed effects model with speaker as one of the random effects, Shih and Lu (2015) examined talker-to-listener distance on tone and tonal perception. The results of their study show the advantage of the mixed effects model with the random effect of speaker. The"separation of fixed and random effects is a highly desirable feature in speech synthesis, making it is possible to build a speaker-independent model using fixed effects and then adapt it to speaker-dependent systems by adding speaker coefficients from the random effects" (Shih and Lu, 2015). The syllable duration of Chinese word can be calculated via the mixed effects model, and compare the model prediction with the observation, and the random effect of speaker provides the possibility to build a speaker-specific duration system.

In order to further investigate the individual difference on duration, the syllable duration data from monosyllabic words (Study 1), disyllabic words (Study 1), trisyllabic words (Study 2) and quadrisyllabic word (Study 3) are pooled together to explore the syllable duration of Chinese words. There are 86,971 syllables measured in these three studies: 8,625 syllables from monosyllabic words; 13,726 syllables from disyllabic words; 9,180 syllables from trisyllabic words; and 55,440 syllables from quadrisyllabic words. Figure 6.3 exhibits the distribution of individual speaker's syllable duration from three studies. Each density plot represents the data from one individual speaker, and three panels represent three Mandarin speaking regions. Within each region, the dash line indicates the overall mean value of syllable duration of the region. Among 10 Beijing Mandarin speakers, speaker bj02 (male) has much shorter syllable duration and speaker bj09 (male) has longer syllable duration in this group. In Figure 6.3 (b), Taiwan Mandarin speakers tw02 (female) and tw08 (female) both have much longer syllable duration than other speakers in the same group. The results in Figure 6.3 (c) show that Malaysia Mandarin speaker have less variability. In Malaysia group, only speaker ms06 (female) and speaker ms08 (male) show slightly longer duration than the other speakers.



(a) Beijing Mandarin Speaker (b) Taiwan Mandarin Speaker (c) Malaysia Mandarin Speaker

Figure 6.3: Individual difference on distribution of syllable duration among three Mandarin region

6.4 Gender Effect

This section addresses the gender effect on syllable duration of Chinese words. Many previous works focus on the relationship between gender and duration measurement on different language (Byrd, 1994b; Whiteside, 1996; Yuan et al., 2006; Heffernan, 2007; Jacewicz et al., 2009; Oh, 2011; Wu and Shih, 2015). For example, Byrd (1994b) reported that women speak slower than men in American English and Whiteside (1996) reported that women have longer sentence durations in British English. In Mandarin, Wu and Shih (2015) found out that female speakers have longer syllable duration than male speakers in both Beijing Mandarin and Taiwan Mandarin.

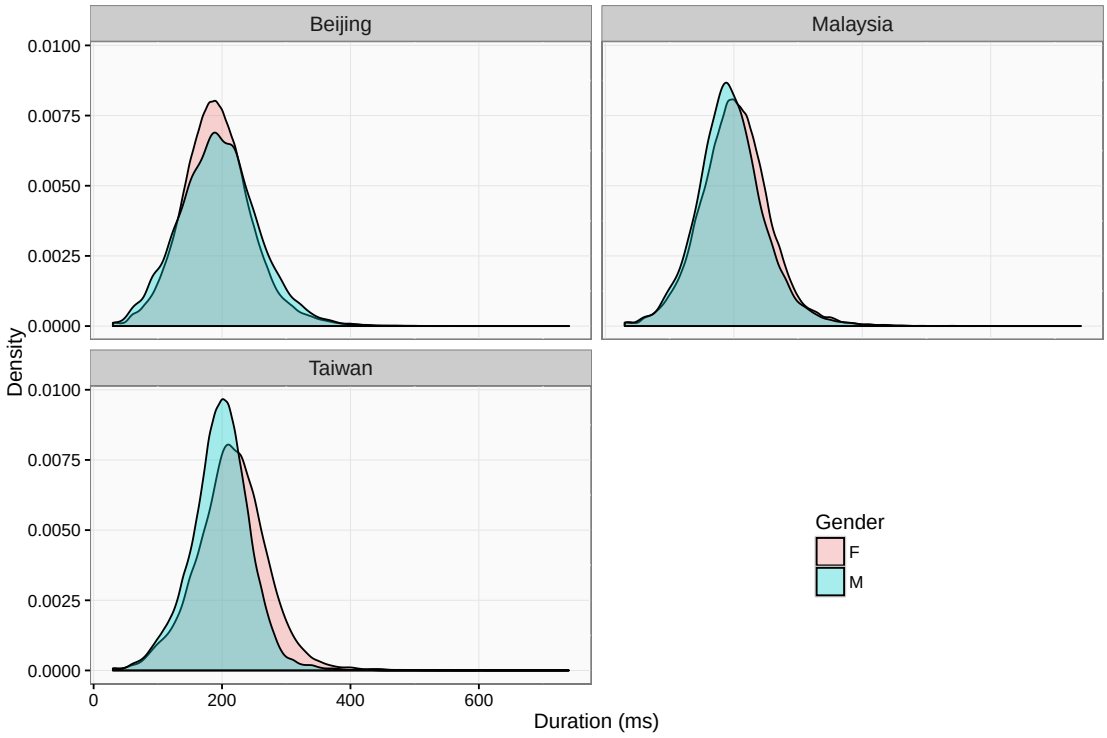


Figure 6.4: Gender difference on syllable duration (pooled all three studies) among three Mandarin regions

Gender is not a significant effect on syllable duration in all three studies in this dissertation. But after pooling the syllable duration data from three studies, Figure 6.4 demonstrates that there are durational differences between two genders. For female and male speakers of

Beijing Mandarin, they have similar mean (female=193ms, male=195ms) and same median (190 ms), but male speakers have larger variance (61ms) than female speakers (55ms). By contrast, both Malaysia Mandarin and Taiwan Mandarin show that female speakers have longer syllable duration than male speakers. The gender divergence of Taiwan Mandarin is larger than the other two regions, and the gap between the mean value of female speakers and the mean value of male speakers is around 18ms.

6.5 Further Research

In this dissertation, three studies build a system of word duration patterns with both linguistic and social factors, and this system can be improved and enhanced in the further research.

6.5.1 Tonal variation

The current research frame of this dissertation concentrates on the influences of non-supersegmental features on word duration patterns. Examining duration variation related to tonal environment is an interesting topic of further research of duration patterns.

Both Mandarin and Chinese dialects are rich in tonal variations such as tone sandhi rules (Shih, 1986; Chen, 1987; Zhang, 1988; Tsay and Myers, 1996; Chen, 2000) and contextually tonal change (Shih, 1988; Xu, 1993). In a particular tonal context, the citation forms of lexical tones are affected by the preceding and following tones. On the one hand, the context may lead to the tonal change on a syllable. For example, Shih (2005) explored the nature of Mandarin tone2 in an environment that can cause tonal reduction, and the result indicates that the reduction occurs on prosodically weak position with a change of syllable duration. On the other hand, in order to fulfill the requirement of accuracy in speech communication, speakers need a certain time to finish the articulation of the target syllable to carry an intelligible tonal information. Xu (1999) examined timing-related issues from

the articulatory perspective on Mandarin tone and pointed out that the ability of a syllable to carry a contour tone is directly related to the duration of its tone-bearing portion.

