PRODUCTION AND PERCEPTION OF KAIFENG MANDARIN TONES

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: The 'bridge scene' in the famous painting *清明上河图* 'along the river during the Qingming festival' by Zeduan Zhang, which depicts the people and the landscape of Kaifeng in the Northern Song dynasty (960–1127). The painting had been redrawn by Hai Ying and the cover picture was taken by Lei Wang with the former's kind permission.

ISBN: 978-94-6093-311-0
NUR 616

PRODUCTION AND PERCEPTION OF KAIFENG MANDARIN TONES

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op

woensdag 19 december 2018

om 16.30 uur precies

door

Lei Wang

geboren op 6 november 1987

te Kaifeng, China

Promotor: Prof. dr. C.H.M. Gussenhoven

Copromotor: Dr. M.A.M. van de Ven

Manuscriptcommissie:

Prof. dr. J.P.M. Fikkert (voorzitter)

Dr. Y. Chen, Universiteit Leiden

Dr. P. Hallé, Université Paris 3 – Sorbonne Nouvelle, Frankrijk

Prof. dr. V.J.J.P. van Heuven, Pannon Egyetem, Hongarije

Prof. dr. J. Zhang, University of Kansas, Verenigde Staten

ACKNOWLEDGEMENTS

This dissertation marks the end of my two-year journey at Radboud Universiteit (RU). Words fail me when I attempt to extend my heartfelt gratitude to my supervisor, Carlos Gussenhoven, without whom this journey could not have started in the first place.

Carlos is always ready for feedback amidst his busy schedule. His theoretical input has inspired me from the very beginning. He effortlessly interpreted my tone data in our first weekly meeting (Wednesday, 11:00), which eventually led to the submission of my first journal article (Chapter 3) about three months later. "Well, let's go to *JIPA*, *Journal of Phonetics* or *Phonetica*…" I had never thought I could make it into professional A-rated journals in my entire research career before I became a guest researcher at RU in 2016/2017. Not only is Carlos a perfect academic figure, but he also remains approachable and considerate. I feel cool, calm and collected, each time when greeted with a warm friendly smile; I shall never forget our pleasant conversations and relaxing get-togethers; I also still remember the food he purchased for me when I first arrived in Nijmegen on the evening of 18-09-2016, Muesli, croissant, instant cupnoodles, egg, apple, milk, juice and canned meat – that is virtually everything one can ever imagine.

I am grateful to my co-supervisor, Marco van de Ven, especially for his detailed and step-by-step assistance in linear mixed effects modelling. Our cheerful conversations nicely occupied the otherwise boring hours it took for the complex models to converge. I would like to thank the members of the manuscript committee, Paula Fikkert, Yiya Chen, Pierre Hallé, Vincent van Heuven and Jie Zhang for their careful reading of the manuscript and for their constructive suggestions. I have also benefited enormously from the comments and feedback from the *JIPA* editors and reviewers, which significantly improved the readability of Chapters 2 and 3.

Special thanks are due to my MA and PhD supervisor in Tongji University, Jie Liang for having developed my research interest in

Chinese dialects and for her generous support in my application of the exchange program to RU in 2016/2017. The program was financially supported by the Chinese Scholarship Council, which is gratefully acknowledged. Special thanks also go to Qi Zhang and Bijun Ling for serving as my guarantors in the program, and to Lu Zhang and Yilai Wang for helping me with the extension of the program to the second year. I would like to thank Kitty Rikken for helping me with the visa and for answering my endless questions.

My life in the Netherlands was made much easier and enjoyable by the constant interaction with friends both in China and in the Netherlands. Qi Zhang, Yuan Zhang and Huangmei Liu are so nice and considerate, who have dominated my WeChat (the most popular social communication network in China) for the past two years. I have enjoyed our talks, academic or otherwise, which have become one of the most unforgettable moments in the Netherlands. Special thanks go to Yuan Zhang for sharing many important references. I am thankful to Sam Zhang for the time spent together playing basketball and hunting for delicious cuisine. Chao Guo is a nice neighbor and an excellent cook. Hamed Rahmani is good company. I would like to thank Luis Miguel Rojas-Berscia, Sophie Villerius, Jinbiao Yang, Chen Shen, Hongling Xiao and other colleagues in CLS for discussions and get-togethers.

I would also like to extend my thanks to my friends and my neighbors in Kaifeng for their abundant help with the data collection, Yuetao Ren, Long Ma, Guohua Fang, Yongchuan Xuan, Junqing Tian, Shi Yan, Chenshu Fang, Linlin Deng, Wuping Zhao, just to mention a few.

Lastly, I am deeply indebted to my family, my father, my mother and my brother for their love, understanding and unselfish support. I love them more than anything else in this world.

Table of Contents

CHAPTER 1

INTRODUCTION

This dissertation is about the phonetics of the lexical tones in Kaifeng Mandarin. This chapter starts out with a general background of the tones in Mandarin dialects (§1.1). After defining tone language in the broad sense, a sketch of Chinese dialect classification is introduced and the tone categories in modern Mandarin dialects are outlined. Next, the concept of tone sandhi and its typological properties are briefly given. §1.2 reviews the phonetic work on Mandarin tones in citation forms as well as in context. After highlighting the various phonetic parameters that constitute the acoustic and perceptual correlates of tone, f0, amplitude/intensity, duration and phonation, two sources of contextual tonal variation, phonological tone sandhi and phonetic coarticulation, are discussed. §1.3 introduces the geographical location of Kaifeng Mandarin and a syllable-based account of phonological description widely available in the sinological literature. The final section provides the motivation and the organization of the dissertation.

1.1 Tones in Mandarin dialects[1]

In China, those who want to lose weight have been told to avoid drinking soup, eating sugar, lying and eating hot food. In many Mandarin dialects, the words for 'soup', 'sugar', 'lying' and 'hot' share the same segmental string [t^haŋ][2] but differ from one another in their tones. For instance, in Kaifeng Mandarin, the tone structure of which constitutes the subject matter of this investigation, the word 'soup' has a rising tone, 'sugar' has a falling tone, 'lying' is pronounced as a high-level tone and 'hot' is realized as a low tone. In fact, more than half of

[1] In this dissertation, 'dialects' and 'languages' are used interchangeably. The word 'Mandarin' or sometimes 'Mandarin Chinese' has been narrowly used to refer to Standard Mandarin or Standard Chinese, the official language of China. In this dissertation, 'Mandarin' is used in a broader sense to refer to Mandarin dialects or Northern dialects.

[2] Prior to a proper phonological analysis (in Chapter 2), it is not clear whether a given sound is phonemic or not. In this chapter, all the transcriptions are included in square brackets.

the world's languages are tonal, in the sense that they have at least some tones with either a lexical or a grammatical function (Hyman 2001; Gussenhoven & Jacobs 2017). As illustrated above, the tones attached to the same segmental string [tʰaŋ] distinguish these four morphemes, hence forming an integral part of the phonological representation of words. Lexical tones in Chinese languages are typically syllabic and contoured. In this, they differ from the tone systems in African languages, which are word-level melodies and register-based and frequently serve grammatical functions (Wan & Jaeger 1998). Tone languages also include 'pitch accent' languages like Japanese, Swedish, Norwegian and some Limburgish dialects. Unlike Chinese, where each syllable can be specified with an independent tone (there are also toneless syllables, e.g., underlying neutral tone syllables), accentual languages usually have a simple tone inventory with limited distribution, i.e., tones are absent in a number of words and are confined to specific syllables in a word (e.g., Yip 2002). Non-tone languages like English, on the other hand, only have intonational tones, tones that are used to convey discoursal meaning or signal prosodic phrasing (e.g., Gussenhoven & Jacobs 2017). Intonational tones also occur in tone languages.

Modern Chinese languages or dialects can be generally classified into seven major families: Mandarin, Wu, Yue, Min, Hakka, Xiang and Gan (Yuan 2001). More meticulously, Li (1989) separates Jin from the Mandarin family, Hui from the Wu family and Ping from the Yue family, yielding ten major families (see Figure 1.1), which view was adopted in the *Language Atlas of China* (Wurm & Li 1987). The Mandarin family, also named 'Northern family', is mostly spoken to the north of the Yangtze river and in parts of Jiangxi, Hubei, Sichuan, Yunnan, Guizhou, Guangxi and Hunan provinces. The Mandarin family is the largest, encompassing three fourths of the Chinese speakers. Most Mandarin dialects have lost the historical checked syllables (syllables closed by a stop and bearing checked or *ru* tone) and maintained a four-tone system, *yinping*, *yangping*, *shang* and *qu*, (often named T1–T4[3] for

[3] In this dissertation, since traditional tone categories (*yinping*, *yangping*, *shang* and *qu*) only make sense for sinologists, both the numerical notation (T1–T4, T0 or superscript following IPA symbols, as in /pa$^{1–4}$/ or /pa^{0}/) and the phonological representations (LH, HL, H and L) are used to signal the tone categories in Kaifeng Mandarin. The numerical notation is mainly used in transcription, following IPA

convenience) which have evolved etymologically from the four tones in Middle Chinese (henceforth MC), traditionally named *ping*, *shang*, *qu* and *ru*. The *yin*/*yang* register split is conditioned by the voicing contrast in the syllable onset (or 'initial' in traditional Chinese phonology). Often, as in most Mandarin dialects, the voice-sensitive tone split into *yin*/*yang* registers from the four MC tones is not always perfectly symmetrical. The complex splitting and merging process of MC tones can be illustrated briefly by relating it to the synchronous tone system in Kaifeng Mandarin in Table 1.1, characteristic of the Mandarin family to which it belongs. According to Zhang et al. (1993: 288), Kaifeng T1 comes from the MC *ping* tone with voiceless initials (syllable onsets) and from the *qu* tone with voiceless and sonorant initials; T2 from the MC *ping* tone with voiced initials and from the *qu* tone with voiced obstruent initials; T3 from the MC *shang* tone with voiceless and sonorant initials; T4 from the MC *shang* tone with voiced obstruent initials and from the *qu* tone with all initial types. Due to the genealogical closeness of the related dialects, they often share a large number of etymons (also called cognates, over 70% according to Wu (2016: 3) in the case of Standard Mandarin and Jinan Mandarin) within each tone category. Often, the same tone category in different dialects is realized by drastically different tone contours synchronically. For instance, *yinping* (e.g., in the word 'soup') is a rise in Kaifeng Mandarin (24) (Zhang et al. 1993; Liu 1997), a high-level tone (55) in Standard Mandarin, a mid-level tone (33) in Luoyang Mandarin (He 1984), a low dipping tone (213) in Xuzhou Mandarin (Su & Lü 1996), and a mid-to-low fall (41) in Tianjin Mandarin (Zhang & Liu 2011). The digits in parentheses (often in superscript following the syllable) used to translate the pronunciations of the tones were invented by Chao (1930), which has been the standard practice in the description of tone systems. The Chao digits differentiate five tone levels: '1' indicates the lowest and '5' the highest. Besides the full lexical tones, many Mandarin dialects have a neutral tone, commonly referred to as 'T0'. Neutral-tone syllables are generally short and prosodically weak and receive tone values from the preceding full-tone syllable (see Chen 2015 for an

---

recommendation (occurring mainly in Chapter 2). The phonological representation is justified in Chapter 3 and will be used in the remaining chapters.

overview of neutral tones in Standard Mandarin and across Chinese dialects).

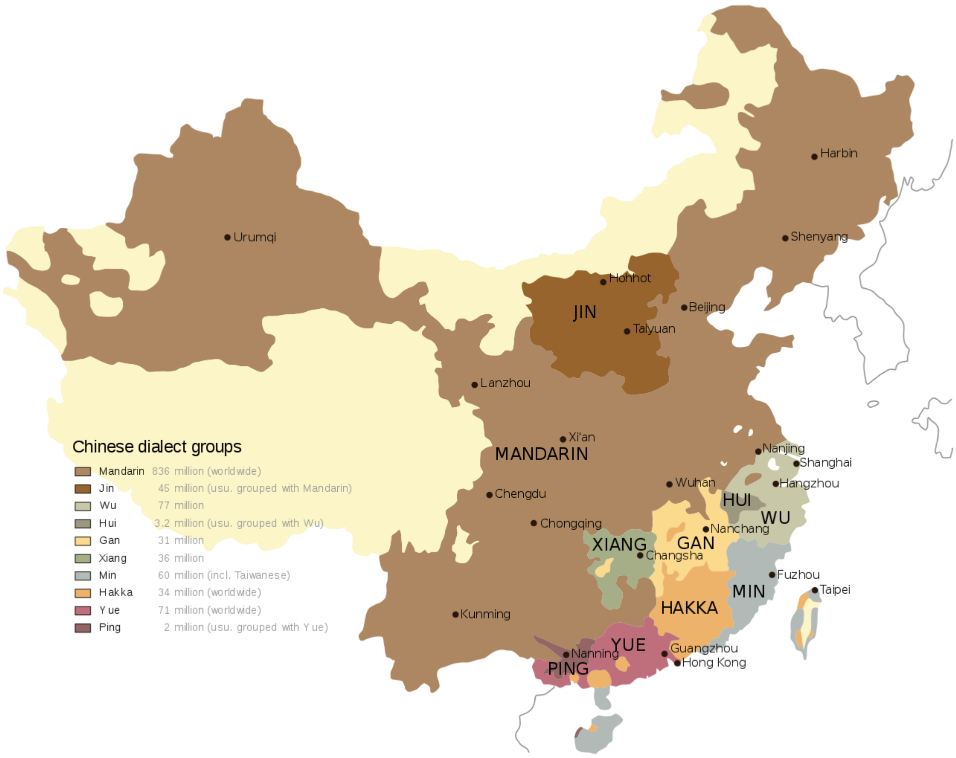


Figure 1.1 Ten major dialect families: Mandarin, Jin, Wu, Hui, Gan, Xiang, Min, Hakka, Yue and Ping in (Wurm & Li 1987), map downloaded on 13-05-2018 from https://en.wikipedia.org/wiki/Language_Atlas_of_China#/media/File:Map_of_sinitic_languages_full-en.svg

Unlike tones produced in isolation, which are usually stable and well-defined, in running speech tones undergo changes to the extent that the canonical shape of the tone may be drastically altered. One source of this contextual tone variation is phonological tone sandhi, a tone alternation process conditioned by the prosodic, phonological and morpho-syntactic environment in which the tone occurs (Chen 2000; Zhang 2007). Diverse tone sandhi phenomena in Chinese languages have been described since the 1980s, to a large extent due to the cumulative publications in the Chinese authoritative journal *Fangyan* (Dialect) by phonetically-trained sinologists or dialectologists (Chen 1996; also see Chen 1993a, 1993b for a review). Based on these impressionistic transcriptions, linguists have probed the phonological generalizations and typological characteristics of tone sandhi (Yue-

Hashimoto 1987; Chen 2000; Zhang 2007, among many others), and how sandhi processes support and are better captured by an insightful phonological representation of tone (Yip 1980, 2002; Bao 1990, 1999).

Tone sandhi in Mandarin dialects, together with most Min and Southern Wu dialects, is known to be 'right dominant', in contrast with 'left-dominant' sandhi, typically found in Northern Wu. The typological division of left-dominant versus right-dominant sandhi depends on the position of the triggering tone in a tone sandhi domain – the grammatical domain within which tone sandhi is applied. Left-dominant sandhi retains the initial tone and changes the non-initial tone, whereas right-dominant sandhi keeps the final tone intact and changes the non-final tone. The most frequently mentioned tone sandhi pattern in Mandarin is Tone 3 sandhi (henceforth T3S) in Standard Mandarin, where T3 (214) becomes T2 (35) before another T3 (214). Thus, [xau214 tɕjou214] 'good wine' surfaces as [xau35 tɕjou214].

Table 1.1 Kaifeng Mandarin tones and their etymological relation with the Middle Chinese tones, adopted and revised from Zhang et al. (1993: 288). The synchronous tone categories *yinping, yangping, shang* and *qu* are named T1–T4, respectively. The phonological representations are included in parentheses.

register	MC onsets	Middle Chinese tones			
		ping	*shang*	*qu*	*ru*
yin	voiceless	T1 (LH)	T3 (H)	T4 (L)	T1 (LH)
yang	sonorant	T2 (HL)	T3 (H)	T4 (L)	T1 (LH)
	voiced obstruent	T2 (HL)	T4 (L)	T4 (L)	T2 (HL)

1.2 Production and perception of Mandarin tones

Production and perception studies of Mandarin tones have been done exclusively on Standard Mandarin, with notably rare exceptions (e.g., Tianjin: Zhang & Liu 2011, Li & Chen 2016; Yuzhou: Zhang & Kong 2014; Jinan: Wu 2015). Tones are generated by the periodic vibration of the vocal folds, the rate of which is acoustically quantified by f0 (fundamental frequency in Hz, i.e., cycles per second) and is perceived to have some pitch.[4] Thus, to a great extent, the acoustic analysis of tones comes down to how f0 varies as a function of time. To the best of

[4] In the remainder of the dissertation, f0 and pitch are used interchangeably.

our knowledge, the earliest acoustic description of Mandarin tones goes back to Chao (1922), who generated tonal contours of several Mandarin dialects (dialects in other families as well) including Kaifeng Mandarin and Standard Mandarin. Despite the inherent limitation of the technique and instrumentation in the 1920s, those pitch contours are not far off the mark. With the advances of acoustic software which makes rigorous examination and manipulation of pitch tracks widely available, it has been repeatedly shown that f0, notably f0 height and f0 contour, are the main acoustic correlates and perceptual cues of tones (see Jongman et al. 2006 for a review and references cited therein). For instance, Peng (2006: 147) displayed a two-dimensional scatter plot with concentration ellipses superimposed (like the acoustic analysis of vowels in an F1/F2 plane) of the four full lexical tones and the neutral tone in Standard Mandarin, with f0 slope as the horizontal axis and f0 height (normalized in five-level scale) as the vertical axis. The Mandarin tones remain distant from each other in the acoustic space. This acoustic information leads to successful tone recognition (83.06%), according to Peng et al. (2004). In addition to f0, the amplitude/intensity profile, duration and phonation also contribute to tone identification (amplitude: Whalen & Xu 1992; duration: Blicher et al. 1990; phonation: Yang 2015; among many others).