Besides four lexical tones, there is a neutral tone in Mandarin that only occurs in a weak syllable without a fixed pitch contour (Chao, 1948, 1968). Syllables with neutral tones in Mandarin words do not occur at the word-initial position and must be preceded by the syllable or syllables with full lexical tone. Neutral tone is regarded as a tonal phenomenon in Chinese, and it is not only related to duration reduction but also related to F0 change (Lin, 1985; Cao, 1986; Chen and Xu, 2006). Since neutral tone is the suprasegmental feature leading to durational reduction, so it is not considered under the current research frame of this dissertation. As a high frequency tonal condition in Mandarin, there are regional differences in neutral tone production. For example, Tseng (2004) reported that Beijing Mandarin show great F0 change from the preceding accented syllable to the syllable with neutral tone, but the neutral tone syllable in Taiwan Mandarin is more independent from the preceding syllable and it tends to be produced as a low entering tone regardless the preceding tone. It is necessary to address the super-segmental features in the next step of word duration pattern research, and the duration reduction on syllable with neutral tone in Chinese multi-syllabic words is a meaningful research topic.

6.5.2 Mandarin Suffixes

Mandarin only has a small number of suffixes, and some of them are related to durational reduction, such as 子 *zi* and 儿 *er*. Both *zi* and 儿 *er* were monosyllabic words meaning "child" in Classical Chinese. *Zi* is typically added to nouns and produced as a light syllable with neutral tone. *Er* suffix are commonly used as a nominal suffix in northern Mandarin speech, especially in Beijing Mandarin. The *er* suffix gives an r-coloring (retroflex) to the preceding vowel of the rime (Chao, 1948, 1968). The production of *er* has regional differences. Speakers of Beijing Mandarin have merged the suffix into the host syllable in the word, but Mandarin speakers in many other regions do not use *er* as frequently as speaker of Beijing

Mandarin, and "some Standard Chinese speakers, especially those in the southern part of China, such as Hong Kong and Shanghai, and in Taiwan and Singapore, either pronounce the suffix as a separate syllable or do not even exhibit the r-suffixation process at all" (Lin, 2007, p183). Wu and Shih (2015) reported the regional differences on duration of *er* suffix by examining the duration of the suffix and its host syllable, and their results indicate that without *er* suffix, Beijing Mandarin, Taiwan Mandarin, and Mandarin spoken in central China have similar syllable duration. But when *er* is added to the host syllable, Taiwan speakers produced the longest duration of the suffix and its host syllable and Beijing speakers produced the shortest duration.

A systematic and well-design research on duration pattern of Mandarin suffix can provide new insight concerning duration variability of words. The variation of duration on suffix among different regions also can be used as a parameter to differentiate varieties of Mandarin.

Chapter 7

Conclusion

This dissertation investigates the cross-regional word duration patterns of Mandarin words among three Mandarin speaking areas—Beijing, Taiwan, and Malaysia. The questions addressed in this dissertation are:

(1) What are the duration patterns of Mandarin words?

(2) What are the linguistic factors and social factors leading to durational variations?

(3) How do they interact with each other?

(4) How to create a dynamic speech context to cover and examine these factors?

(5) How to use the statistical model to reveal duration patterns?

Three speech production studies are conducted to explore the duration patterns of words with different length and internal structures.

This dissertation brings diverse methodologies on speech data collection, annotation, and corpus construction. On the one hand, this dissertation adopted the experimental research approach to collect the speech data with linguistic and social factors. On the other hand, in order to reduce the possibility of participants' awareness of research purpose and collect natural speech data, Study 1 used six stories to examine monosyllabic and disyllabic words, and Study 2 and 3 used unique carrier sentence for each target trisyllabic and quadrisyllabic word in sentence reading task. And then speech data was annotated via corpus research approach. The combination of different research approaches provides an efficient way to examine the selected factors in a speech database covering factors in a natural speech context.

Three studies reveal the general duration patterns of Mandarin Words. First of all, all the multi-syllabic words demonstrate the disyllabic long-short metrical form, which supports the

theoretical analysis of the trochaic metrical structure of Mandarin words. Second, linguistic factors—syllable structure, positions (syllable position, word position, and sentence position), word frequency, word category, word internal structure, particle attachment, speech rate of sentence have significant effects on syllable duration. Thirdly, social factor—region interacts with multiple linguistic factors (word structure, syllable position, and particle attachment) and play an important role in duration prediction. There indeed have regional differences in rhythmic contrast among different Mandarin speaking regions. For example, Beijing Mandarin speakers are more sensitive to the length change of linguistic unit and show stronger rhythmic contrast than speakers from Taiwan and Malaysia Mandarins. Finally, the results of word duration patterns also show that the frequency effect is not observed in the reading of quadrisyllabic idioms which is a relic of Classical Chinese commonly used in Modern Chinese. When Mandarin speakers speak Modern Chinese, the duration pattern is affected by lexical structures and frequency, among others. When they speak classical Chinese, frequency no longer has an effect on duration.

The investigation of duration patterns in this dissertation provided a detailed description of the whole image on word duration in Mandarin. This dissertation also provides the foundation for further research on duration pattern related super-segmental feature. A comprehensive understanding of duration pattern with linguistic and social factors is helpful to improve the quality of durational models used in speech technology. The result of duration pattern also can be used in language teaching and testing.

References

Allen, J., Hunnicutt, M. S., Klatt, D. H., Armstrong, R. C., and Pisoni, D. B. (1987). *From text to speech: The MITalk system*. Cambridge University Press.

Allen, P. A., Wallace, B., and Weber, T. A. (1995). Influence of case type, word frequency, and exposure duration on visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4):914.

Baayen, R. H. (2001). *Word frequency distributions*, volume 18. Springer Science & Business Media.

Balota, D. A. and Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, 24(1):89–106.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). Linear mixed-effects models using eigen and s4. R package version 1.0-5.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, pages 633–644.

Booth, A. D. (1967). A law of occurrences for words of low frequency. *Information and Control*, 10(4):386–393.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Browman, C. P. and Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, pages 341–376.

Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.

Browman, C. P. and Goldstein, L. (1995). Dynamics and articulatory phonology. *Mind as motion*, pages 175–193.

Browman, C. P. and Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3(01):219–252.

Brown, G. K. (2007). Making ethnic citizens: The politics and practice of education in Malaysia. *International Journal of Educational Development*, 27(3):318–330.

Byrd, D. (1994a). *Articulatory Timing in English Consonant Sequences. UCLA Ph. D.* PhD thesis, Dissertation Appeared as UCLA Working Papers in Phonetics 86..(1994b). Palatogram Reading as a Phonetic Skill: A Short Tutorial. Journal of the 1nternational Phonetic Association 24, 21-34.

Byrd, D. (1994b). Relations of sex and dialect to reduction. *Speech Communication*, 15(1-2):39–54.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24(2):209–244.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57(1):3–16.

Cao, J. (1986). The analysis of syllable with neutral tone in Mandarin 普通话轻声音节特征分析. *Applied Acoustics* 应用声学, 5(4):1–6.

Carlson, R. and Granström, B. (1973). Word accent, emphatic stress, and syntax in a synthesis-by-rule scheme for Swedish. *Speech Transmission Lab QPSR*, 1973:2–3.

Chang, Y. (2008). *The language marketing strategy of "Speak Mandarin Campaign" in Singapore*. PhD thesis, National Taitung University.

Chang, Y.-h. S. and Shih, C. (2015). Place contrast enhancement: the case of the alveolar and retroflex sibilant production in two dialects of Mandarin. *Journal of Phonetics*, 50:52–66.

Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken Chinese*. Harvard University Press.

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

Chen, F., Li, A., Wang, H., Wang, T., and Fang, Q. (2004). Acoustic analysis of friendly speech. In *ICASSP (1)*, pages 569–572.

Chen, M. Y. (1987). The syntax of Xiamen tone sandhi. *Phonology*, 4(01):109–149.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*, volume 92. Cambridge University Press.

Chen, S.-H., Lai, W.-H., and Wang, Y.-R. (2003). A new duration modeling approach for Mandarin speech. *IEEE Transactions on Speech and Audio Processing*, 11(4):308–320.

Chen, Y. (2006). Emphasis, syllable duration, and tonal realization in Standard Chinese. *Speech prosody 2006*, pages 382–85.

Chen, Y. and Xu, Y. (2006). Production of weak elements in speech–evidence from F0 patterns of neutral tone in Standard Chinese. *Phonetica*, 63(1):47–75.

Cheng, K.-S., Young, G. H., and Wong, K.-F. (1999). A study on word-based and integral-bit chinese text compression algorithms. *Journal of the Association for Information Science and Technology*, 50(3):218.

Chien, L.-F. (1997). Pat-tree-based keyword extraction for Chinese information retrieval. In *ACM SIGIR Forum*, volume 31, pages 50–58. ACM.

Cho, T., Jun, S.-A., and Ladefoged, P. (2000). Acoustic and aerodynamic correlates to Korean stops and fricatives. *UCLA Working Papers in Phonetics*, pages 109–148.

Clopper, C. G. and Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics*, 49:1–15.

Cooper, W. E., Paccia, J. M., and Lapointe, S. G. (1978). Hierarchical coding in speech timing. *Cognitive Psychology*, 10(2):154–177.

Cucchiarini, C., Strik, H., and Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, 107(2):989–999.

Dai, Y., Loh, T. E., and Khoo, C. S. (1999). A new statistical formula for Chinese text segmentation incorporating contextual information. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 82–89. ACM.

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

de Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics*, 32(4):493–516.

de Jong, K. and Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, 30(1):53–75.

Deng, D., Shi, F., and Lu, S. (2006). The contrast on tone between Putonghua and Taiwan Mandarin 普通话与台湾国语声调的对比分析. *Journal of Acoustics* 声学学报, 31(6):536.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*, volume 9. Walter de Gruyter.

Duanmu, S. (1994). Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology*, 11(01):1–24.

Duanmu, S. (1999). Stress and the development of disyllabic words in chinese. *Diachronica*, 16(1):1–35.

Duanmu, S. (2000). Prosody in Chinese 汉语的节奏. *Contemporary Linguistics* 当代语言学, 4:203–209.

Duanmu, S. (2007). *The phonology of Standard Chinese*. Oxford University Press.

Duanmu, S. (2009). *Syllable structure: The limits of variation*. Oxford University Press.

Ehrman, M. E., Leaver, B. L., and Oxford, R. L. (2003). A brief overview of individual differences in second language learning. *System*, 31(3):313–330.

Eisler, F. G. (1968). *Psycholinguistics: Experiments in spontaneous speech*. Academic Press.

Ellermeier, W. and Zimmer, K. (1997). Individual differences in susceptibility to the "irrelevant speech effect". *The Journal of the Acoustical Society of America*, 102(4):2191–2199.

Feng, L. (1985). Duration of initials, finals and tones in Beijing dialect 北京话语流中声韵调的时长. In *Working paper in experiment phonetics* 北京语音实验录. Peking University Press.

Feng, S. (1998). Discussion on natural foot in Chinese 论汉语的自然音步. *Chinese Linguistics* 中国语文, 1:40–47.

Feng, S. (2005). *Prosodic research on Chinese* 汉语韵律语法研究. Peking University Press.

Forster, K. I. and Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, 12(6):627–635.

Fosler-Lussier, E. and Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication*, 29(2):137–158.

Fromkin, V. A. (1973). The non-anomalous nature of anomalous utterances. *Speech errors as linguistic evidence*, pages 215–242.

Gaitenby, J. (1965). The elastic word. *Haskins Laboratories Status Report on Speech Research*, 2:3–1.

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of learning and motivation*, 9:133–177.

Gravano, A. (2009). *Turn-taking and affirmative cue words in task-oriented dialogue*. PhD thesis, Columbia University.

Gravano, A. and Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25(3):601–634.

Ha, L. Q., Sicilia-Garcia, E. I., Ming, J., and Smith, F. J. (2002). Extension of Zipf's law to words and phrases. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pages 1–6. Association for Computational Linguistics.

Han, M. S. and Weitzman, R. S. (1970). Acoustic features of Korean/p, t, k/,/p, t, k/and/ph, th, kh. *Phonetica*, 22(2):112–128.

He, K. and Li, D. (1987). *Three thousand most commonly used words in Modern Chinese* 现代汉语三千常用词表. Beijing: Beijing Shifan Daxue Chubanshe.

Heffernan, K. M. (2007). *Phonetic distinctiveness as a sociolinguistic variable*. PhD thesis, University of Toronto.

Ho, A. T. (1977). Intonation variation in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative. *Phonetica*, 34(6):446–457.

House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1):105–113.

Howes, D. H. and Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, 41(6):401.

Jacewicz, E., Fox, R. A., O'Neill, C., and Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, 21(02):233–256.

Jacewicz, E., Fox, R. A., and Salmons, J. (2007). Vowel duration in three american english dialects. *American Speech*, 82(4):367–385.

Jin, S. (1996). *A Acoustic Study of Sentence Stress in Mandarin Chinese.* PhD thesis, Ohio State University.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological Studies in Language*, 45:229–254.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, pages 286–319.

Kim, C.-W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to korean stops). *Word*, 21(3):339–359.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America*, 54(4):1102–1104.

Klatt, D. H. (1974). Review of speech synthesis. *Journal of the Acoustical Society of America*, 55.

Klatt, D. H. (1975a). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research*, 18(4):686–706.

Klatt, D. H. (1975b). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3(3):129–140.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221.

Kochanski, G., Shih, C., and Jing, H. (2003). Quantitative measurement of prosodic strength in Mandarin. *Speech Communication*, 41(4):625–645.

Kormos, J. and Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2):145–164.

Lai, W.-H. and Chen, S.-H. (2001). A novel syllable duration modeling approach for Mandarin speech. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 1, pages 93–96. IEEE.

Lee, S., Byrd, D., and Krivokapić, J. (2006). Functional data analysis of prosodic effects on articulatory timing. *The Journal of the Acoustical Society of America*, 119(3):1666–1671.