Beyond citation tones, tones in connected speech are subject to systematic deviations from their canonical forms in isolation. Among the various sources of tonal variations (e.g., those documented in Xu 2001), one is phonological tone sandhi, mentioned in §1.1; another source is phonetic tonal coarticulation. Being phonological in nature, tone sandhi has been assumed to be categorical and often neutralizing (e.g., Chao 1968) and forms part of the tacit phonological knowledge of the speaker. There is little doubt that this phonological knowledge is actually applied online (Myers & Tsay 2003). However, the categorical and neutralizing nature of tone sandhi is only evidenced in speech perception, i.e., native speakers' inability to discriminate the sandhi tone from the original tone (Wang & Li 1967; Peng 2000), while in speech production they often show instrumentally detectable systematic differences (Zee 1980; Xu 1993; Peng 2000; Cheng et al. 2013; Yuan & Chen 2014), indicating incomplete neutralization (Port & O'Dell 1985). In T3S, the sandhi T2 has been collectively shown to be lower and flatter than the original T2.

Phonetic tonal coarticulation in Mandarin is detailed in Xu (1994, 1997). More recently, Zhang & Liu (2011) and Li & Chen (2016) examined tonal coarticulation in Tianjin Mandarin. These studies revealed some cross-linguistic properties of tonal coarticulation (also see Chen 2012) for a review). First, tonal coarticulation is bidirectional, involving both carry-over effects of the preceding tone on the following tone and anticipatory effects of the following tone on the preceding tone. Second, carry-over effects are assimilatory in nature, while anticipatory effects are dissimilatory and most likely triggered by a low tone. Third, carry-over effects have a greater magnitude than anticipatory effects. Xu (1994) shows that perceptually, listeners to some extent compensate for the considerable contextual variation in the realization of the second tone of a tritonal sequence in Standard Mandarin.

1.3 Kaifeng Mandarin

Figure 1.2 A map of Kaifeng. The municipality of Kaifeng includes the City of Kaifeng (in white), Xiangfu, Lankao, Qi, Tongxu and Weishi.

Kaifeng [k^hai^1 $fəŋ^0$] (开封), a prefecture-level city, is located at 34˚48′N 114˚18′E in east central Henan province, People's Republic of China. As shown in Figure 1.2, the municipality of Kaifeng consists of the City of Kaifeng, Xiangfu,[5] Lankao, Qi, Tongxu and Weishi, covering an area of 6,246 square kilometers, with five million inhabitants.[6] There are slight differences in the phonology among these regional dialects (see Zhang et al. 1993). In this dissertation, Kaifeng Mandarin is confined

[5] It was formerly known as Kaifeng County.
[6] From the government website of Kaifeng http://www.kaifeng.gov.cn, last consulted February 14, 2017

to the language spoken in the City of Kaifeng and the subjects and informants involved in the production and perception experiments all came from this area.

Kaifeng Mandarin belongs to Zhongyuan Mandarin, one of the eight subgroups of the Mandarin family, together with Northeast, Beijing, Jilu, Jiaoliao, Lanyin, Jianghuai, and Southwest (Wurm & Li 1987; He 2005), as shown in Figure 1.3.

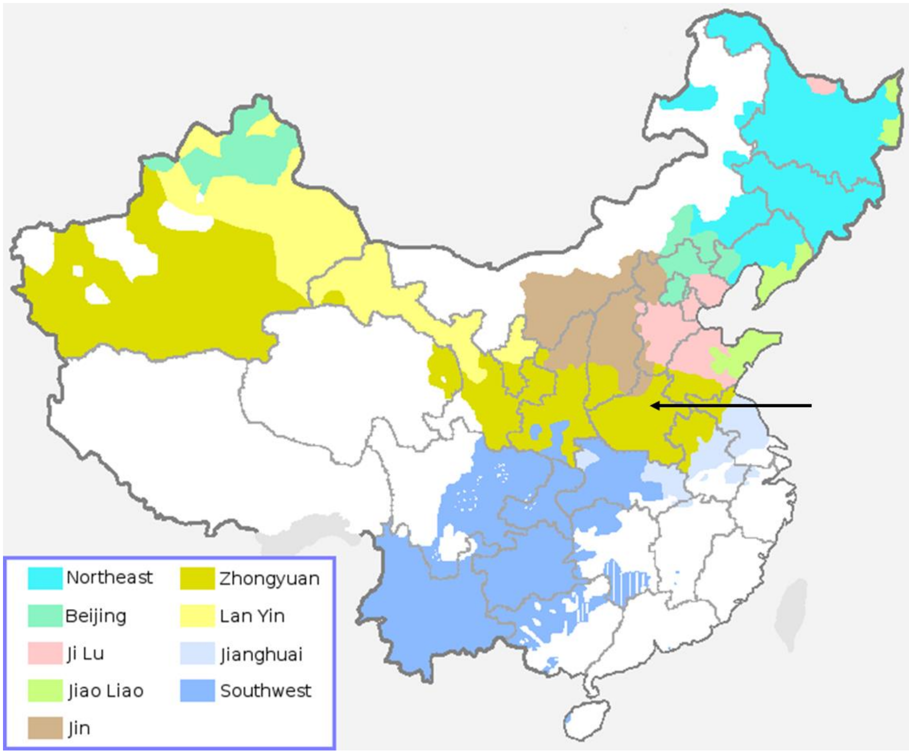


Figure 1.3 Eight subgroups in the Mandarin language family: Northeast, Zhongyuan, Beijing, Lanyin, Jilu, Jianghuai, Jiaoliao and Southwest, together with the Jin group, which is considered a part of the Mandarin language family in Yuan (2001), but distinct from it in Wurm & Li (1987). The geographical location of Kaifeng is indicated by an arrow. The map was downloaded on 13-05-2018 from https://en.wikipedia.org/wiki/Mandarin_Chinese#/media/File:Mandarin_subgroups_and_Jin_group.png

In the sinological literature, descriptions of the segmental structure start with the listings of 'initials' and 'finals' and are essentially syllable-based, quite different from the analyses in the western tradition, which are phoneme-based (Yi & Duanmu 2015). The initial corresponds to the onset in general linguistic terms and the final, also called 'rhyme', is the remainder of the syllable without the initial, which maximally consists of three elements, a medial (filled by a high vowel/glide), a nucleus (filled by a nucleus vowel) and an ending, either vocalic (when filled by a glide) or consonantal (when filled by a nasal) (see e.g., Cheng 1973). Only the nucleus is the obligatory element. When the initial is absent, the syllable is said to have a 'zero initial' (indicated by 'Ø' in the initial inventory). This different analytical preference (i.e., syllable-based vs phoneme-based) 'may have resulted from their respective writing systems' (Yi & Duanmu 2015: 819), in that western languages are written alphabetically, while Chinese orthography is syllable-based. Consequently, for Chinese speakers, syllables are apparently closer to awareness than segments. The analysis of Chinese syllables into initials and finals is enshrined in the *Fanqie* system, in which pronunciations of characters are indicated by a concatenation of two characters, the first of which matches the initial and the second matches the final and the tone. The *Fanqie* system was applied in many old Chinese phonology books such as *Qieyun*,[7] which appeared in 601 AD.

Table 1.2 Initial inventory in Kaifeng Mandarin based on Liu (1997: 271), and additional sounds reported in Zhang et al. (1993: 61) are included in parentheses. In Liu (1997) and Zhang et al. (1993), aspiration is indicated by a single opening quote, as in [p‘]. The non-IPA symbol [ȵ] in Zhang et al. (1993) stands for an alveolo-palatal nasal. [Ø] stands for zero initial, occurring in vowel/glide-initiated syllables.

p	t	ts	tʂ	tɕ	k	Ø
pʰ	tʰ	tsʰ	tʂʰ	tɕʰ	kʰ	
f		s	ʂ	ɕ	x	
	l		ʐ		(ɣ)	
m	n			(ȵ)		

Initial and final inventories in Kaifeng Mandarin are detailed in Liu (1997) and Zhang et al. (1993) and are summarized in Table 1.2 and Table 1.3, respectively. As is often the case, dialectologists may transcribe the sounds differently. In addition to the alternative transcriptions documented in some finals in Table 1.3, Zhang et al. (1993) listed two more sounds [ȵ ɣ] in the initial inventory and one additional [yɛ] in the final inventory. Compared with the initial

[7] *Qieyun* is a pronunciation guidebook written by Lu Fayan in the Sui dynasty. The *Fanqie* system was used to indicate the pronunciation of the characters.

inventory, the final inventory is more complicated. The finals in Table 1.3 are grouped into four columns, named *kai* 'open', *qi* 'even', *he* 'closed' and *cuo* 'protrusive' in traditional Chinese phonology (e.g., Wang 1993: 20), depending mainly on the distribution of the medial [i u y]. A final is *kai*, if it does not have a medial and the nucleus is not [i u y]. A final is *qi*, if it starts with [i], either as the medial or nucleus. Likewise, [u]- and [y]-initiated finals are aptly named *he* and *cuo*, respectively.

Table 1.3 Final inventory in Kaifeng Mandarin based on Liu (1997: 271), and alternative transcriptions of the same sound and additional sound reported in Zhang et al. (1993: 63) are included in parentheses. Non-IPA symbols [ʅ ɿ] stand for 'apical vowels' (see Chapter 2 for more detailed discussion).

kai	*qi*	*he*	*cuo*
ʅ	i	u	y
ɿ			
a	ia	ua	
ɚ (ər)			
ɤ		uɤ (uo)	yɤ (yo)
ɛ	iɛ	uɛ	(yɛ)
ai		uai	
ei		uei	
au (ao)	iau (iao)		
ou	iou		
an	ian (iɛn)	uan	yan (yɛn)
ən	in	uən	yn
aŋ	iaŋ	uaŋ	
əŋ	iŋ	uəŋ	yŋ
oŋ (uŋ)			

Kaifeng Mandarin has two cases of suffix-triggered adjustments of the final, which are detailed in Liu (1997). The first case of these vocalic adjustments is triggered by the nominal suffix [u], also termed *zi* because of their functional equivalence. The host syllable is merged with [u] and creates a new syllable, for instance, [ʂa^1] 'sand'+[u]→[ʂau^1]. However, according to Zhang et al. (1993) and based on our own observation, in modern Kaifeng Mandarin, the old [u]-suffix has almost entirely been replaced by the suffix *zi*; the latter does not trigger any syllable merger or vocalic adjustment. The second case is triggered by the retroflex suffix [ɚ], a diminutive. Unlike the [u]-suffix, the retroflex suffixation has remained productive. The regularities are shown in Table 1.4, following Liu (1997: 273). A more concise generalization as well as some recent changes are documented in § 2.3.3 (Chapter 2).

Table 1.4 Retroflex suffixation in Kaifeng Mandarin, adopted from Liu (1997: 273)

a→ɐr	au→aur	ei→ər	ən→ər
ia→iɐr	iau→iaur	uei→uər	in→iər
ua→uɐr	ou→our	an→ɐr	uən→uər
u→ur	iou→iour	ian→iɐr	yn→yər
ai→ɐr	uan→uɐr	i→iər	y→yər
uai→uɐr	yan→yɐr	ɿ→ər	ʅ→ər
ɤ→ɐr	ɛ→ɐr	aŋ→ɤ̃r	iŋ→iɤ̃r
uɤ→uɐr	iɛ→iɐr	iaŋ→iɤ̃r	yŋ→yɤ̃r
yɤ→yɐr	əŋ→ɤ̃r	uaŋ→uɤ̃r	oŋ→uɤ̃r

The tone inventories reported in Liu (1997) and Zhang et al. (1993) are listed in Table 1.5. They differ in the analysis of the T4 (whether it is a dipping or falling tone), which motivates the study in Chapter 3.

Table 1.5 Tone inventory in Kaifeng Mandarin based on Liu (1997: 271), alternative transcription in Zhang et al. (1993: 62) on Kaifeng T4 is indicated in parenthesis.

T1 (*yinping*)	T2 (*yangping*)	T3 (*shang*)	T4 (*qu*)
24	41	55	312 (31)

1.4 Motivation and outline

An improved understanding of Mandarin tone systems has been developed in the past decades, due to the collective work by phonetically-trained dialectologists and theoretical linguists who base phonological generalizations on their data. However, as noted by Zhang (2010: 1140), the impressionistic transcription of tones by dialectologists, however careful these may be, is not a generally adequate basis for accurate phonological interpretations. Moreover, whenever transcriptions differ, phonological interpretations may vary as a result. For instance, Tianjin Mandarin tones have been analyzed by many different sources. Among them are Li & Liu (1985) and Shi

(1990), summarized in Table 1.6. Adopting Li & Liu's (1985) transcription, tone sandhi patterns T1→T3/_T1 and T4→T2/_T1 can be interpreted as a result of contour metathesis/OCP, as shown in Bao (1999: 60). On the other hand, if Shi's (1990) system is chosen, the former pattern can be analyzed as dissimilation of a sequence of two low-toned word melodies and the latter as a tone absorption process conditioned by two adjacent phonological low tones, as shown in Chen (2000: 105). Therefore, detailed acoustic studies of tones both within and across Chinese languages are warranted to make more reliable linguistic data available for further examination.

Table 1.6 The four tones in Tianjin Mandarin

	T1	T2	T3	T4
Li & Liu 1985	21	45	213	53
Shi 1990	11	55	24	53

The situation in the analysis of Tianjin tones also holds for Kaifeng Mandarin. Kaifeng T4 is transcribed as 312 in Liu (1997), representing a concave shape, but as 31 in Zhang et al. (1993), a low fall (Table 1.6). A choice between these two will inevitably affect the overall phonetic and phonological analysis. The situation is aggravated by the fact that, unlike Standard Mandarin and Tianjin Mandarin, whose tone structures have been well documented, Kaifeng Mandarin lacks reliable descriptions of tonal processes like tone sandhi and the treatment of neutral tone. Furthermore, as will be shown in Chapter 2, a syllable-based account on the segmental phonology of Kaifeng Mandarin (i.e., in terms of initials and finals) misses important phonological generalizations, like the distribution of the nasal coda, which only occurs after monophthongs. Also, many forms reported previously are no longer attested in the synchronic phonology of Kaifeng Mandarin. Consequently, a more thorough and in-depth investigation into the language, both descriptively and analytically, is the focus of this dissertation. As outlined below, the synchronous phonology is introduced in Chapter 2, while the tone structure, the main focus of this dissertation, is elaborated in Chapters 3–5, while concluding remarks appear in Chapter 6.

Chapter 2 reanalyzes the synchronic phonological structure of present-day Kaifeng Mandarin, which is backed up by relevant phonetic data. More importantly, the segmental structure is based on a phonemic analysis, i.e., in terms of consonants and vowels. The phonetic realizations are mainly based on the recordings from two male speakers. Some recent changes and variations in the synchronic system are also reported. Finally, in the tone section, the phonetic properties of citation tones, sandhi tones and neutral tones are presented.

Chapter 3 aims to report and explain the variation that has appeared in the different transcriptions of Kaifeng T4 (L) in the literature, by means of a multi-speaker acoustic study. Thirty-five tonal minimal quadruplets were designed and they were produced twice by ten native speakers of Kaifeng Mandarin. The tones were analyzed in terms of duration and f0. The f0 contours were extracted, manually corrected and z-normalized so that between-subject differences in the f0 realizations can be visualized. The inter- and intra-speaker variations of tone contours were further analyzed in terms of f0 height and f0 slope.

Chapter 4 evaluates the discriminative power between two within-L variations reported in Chapter 3, the dipping L and the falling L, in the context of three L vs non-L (LH, HL and H) contrasts, adopting a tone identification test. The aim is to spell out why the Kaifeng speech community has multiple enhancement forms of the L-tone, assuming that the best enhancement form should best discriminate L from the rival tones. A second goal is to show how syllable lengthening contributes to the perception of L. Thus, six ten-stepped tone continua (2 Ls vs 3 non-Ls) were created and added proportionally in a sonorant-initiated syllable. The syllables had two durational conditions and were pronounced by two different voices (one male voice and one female voice). Forty listeners completed an identification task by selecting one of the toned characters on the computer screen, based on the recordings they were exposed to. Their binary responses were analyzed using a linear mixed effect logistic regression model for each of the contrasts in order to examine the effect of the L-tone variant, duration and speaker voice.