Lee, W.-S. and Zee, E. (2003). Standard Chinese(Beijing). *Journal of the International Phonetic Association*, 33(1):109–112.

Lehiste, I. (1975). Some factors affecting the duration of syllabic nuclei in english. In *Proceedings of the First Salzburg Conference on Linguistics*, pages 81–104.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of phonetics*.

Levelt, W. J. (1993). *Speaking: From intention to articulation*, volume 1. MIT press.

Li, A. and Wang, H. (2004). Friendly speech analysis and perception in Standard Chinese. In *INTERSPEECH*.

Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on information theory*, 38(6):1842–1845.

Li, Y. (2014). Hierarchical change of chinese. In *The 22nd Annual Conference of the International Association of Chinese Linguistics*. University of Maryland.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2):249–336.

Lin, M.-Y., Chiang, T.-H., and Su, K.-Y. (1993). A preliminary study on unknown word problem in Chinese word segmentation. In *Proceedings of Rocling VI*, pages 119–137.

Lin, T. (1985). Preliminary experiments on the exploration of the nature of Mandarin neutral tone 探讨北京话轻声性质的初步实验. In *Working paper in experiment phonetics* 北京语音实验录, pages 1–26. Peking University Press, Beijing.

Lin, X. (2016). Implicit EMI policies: Taking Malaysia, Singapore and Thailand as examples 英语作为教学媒介语隐性政策分析: 以马来西亚、新加坡和泰国为例. *Language police* 语言战略研究, 1(1).

Lin, Y.-H. (2007). *The Sounds of Chinese*, volume 1. Cambridge University Press.

Lindblom, B. and Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Auditory Analysis and Speech Perception (London, 1975)*, pages 387–96.

Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3):384–422.

Lisker, L. and Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and speech*, 10(1):1–28.

Loukina, A. and Kochanski, G. (2010). Patterns of durational variation in British dialects. In *PAC colloquium*.

Lu, B. and Duanmu, S. (1991). A case study of the relation between rhythm and syntax in Chinese. In *The Third North America Conference on Chinese Linguistics*.

Lucero, J. C., Munhall, K. G., Gracco, V. L., and Ramsay, J. O. (1997). On the registration of time and the patterning of speech movements. *Journal of Speech, Language, and Hearing Research*, 40(5):1111–1117.

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication Theory*, 84:486–502.

McFarland, D. H. and Smith, A. (1992). Effects of vocal task and respiratory phase on prephonatory chest wall movements. *Journal of Speech, Language, and Hearing Research*, 35(5):971–982.

Meyer, D. E. and Gordon, P. C. (1985). Speech production: Motor programming of phonetic features. *Journal of Memory and Language*, 24(1):3–26.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. *Perspectives on the study of speech*, pages 39–74.

Miller, J. L., Aibel, I. L., and Green, K. (1984). On the nature of rate-dependent processing during phonetic perception. *Perception & psychophysics*, 35(1):5–15.

Monsell, S., Doyle, M., and Haggard, P. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, 118(1):43.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, 46(2):135–147.

Newman, J. (1988). Singapore's Speak Mandarin Campaign. *Journal of Multilingual & Multicultural Development*, 9(5):437–448.

Norman, J. (1988). *Chinese*. Cambridge University Press.

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39(1):59–67.

Oldfield, R. C. and Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4):273–281.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The journal of the Acoustical Society of America*, 54(5):1235–1247.

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.

Pennebaker, J. W. and King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296.

Peterson, G. E. and Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, 32(6):693–703.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, 118(4):2561–2569.

Pulleyblank, E. G. (2010). *Outline of classical Chinese grammar*. UBC Press.

Qi, S. and Zhang, J. (1982). The analysis of consonant duration in Mandarin 汉语普通话辅音音长分析. *Journal of Acoustics* 声学学报, 1.

Ramsey, S. R. (1987). *The languages of China*. Princeton University Press.

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

Ranganath, R., Jurafsky, D., and McFarland, D. A. (2013). Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates. *Computer Speech & Language*, 27(1):89–115.

Rayner, K. and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3):191–201.

Ren, H. (1985). Linguistically conditioned duration rules in a timing model for Chinese. *UCLA Working Papers in Phonetics*, 62(1).

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes*, 14(4):423–441.

Riley, M. D. (1991). Tree-based modelling for speech synthesis. In *The ESCA Workshop on Speech Synthesis*.

Riney, T. J., Takagi, N., Ota, K., and Uchida, Y. (2007). The intermediate degree of vot in japanese initial voiceless stops. *Journal of Phonetics*, 35(3):439–443.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. *Linguistic Controversies*, pages 73–79.

Rochet, B. L. and Yanmei, F. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, 19(4):105–106.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, pages 295–342.

Shen, X. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech*, 36(4):415–433.

Shi, Y. and Li, C. N. (2002). The establishment of the classifier system and the grammaticalization of the morphosyntactic particle 'de' in Chinese. *Language sciences*, 24(1):1–15.

Shih, C. (1986). *The prosodic domain of tone sandhi in Chinese*. PhD thesis, University of California, San Diego.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3:83–109.

Shih, C. (2005). Understanding phonology by phonetic implementation. In *INTERSPEECH*, pages 2469–2472.

Shih, C. and Ao, B. (1997). Duration study for the bell laboratories Mandarin text-to-speech system. In *Progress in speech synthesis*, pages 383–399. Springer.

Shih, C. and Lu, H.-Y. D. (2015). Effects of talker-to-listener distance on tone. *Journal of Phonetics*, 51:6–35.

Shriberg, E. (2001). Toerrrr'is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31(1):153–164.

Singmann, H., Bolker, B., and Westfall, J. (2015). Afex: analysis of factorial experiments. R package version 0.13–145.

Solomon, R. L. and Howes, D. H. (1951). Word frequency, personal values, and visual duration thresholds. *Psychological Review*, 58(4):256.

Sproat, R., Gale, W., Shih, C., and Chang, N. (1996). A stochastic finite-state word-segmentation algorithm for Chinese. *Computational linguistics*, 22(3):377–404.

Sproat, R. and Shih, C. (2002). Corpus-based methods in Chinese morphology. *Tutorial at the 19th COLING*.

Sui, Y. (2013). *Phonological and phonetic evidence for trochaic metrical structure in Standard Chinese*. PhD thesis, University of Pennsylvania.

Sun, C. (2006). *Chinese: A linguistic introduction*. Cambridge University Press.

Tsay, J. and Myers, J. (1996). Taiwanese tone sandhi as allomorph selection. In *Annual Meeting of the Berkeley Linguistics Society*, volume 22, pages 395–405.

Tseng, C.-C. (2004). Prosodic properties of intonation in two major varieties of Mandarin Chinese: Mainland China vs. Taiwan. In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*.

Tuller, B. and Fowler, C. A. (1980). Some articulatory correlates of perceptual isochrony. *Perception & Psychophysics*, 27(4):277–283.

Tuller, B. and Fowler, C. A. (1981). The contribution of amplitude to the perception of isochrony. *Haskins Laboratories Status Report on Speech Research*, pages 245–250.

Umeda, N. (1975). Vowel duration in american english. *The Journal of the Acoustical Society of America*, 58(2):434–445.

Umeda, N. (1977). Consonant duration in American English. *The Journal of the Acoustical Society of America*, 61(3):846–858.

van Santen, J. P. (1992). Contextual effects on vowel duration. *Speech Communication*, 11(6):513–546.

van Santen, J. P. (1994). Assignment of segmental duration in text-to-speech synthesis. *Computer Speech & Language*, 8(2):95–128.

van Santen, J. P. and Olive, J. P. (1990). The analysis of contextual effects on segmental duration. *Computer Speech & Language*, 4(4):359–390.

van Santen, J. P. and Shih, C. (2000). Suprasegmental and segmental timing models in Mandarin Chinese and American English. *The Journal of the Acoustical Society of America*, 107(2):1012–1026.

Vanderplank, R. (1993). Pacing and spacing as predictors of difficulty in speaking and understanding English. *ELT Journal*, 47(2):117–125.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., and Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and speech*, 40(1):47–62.

Wagner, P. S. and Dellwo, V. (2004). Introducing yard (yet another rhythm determination) and re-introducing isochrony to rhythm research. In *Speech Prosody 2004, International Conference*.

White, L. and Mattys, S. L. (2007). Rhythmic typology and variation in first and second languages. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 282:237.

Whiteside, S. P. (1996). Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. *Journal of the International Phonetic Association*, 26(01):23–40.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91(3):1707–1717.

Winkworth, A. L., Davis, P. J., Adams, R. D., and Ellis, E. (1995). Breathing patterns during spontaneous speech. *Journal of Speech, Language, and Hearing Research*, 38(1):124–144.

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition*, 7(6):411–419.

Wu, C.-h. (2008). Filled pauses in L2 Chinese: A comparison of native and non-native speakers. In *Proceedings of the 20th North American Conference on Chinese Linguistics*, volume 1, page 213.

Wu, C.-h. (2011). *The evaluation of second language fluency and foreign accent*. PhD thesis, University of Illinois at Urbana-Champaign.

Wu, D. and Shih, C. (2015). Dialectal duration variations reveal historical sound change. *The Journal of the Acoustical Society of America*, 137(4):2415–2415.

Wu, Z. and Tseng, G. (1993). Chinese text segmentation for text retrieval: Achievements and problems. *Journal of the American Society for Information Science*, 44(9):532.

Xu, Y. (1993). *Contextual tonal variation in Mandarin Chinese*. PhD thesis, University of Connecticut.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27(1):55–105.

Xu, Y. and Wang, M. (2009). Organizing syllables into groups—evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics*, 37(4):502–520.

Yip, M. J. (1980). *The tonal phonology of Chinese*. PhD thesis, Massachusetts Institute of Technology.

Yuan, J. and Liberman, M. (2008). Speaker identification on the scotus corpus. *Journal of the Acoustical Society of America*, 123(5):3878.

Yuan, J. and Liberman, M. (2009). Investigating /l/ variation in English through forced alignment. In *Interspeech*, volume 2009.

Yuan, J., Liberman, M., and Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *INTERSPEECH*.

Zhan, W., Chang, B., Duan, H., and Zhang, H. (2006). Recent developments in Chinese corpus research. In *The 13th NIJL International Symposium, Language Corpora: Their Compilcation and Application. Tokyo, Japan*, pages 3–6.

Zhang, H. (2014). Nlpir: Natural language processing and information retrieval sharing platform http://www.nlpir.org.

Zhang, S. (1984). On some problems with Chinese phonetic writing orthography 谈汉语拼音正词法的几个问题. *Wenzi Gaige* 文字改革, 1.

Zhang, Z. (1988). *Tone and Tone Sandhi in Chinese. OSU Ph. D.* PhD thesis, Ohio State University.

Zipf, G. K. (1932). *Selective studies and the principle of relative frequency in language.* Harvard University Press.

Appendix A: Materials (stories) in Study 1

Story 1

大山里有一片森林，森林里有一头驴子。这头驴子喜欢披着狮子的皮到处游荡，叫喊着吓唬森林里的小动物。有一次，驴子在路上看见了狼，便想吓唬一下它。狼正巧以前听到过驴子的叫声，所以狼不害怕，便对驴子说:“如果我听不出你的声音，恐怕我也会害怕一头驴子。”

Story 2

一只狐狸走在果园里，她非常地渴。正在这个时候，她看见葡萄架上挂着一串串晶莹的葡萄。狐狸开始不停地流口水，很想摘下来吃，但是又摘不到。狐狸看了几眼，无奈地走了，她边走边安慰自己说:“葡萄还没有成熟，肯定是酸的。我不是一只吃不到葡萄的狐狸，我只是不想吃酸葡萄。”

Story 3

一个很冷的冬天，农夫在雪地里发现了一条冻僵的蛇。农夫可怜它，就解开自己的衣服，把蛇放在了怀里，用身体去温暖它。蛇很快就苏醒了，它立刻露出了残忍的本性，狠狠地咬了农夫。中了毒的农夫在临死之前说:“我竟然去可怜蛇，就应该遭到这种报应啊。”

Story 4

松鼠在四处找吃的，它看见一棵大树的树洞里面有猎人藏好的香肠，马上钻了进去。香肠好吃得不得了，松鼠吃的肚子涨得鼓鼓的。树洞太小了，松鼠费了很大力气，结果却怎么也钻不出来，于是只能坐在树洞里叹气。一只兔子恰巧经过树洞，听到松鼠的呻吟，便跑过去问它:“松鼠，你怎么了?”松鼠把事情的经过讲了一遍，兔子了解了缘由后，便对松鼠

说:“哈哈，偷吃香肠的笨松鼠，你就乖乖待在里面吧，等身材恢复了，就容易出来了。”松鼠听了狠狠地骂道:“死兔子 !”

Story 5

村子旁边有一个池塘，池塘里有一条叫做瘦小姐的可爱的小鱼，她总是开心地生活着，追赶着水面上的蝴蝶和蜻蜓。池塘里还有一条叫做胖先生的凶狠的大鱼，他来自一个大家族，为了显示自己的力量，大鱼每天总要袭击瘦小姐。年轻的渔夫每天都来池塘钓鱼，总是没有什么收获。有一天，渔夫决定买一张渔网来捕鱼。天亮时，他悄悄地来到池塘，然后撒下了渔网。小鱼和大鱼从来都没有见过渔网，所以当渔夫收网时，他们都被网住了。瘦小姐挣扎着从网眼里逃了出来，胖先生这个时候在想：如果我像小鱼一样，渔网就不会困住我了，我真是一个倒霉的胖先生。

Story 6

北风和太阳正在争论谁比较有本领。他们正好看到有个人走过，那个人穿着一件斗篷。于是他们约定，谁可以让那个人脱掉斗篷，就算谁的本领大。北风第一个上场，他开始拼命地吹。他吹得越厉害，那个人就把斗篷裹得越紧。骄傲的北风没有办法，只好放弃了，躲在一旁，挑衅地看着太阳。太阳第二个上场，他用自己的光和热晒了一下，那个人就脱掉了斗篷。太阳战胜了北风。