Chapter 5 presents two cases of incomplete neutralization in the tonal phonology of Kaifeng Mandarin. In the first case, identical syllabic level tones are dissimilated by an OCP tone insertion rule, so that H.H and L.L surface as HL.H and LH.L, respectively. A production experiment was conducted in which ten native speakers read minimal word pairs in a random order. The second case is the neutralization of multiple underlying contrasts, i.e., neutral tones. Thirteen speakers were recruited to read two near-minimal word pairs. The supposedly neutralized tones were statistically compared in terms of f0 height, f0 slope and rhyme duration.

Chapter 6 summarizes the findings, provides the general discussion and outlines future research.

References

Bao, Z. (1990). On the nature of tone. Ph.D. dissertation, MIT.

Bao, Z. (1999). *The structure of tone*. Oxford: Oxford University Press.

Blicher, D., Diehl, R. & Cohen, L. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18, 37–49.

Chao, Y.-R. (1922). Methods of experimental study of tone in Chinese speech. *科学* [Science], 7(9), 871–882.

Chao, Y.-R. (1930). A system of tone-letters. *Le Maître Phonétique*, 45, 24–27.

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

Chen, M. (1996). Tonal geometry – A Chinese perspective. In C-T. J. Huang, Y.-H. A. Li. (Eds.), *New horizons in Chinese linguistics*, Vol. 36, 21–48. Dordrecht: Springer. DOI: 10.1007/978-94-009-1608-1_2

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

Chen, Y. (2012). Tonal variation. In A. C. Cohn, C. Fougeron, K. Huffman (Eds.), *The Oxford handbook of laboratory phonology*, 103–114. Oxford: Oxford University Press.

Chen, Y. (2015). Neutral tone. In Rint Sybesma (Ed.), *Encyclopedia of Chinese language and linguistics*, http://dx.doi.org/10.1163/2210-7363_ecll_COM_000162 [First pulished online: 2015, consulted online on 09 May 2018]. Leiden: Brill.

Chen, Z. (1993a). 汉语方言连读变调研究综述 [An overview of tone sandhi patterns in Chinese dialects]. *语文研究* [Studies of the Chinese language], 47(2), 63–66.

Chen, Z. (1993b). 汉语方言连读变调研究综述（续）[An overview of tone sandhi patterns in Chinese dialects (part 2)]. *语文研究* [Studies of the Chinese language], 48(3), 55–60.

Cheng, C., Chen, J. & Gubian, M. (2013). Are Mandarin Sandhi Tone 3 and Tone 2 the same or different? The Results of Functional Data Analysis. In *27th Pacific Asia Conference on Language, Information, and Computation*, Taipei, Taiwan, 296–301.

Cheng, C.-C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton

Gussenhoven, C. & Jacobs, H. (2017). *Understanding phonology*, fourth edition. Abingdon, UK: Routledge.

He, W. (1984). 洛阳方言记略 [Luoyang dialect – a précis]. *方言* [Dialect], 4, 278–299.

He, W. (2005). 中原官话分区（稿）[Classification of Zhongyuan Mandarin (a draft)]. *方言* [Dialect], 2, 136–140.

Hyman, L. (2001). Tone systems. In M. Haspelmath, E. König, W. Oesterreicher, W. Raible (Eds.), *Language typology and language universals: An international handbook*, Vol. 2, 1367–1380. Berlin & New York: De Gruyter.

Jongman, A., Wang, Y., Moore, C. B. & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In P. Li, L. H. Tan, E. Bates, O. J. L. Tzeng (Eds.), *The Handbook of East Asian Psycholinguistics Vol. 1: Chinese*, 209–217. Cambridge, UK: Cambridge University Press.

Li, Q. & Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics*, 54, 123–150. DOI: 10.1016/j.wovn.2015.10.002

Li, R. (1989). 汉语方言的分区 [The classification of Chinese dialects]. *方言*[Dialect], 4, 241–259.

Li, X. & Liu, S. (1985). 天津方言的连读变调 [Tone sandhi in Tianjin dialect]. *中国语文*[Studies of the Chinese language], 1, 76–80.

Liu, D. (1997). 开封方言记略 [Kaifeng dialect – A précis]. *方言* [Dialect], 4, 271–285.

Myers, J. & Tsay, J. (2003). Investigating the phonetics of Mandarin tone sandhi. *Taiwan Journal of Linguistics*, 1(1), 29–68.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, 34, 134–154.

Peng, G., Zheng, H. & Wang, W. S.-Y. (2004). Tone recognition for continuous Chinese speech: A comparative study of Mandarin and Cantonese. *Proceedings of the 4th International Symposium on Chinese Spoken Language Processing*, Hong Kong, 233–236.

Peng, S. H. (2000). Lexical versus "phonological" representations of Mandarin sandhi tones. In M. B. Broe, J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon*, 152–279. Cambridge, England: Cambridge University Press.

Port, R. & O'Dell, M. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13, 455–471.

Shi, F. (1990). *汉语和侗台语的声调格局* [Tone patterns in Chinese and Kam-Tai languages]. PhD dissertation, Nankai University.

Su, X. & Lü, Y. (1996). *徐州方言字典* [A dictionary of Xuzhou dialect]. Nanjing: Jiangsu Education Press.

Wan, I.-P. & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology*, 15, 417–461.

Wang, J. Z. (1993). The geometry of segmental features in Beijing Mandarin, Ph.D. dissertation, University of Delware.

Wang, W. S.-Y. & Li, K. P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, 10, 629–236.

Whalen, D. & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25–47.

Wu, J. (2015). *Tonal bilingualism: The case of two closely related Chinese dialects*. Utrecht: LOT.

Wu, J. (2016). Predicting tonal realizations in one Chinese dialect from another. *Speech Communication*, 76, 1–27.

Wurm, S. A. & Li, R. (Eds.) (1987). *Language Atlas of China*. Hong Kong: Longman.

Xu, Y. (1993). Contextual tonal variation in Mandarin Chinese. PhD dissertation, University of Connecticut, New Haven.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95(4), 2240–2253.

Xu, Y. (1997). Contextual tonal variation in Mandarin. *Journal of Phonetics*, 25, 61–83.

Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics, monograph series*, 17, 1–31.

Yang, R. (2015). The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics*, 43(1), 453–472.

Yi, L. & Duanmu, S. (2015). Phonemes, features and syllables: Converting onset and rime inventories to consonants and vowels. *Language and linguistics*, 16(6), 819–842.

Yip, M. (1980). The tonal phonology of Chinese. Ph.D. dissertation, MIT.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yuan, J. (2001). *汉语方言概要*[An outline of Chinese dialects], second edition. Beijing: Language & Culture Press.

Yuan, J. & Chen, Y. (2014). 3rd tone sandhi in Standard Chinese: A corpus approach. *Journal of Chinese Linguistics*, 42, 218–234.

Yue-Hashimoto, Anne O. (1987). Tone sandhi across Chinese dialects. Chinese Language Society of Hong Kong (Eds.), *Wang Li memorial volumes, English volume*, 445–474. Hong Kong: Joint Publishing Co.

Zhang, J. (2007). A directionality asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*, 16(4), 259–302.

Zhang, J. (2010). Issues in the analysis of Chinese tones. *Language and Linguistics Compass*, 4/12, 1137–1153.

Zhang, J. & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, 68, 161–191.

Zhang, Q., Chen, T. & Cheng, Y. (1993). *河南方言研究* [A study of Henan dialects]. Kaifeng: Henan University Press.

Zhang, R. & Kong, J. (2014). 河南禹州方言声调的声学及感知研究 [An acoustic and perceptual study of Yuzhou Henan dialect tones]. *方言* [Dialect], 3, 206–214.

Zee, E. (1980). A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers*, 49, 98–116.

CHAPTER 2

KAIFENG MANDARIN: THE PHONOLOGY

2.1 Introduction

Kaifeng Mandarin belongs to the Zhengkai subgroup (郑开片), one of the eight subgroups in Zhongyuan Mandarin (中原官话), according to He (2005). The earliest description of Kaifeng Mandarin dates back to Karlgren (1915–1926), who transcribed the pronunciation of thousands of Chinese characters, which was followed by Zhang, Chen & Cheng (1993) and Liu (1997), who provided listings of initials (i.e., onsets), finals (i.e., rhymes) and tones. Compared to Zhang et al. (1993) (though Zhang et al. (1993) predates Liu (1997)), Liu's (1997) description is more conservative and hence better reflects the language of the older generation reported in Karlgren (1915–1926). Some earlier forms reported in Liu (1997) are no longer included in Zhang et al. (1993), like phonemic alveolo-palatals, which now appear as allophones [tɕ $tɕ^h$ ɕ] of alveolars /ts ts^h s/ before /i j y ɥ/, and the nominal suffix /u/, functionally equivalent to the modern suffix *zi* (子) 'son'. With the promotion of Standard Chinese, virtually all young speakers are able to speak both Kaifeng Mandarin and Standard Mandarin fluently. Older speakers have no difficulty in understanding Standard Mandarin, but their speech is noticeably Kaifeng-accented for most of them. The influence of Standard Mandarin on Kaifeng Mandarin is shown at the lexical level in that Standard Chinese cognates coexist with dialectal pronunciation (see § 2.3.2 and § 2.5.1).

The present description focuses on the phonology of present-day Kaifeng Mandarin as spoken in the city of Kaifeng. Phonetic realizations are also provided. For consonants and vowels, most acoustic analyses are based on recordings from a 31-year-old male speaker, who was born and raised in the inner city of Kaifeng and lived mostly in Shunhehuizu district. Materials used to analyze tones are from another male speaker (age: 44). Our analyses differ from previous analyses on the following points:

1. phonemic alveolo-palatals now become allophones of alveolar /ts tsʰ s/ due to a recent diachronic change;
2. fricative [ʐ] and apical vowels [ɿ ʅ] are allophones of the retroflex approximant /ɻ/;
3. pre-vocalic semi-vowels, the 'medial' of traditional analyses, are the consonantal approximants /j w ɥ/, not the pre-nuclear high vowels /i u y/;
4. besides four closing diphthongs, two opening diphthongs /eɛ/ and /ɤʌ/ are presented which have traditionally been described as monophthongs;
5. a phonological generalization is provided for the vocalic adjustments by the retroflex suffix and the resulting retroflex vowels are analyzed as phonemes;
6. an analysis is offered of tone sandhi and neutral tone.

2.2 Consonants

Kaifeng Mandarin has 22 phonemic consonants, as shown in the chart. The central approximants, two of which have a complex articulation place, are listed separately.

	Bilabial	Labiodental	Alveolar	Retroflex	Velar
Plosive	pʰ p		tʰ t		kʰ k
Affricate			ʦʰ ts	tʂʰ tʂ	
Nasal	m		n		ŋ
Fricative		f	s	ʂ	x
Lateral approximant			l		

Central approximants

Retroflex	Palatal	Labial-palatal	Labial-velar
ɻ	j	ɥ	w

The four tone categories are indicated by superscript 1–4, following IPA recommendations. For the pronunciation of the tones, see § 2.5.

pʰ	pʰu^1	铺	‘to pave’	ʦʰ	tsʰu^1	粗	‘thick’
p	pu^1	不	‘no’	ts	tsu^1	租	‘to rent’
tʰ	tʰu^1	秃	‘bald’	tʂʰ	tʂʰu^1	初	‘beginning’
t	tu^1	都	‘capital’	tʂ	tʂu^1	竹	‘bamboo’
kʰ	kʰu^1	哭	‘to cry’				
k	ku^1	骨	‘bone’	m	mu^1	木	‘wood’
				n	kan^1	干	‘dry’
f	fu^1	福	‘blessing’	ŋ	kaŋ^1	钢	‘steel’
s	su^1	酥	‘crispy’				
ʂ	ʂu^1	书	‘book’	ɻ	ɻan^3	染	‘to dye’
x	xu^1	呼	‘to breathe out’	j	jan^3	眼	‘eye’
				ɥ	ɥan^3	远	‘far’
l	lu^1	鹿	‘deer’	w	wan^3	晚	‘late’

All consonants except /ŋ/ can occur in the onset, while in syllable-final positions only /n ŋ/ occur. /ɻ/ is the only syllabic consonant.

2.2.1 Articulation place

It was established by proprioceptive and visual observation that /t tʰ n/ are apicolaminal denti-alveolar, while /ts tsʰ s/ are laminal denti-alveolar and /l/ is apical denti-alveolar. /tʂ tʂʰ ʂ ɻ/[8] are apical post-alveolar or retroflex.[9]

The laminal-anterodorsal alveolo-palatals [tɕ tɕʰ ɕ][10] are allophones of /ts tsʰ s/ before /i j y ɥ/. Karlgren (1915–1926) and Liu (1997) presented the alveolo-palatal and alveolar consonants as distinct, but their

[8] /tʂʰ tʂ/ are transcribed as /ʈʂʰ ʈʂ/ by Karlgren (1915–1926).

[9] Lee (1999) and Lee & Zee (2003, 2014) distinguish between retroflex and apical post-alveolar consonants. Retroflex consonants have the tip of the tongue curled back so that a constriction is formed between the alveolar ridge or the forward part of the hard palate and the underside of the tongue (cf. Hindi, Ladefoged & Johnson 2011). We will use ‘retroflex’ in a broad sense so as to cover constrictions between the tip of the tongue as well as the underside of the tongue blade and the roof of the mouth. Phonologically, Mandarin retroflexes exhibit the common co-occurrence restriction between retroflexes and high front vowels (Hamann 2003).

[10] [tɕʰ tɕ] are transcribed as [ȶɕʰ ȶɕ] by Karlgren (1915–1926).

contrastive status is no longer attested in the present-day dialect due to a recent diachronic change. The alveolo-palatal obstruents are in complementary distribution with alveolar /ts ts^h s/, retroflex /tʂ $tʂ^h$ ʂ/ and velar /k k^h x/, none of which appear before /i j y ɥ/. Historically, most instances of the alveolo-palatal obstruents go back to the alveolar ones, while some come from the velar obstruents. Their assignment as allophones of /ts ts^h s/ is supported by the phonetic similarity between the alveolar and alveolo-palatal series. To be sure, earlier forms like /$kjeɛ^1$/ (隔) 'to separate' (Liu 1997) now appear as /$keɛ^1$/, without any palatalization effect on /k/. This lack of palatalization is also shown by some nouns with the /u/ suffix like /$xjau^2$/ (孩) 'kid' and /k^hjau^4/ (筷) 'chopsticks', in which /j/ appears in the speech of older speakers. A third reason for choosing alveolars as the source of the alveolo-palatals is that this allophonic rule generalizes to /n/, which palatalizes before /i j y ɥ/ (see § 2.2.3).[11] Examples of the alveolo-palatals are given below.

ts^h	[$tɕ^hi^1$]	七	'seven'	[$tɕ^hja^1$]	掐	'to pinch'
ts	[$tɕi^1$]	鸡	'chicken'	[$tɕja^1$]	家	'home'
s	[$ɕi^1$]	西	'west'	[$ɕja^1$]	虾	'shrimp'
ts^h	[$tɕ^hy^1$]	区	'district'	[$tɕ^hɥeɛ^1$]	缺	'lack'
ts	[$tɕy^1$]	菊	'chrysanthemum'	[$tɕɥeɛ^1$]	撅	'to lift'
s	[$ɕy^1$]	虚	'weak'	[$ɕɥeɛ^1$]	雪	'snow'

2.2.2 Laryngeal contrast

Plosives and affricates show a one-way aspiration contrast. In connected speech, unaspirated plosives and affricates are likely to be voiced in sonorant contexts, especially in fast speech. This inter-sonorant voicing typically occurs in the second syllable of a disyllabic word, where the left sonorant is either a vowel or a nasal, as illustrated below.

[11] This argument was also used by Duanmu (2007) in support of his non-phonemic analysis of the alveolo-palatals in Standard Mandarin.

p	[tɕjou^{3} bei^{1}]	酒杯	‘cup’
t	[fan$^{4\to1}$ djan^{4}]	饭店	‘restaurant’
k	[kwaŋ^{3} gau^{4}]	广告	‘advertisement’
ts	[ɕjeɛ^{3} dzɹ̪̩^{4}]	写字	‘to write characters’
tʂ	[ɕin^{4} dzɻ̩^{3}]	信纸	‘stationery’

2.2.3 Sonorants

Syllable-initial /n/ is palatalized and realized as laminal-anterodorsal alveolo-palatal [ɲ̟][12] before /i j y ɥ/.

n	[ɲ̟i^{3}]	你	‘you’	[ɲ̟jeɛ^{1}]	捏	‘to pinch slightly’
n	[ɲ̟y^{3}]	女	‘female’	[ɲ̟ɥeɛ^{1}]	虐	‘to abuse’

Syllable-final /n ŋ/ lack complete oral closures when followed by syllables beginning with a vowel or a central approximant, with concomitant nasalization of the entire rhyme.

Besides functioning as an onset, /ɻ/ can form the syllable nucleus, either without preceding onset or after alveolar and post-alveolar sibilants. Syllabic /ɻ/ is realized as a denti-alveolar approximant [ɹ̪̩] after /ts tsʰ s/ and as a retroflex approximant [ɻ̩] elsewhere.

ɻ	[tsʰɹ̪̩^{1}]	呲	‘to show one’s teeth’	ɻ	[tʂʰɻ̩^{1}]	吃	‘to eat’	ɻ	[ɻ̩^{1}]	日	‘day’
ɻ	[tsɹ̪̩^{1}]	资	‘money’	ɻ	[tʂɻ̩^{1}]	知	‘to know’				
ɻ	[sɹ̪̩^{1}]	丝	‘silk’	ɻ	[ʂɻ̩^{1}]	湿	‘wet’				

[12] [ɲ̟] was reported in Karlgren (1915–1926) and Zhang et al. (1993), not in Liu (1997).

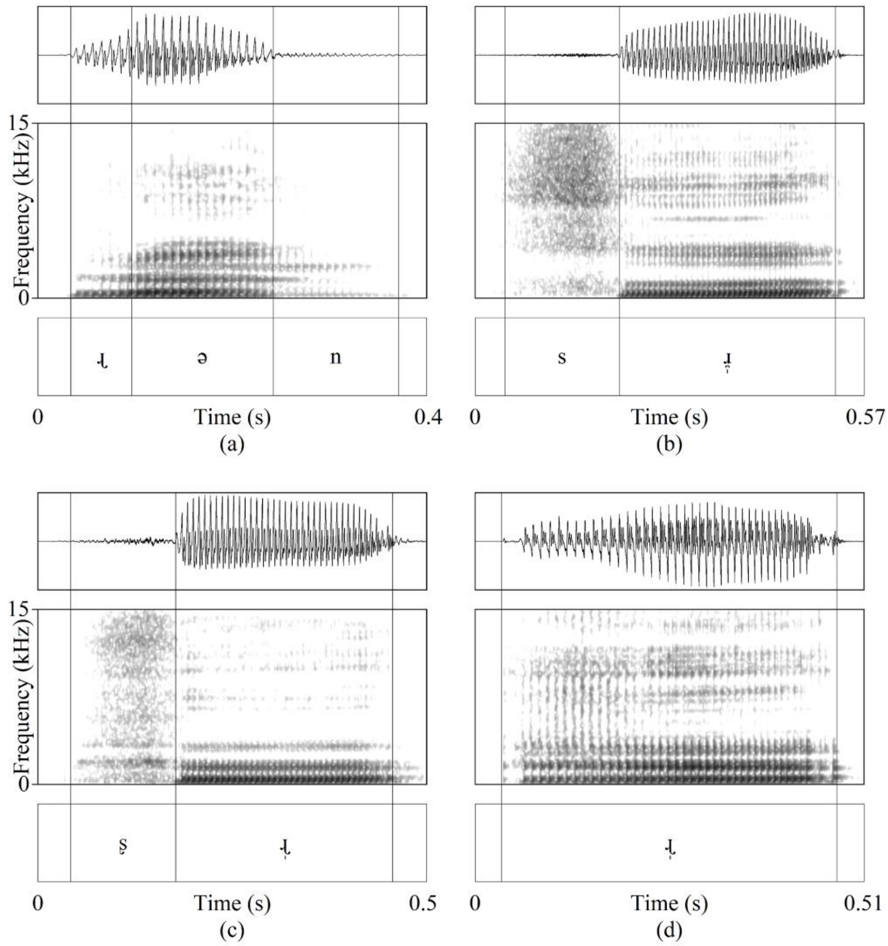


Figure 2.1 Waveform and spectrogram for onset /ɻ/ (panel a), post-sibilant syllabic /ɻ̩/ (panels b–c) and onsetless syllabic /ɻ̩/ (panel d). The upper limit of the spectrogram is set to 15 KHz to better examine the friction noise of /ɻ̩/.

Our analysis follows Lee (1999) and Lee-Kim (2014). There have been two alternative analyses. First, in Karlgren (1915–1926), Zhang et al. (1993) and Liu (1997), onset /ɻ/ is transcribed as a retroflex fricative /ʐ/. While there may on occasion be weak friction, the spectrogram shown in panel (a) of Figure 2.1 shows no friction for onset /ɻ/ in /ɻən^2/ (人) 'person', a representative pronunciation. The same sources transcribe syllabic /ɻ̩/ as [ɿ] after the alveolar sibilants /ts ts^h s/ and as [ʅ] after the retroflex sibilants /tʂ tʂ^h ʂ ʐ/. These symbols were introduced by Karlgren for what he described as 'apical vowels' and have been widely used by Chinese dialectologists. A second alternative for Standard Chinese is to analyze the ‘apical vowels’ as syllabic fricatives, in view of the fact that they are a voiced prolongation of the preceding sibilant (Chao 1948: 22), transcribed [z̩] and [ʐ̩] in Duanmu (2000, 2007). The approach for Standard Chinese by Lee (1999) is confirmed by Lee-Kim (2014), who argued that these ‘apical vowels’ are in fact syllabic approximants in that they are typically pronounced without friction and their F2 patterns are better captured by an acoustic model of sonorant consonants. This consideration also holds for Kaifeng Mandarin. Panels (b) and (c) of Figure 2.1 display the spectrograms of [sɹ̪̩^{3}] (死) ‘to die’ and [ʂɻ̩^{3}] (史) ‘history’ produced by the informant, where the friction noise is absent for the whole vocalic segment, suggesting that these post-sibilant segments are not syllabic fricatives. Moreover, if these segments are vowels, [ɻ̩] should have a lower F2 than [ɹ̪̩] due to its more backward constriction than the latter. However, contrary to the prediction, the reverse pattern is found, with the retroflex approximant [ɻ̩] having a higher F2 than the denti-alveolar approximant [ɹ̪̩]. Figure 2.2 presents formant ellipses for the two approximants. Each sound was repeated six times by ten male speakers (panel a) and ten female speakers (panel b). Each token is realized as a point in the F1/F2 plane, with concentration ellipses added (data coverage = 86.5%). It is revealed that for both male and female speakers, realizations of [ɻ̩] and [ɹ̪̩] slightly overlap and distributionally, [ɻ̩] is located on the left-hand side of [ɹ̪̩], indicating that [ɻ̩] (Male: 1667 Hz, Female: 1959 Hz) has a greater F2 than [ɹ̪̩] (Male: 1262 Hz, Female: 1551 Hz). Its statistical significance was confirmed by paired-samples *t*-tests using the mean F2 values (bark) of [ɻ̩] and [ɹ̪̩] for each speaker [Male: $t(9) = 7.99$, $p < .001$; Female: $t(9) = 8.59$, $p < .001$]. An explanation provided by Lee-Kim (2014: 276–277) is that F2 for [ɹ̪̩] and [ɻ̩] is not attributed to the front cavity as in the acoustic model for vowels but is instead associated with the back cavity due to the resulting short front cavity caused by the coronal articulator. Thus, a lower F2 for [ɹ̪̩] than [ɻ̩] is a result of a larger back cavity of the former than the latter.

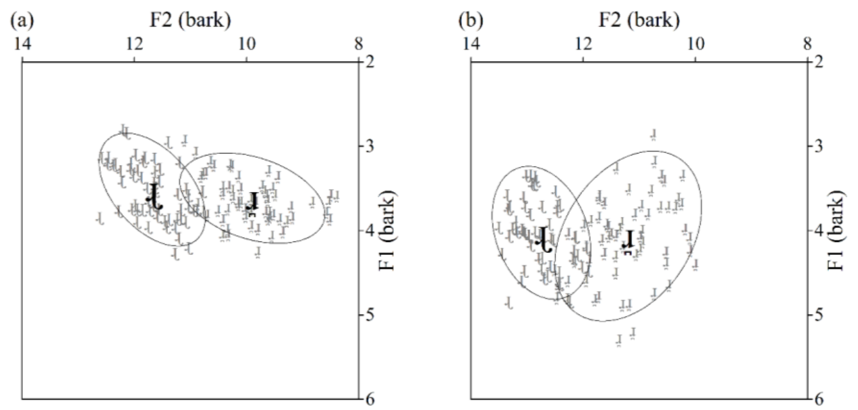


Figure 2.2 Formant ellipses for denti-alveolar approximant [ɹ̪̩] and retroflex approximant [ɻ̩]. Data were obtained from ten male speakers (panel a) and ten female speakers (panel b) (age: mean = 56, sd = 8.2). Each speaker has six repetitions per sound. Each repetition is represented as a plot symbol in the F1/F2 plane (in bark). Sigma ellipses (number of sigmas = 2) were superimposed in Praat (Boersma & Weenink 1992–2015), covering 86.5% of the data points.

Finally, the onsetless syllabic /ɻ̩/ (panel d in Figure 2.1) is noisier than post-sibilant syllabic /ɻ̩/ (panels b and c in Figure 2.1). To compare the noise level in onsetless /ɻ̩/ and post-sibilant [ɹ̪̩] and [ɻ̩], the mean harmonics-to-noise ratio (HNR) for each sonorant rhyme (each containing nine tokens/repetitions) was obtained. A one-way ANOVA showed that the effect of *sonorant type* (onsetless [ɻ̩], post-sibilant [ɹ̪̩] and [ɻ̩]) has a significant effect on HNR ($F(2, 24) = 10.28$, $p = .001$). Post-hoc LSD tests indicated that HNR for onsetless syllabic [ɻ̩] (17.44 dB) is significantly smaller than post-sibilant [ɹ̪̩] (19.25 dB) and [ɻ̩] (20.71 dB). This finding is evidence against the analysis by Chao (1948) and Duanmu (2000, 2007) whereby the fricatives are treated as syllabic segments in the rhyme.

2.3 Vowels

In the sinological literature, vowels have traditionally been grouped into a complex ‘final’ paradigm maximally consisting of three elements, (G)VX (see Table 1.3). In Kaifeng Mandarin, G stands for the glide /j w ɥ/; V is the nuclear vowel and X is either a vocalic element (V) or a nasal coda (C). Analyses of vowel phonemes therefore depend on how the final is subclassified (see also § 2.4). Yi & Duanmu (2015) summarize three views. The first is by You, Qian & Gao (1980), who separated GVX into G, analyzed as the vowel phonemes /i u y/, and VX, where VX includes monophthongs, diphthongs and vowels followed by a nasal. One reason for separating G from V is that G and V have independent phonetic targets, while monophthongs and diphthongs have only one target. This view is supported by an articulatory and acoustic study of Ningbo diphthongs by Hu (2013), who claimed that the onset and offset of 'rising diphthongs' (e.g., /ia/) are comparable in formant values to corresponding monophthongs /i/ and /a/, whereas 'falling diphthongs' (e.g., /ai/) start roughly from the target of monophthongal /a/ and end up with a more variable offset. When X corresponds to a nasal, the one-target characterization is defended by You et al. (1980: 329) on the basis of the vocoid nature of a syllable-final nasal compared to an onset nasal: '[the syllable-final nasal in /an/] lacks the occlusion and release phases in an onset nasal as in /na/ and is only realized as a nasalization feature on the rhyme when followed by certain syllables'. The second view, reflected in Lee & Zee (2003), separates the nasal coda from the preceding GV, ending up with monophthongs without a preceding glide, GV rising diphthongs, VV falling diphthongs and GVV triphthongs. The third view, defended by Duanmu (2007), decomposes the finals into three independent elements G plus VX. The separation of G from VX is based on the fact that poetic rhyme ignores G (cf. Wang 1963). A further decomposition of VX into V and X puts the second element of diphthongs and the second half of long vowels on a par with the coda nasal, which expresses the temporal near-equivalence of VV and VN. Our analysis follows Duanmu (2007), except that we treat monophthongs as well as diphthongs as single segments. Further evidence for the non-existence of GV diphthongs and GVV triphthongs comes from retroflex suffixation, which leaves any pre-nuclear approximant non-retroflexed, as well as from phonetic evidence showing G and V have independent phonetic targets (e.g., Hu 2013). Table 2.1 presents a comparison of the above analyses of Mandarin vowels.

Table 2.1 Classifications of vowels in Standard Mandarin (revised from Yi & Duanmu 2015: 828)

Final	iau	uan	ia	an
You et al. (1980)	i, a^u	u, a^n	i, a	a^n
Lee & Zee (2003)	iau	ua, n	ia	a, n
Duanmu (2007)	j, a, u	w, a, n	j, a	a, n
Current analysis	j, au	w, a, n	j, a	a, n

In addition to the five monophthongs and six diphthongs, Kaifeng Mandarin contains six retroflex vowels /ɯ˞ oɯ˞ ɤʌ˞ aɯ˞ ɚ ɐ˞/. In most cases, the retroflexion comes in the shape of a suffix on plain rhymes, whereby neutralizations leave this six-way contrast.

Onsetless syllables may start from a weakly released glottal closure or begin with weak homorganic friction. The high vowels /i u ɯ˞ y/ may be pronounced as [ji], [wu], [wɯ˞] and [ɥy], respectively, as in [ji^1] (衣) 'clothes', [wu^3] (五) 'five', [wɯ˞^{1}] (屋儿) 'house' and [ɥy^2] (鱼) 'fish', while non-high vowels /a ə ai au ou eɛ ɤʌ oɯ˞ ɤʌ˞ aɯ˞ ɚ ɐ˞/ may start off from a velar approximant [ɰ], as in [ɰan^1] (安) 'safe', [ɰən^1] (恩) 'grace', [ɰai^2] (挨) 'to suffer', [ɰau^3] (袄) 'padded jacket', [ɰou^3] (藕) 'lotus root', [ɰeɛ^1] (挨) 'next to', [ɰɤʌ^1] (恶) 'disgusting', [ɰoɯ˞^{3}] (偶儿) 'spouse', [ɰɤʌ˞^{2}] (蛾儿) 'moth', [ɰaɯ˞^{1}] (嗷儿) 'to shout', [ɰɚ^2] (儿) 'son' and [ɰɐ˞^{4}] (二) 'two'.[13]

2.3.1 Monophthongs

The monophthongs /i y u a/ are phonetically long in open syllables and short when followed by the coda nasal. /ə/ only occurs in closed syllables as a short monophthong.

i	i^3	椅	'chair'	in^3	瘾	'addiction'	iŋ^3	影	'shadow'
y	y^3	雨	'rain'	yn^3	允	'fair'	yŋ^3	勇	'courage'
u	u^3	舞	'dancing'				kuŋ^1	工	'work'[14]
a	sa^1	仨	'three'	an^1	安	'safe'	kaŋ^1	钢	'steel'
ə				ən^1	恩	'grace'	kəŋ^1	耕	'to plough'

[13] The velar approximant [ɰ] is transcribed as a velar fricative [ɣ] in Karlgren (1915–1926) and Zhang et al. (1993). Kaifeng Mandarin lacks onsetless /ei/, which provisionally may also start with the approximant [ɰ].
[14] /uŋ/ is transcribed as /oŋ/ in Liu (1997). However, in retroflex suffixation, /uŋ/ is realized as /wɤʌ˞/. Thus, it is more reasonable to transcribe it as /uŋ/.

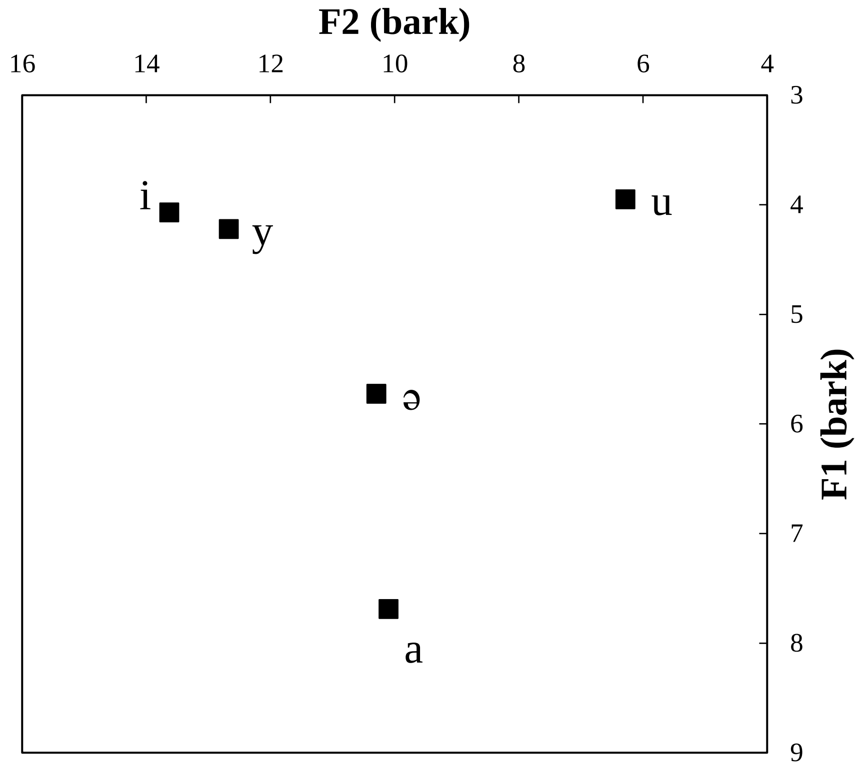


Figure 2.3 Formant values (bark) of the monophthongs. Mean F1 (vertically, Bark) plotted against F2 (horizontally, Bark) measured at the temporal midpoint.

Except /u/, which can only be followed by a velar nasal /ŋ/, all monophthongs can be followed by an alveolar or velar nasal. Figure 2.3 shows the formant values of the monophthongs. For each vowel, F1 and F2 values were obtained by averaging ten repetitions of the key words recorded by the informant. Formant values (in bark) were measured at the midpoint of the vocalic segments for /i y u a/. For /ə/, formant values were taken from the midpoint of the vowel portion in syllables closed by the alveolar nasal.[15]

[15] Measuring the formants of /ə/ from /ən/ is not reliable because of the nasalization effect of the nasal coda. The estimated formant values of /ə/ are only for presentation purposes, i.e., to reveal the distribution of the monophthongs in the vowel chart.