Appendix B: Sentences in Study 2

No.	Word	Sentence
01	小行星:	记者正在报道小行星撞地球的新闻。
02	旅行家:	明代旅行家徐霞客曾经到过徽州。
03	紫水晶:	她左手带着一枚紫水晶戒指。
04	山水画:	李可染是中国现代山水画大师。
05	古文明:	黄河流域是华夏古文明的发源地。
06	天文学:	古希腊在天文学上处于领先地位。
07	红辣椒:	重庆是红辣椒的重要产地。
08	酸辣汤:	她在自己的酸辣汤里面加了一些盐。
09	慢镜头:	比赛的慢镜头显示运动员抢跑了。
10	眼镜蛇:	这种生长在亚洲的眼镜蛇有剧毒。
11	电冰箱:	厨房里的电冰箱坏掉了。
12	破冰船:	探险队的破冰船到达了南极。
13	土皇帝:	山西的土皇帝阎锡山参加了抗日。
14	保皇党:	康有为的保皇党思想十分顽固。
15	大前天:	新闻稿是大前天早上发出去的。
16	史前人:	考古学家发现了史前人的遗迹。
17	肝功能:	患者出现了肝功能衰竭的症状。
18	成功率:	提高投资成功率会吸引更多客户。
19	厚脸皮:	记者是一项需要厚脸皮的职业。
20	黄脸婆:	县长家的黄脸婆待人十分凶悍。