2.3.2 Diphthongs

There are four closing diphthongs, /ai ei au ou/, and two opening diphthongs, /eɛ ɤʌ/.[16]

ai	ai³	矮	'short'	eɛ	xeɛ¹	黑	'black'
au	au³	袄	'padded jacket'	ɤʌ	xɤʌ¹	喝	'to drink'
ei	kei³	给	'to give'				
ou	kou³	狗	'dog'				

Figure 2.4 shows the formant values of the diphthongs. Each diphthong is indicated by a black arrow pointing from the centroid measured at 20% of vowel duration (in bark) to the centroid measured at 80% of vowel duration in the F1/F2 plane. The same measurements for the syllable-final monophthongs /i y u a/ are indicated by a red arrow to show the spectral changes of the monophthongs. Strikingly, the diphthongs show a longer vector length than these monophthongs. The formant trajectories of /eɛ ɤʌ/ in particular show their diphthongal nature, in line with their analysis here. The distinction between monophthongs and diphthongs is based on the phonological distribution of the nasal coda. Specifically, the monophthongs /i y u a ə/ can be closed by a nasal, while the diphthongs /ai au ei ou eɛ ɤʌ/ cannot.

The diphthong /eɛ/ and /ɤʌ/ only contrast after velars. After coronals, the dialect uses /ɤʌ/ instead of /eɛ/ only as a Mandarin-inspired alternative pronunciation, as in the case of [tʂɤʌ⁴] (这) 'this', which is used interchangeably with the Kaifeng form [tʂeɛ⁴]. In addition, there is also a contrast in onsetless cases, as in /ɤʌ¹/ (恶) 'disgusting' and /eɛ¹/ (挨) 'next to'.

[16] /eɛ/ is transcribed as /ɛ/ by Karlgren (1915–1926), Zhang et al. (1993) and Liu (1997), representing a low front monophthong. /ɤʌ/ is transcribed as /ɯ/ by Karlgren (1915–1926), as /ɤ/ by Zhang et al. (1993) and Liu (1997), representing a high back monophthong.

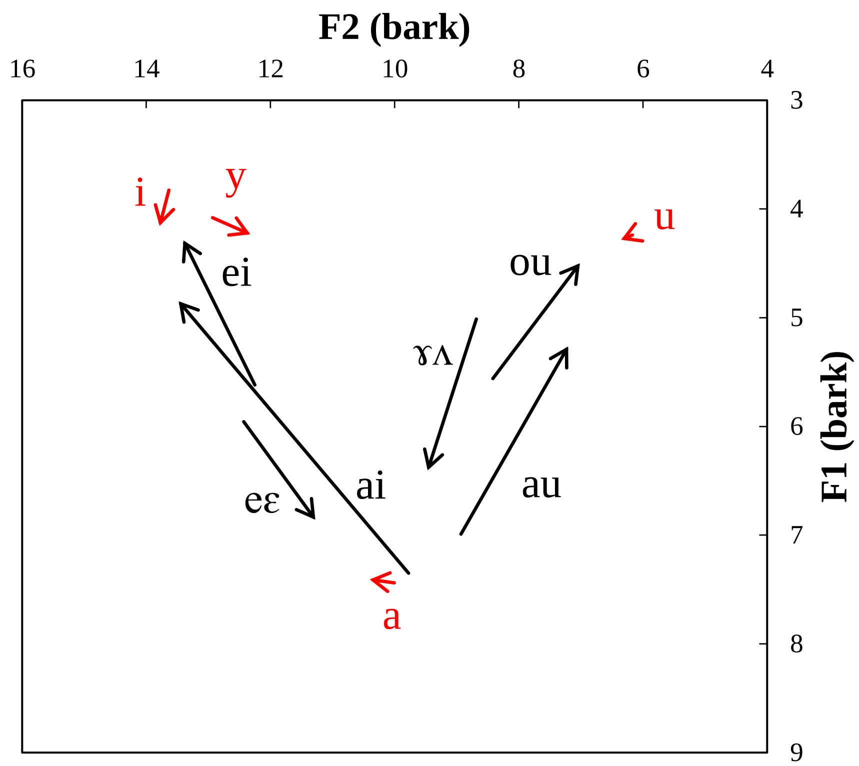


Figure 2.4 Formant values (bark) of the diphthongs. Each diphthong is indicated by a black arrow pointing from the centroid measured at 20% of vowel duration to the centroid measured at 80% of vowel duration in the F1/F2 plane. For comparison, formant frequencies measured at 20% and 80% of vowel duration for the monophthongs /i y u a/ are indicated by a red arrow.

2.3.3 Retroflex vowels

Apart from /ɐ˞/ and /ɚ/, which also occur in underived words, retroflex vowels arise from suffixation of /ɚ^{2}/ (儿) 'child' to nominal hosts, a diminutive formation known as *Erhua* (儿化). More rarely, the suffix can be added to adverbs and verbs, as in /man^{4} mɐ˞^{4} tʂʅ^{1}/ (慢慢儿吃) 'to eat slowly' and /wɐ˞^{2}/ (玩儿) 'to play'. According to Lu (1995), retroflex suffixation surfaces in three forms in modern Chinese dialects. In the first type, the suffix retains its status as a monosyllabic morpheme, in the second it loses its syllabicity and is realized as a retroflex coda consonant, while in the third case, which is applicable to Kaifeng Mandarin, it is a retroflex feature on the vowel of the host. In general, high front vowels are retracted and lowered, because of the incompatibility of their tongue gestures and retroflexion (Chao 1968, Hamann 2003). Coda /n ŋ/ are deleted. In the case of /ŋ/, the vowel is optionally nasalized. Ignoring this optional feature, there are six retroflex vowels, /u˞ ou˞ ɤʌ˞ au˞ ə˞ ɐ˞/.

u˞	ku˞¹	箍儿	'hoop'	ou˞	kou˞¹	勾儿	'hook'
ə˞	kə˞¹	根儿	'root'	ɤʌ˞	kɤʌ˞¹	歌儿	'song'
ɐ˞	kɐ˞¹	肝儿	'liver'	au˞	kau˞¹	羔儿	'lamb'

There are a number of arguments for this non-abstract, phonemic analysis of retroflex vowels. First, if (underived) [ə˞] were to be analyzed as an allophone of /ə/ occurring in open syllables, as in Lee & Zee (2003) in the case of Standard Mandarin, no derivation of [ɐ˞] could be made on the basis of /a/. Second, many words in Kaifeng Mandarin such as /ɤʌ˞²/ (蛾儿) 'moth' are only used in retroflexed forms, meaning that the underlying form would have to be chosen arbitrarily. Third, retroflex and plain vowels contrast on the surface, as in /kɤʌ˞¹/ (歌儿) 'song' and /kɤʌ¹/ (割) 'to cut'. Differently from the retroflex approximant [ɻ̩], for /ə˞/ the tongue tip is raised towards a point on the hard palate immediately behind the alveolar ridge. Often, the rhotacization sets in at some point during the vowel, such that it retains its plain quality at the beginning and becomes rhotacized only in the second half. The acoustic difference between the retroflexed schwa /ə˞/ and the retroflex approximant /ɻ̩/ is shown in Figure 2.5. Unlike /ɻ̩/, which has stationary F1, F2 and F3, /ə˞/ is characterized by a decrease in F1 and F3, and a weak increase in F2. According to Lee & Zee (2014), pharyngealization occurs as a by-product of the retraction of the tongue body in the pronunciation of retroflex vowels, which also applies to Kaifeng Mandarin.

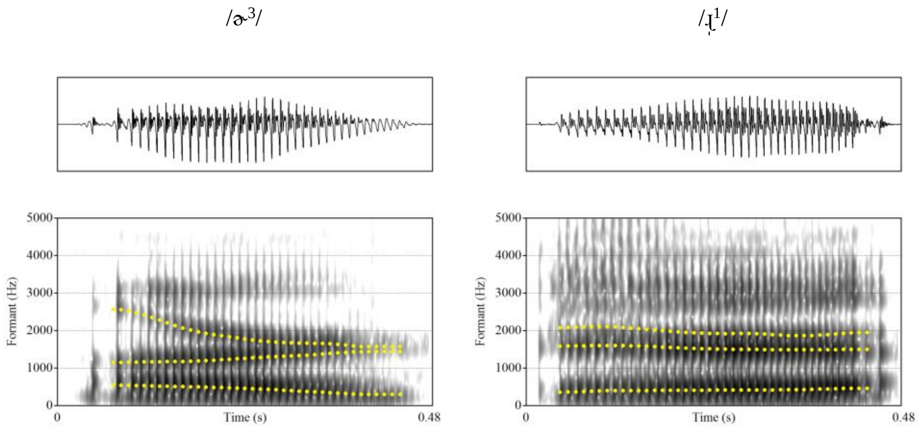


Figure 2.5 Spectrograms and formant tracks of /ɚ^{3}/ and /ɻ̩^{1}/ of the informant. The first three formants are indicated by yellow speckles superimposed on the spectrogram.

The neutralizations of retroflex vowels and language examples are documented in Liu (1997) (see Table 1.5), and are reanalyzed below.

The diphthong /ei/ and syllabic /ɻ̩/ are replaced by /ɚ/. Retroflexion of the monophthongs /i y/ result in /jɚ ɥɚ/, respectively. Below, an example of the underived retroflex vowel is also included.

Stem vowel	Retroflex vowel	Key word	Orthography	Gloss
i	jɚ	p^{h}jɚ^{2}	皮儿	‘skin’
y	ɥɚ	ɥɚ^{2}	鱼儿	‘fish’
ei	ɚ	pɚ^{4}	辈儿	‘generation’
ɻ̩	ɚ	tsɚ^{4}	字儿	‘character’
ɻ̩	ɚ	tʂɚ^{1}	枝儿	‘branch’
	ɚ	ɚ^{3}	耳	‘ear’

The mid or low ending vowels /eɛ ai a/ appear as /ɐ˞/. Below, the only example of the underived retroflex vowel /ɐ˞/ is also included.

Stem vowel	Retroflex vowel	Key word	Orthography	Gloss
eɛ	ɐ˞	tʂʰɐ˞¹	车儿	‘vehicle’
ai	ɐ˞	kɐ˞⁴	盖儿	‘cover’
a	ɐ˞	pɐ˞⁴	把儿	‘handle’
	ɐ˞	ɐ˞⁴	二	‘two’

The back vowels /u ou ɤʌ au/ remain distinct as /ɯ˞ oɯ˞ ɤʌ˞ aɯ˞/.

Stem vowel	Retroflex vowel	Key word	Orthography	Gloss
u	ɯ˞	kɯ˞¹	箍儿	‘hoop’
ou	oɯ˞	koɯ˞¹	勾儿	‘hook’
ɤʌ[17]	ɤʌ˞	kɤʌ˞¹	歌儿	‘song’
au	aɯ˞	kaɯ˞¹	羔儿	‘lamb’

In retroflex suffixation, coda /n/ is deleted, leaving the retroflex vowels as they appear in open syllables. Deletion of /ŋ/ is incomplete, in the sense that the retroflex vowel, which is always /ɤʌ˞/, may be nasalized. The sample recordings provided by our informant do not have this nasalization.

Stem vowel	Retroflex vowel	Key word	Orthography	Gloss
in	jɚ	jɚ⁴	印儿	‘mark’
yn	ɥɚ	tsʰɥɚ²	裙儿	‘skirt’
an	ɐ˞	kɐ˞¹	肝儿	‘liver’
ən	ɚ	mɚ²	门儿	‘door’
iŋ	jɤʌ˞	mjɤʌ˞²	名儿	‘name’
yŋ[18]	ɥɤʌ˞			
uŋ	wɤʌ˞	kʰwɤʌ˞⁴	空儿	‘free’
aŋ	ɤʌ˞	tʂʰɤʌ˞²	肠儿	‘intestine’
əŋ	ɤʌ˞	ʂɤʌ˞²	绳儿	‘rope’

[17] According to Liu (1997), suffixed /ɤʌ/ is realized as [ɐ˞] in the older generation. At present, this form can still be found in the speech of some older speakers. For example, the retroflexed form of /tsɿ̩⁴ kɤʌ¹/ (自个) ‘self’ is /tsɿ̩⁴ kɐ˞¹/.

[18] Suffixed /yŋ/ was reported in Liu (1997). However, we are unable to find an example of that kind.

2.4 Syllable Structure

The syllable structure of Kaifeng Mandarin is (C)(G)VX, where G stands for /j w ɥ/ and C for other consonants. The obligatory elements VX can be filled by a long monophthong, a syllabic approximant, a diphthong or a short monophthong followed by a nasal coda, as in /i^{1}/ (衣) 'clothes', /ʅ^{1}/ (日) 'day', /ou^{3}/ (藕) 'lotus root' and /an^{3}/ (俺) 'I'.

Table 2.2 CG combinations in Kaifeng Mandarin

	Bilabial	Labiodental	Alveolar	Retroflex	Velar
j	Yes	No	Yes	No	No
w	No	No	Yes	Yes	Yes
ɥ	No	No	Yes (*t^{h}ɥ, *tɥ)	No	No

The affiliation of G to either the syllable onset or the syllable rhyme is controversial (see van de Weijer & Zhang 2008 for a review). A well-known argument for not considering it part of the rhyme is that it plays no role in poetic rhyme in modern Chinese dialects. That is, syllables like /ma/, /tsja/, /kwa/ rhyme (Wang 1963), ruling out GV diphthongs or GVX triphthongs. Adopting standard onset-rhyme theory, the prevocalic approximant must therefore be in the onset. Under that assumption, two analyses have been proposed. In the first, G is the second member of a CG cluster (Bao 1990, Wang 1999), while in the second it is incorporated into the C as a secondary articulation, i.e., C^{G} (Duanmu 1990, 2000, 2007, 2008; Wang 1993). Two additional analyses exist which avoid a choice between onset and rhyme. First, a treatment of G as a constituent of a linear flat syllable structure without sub-syllabic constituency has been proposed by Yip (2003) and, second, an intermediate node for G between C and V in an X-bar structure has been proposed by van de Weijer & Zhang (2008). Regardless of the analysis, there are co-occurrence restrictions on C and G. Specifically, the labiodental consonant /f/ does not combine with any of these glides, while the bilabial consonants do not combine with /w ɥ/ and the velar and retroflex consonants do not combine with /j ɥ/. It is worth noting that bilabial consonants are labialized before /ɤʌ/, which feature we treat as allophonic, as in /pɤʌ^{1}/ [p^{w}ɤʌ] (剥) 'to skin', /p^{h}ɤʌ^{1}/ [p^{hw}ɤʌ] (坡) 'slope' and /mɤʌ^{1}/ [m^{w}ɤʌ^{1}] (摸) 'to touch'. There are two more gaps which are probably accidental. /t^{h}/ and /t/ cannot combine with /ɥ/, but the other alveolars /n l ts^h ts s/ can. The CG combinations are summarized in Table 2.2.

In addition, there are co-occurrence restrictions between /j w ɥ/ and following vowels. The approximant cannot precede /i(n/ŋ) u(ŋ) y(n/ŋ)/. /ən/, /əŋ/, /ai/ and /ei/ can only be preceded by /w/, as in /swən^1/ (孙) 'grandson', /wəŋ^1/ (翁) 'old man', /ʂwai^4/ (帅) 'handsome' and /tʂ^{h}wei^1/ (吹) 'to brag'. While /an/ can be preceded by glides /j w ɥ/, as in /ljan^3/ (脸) 'face', /lwan^4/ (乱) 'messy' and /sɥan^3/ (选) 'to select', /a/ and /aŋ/ can only be preceded by /j w/, as in /lja^3/ (俩) 'two', /xwa^1/ (花) 'flower', /ljaŋ^2/ (凉) 'cold' and /tʂwaŋ^4/ (壮) 'strong'. /eε/[19] can be preceded by /j/ and /ɥ/, as in /t^{h}jeε^1/ (铁) 'iron' and /sɥeε^1/ (雪) 'snow'. /ɤʌ/ can only be preceded by /w/ and /ɥ/, as in /twɤʌ^1/ (多) 'many' and /tsɥɤʌ^1/ (脚) 'foot'. /ɤʌ/ in /wɤʌ/ and /ɥɤʌ/ is slightly centralized and rounded. /au/[20] and /ou/ can only be preceded by /j/, as in /pjau^3/ (表) 'watch' and /tjou^1/ (丢) 'to lose'. Retroflex vowels /ɯ˞ oɯ˞ aɯ˞/ follow the same rules as /u ou au/, while /ɚ/, /ɐ˞/ and /ɤʌ˞/ can be preceded by all three approximants, as in /jɚ^4/ (印儿) 'mark', /wɚ^4/ (味儿) 'taste', /ɥɚ^2/ (鱼儿) 'fish', /jɐ˞^{1}/ (叶儿) 'leaf', /wɐ˞^{1}/ (弯儿) 'turn', /ɥɐ˞^{4}/ (院儿) 'yard', /mjɤʌ˞^{2}/ (名儿) 'name', /k^{h}wɤʌ˞^{4}/ (空儿) 'free', /tsɥɤʌ˞^{1}/ (角儿) 'angle'. The GVX combinations are summarized in Table 2.3.