21 咸豆浆: 上海人早餐喜欢咸豆浆配油条。

22 红豆冰: 街边的红豆冰是夏天的一等美味。

23 年产量: 粮食年产量打破了历史记录。

24 生产线: 手机生产线的工人没有假期。

25 大客车: 他有大客车的驾驶执照。

26 会客室: 经理在会客室偷偷抽烟。

27 纸尿裤: 婴儿穿的纸尿裤正在降价。

28 糖尿病: 胰岛素是治疗糖尿病的常用药。

29 市中心: 购物广场坐落在市中心的繁华地段。

30 眼中钉: 继母一直想把她这颗眼中钉铲除掉。

31 农作物: 水污染导致大量农作物死亡。

32 工作日: 公司要求工作日必须穿正式服装。

33 小心眼: 县长本来就是一个小心眼的人。

34 向心力: 这种弯道设计可使向心力减弱。

Appendix C: Sentences in Study 3

1 青红皂白: 老板不分青红皂白就骂了员工一顿。

2 衣食住行: 秘书替他做好了衣食住行的安排。

3 吃喝嫖赌: 他有富家少爷吃喝嫖赌的种种恶习。

4 琴棋书画: 她是一个琴棋书画样样精通的天才。

5 亭台楼阁: 红楼梦对亭台楼阁做了详细描写。

6 妖魔鬼怪: 聊斋里面到处都是妖魔鬼怪的描写。

7 老弱病残: 战士们准备杀掉那些老弱病残的马匹。

8 柴米油盐: 她老公从来不关心柴米油盐的琐事。

9 兄弟姐妹: 她家里兄弟姐妹都在学校教书。

10 喜怒哀乐: 她对自己的喜怒哀乐从来不掩饰。

11 叮呤当啷: 他在车库里叮呤当啷的修理着割草机。

12 噼里啪啦: 一群鸭子噼里啪啦地跳进了河里。

13 叽哩呱啦: 一群外国人在机场叽哩呱啦地讲着什么。

14 稀里哗啦: 她躲在家里稀里哗啦地哭了一场。

15 滴里嘟噜: 他滴里嘟噜的解释了一大堆。

16 当家做主: 买房子后她终于有了当家做主的感觉。

17 门当户对: 中国人讲究男女双方门当户对的婚姻。

18 对酒当歌: 他羡慕那种能够对酒当歌的人生境界。

19 旗鼓相当: 她在公司里没有旗鼓相当的竞争对手。

20 上行下效: 公司里上行下效的风气十分不好。

21 天上人间: 上帝创造了天上人间的一切事物。

22 成千上万: 每年有成千上万的游客到芝加哥旅游。

23 箭在弦上: 谈判进入了箭在弦上的关键时刻。

24 尔虞我诈: 他十分厌恶这个尔虞我诈的社会。

25 莞尔一笑: 他喜欢那个女生莞尔一笑的样子。

26 温文尔雅: 她欣赏那个男人温文尔雅的气度。

27 新婚燕尔: 这对新婚燕尔的小夫妻准备买房。

28 来龙去脉: 事情的来龙去脉终于都说清楚了。

29 南来北往: 火车站里到处都是南来北往的旅客。

30 心血来潮: 她心血来潮决定开始学习钢琴。

31 古往今来: 他是古往今来最伟大的历史学家。

32 地广人稀: 明尼苏达州是典型的地广人稀的地区。

33 落地生根: 第一代移民都经历过落地生根的艰难。

34 天经地义: 这世界上没有什么事情是天经地义的。

35 死心塌地: 她这辈子死心塌地的爱着自己的丈夫。

36 国破家亡: 他经历了国破家亡的时代剧变。

37 卖国求荣: 他们坚决不当卖国求荣的叛徒。

38 皇亲国戚: 他像皇亲国戚一样受到了特殊待遇。

39 闭关锁国: 清朝的闭关锁国导致了鸦片战争。

40 家喻户晓: 蝙蝠侠是家喻户晓的大英雄。

41 万家灯火: 城市里的万家灯火让人觉得温暖。

42 万贯家财: 他将万贯家财全都挥霍掉了。

43 小户人家: 她是小户人家出身的女孩子。

44 事半功倍: 这种设计可以达到事半功倍的效果。

45 无事生非: 他绝不是那种无事生非的孩子。

46 置身事外: 你想置身事外已经不可能了。

47 见机行事: 上司很喜欢那种懂得见机行事的下属。

48 和睦相处: 她不是那种能与其他员工和睦相处的人。

49 风和日丽: 五月是个风和日丽的好时节。

50 一团和气: 她很喜欢一团和气的工作环境。

51 心平气和: 他从来不会和员工心平气和地讲话。

52 何足挂齿: 这种何足挂齿的小事不用放在心上。

53 有何面目: 一个杀人犯有何面目活在世上。

54 曾几何时: 那口老井曾几何时是村中唯一的水源。

55 无可奈何: 数学老师无可奈何地点了点头。

56 知书达理: 男人都希望能娶到知书达理的妻子。

57 明知故问: 他明知故问地去试探朋友的诚意。

58 迷途知返: 很多迷途知返的青少年来当志愿者。

59 未卜先知: 小男孩有未卜先知的超能力。

60 之乎者也: 他是一个满嘴之乎者也的老先生。

61 言之有理: 他能给出一个言之有理的说明。

62 后起之秀: 教授对学术界的后起之秀十分关注。

63 堂而皇之: 色情广告堂而皇之的出现在网页上。

64 儿女情长: 他的人生有太多儿女情长的牵绊。

65 生儿育女: 婆婆总认为生儿育女是女人的责任。

66 视为儿戏: 这就是将法律视为儿戏的后果。

67 而立之年: 他到了而立之年还一事无成。

68 铤而走险: 新员工决定铤而走险去赌一次。

69 从一而终: 她在婚姻上坚守从一而终的信念。

70 以貌取人: 面试官常常犯以貌取人的错误。

71 难以置信: 这是一个让人难以置信的实验结果。

72 拭目以待: 消费者拭目以待公司的新产品上市。

73 忘乎所以: 他很少流露出那种忘乎所以的轻狂。

74 于心无愧: 他想当一个于心无愧的律师。

75 安于现状: 她不是一个安于现状的家庭主妇。

76 青出于蓝: 他知道不是每个学生都有青出于蓝的本事。

77 余音绕梁: 这个时代再也找不出余音绕梁的经典之作。

78 茶余饭后: 他的婚外情成了同事们茶余饭后的新话题。

79 心有余悸: 她摆脱不掉车祸后那种心有余悸的感觉。

80 游刃有余: 她不喜欢那种做起来游刃有余的工作。

81 与众不同: 他从小就是一个与众不同的小孩。

82 无与伦比: 爱情给了他无与伦比的学习动力。

83 生死与共: 他们两个是战场上生死与共的兄弟。

84 雨过天晴: 她相信总有雨过天晴的那一天。

85 血雨腥风: 战争中的血雨腥风使她变得更加成熟。

86 风调雨顺: 村民们希望能有一个风调雨顺的收获季节。

87 倾盆大雨: 一场倾盆大雨导致城市交通瘫痪。

88 错上加错: 妈妈不会原谅这种错上加错的行为。

89 话里有话: 女朋友话里有话地暗示着他求婚。

90 人无完人: 女生找男朋友首先要明白人无完人的道理。

91 天外有天: 有些人根本就不懂天外有天的道理。

92 微乎其微: 油价上涨对汽车价格的影响是微乎其微的。

93 年复一年: 他年复一年地盼望着公司给他升职加薪。

94 痛定思痛: 他在痛定思痛之后决心重新振作起来。

95 为所欲为: 老师不喜欢那种在课堂上为所欲为的学生。

96 贼喊捉贼: 她和房东玩起了贼喊捉贼的把戏。

97 精益求精: 设计上的精益求精未必能得到良好的结果。

98 百依百顺: 她不可能成为一个百依百顺的妻子。

99 合情合理: 这种合情合理的要求应该得到满足。

100 绘声绘色: 她绘声绘色的给小朋友讲着故事。

101 有勇有谋: 排长喜欢那种有勇有谋的新兵。

102 随时随地: 家长应该随时随地注意小朋友的安全。

103 无影无踪: 他就这么无影无踪地从世间消失了。

104 任劳任怨: 她不想当一个任劳任怨的家庭主妇。

105 多才多艺: 父母想把她培养成多才多艺的女孩。

106 没完没了: 她妈妈没完没了地唠叨着结婚的事情。

107 愣头愣脑: 新生愣头愣脑地进了化学实验室。

108 不懂装懂: 她没有意识到不懂装懂会害了自己。

109 爱理不理: 她总是对客户摆出一副爱理不理的样子。

110 心服口服: 他终于获得了让人心服口服的实验结果。

111 知法犯法: 法官吸毒就是知法犯法的典型案例。

112 将计就计: 我们不如将计就计送个人情给她。

113 讨价还价: 留学生终于学会了讨价还价的技能。

114 倚老卖老: 他喜欢倚老卖老地教育新同事。

115 就事论事: 我们需要就事论事地考虑这个问题。

116 自然而然: 事情就那么自然而然的过去了。

117 在商言商: 他是那种在商言商的生意人。

118 得意洋洋: 他得意洋洋地宣布了比赛结果。

119 洋洋得意: 他讨厌冠军队那种洋洋得意的样子。

120 漫漫长夜: 一个笑容能让漫漫长夜变得温暖。

121 长夜漫漫: 失眠会让人有种长夜漫漫的感觉。

122 天网恢恢: 他知道天网恢恢难以逃脱法律制裁。

123 恢恢天网: 警察早已布下了恢恢天网抓捕他。

124 含情脉脉: 女朋友喜欢含情脉脉地望着他。

125 脉脉含情: 张小姐有一双脉脉含情的眼睛。

126 大名鼎鼎: 他就是大名鼎鼎的战斗英雄。

127 鼎鼎大名: 他的鼎鼎大名在小镇里无人不晓。

128 文质彬彬: 她一直想找一个文质彬彬的男朋友。

129 彬彬有礼: 他给面试官留下了彬彬有礼的印象。

130 两手空空: 你不能两手空空地去别人家拜访。

131 空空如也: 她一个人生活在一座空空如也的房子里面。

132 风度翩翩: 他就是一个风度翩翩的绅士。

133 翩翩起舞: 她在台上翩翩起舞的时候特别美。

134 情意绵绵: 男朋友给她写了一封情意绵绵的长信。

135 绵绵不绝: 他的作品表达了绵绵不绝的乡愁。

136 无所事事: 他辞职后过着无所事事的日子。

137 事事躬亲: 经理不是那种事事躬亲的人。

138 安安静静: 她不是那种安安静静的学生。

139 结结巴巴: 犯人结结巴巴地描述了犯罪过程。

140 婆婆妈妈: 她最讨厌那种婆婆妈妈的男人。

141 大大方方: 她大大方方地讲完了自己的研究报告。

142 兢兢业业: 老板喜欢那种兢兢业业的员工。

143 匆匆忙忙: 她晚上匆匆忙忙地离开了公司。

144 沸沸扬扬: 他的升职引起了大家沸沸扬扬的议论。

145 堂堂正正: 父母从小就教育孩子要堂堂正正做人。

146 日日夜夜: 她日日夜夜照顾着生病的女儿。

147 口口声声: 总经理口口声声说一定会帮忙。

148 生生世世: 世界上根本就没有生生世世的爱情。

149 密密麻麻: 他做了一张密密麻麻的考前计划表。

150 恍恍惚惚: 她从梦中恍恍惚惚地醒过来。

151 吞吞吐吐: 老员工吞吞吐吐不愿讲出实情。

152 详详细细: 他写了一份详详细细的事故报告

153 自欺欺人: 他讨厌这种自欺欺人的行为。

154 一衣带水: 中国和朝鲜是一衣带水的邻邦。

Appendix D: Idioms

1. Repeated Characters

Group1

1) 错上加错 (*cuo shang jia cuo*): heaped error on error;

2) 话里有话 (*hua li you hua*): the words mean more than they say;

3) 人无完人 (*ren wu wan ren*): no one is perfect;

4) 天外有天 (*tian wai you*): no limit in the universe;

5) 微乎其微 (*wei hu qi wei*): a million to one;

6) 年复一年 (*nian fu yi nian*): year after year;

7) 痛定思痛 (*tong ding si tong*): draw a lesson from a bitter experience;

8) 为所欲为 (*wei suo yu wei*): act wilfully;

9) 贼喊捉贼 (*zei han zhuo zei*): accuse sb. of the theft and try to sneak away oneself;

10) 精益求精 (*jing yi qiu jing*): constantly strive for perfection;

Group 2

1) 百依百顺 (*bai yi bai shun*): obey in every way;

2) 合情合理 (*he qing he li*): be perfectly logical and reasonable;

3) 绘声绘色 (*hui sheng hui se*): a lively description;

4) 有勇有谋 (*you yong you mou*): intelligent and courageous;