Table 2.3 GVX combinations in Kaifeng Mandarin

	i(n/ŋ), u(ŋ), y(n/ŋ), ɯ˞	ən/ŋ, ai, ei	a(ŋ)	an, ɐ˞, ɚ, ɤʌ˞	eε	ɤʌ	au, aɯ˞, ou, oɯ˞
j	No	No	Yes	Yes	Yes	No	Yes
w	No	Yes	Yes	Yes	No	Yes	No
ɥ	No	No	No	Yes	Yes	Yes	No

[19] In the speech of the older generation, /eε/ can be preceded by the glide /w/, as in /kweε^1/ (国) 'nation'. At present, however, /weε/ becomes marginal and is only kept for some older speakers.

[20] In some marginal syllables, /au/ can be preceded by /w/ and /ɥ/, as exemplified by some output forms of the old /u/ suffixation, i.e., /tʂwau^1/ (桌) 'desk' and /ɥau^1/ (月) 'confinement' (Liu 1997).

2.5 Tones

Kaifeng Mandarin has four distinctive tones T1–T4 and one neutral tone T0. Below, the phonetic properties of citation tones, sandhi tones and neutral tones are presented, using data from one male speaker as illustration (ZWP, age 44). The f0 contours were measured in Praat (Boersma & Weenink 1992–2015), using the script *Prosodypro* (Xu 2013). *Prosodypro* allows for visual inspection and manual correction of vocal pulse cycles, and smoothing of f0.

2.5.1 Citation tone

T1–T4, traditionally labeled *yinping*, *yangping*, *shang* and *qu*, are transcribed in Chao digits by Zhang et al. (1993) and Liu (1997) as 24, 41, 55, 31(2).

T1	*Yinping*	ta^1	搭	'to build'
T2	*Yangping*	ta^2	达	'to arrive'
T3	*Shang*	ta^3	打	'to hit'
T4	*Qu*	ta^4	大	'big'

The four tone categories in Kaifeng Mandarin are cognate to those in Standard Mandarin. These two varieties share a large number of etymological equivalents within each category. When the tone category of a word is not equivalent in these two varieties, young bilingual speakers' assignment of the Kaifeng tone category may be biased by its tone class in Standard Mandarin, especially when the word is pronounced without a context. For instance, the same word /lu^1/ (鹿) 'deer' is *yinping* in Kaifeng Mandarin, but *qu* in Standard Mandarin. Young speakers may mis-assign Kaifeng *qu* (T4) tone to this word, pronouncing it as /lu^4/. At present, both native and Mandarin-inspired pronunciations are equally acceptable as confirmed by multiple consultants.

Figure 2.6 displays the f0 contours of the four tones. Data were from two repetitions of 35 tonal minimal quadruplets, giving 70 tokens per tone. Twenty-one equidistant f0 sampling points were extracted from the syllable rhyme. For each tone, tokens of the same tone were point-to-point averaged and their mean rhyme duration was used as the time axis in each case. T1 starts from low f0, followed by a slight initial fall in the first quarter of the tone and a rise to mid pitch. T2 starts at the f0 ceiling and falls steeply to low pitch. T3 is a mid-level, while T4 is a mid-to-low fall. As is shown in Chapter 3, T4 in isolation is realized as a dipping contour by some speakers, while in connected speech, T4 is usually realized as a low falling contour in non-prepausal positions. In the case of T1, T2 and T4, male as well as female speakers may have creaky phonation at the point where a low pitch target is expected, i.e., in the initial portion of T1, in the final portion of T2 and in the center of a dipping T4 or in the final portion of a low falling T4. In terms of rhyme duration, T4 (235 ms) is the longest and T3 (175 ms) is the shortest, while T1 (201 ms) and T2 (210 ms) are in between.

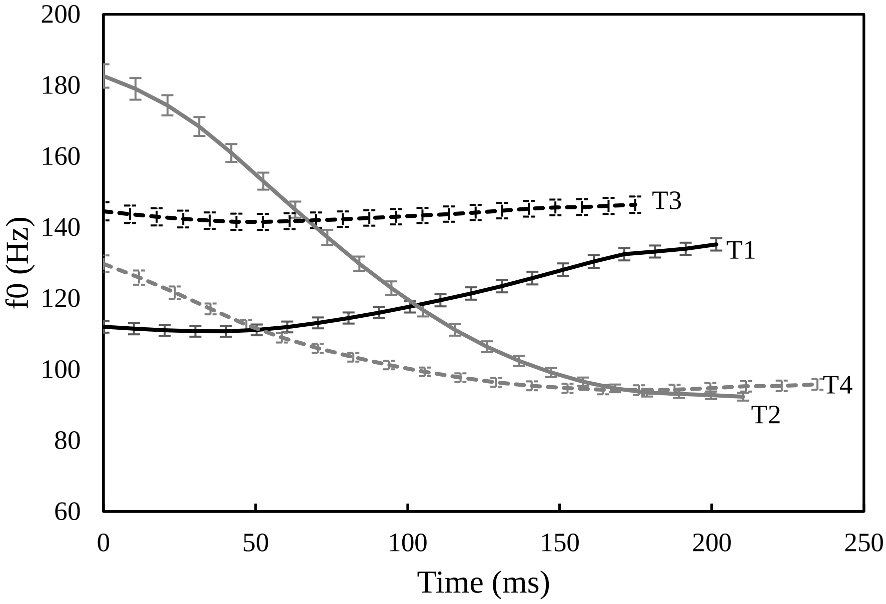


Figure 2.6 Pitch contours of T1–T4, obtained by averaging two repetitions of 35 minimal quadruplets. The vertical axis represents f0 in Hertz; the horizontal dimension shows time (ms). Twenty-one equidistant f0 measurement points were taken from the syllable rhyme for each token. Error bars stand for ±2 standard errors.

2.5.2 Tone sandhi in disyllables

Kaifeng Mandarin has two unexceptional tone dissimilation rules.[21] By the first of these, T3 becomes T2 when followed by another T3. This sandhi process is cognate to the 'third tone sandhi' in Standard Mandarin '*shang* + *shang* → *yangping* + *shang*', a historical rule dating back to the 16th century (Mei 1977). By the second rule, T4 becomes T1 when followed by another T4, as shown below.

T3→T2/_T3	xwɤʌ$^{3\to2}$ pa^{3}	火把	'torch'
T4→T1/_T4	ts^{h}i$^{4\to1}$ tian^{4}	气垫	'air cushion'

Tone sandhi patterns in Chinese dialects are either left-dominant or right-dominant, depending on the position of the triggering tone (Chen 2000; Zhang 2007, 2010). In left-dominant sandhi, which typically occurs in Northern Wu, the non-initial tonal melodies are deleted, while the leftmost tone extends over the whole sandhi domain. By contrast, in right-dominant sandhi, typically found in Min, Southern Wu and Mandarin dialects, the rightmost tone is retained and non-final tones are assigned a generic default tone or a tone specifically triggered by the retained tone. As a result, right-dominant sandhi will typically lead to neutralizations between derived forms and tonally equivalent underlying forms. This is the case in Kaifeng Mandarin. For example, pronunciations of /xwɤʌ$^{3\to2}$ pa^{3}/ (火把) 'torch', with an underlying tone sequence T3T3, and /xwɤʌ^{2} pa^{3}/ (活靶) 'target', underlyingly T2T3, are tonally identical, and so are /ts^{h}i$^{4\to1}$ tian^{4}/ (气垫) 'air cushion', underlyingly T4T4, and /ts^{h}i^{1} tian^{4}/ (漆店) 'paint shop', underlyingly T1T4.

[21] There is one additional sandhi rule: T3→T2/_T4. In modern Kaifeng Mandarin, this only applies to a small number of words, e.g., /pan$^{3\to2}$ təŋ^{4}/ (板凳) 'stool' and is optional in all cases. For some older speakers, this rule can be more general, i.e., applicable to a larger number of words and even to nonsense words.

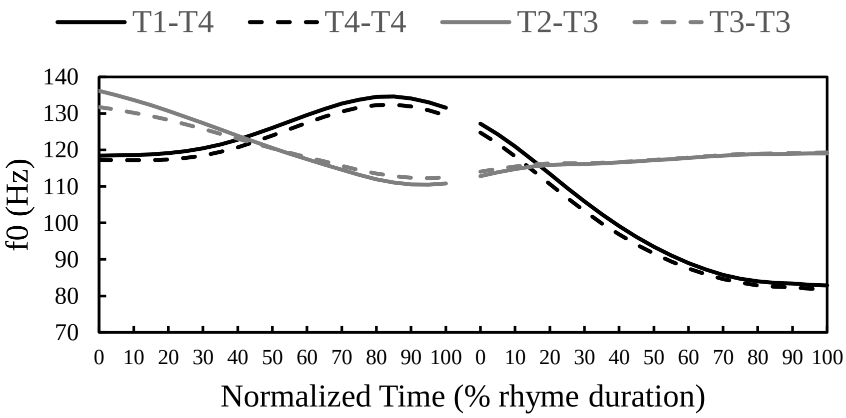


Figure 2.7 Tone sandhi in Kaifeng Mandarin. Each pitch curve (Hz) was obtained by averaging over four repetitions of 14 disyllabic words. Time is expressed as the percentage into the rhyme duration.

Time-normalized f0 contours of the disyllables are provided in Figure 2.7. Test words included two sets of 14 phonologically identical word pairs. With four repetitions per word, this produced a corpus of 224 words (14 pairs × 2 putative contrasts per pair × 2 sets × 4 repetitions). There are small phonetic differences between the derived and underlying tone patterns. Specifically, the derived T2 would appear to have a flatter f0 contour than the original T2 and the derived T1 a lower f0 contour than the original T1. Thus, it is probably true to say that neither of these two sandhi rules yields a complete neutralization acoustically and as such both represent cases of incomplete neutralization (Port & O'Dell 1985) (see Chapter 5 for detailed analysis).

2.5.3 Neutral tone

In addition to the four lexical tones, Kaifeng Mandarin has a neutral tone, occurring in grammatical morphemes, in post-verbal object pronouns, in the second syllable of some lexical items and noun/verb reduplicated forms, among other morphemes. Below, the neutral tones are indicated by superscript 0; deletion of underlying tones and assignment of neutral tones are indicated by an arrow. Syllables with neutral tone are either bound morphemes, like the possessive/nominalizer marker 嘞 /leɛ^0/ as occurring in 紫嘞 /tsɻ̩^{3} leɛ^0/ 'something purple', in which case they have no underlying tone, or free morphemes that lose their underlying tone in context, like the personal pronoun 他 /tʰa³/ 'him'.

Suffixation	grammatical morpheme	紫嘞	tsɿ³ leɛ⁰	'something purple'
Tone deletion	object pronoun	问他	wən⁴ tʰa³→⁰	'to ask him'
	lexical item	抹布	ma¹ pu⁴→⁰	'dishcloth'
	reduplication	炸炸	tʂa² tʂa²→⁰	'to fry'

Underlying toneless syllables and tone-deleted syllables are realized identically. Their f0 realizations are conditioned by the preceding full lexical tone.

Lexical tone	Orthography	IPA	Gloss
T1	姑夫	ku¹ fu⁰	father's sister's husband
T2	姨夫	i² fu⁰	mother's sister's husband
T3	姐夫	tsjeɛ³ fu⁰	elder sister's husband
T4	妹夫	mei⁴ fu⁰	younger sister's husband

The f0 contours of these neutral-tone syllables are illustrated in Figure 2.8. After T1, the f0 of the neutral tone remains a mid-level. After T2, the f0 of the neutral tone starts with a low f0 onset and falls even lower. After T3, the f0 of the neutral tone starts high and falls steeply to low. Finally, after T4, the f0 of the neutral tone starts fairly low and then rises weakly. The pitch range of neutral-tone syllables is largely maintained because they are not shortened in Kaifeng Mandarin to the same extent as in Standard Mandarin. Based on the corpus of this male speaker (325 disyllabic words with neutral tone in the second syllable read in isolation), the rhyme duration of the initial full-tone syllable (134 ms, sd = 32.93) is even shorter than that the neutral-tone syllable (160 ms, sd = 27.25), giving a duration ratio of 0.84, which is very much different from Standard Mandarin where the initial full-tone syllable is roughly twice as long as the neutral-tone syllable (Lee & Zee 2008, with the test words read in isolation).

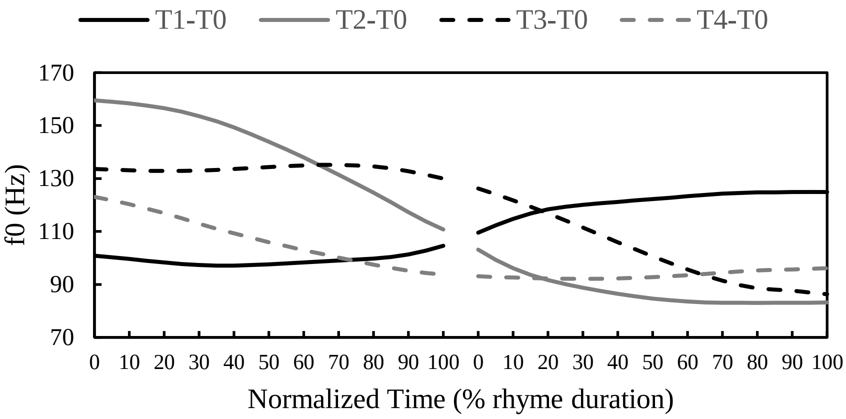


Figure 2.8 f0 contours of neutral-tone rhymes in different prosodic conditions. Each curve was obtained by averaging multiple disyllables containing neutral tone (after T1, n = 70; after T2, n = 82; after T3, n = 66; after T4, n = 107). Time is expressed as the percentage into the rhyme duration.

When a T3T3 combination satisfies both tonal deletion in the second syllable (so as to create a derived neutral tone in the second syllable) and tonal dissimilation in the first syllable due to sandhi (T3→T2/_T3), the sandhi rule applies first. That is, the neutral tone is pronounced as after T2, not as after T3, i.e., T3T3→T2T3→T2T0 (not T3T3→T3T0).

Orthography	IPA	Gloss	Type
老虎	lau$^{3\to2}$ xu$^{3\to0}$	'tiger'	lexical item
打他	ta$^{3\to2}$ tʰa$^{3\to0}$	'to hit him'	object pronoun
哥哥	kɤʌ$^{3\to2}$ kɤʌ$^{3\to0}$	'elder brother'	reduplication

By contrast, in T4T4, only deletion of the second T4 is applied, giving T4T4→T4T0 (not T4T4→T1T4→T1T0). This is shown below.

Orthography	IPA	Gloss	Type
挂面	kwa^{4} mjan$^{4\to0}$	'dried noodles'	lexical item
弟弟	ti^{4} ti$^{4\to0}$	'younger brother'	reduplication

2.6 Transcription

A broad phonemic transcription is provided of the North Wind and Sun fable using the vowel and consonant symbols introduced here. Neutral tones, derived or inherent, are all indicated by a superscript 0. Sandhi tones are marked by an arrow. Major phrases are marked by | and utterances are marked by ‖.

tʰai^{4} jaŋ^{0} kən^{1} pei^{1} fəŋ^{1} leɛ^{0} ku^{4} ʂɻ̩^{0}

jou^{3} i^{1} xwei^{2} | pei^{1} fəŋ^{1} kən^{1} tʰai^{4} jaŋ^{0} tʂəŋ^{4} kei^{3} nɐ˞^{4} pi^{3} ʂei^{2} leɛ^{0} pən^{3} ʂɻ̩^{0} ta^{4} ‖ tʂəŋ^{4} ʂwɤʌ^{1} tʂu^{0} leɛ^{0} ʂɻ̩^{2} xou˞^{4} | lai^{2} la^{0} ɥɤʌ^{2} kwɤʌ$^{4\to1}$ lu^{4} leɛ^{0} | ʂən^{1} ʂaŋ^{0} tʂʰwan^{1} la^{0} kɤʌ^{4} au^{3}‖ tʰa$^{3\to2}$ lja^{3} tou^{4} ʂaŋ^{1} ljaŋ^{0} xau^{3} la^{0} | ʂei^{2} sjan^{1} ɻaŋ^{4} tʂeɛ^{4} kɤʌ^{0} kwɤʌ$^{4\to1}$ lu^{4} leɛ^{0} pa^{3} tʰa^{3} leɛ^{0} au^{3} tʰwɤʌ^{1} lou^{0} | tou$^{4\to1}$ swan^{4} ʂei^{2} leɛ^{0} pən^{3} ʂɻ̩^{0} ta^{4}‖ pei^{1} fəŋ^{1} tou^{4} kei^{3} nɐ˞^{4} pjeɛ^{1} tʂu^{0} | ʂɻ̩^{3} tsjɚ^{4} leɛ^{0} kwa^{1}‖ tʰa^{3} kwa^{1} leɛ^{0} ɥeɛ^{1} tsʰi^{3} tsjɚ^{4}| na$^{4\to1}$ xwɤʌ^{4} tou^{4} pa^{3} tʰa^{3} leɛ^{0} au^{3} kwɤʌ^{3} leɛ^{0} ɥeɛ^{1} tsin^{3} ‖ mɤʌ^{1} ljau˞^{0}| pei^{1} fəŋ^{1} mou^{3} fɐ˞^{1} la^{0} | tou^{4} la^{1} tau^{3} la^{0} ‖ təŋ^{3} la^{0} xwɚ^{4}| tʰai^{4} jaŋ^{0} tsʰwan^{1} tʂʰu^{0} lai^{0} i^{2} ʂai^{4} | na$^{4\to1}$ xwɤʌ^{4} kan$^{3\to2}$ tsin^{0} pa^{3} tʰa^{3} leɛ^{0} au^{3} tʰwɤʌ^{1} la^{0} ‖ pei^{1} fəŋ^{1} pu^{1} tʂʰəŋ^{2} ɻən^{4} pu^{1} tʂuŋ^{1} la^{0} | xai^{2} ʂɻ̩^{4} tʰai^{4} jaŋ^{0} leɛ^{0} pən^{3} ʂɻ̩^{0} ta^{4}‖

Orthographic version

太阳跟北风嘞故事

有一回，北风跟太阳正给那儿比谁嘞本事大。正说住嘞时候儿，来啦哟过路嘞，身上穿啦个袄。他俩都商量好啦，谁先让这个过路嘞把他嘞袄脱喽，都算谁嘞本事大。北风都给那儿憋住，使劲儿嘞刮。他刮嘞越起劲儿，那货都把他嘞袄裹嘞越紧。末了儿，北风某法儿啦，都拉倒啦。等啦会儿，太阳蹿出来一晒，那货赶紧把他嘞袄脱啦。北风不承认不中啦，还是太阳嘞本事大。

References

Bao, Z. (1990). Fanqie languages and reduplication. *Linguistic Inquiry*, 21(3), 317–350.

Boersma, P. & Weenink, D. (1992–2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved from http://www.praat.org/

Chao, Y.-R. (1948). *Mandarin primer*. Cambridge: Harvard University Press.

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

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

Duanmu, S. (1990). A formal study of syllable, tone, stress and domain in Chinese languages. Ph.D. dissertation, MIT.

Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford: Oxford University Press.

Duanmu, S. (2007). *The phonology of Standard Chinese* (2nd edition). Oxford: Oxford University Press.

Duanmu, S. (2008). *Syllable structure: The limits of variation*. Oxford: Oxford University Press.

Hamann, S. R. (2003). *The phonetics and phonology of retroflexes*. Utrecht: LOT.

He, W. (2005). 中原官话分区(稿) [Classification of Zhongyuan Mandarin (a draft)]. *方言* [Dialect], 2, 136–140.

Hu, F. (2013). 降峰双元音是一个动态目标而升峰双元音是两个目标：宁波方言双元音的声学与发音运动学特征 [Falling diphthongs have one dynamic target but rising diphthongs have two static targets: On the diphthong production in Ningbo Chinese]. *语言研究集刊* [Bulletin of linguistic studies], 1, 12–37.

Karlgren, B. (1915–1926). *Études sur la phonologie chinoise*. Leyde: E.-J. Brill.

Ladefoged, P. & Johnson, K. (2011). *A course in Phonetics* (6th edition). Wadsworth: CENGAGE Learning.

Lee, W.-S. (1999). An articulatory and acoustical analysis of the syllable-initial sibilants and approximants in Beijing Mandarin. *Proceedings of the International Congress of Phonetic Sciences 1999*, San Francisco, 413–416.

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

Lee, W.-S. & Zee, E. (2008). Prosodic characteristics of the neutral tone in Beijing Mandarin. *Journal of Chinese Linguistics*, 36(1), 1–28.

Lee, W.-S. & Zee, E. (2014). Chinese Phonetics. In C. T. J. Huang, Y. H. A. Li, A. Simpson. (Eds.), *The handbook of Chinese linguistics*, 369–399. Somerset: Wiley.

Lee-Kim, S.-I. (2014). Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association*, 44(3), 261–282.

Liu, D. (1997). 开封方言记略 [Kaifeng dialect – A précis]. *方言* [Dialect], 4, 271–285.

Lu, Y. (1995). *普通话的轻声和儿化* [Neutral tone and retroflex suffixation in Mandarin]. Beijing: Commercial Press.

Port, R. & O'Dell, M. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13, 455–471.

Mei, T.-L. (1977). Tones and tone sandhi in 16th century Mandarin. *Journal of Chinese Linguistics*, 5, 237–260.

van de Weijer, J. & Zhang, J. (2008). An X-bar approach to the syllable structure of Mandarin. *Lingua*, 118(9), 1416–1428. doi:10.1016/j.lingua.2007.09.006

Wang, H. (1999). *汉语非线性音系学* [Chinese Non-linear Phonology]. Beijing: Beijing University Press.

Wang, J. Z. (1993). The geometry of segmental features in Beijing Mandarin. Ph.D. dissertation, University of Delaware.

Wang, L. (1963). *汉语音韵* [Traditional Chinese phonology]. Reprinted 2003. Beijing: Zhonghua Shuju.

Xu, Y. (2013). ProsodyPro – A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France, 7–10.

Yi, L. & Duanmu, S. (2015). Phonemes, features, and syllables: Converting onset and rime inventories to consonants and vowels. *Language and Linguistics*, 16(6), 819–842.

Yip, M. (2003). Casting doubt on the Onset-Rime distinction. *Lingua*, 113(8), 779–816. doi:10.1016/S0024-3841(02)00130-4

You, R., Qian, N. & Gao, Z. (1980). 论普通话的音位系统 [Phonemic system of Standard Mandarin]. *中国语文* [Studies of the Chinese language], 5, 328–334.

Zhang, J. (2007). A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*, 16(4), 259–302. doi:10.1007/s10831-007-9016-2

Zhang, J. (2010). Issues in the analysis of Chinese tone. *Language and Linguistics Compass*, 4(12), 1137–1153. doi:10.1111/j.1749-818x.2010.00259.x

Zhang, Q., Chen, T. & Cheng, Y. (1993). *河南方言研究* [A study of Henan dialects]. Kaifeng: Henan University Press.

CHAPTER 3

HOW TO PRONOUNCE A LOW TONE: A LESSON FROM KAIFENG MANDARIN[22]

3.1 Introduction

Chinese languages are known for their lexical tones, the syllabic tone melodies which form part of the phonological representations of words. These tone categories are conventionally represented in Chao digits with five distinctive pitch levels, ranging from 1, the pitch floor, to 5, the pitch ceiling (Chao 1968). Standard Mandarin, for instance, contrasts such tone melodies, [55, 35, 214, 51], designated Tone 1–Tone 4, respectively. Many studies have shown that their phonetic shapes can be quite variable, which is particularly true for Tone 3, a low tone. For one thing, Tone 3 is realized as [21] when followed by a non-low tone and as [214] in pre-pausal position. Moreover, when followed by another Tone 3, it is realized as [35] (e.g., Chao 1968). Phonologically, [21] is commonly represented as L, while [214] and [35] often give rise to a trailing H-tone in the representation (Duanmu 2007: 236–237). By contrast, the high tone (Tone 1) is the most stable one among the four tones. This asymmetry between the H and L tone melodies of Standard Mandarin has been addressed in a number of studies which together suggest that the L is more marked, more variable and less salient than H. First, van de Weijer & Sloos (2014) summarized the order of acquisition of Mandarin tones as Tone 1 > Tone 4 > Tone 2 > Tone 3, indicating that Tone 1 is the easiest to acquire and Tone 3 the hardest. Second, a brain-damaged Mandarin speaker was observed to produce only high level tones (Liang & van Heuven 2004), suggesting that H is the default tone. Third, typological surveys of Chinese languages show that H tone melodies are substantially more frequent than L tone melodies (Jiang 1999) and less often subject to

---

[22] This manuscript has been accepted by *Journal of the International Phonetic Association* as Wang, L., Gussenhoven, C. & Liang, J. (to appear) How to pronounce a low tone: A lesson from Kaifeng Mandarin.

sandhi processes (Lu 2001). Lastly, nearly level pitch is more likely to be parsed as H by Chinese listeners, unless it is very low (Whalen & Xu 1992), suggesting that L tones are less salient than H tones and as a result need some form of enhancement.

To a large extent, the apparently adverse conditions that beset the phonetic realization of Mandarin Chinese Tone 3 also apply to intonational L-tones. Significantly, L-tones in accented syllables, ‘starred’ L-tones, tend to be pronounced with more care than other L-tones and resemble the L-tone of Mandarin. Pierrehumbert (1980: 68) observes that while more prominent realizations of H* will have higher f0, those for L* should have decreased f0, as illustrated in her Figure 2.9. Perceptual confirmation for Dutch was provided by Gussenhoven & Rietveld (2000), who found that L* needs to have lower pitch and H* higher pitch for perceived surprise on the part of the speaker to increase, a result attributed to increased pitch range in the signaling of surprise. Earlier, Leben (1976) described the realization of English L* as a ‘dip’. Importantly, as noted by Pierrehumbert (1980: 69), lowering of L* is limited by a nearby low baseline and pitch range expansion will thus quickly lead to saturation at the low end. As a result, intonational L* may be enhanced by other features. A fall from mid-pitch preceding the low pitch, for instance, has been reported for European Portuguese (Vigário & Frota 2003), Stockholm Swedish (Bruce 1977), and the Borgloon variety of Limburgish (Peters 2007; see also Gussenhoven 2007: 263). Quite analogously to the analyses of Tone 3 (see below), there may be uncertainty over the attribution of the pre-L* fall to a H-tone, as opposed to allowing a phonetic implementation rule to create it (for Swedish, see also Riad 1998: 225). The interpretation of L*-lowering as a form of hyperarticulation is supported by the fact that non-starred L-tones may lack these enhancing features and instead be subject to undershoot, as illustrated by Arvaniti (2016: 11), who notes that the L-tones of the Romani interrogative contour L* HL% on phrase-final syllable in [laˈtʃʰo] ‘nice’ behave differently. The focal L* shows an f0 dip on a considerably lengthened rhyme, but the final L% is ‘undershot’, a case of contour truncation. This stands to reason, since L* is associated with an accented syllable and as such represents a tonal basis for emphasis. Going by the evidence provided by Chinese lexical Ls and focus-marking L*s, it would thus appear that adding f0 flanks (i.e., pitch movements towards and after the low target) and lengthening the syllable rhyme are effective ways of enhancing L-tones.

The Kaifeng dialect of Mandarin, spoken in east central Henan, China, provides an excellent opportunity to investigate the nature of L-enhancements further. Earlier impressionistic descriptions of the dialect (Zhang, Chen & Cheng 1993; Liu 1997) distinguish four citation tones, represented as [24, 41, 55, 31(2)] (see Table 1). The first three tones represent a rising, a falling and a high level tone respectively, but descriptions of the fourth tone vary. Zhang et al. (1993) regard it as a falling tone starting from the mid-range and ending at the pitch floor, transcribing it as [31], whereas Liu (1997) considers it a dipping tone, transcribing it as [312]. The choice between these two transcriptions will inevitably affect the overall phonetic and phonological analysis. Below, we will provisionally take the fourth tone to be L, assuming a symmetrical four-way system LH, HL, H and L, adopting the representations of tonal phonology (Goldsmith 1976; Duanmu 2007).

The tonal categories in modern Chinese dialects are defined etymologically and referred to in the traditional nomenclature of the four tones in Middle Chinese, *ping*, *shang*, *qu* and *ru*. Each category is sub-classified into *yin* and *yang* tones, depending on the voicing status of the onset consonant (Chen 2000). For ease of reference, each tonal category is assigned a numerical label in the sinological literature. For example, the four tonal categories in Standard Chinese, *yinping*, *yangping*, *shang* and *qu* are labelled T1 to T4, respectively. Often, as in the case of Kaifeng Mandarin, the phonetic shape of the tones used on the words in these etymological classes differ from those in Standard Chinese. For instance, Standard Chinese T4 is a fall from high to low, while T4 in Kaifeng Mandarin is phonologically equivalent to Standard Chinese T3 (see Table 3.1). The phonological analysis of [24, 41, 55] as LH, HL and H, respectively, would seem to be unproblematic. The proposed L for Kaifeng T4 is supported by observations by Chao (1922) and Wang & Liang (2015). In Chao (1922), its pitch curve as obtained by rudimentary instrumentation resembled the pitch curve of the low tone (T3) in Beijing Mandarin. A pilot study by Wang & Liang (2015) suggests that Kaifeng Mandarin L has a low dipping pronunciation with a weakly rising trend in the final portion of the syllable.

Table 3.1 A comparison of traditional tone categories, tone labels and phonological tones in Standard Chinese and Kaifeng Mandarin.

Traditional tone category		*Yinping*	*Yangping*	*Shang*	*Qu*
Tone labels		T1	T2	T3	T4
Phonological tones	Standard Chinese	H	LH	L	HL
	Kaifeng Mandarin	LH	HL	H	L

The goal of this investigation is, first, to provide detailed acoustic information on the realization of Kaifeng Mandarin L and, second, to evaluate the variation that is encountered in terms of possible enhancements of a low tone.

3.2 Method

3.2.1 Design

Kaifeng Mandarin syllables have a single optional consonant in the onset, while the rhyme has an obligatory vowel and an optional coda nasal [n] or [ŋ], which may appear after monophthongs. A corpus of 35 minimal tonal quadruplets was prepared, shown in Table 3.2. All of them have an onset consonant, of which 18 are aspirated plosives or affricates, six unaspirated plosives or affricates, seven fricatives and four sonorants; seven quadruplets have a nasal coda and four have diphthongal rhymes, while the remaining 24 have monophthongal vowels without a coda. Vowel height is balanced by including both high and low vowels in the rhyme, by the side of three quadruplets with retroflex approximants in the rhyme (Lee-Kim 2014).

Table 3.2 Stimuli used in the audio recording including 140 characters varying in tone, segmental makeup and syllable structure.

IPA	LH	HL	H	L	IPA	LH	HL	H	L
pa	八	拔	把	爸	ʂu	书	熟	叔	树
ma	抹	麻	马	骂	xu	呼	湖	虎	互
ta	搭	达	打	大	ly	绿	驴	吕	滤
tʂʅ	知	直	纸	智	tɕy	菊	局	举	锯
tʂʰʅ	尺	迟	齿	翅	tɕʰy	区	渠	娶	去
ʂʅ	湿	石	史	事	ɕy	需	徐	许	婿
pi	逼	鼻	鄙	弊	tsʰai	猜	财	彩	菜
pʰi	批	皮	匹	屁	pʰau	抛	袍	跑	炮
ti	低	敌	抵	弟	xau	蒿	豪	好	号
tʰi	梯	题	体	替	tʂʰou	抽	仇	丑	臭
li	力	厘	礼	例	tʰan	贪	谈	毯	探
tɕi	积	集	几	计	pʰin	拼	贫	品	聘
tɕʰi	欺	旗	起	气	tʰaŋ	汤	糖	躺	烫
pʰu	铺	葡	谱	瀑	tʂʰaŋ	昌	常	场	唱
fu	福	服	府	父	tɕʰiŋ	清	晴	请	庆
tʰu	秃	图	土	兔	ɕiŋ	星	行	醒	幸
lu	鹿	卢	鲁	路	tʰuŋ	通	铜	统	痛
tʂʰu	出	除	础	处					

3.2.2 Participants

Ten native speakers of Kaifeng Mandarin participated in the experiment. All of them were born and raised in Kaifeng and lived in the inner city at the time of the experiment. None of them reported any speaking or hearing deficiencies. The age and gender distribution is as follows: 27–30 (M = 4), 31–50 (M = 1, F = 1), 51–60 (M = 2, F = 2). They use Kaifeng Mandarin in their daily communication and nine of them can speak Standard Chinese at varying proficiency levels (see Table 3.3 for more detailed information). Among the ten speakers, ZBR cannot speak Standard Chinese at all. Four young speakers (XYC, FGH, TJQ, QHC) are native bilinguals of both Kaifeng Mandarin and Standard Chinese. The Standard Chinese proficiency for the other older speakers is either near-native or recognizably Kaifeng-accented.

Table 3.3 Ten participants with information of gender, age, and their self-reported proficiency in Standard Chinese.

Initial	Gender	Age	Standard Mandarin	Initial	Gender	Age	Standard Mandarin
LYH	F	38	near-native	TJQ	M	28	native
WCX	F	51	accented	XYC	M	29	native
DLL	F	54	accented	ZWP	M	44	near-native
QHC	M	27	native	CJY	M	51	near-native
FGH	M	28	native	ZBR	M	60	no

3.2.3 Procedure

The audio recordings were conducted in quiet rooms, where the subjects were seated in front of a laptop. The reading materials were visually presented in random order, one per slide, at the centre of the screen with the help of Prorec software (Huckvale 2014). The participants read each word once and were asked to repeat it if they were uncertain about the quality of their first pronunciation. Each word was presented twice, resulting in 280 slides (140 syllables*2 repetitions). After a participant had read a word, the experimenter would click a button to make the slide with the next character appear. The audio data were digitized via a Samson C01U Pro USB condenser microphone placed at 10 cm distance from the speaker's mouth with a microphone stand. The starting and ending time of each presentation slide were collected for automatic segmentation of the recordings of each of the 280 slides.