5) 随时随地 (*sui shi sui di*): whenever and wherever possible;

6) 无影无踪 (*wu ying wu zong*): disappear completely;

7) 任劳任怨 (*ren lao ren yuan*): bear hardship without complaint;

8) 多才多艺 (*duo cai duo yi*): gifted in many ways;

9) 没完没了 (*mei wan mei liao*): endless;

10) 愣头愣脑 (*leng tou leng nao*): rash; impetuous;

Group 3

1) 不懂装懂 (*bu dong zhuang dong*): pretend to know when one does not know;

2) 爱理不理 (*ai li bu li*): attend to sb. halfheartedly;

3) 心服口服 (*xin fu kou fu*): be sincerely convinced;

4) 知法犯法 (*zhi fa fan fa*): know the law but break it;

5) 将计就计 (*jiang ji jiu ji*): turn sb.'s trick to one's own use;

6) 讨价还价 (*tao jia huan jia*): bargain with sb. for a supply of sth.;

7) 倚老卖老 (*yi lao mai lao*): pride oneself on being a veteran;

8) 就事论事 (*jiu shi lun shi*): take the matter on its merits;

9) 自然而然 (*zi ran er ran*): come very naturally;

10) 在商言商 (*zai shang yan shang*): business is business;

Group 4

1) 天网恢恢 (*tian wang hui hui*): justice has long arms;

2) 含情脉脉 (*han qing mo mo*): exuding tenderness and love through eyes;

3) 大名鼎鼎 (*da ming ding ding*): be very famous;

4) 文质彬彬 (*wen zhi bin bin*): the ornamental and the combined plain properties;

5) 两手空空 (*liang shou kong kong*): be left with nothing whatsoever;

6) 风度翩翩 (*feng du pian pian*): be dapper in appearance;

7) 情意绵绵 (*qing yi mian mian*): inextricable feeling between the lovers;

8) 无所事事 (*wu suo shi shi*): have nothing to do;

9) 得意洋洋 (*de yi yang yang*): feel oneself highly flattered;

10) 长夜漫慢 (*chang ye man man*): long night;

Group 5

1) 恢恢天网 (*hui hui tian wang*): justice has long arms;

2) 脉脉含情 (*mo mo han qing*): exuding tenderness and love through eyes;

3) 鼎鼎大名 (*ding ding da ming*): be very famous;

4) 彬彬有理 (*bin bin you li*): behaving in a refined and civil manner;

5) 空空如也 (*kong kong ru ye*): absolutely empty;

6) 翩翩起舞 (*pian pian qi wu*): rise and dance in a happy mood;

7) 绵绵不绝 (*mian mian bu jue*): remain unbroken;

8) 事事躬亲 (*shi shi gong qin*): take care of every single thing personally;

9) 洋洋得意 (*yang yang de yi*): feel oneself highly flattered;

10) 漫漫长夜 (*man man chang ye*): long night;

Group 6

1) 安安静静 (*an an jing jing*): peaceful and serene;

2) 结结巴巴 (*jie jie ba ba*): hesitating in speaking;

3) 婆婆妈妈 (*po po ma ma*): womanishly fussy; garrulous;

4) 大大方方 (*da da fang fang*): very natural and poised;

5) 兢兢业业 (*jing jing ye ye*): bustle up; be jammed for time;

6) 匆匆忙忙 (*cong cong mang mang*): be jammed for time;

7) 沸沸扬扬 (*fei fei yang yang*): raise a babel of criticism of ...;

8) 堂堂正正 (*tang tang zheng zheng*): dignified and imposing;

9) 日日夜夜 (*ri ri ye ye*): day and night;

10) 口口声声 (*kou kou sheng sheng*): keep on proclaiming [saying];

2. Identical Characters on Four Syllable positions

当 *dang*	当家做主 *dang jia zuo zhu*	be the master of one's own affairs
	门当户对 *men dang hu dui*	marry into a proper family
	对酒当歌 *dui jiu dang ge*	sing while drinking
	旗鼓相当 *qi gu xiang dang*	be equal in match or contest of strength
上 *shang*	上行下效 *shang xing xia xiao*	follow the example of the superiors
	天上人间 *tian shang ren jian*	immeasurably vast difference
	成千上万 *cheng qian shang wan*	tens of thousands of
	箭在弦上 *jian zai xian shang*	like an arrow on the bowstring
尔 *er*	尔虞我诈 *er yu wo zha*	deceive and blackmail each other
	莞尔一笑 *wan er yi xiao*	give a soft smile
	温文尔雅 *wen wen er ya*	cultured in manners
	新婚燕尔 *xin hun yan er*	newly married
来 *lai*	来龙去脉 *lai long qu mai*	origin and development
	南来北往 *nan lai bei wang*	coming and going in great number
	心血来潮 *xin xue lai chao*	(have) a sudden inspiration
	古往来来 *gu wang jin lai*	from ancient to modern times
地 *di*	地广人稀 *di guang ren xi*	thinly populated land
	落地生根 *luo di sheng gen*	bryophyllum
	天经地义 *tian jing di yi*	an unquestionable moral truth
	死心塌地 *si xin ta di*	be hell-bent on

国 *guo*	国破家亡	*guo po jia wang*	the country ruined and the people starving
	卖国求荣	*mai guo qiu rong*	betray one's country to obtain a promotion
	皇亲国戚	*huang qin guo qi*	members of the imperial house
	闭关锁国	*bi guan suo guo*	cut off one's country from the outside world
家 *jia*	家喻户晓	*jia yu hu xiao*	be known to every household
	万家灯火	*wan jia deng huo*	a myriad twinkling lights (of a city)
	万贯家财	*wan guan jia cai*	vast wealth
	小户人家	*xiao hu ren jia*	impoverished family
事 *shi*	事半功倍	*shi ban gong bei*	get twofold results with half the effort
	无事生非	*wu shi sheng fei*	make trouble out of nothing
	置身事外	*zhi shen shi wai*	detach oneself from
	见机行事	*jian ji xing shi*	act according to circumstances
和 *he*	和睦相处	*he mu xiang chu*	live amicably with sb.
	风和日丽	*feng he ri li*	a peaceful, suuny day
	一团和气	*yi tuan he qi*	a prevailing mood of harmony
	心平气和	*xin ping qi he*	be in a calm mood
知 *zhi*	知书达理	*zhi shu da li*	educated and reasonable
	明知故问	*ming zhi gu wen*	ask while knowing the answer
	迷途知返	*mi tu zhi fan*	realize one's errors and mend one's ways
	未卜先知	*wei bu xian zhi*	know without seeking divination
余 *yu*	余音绕梁	*yu yin rao liang*	the tune lingered in the room
	茶余饭后	*cha yu fan hou*	at one's leisure
	心有余悸	*xin you yu ji*	be still in a state of shock
	游刃有余	*you ren you yu*	handle a butcher's cleaver skillfully
雨 *yu*	雨过天晴	*yu guo tian qing*	sunshine followed the rain
	血 雨腥风	*xue yu xing feng*	a foul wind and a rain of blood
	风调雨顺	*feng tiao yu shun*	favourable climatic weathers
	倾盆大雨	*qing pen da yu*	a heavy downpour