3.2.4 f0 extraction

For each speaker, 280 recorded words were automatically segmented in Prorec according to the starting and ending time of each slide. Phonetic annotation and f0 extraction of these 2,800 speech files were carried out in Praat (Boersma & Weenink 1992–2015). Rhymes for every syllable were identified in the waveform and spectrogram and boundaries were manually drawn and labeled (see Figure 3.1). With the aid of a Praat script (Xu 2013), vocal pulse marking was checked and manually corrected. The raw f0 measurements were smoothed with a trimming algorithm (see Appendix 1 in Xu 1999) and output files with time-normalized f0, duration, maximum f0, minimum f0 and f0 velocity were automatically generated. 149 recordings, 5.32% of the total, were excluded from further analysis due to mispronunciations, while 15 stimuli (13 Ls and 2 HLs) had so much jitter that they could not be included in the f0 analysis. Their phonetic properties will be discussed in § 3.3.5. From the 2,636 remaining speech files, 21 equidistant f0 measurements (in Hz) were taken from the rhymes.

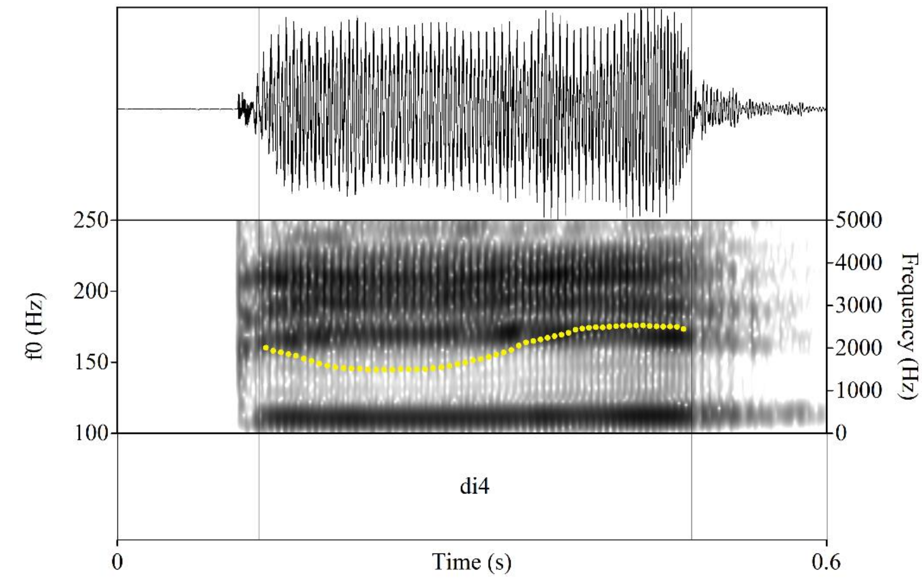


Figure 3.1 Phonetic annotation of a recorded word. The rhyme of each syllable is manually labeled in the Praat textgrid.

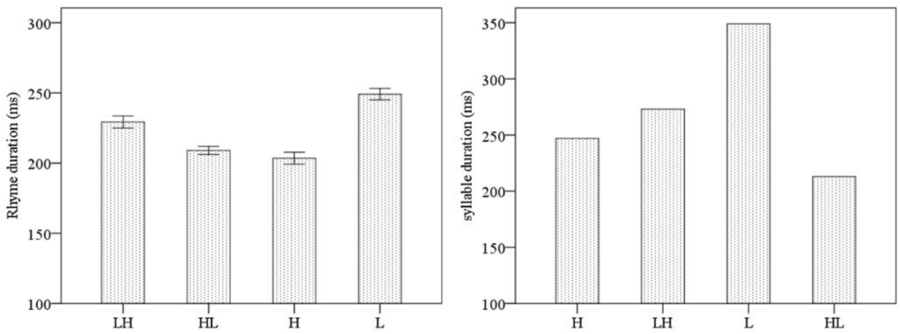
3.2.5 f0 normalization

In order to eliminate speaker pitch range differences (mean pitch range 10.39, sd = 1.22 in ST with a reference value of 100 Hz), we transformed the raw f0 data into speaker-specific z-scores in three steps. First, for each speaker, values were averaged point-by-point across the syllables in each of the four tone categories. Second, the grand average $f0_{mean}$ and standard deviation $f0_{sd}$ of the f0 measurements per speaker over the dataset obtained in step 1 (21 averaged points * 4 tone categories) were calculated. Third, raw f0 measurements for all the stimuli of the speaker $f0_x$ (in Hertz) were transformed into speaker-specific z-scores, using $z = (f0_x - f0_{mean})/f0_{sd}$ (Rose 1987). After the normalization, the z-normalized f0 values were used for graphical and statistical purposes.

3.3 Results

3.3.1 Rhyme duration

In terms of rhyme duration, Tone L is the longest (249 ms, sd = 52.56), followed by LH (229 ms, sd = 55.99), while HL and H are the shortest (209 ms, sd = 37.31 and 204 ms, sd = 54.32, respectively). A one-way ANOVA yielded a significant effect of tone on rhyme duration ($F(3, 2632) = 111.62$, $p < .001$). Dunnett's T3 post hoc tests revealed that significant differences were found between the members of each pair of tones except between HL and H.



(a) (b)

Figure 3.2 Durations of four tone categories ordered according to their etymological class in Kaifeng Mandarin (panel a) and Standard Chinese (panel b, data from Xu 1997). The data in panel (a) are from ten speakers (N = 2636; LH = 663, HL = 652, H = 656, L = 665). Bars are ±2 standard errors.

Figure 3.2 presents the etymological tone classes and the phonological tones in the two dialects together with their mean durations. Because the phonological counterpart of Kaifeng Mandarin L in Standard Mandarin, T3, a low tone, is the longest of its set of four tones, the longer duration of Kaifeng Mandarin L is associated with its phonology, not with its etymological tone class. Standard Mandarin T4, the etymological equivalent of Kaifeng Mandarin L, is in fact the shortest of the four tones in that dialect and is phonologically HL.

3.3.2 f0 contours

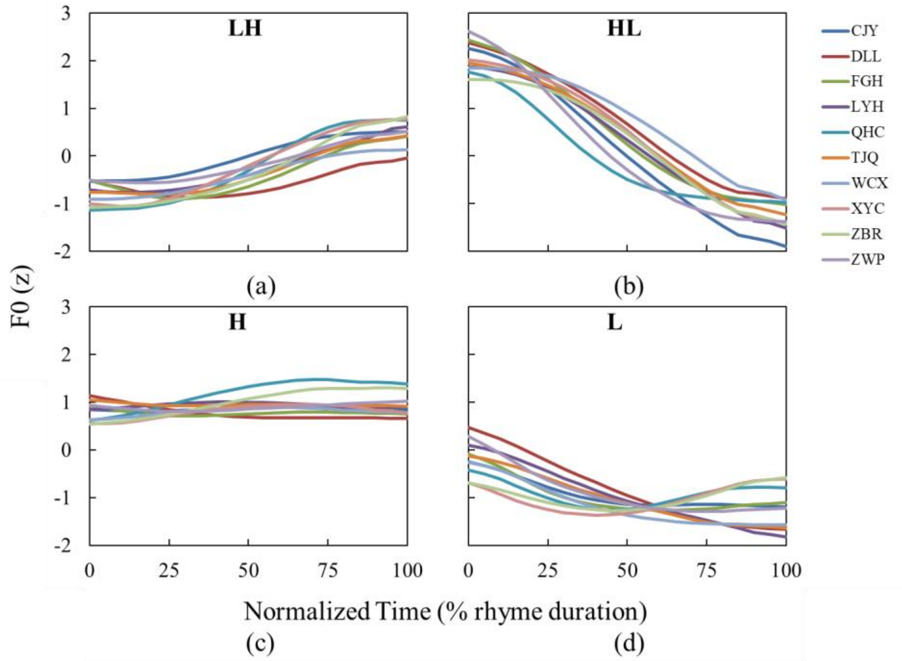


Figure 3.3 Tonal contour from ten speakers (averaged across tokens). The four tones are grouped in separate panels (a–d).

The mean f0 contours of the four tones are shown in panels (a–d) in Figure 3.3. In each panel, the ten colored curves were obtained by averaging over the maximally 70 tokens per tone per speaker and are plotted as functions of normalized time (% rhyme duration). Figure 3.3 suggests there may be more variation in the tone shape of L than in LH, HL and H, especially in the second half of the f0 contour. LH (panel a) starts with a low f0 onset, and rises to the mid-range. Four speakers (CJY, QHC, WCX and ZBR) have rising contours, while the other six speakers have weakly dipping contours, whereby the f0 curves fall slightly during the initial 25% of the rhyme duration and rise thereafter. All ten speakers produced both consistently rising and weakly dipping realizations of LH, with 33% of the 663 LH melodies exhibiting rising contours, the rest being weakly dipping. HL (panel b) starts with the highest f0 onset and falls sharply to the end of the syllable. The shape of the contours is consistent across the ten speakers, but some variation occurs in the duration of the target for the L-tone in HL, as depicted in one extreme case produced by QHC, where the fall begins in the first half of the rhyme. Finally, H (panel c) has a high onset and remains fairly constant throughout the rhyme. Two speakers (QHC and ZBR) have a slightly rising contour. § 3.3.3 presents the way in which the variation in L deviates from that in the other tone categories.

3.3.3 L-specific variation

A widely employed method for revealing the ways in which the data for the four tones represented in Figure 3.3 differ from each other is by determining their f0 shape and f0 height (Jongman et al. 2006), more specifically their f0 slope and mean f0 (z-score). Accordingly, mean f0 was obtained by averaging the 21 points, while f0 slope was defined as the coefficient of the linear regression line for the 21 points of each tone with the corresponding real timestamps as the independent variable. Following Peng (2006), we plot the mean f0 against the overall f0 slope to access the variability of each token. Figure 3.4 is a scatterplot for the citation tones in Kaifeng, with each token represented as a plot symbol. As expected, realizations of H have higher f0 (0.90, sd = 0.45), while realizations of L are lower (–0.97, sd = 0.29), with plot symbols located above and below the horizontal axis, respectively. Mean f0 of HL (0.33, sd = 0.42) is generally greater than that of LH (–0.28, sd = 0.39). Results of a one-way between-subjects ANOVA show that there was a significant effect of tonal category on mean f0 ($F(3, 2632) = 2794.58$, $p < .001$) and post-hoc comparisons using the Dunnett's T3 test indicated that mean f0 of all four tones differed significantly from each other. As for the horizontal dispersion, HL and LH are discretely located on the left and the right of the central vertical axis (HL: –19.74, sd = 5.10; LH: 7.25, sd = 3.42). The relatively flat slopes for LH are due to the initial fall of the weakly dipping contours (see § 3.3.2). The areas for the relatively flat slopes of H (1.14, sd = 3.30) and LH overlap somewhat. An overall analysis of variance yielded a significant tone effect on f0 slope ($F(3, 2632) = 4898.29$, $p < .001$); Dunnett's T3 post hoc tests revealed significant differences in f0 slope between each pair of tone categories. To examine the variation of the contour shapes, ANOVAs were performed for each of the four tones separately, with f0 slope as the dependent variable and speaker as the independent variable. There was a significant speaker effect on f0 slope for each of the tones (LH: $F(9, 653) = 95.17$, $p < .001$, $\eta^2 = .567$; HL: $F(9, 642) = 50.94$, $p <$

.001, $\eta^2 = .417$; H: $F(9, 646) = 114.58$, $p < .001$, $\eta^2 = .615$; L: $F(9, 655) = 288.39$, $p < .001$, $\eta^2 = .798$). Generally, therefore, inter-speaker variation in f0 slope is greater than intra-speaker variation, but the effect size is largest for L. It indicates that 80% of the total variation in f0 slope of L is accounted for by variation between speakers, as opposed to 40–60% for the other tones.

As can be seen in Figure 3.4, L is the only tone whose range of realizations reaches from a slope as steeply falling as that of HL through a slope as flat as that of H to a slope as slightly rising (or dipping) as that of LH. Further inspection of the data shows that the 80% explained variation for L in the ANOVA is due to the existence of three speaker-specific f0 slopes for the second half of the contour in the case of L (see panel d of Figure 3.3). That is, dipping, a contour shape involving flanking movements on both sides of the low targets, is utilized only by speakers XYC, QHC and ZBR, while low falling is typically seen in speakers DLL, TJQ and LYH, where f0 gradually falls towards the low pitch target. A fall with a sustained low stretch after the low point of the fall is used by WCX, FGH, ZWP and CJY. Such speaker-specific variation has no counterpart in the other three tone categories. While it is true that the contours for H include rising as well as falling slopes and those for LH include rising slopes as well as almost flat shapes due to the existence of an initial falling slope, these variations are primarily caused by variation *within each of the speakers*, rather than by systematic variation between the speakers. This intra-speaker variation was captured by the standard deviation of f0 slope calculated per subject per tone. The averaged standard deviation for L is indeed the smallest (0.275), followed by HL (0.348), LH (0.356) and H (0.426). A one-way within-subjects ANOVA showed that the effect of tone was significant ($F(3, 27) = 17.646$, $p < .001$), indicating that the four tones differ in intra-speaker variation. Paired sample *t*-tests indicated that intra-speaker variation for L was significantly smaller than each of the other three tones (L-HL: $t(9) = 6.148$, $p < .001$; L-LH: $t(9) = 4.06$, $p = .003$; L-H: $t(9) = 6.597$, $p < .001$). In the next subsection, the smaller intra-speaker variation in the realization of L is considered in more detail in order to establish how consistent the three speaker subgroups are in the realization of the subtypes.

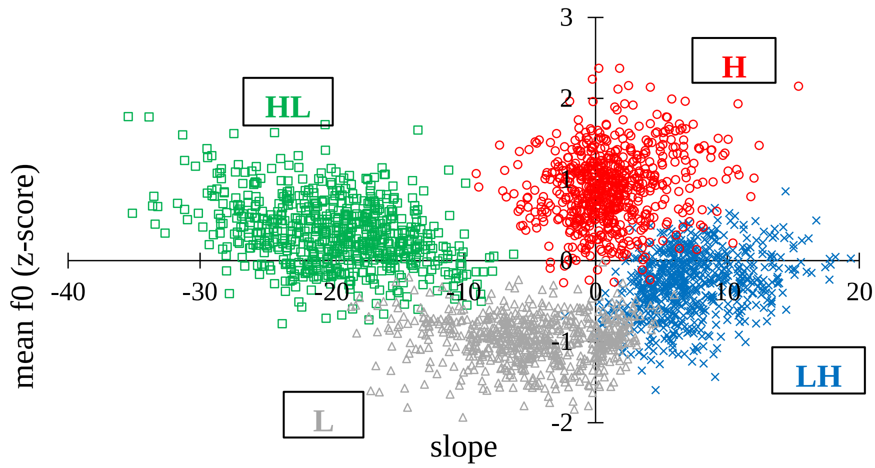


Figure 3.4 Scatter plot for the citation tones in Kaifeng Mandarin with the horizontal axis showing the slope and the vertical axis showing the mean f0. Ten speakers; N = 2636, LH = 663, HL = 652, H = 656, L = 665.

3.3.4 Variation of the L tone

In the case of L in particular, f0 slope is not a realistic measure of f0 shape. As a result of the reduction of L to a single slope in the above analysis, the data for L overlap somewhat with those for HL and LH. In order to represent the variation for L more realistically, separate regressions were carried out over the first and second sets of 11 sampling points, respectively, such that two coefficients were obtained for each realization. A dipping realization is defined by a negative slope for the first half of the contour and a positive slope for the second half, a falling contour by a sequence of two negative slopes, while a falling plus a low level realization will be assigned a negative slope for the first half and a flatter slope for the second half (whether positive or negative). Panels (a–c) in Figure 3.5 exemplify how these data characterize the dipping, falling and falling+lengthening subtypes on the basis of selected tokens.

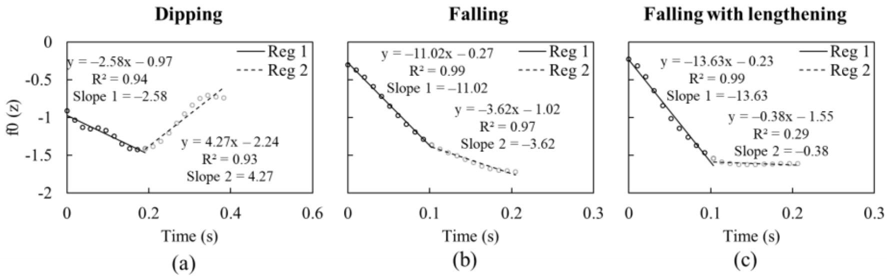


Figure 3.5 Examples of the dipping (a), falling (b) and falling+lengthening (c) subtypes of L, with two linear regression lines superimposed on the f0 sampling points. Slope 1 (solid line) and Slope 2 (dashed line) are the coefficients of the first and last 11 sampling points with their corresponding real time points as the independent variable.

Figure 3.6 shows scatterplots of the realization of the three subtypes of L for the dipping, falling and falling+lengthening subtypes of L in panels (a–c). Each speaker was categorically assigned to a slope subtype (dipping for XYC, QHC and ZBR; falling for DLL, TJQ and LYH; lengthened falling for WCX, FGH, ZWP and CJY), based on the tone shapes in Figure 3.3. Data points for each subtype turn out to fall within a unique band along the horizontal axis, between 0 and 10 for the dipping subtype (panel a), between –10 and 0 for the falling subtype (panel b), and between –5 and 5 for the lengthened falling subtype (panel c). With just one exception, all the data points of the dipping subtype are located on the right of the central vertical axis, indicating that virtually all tokens of L produced by XYC, QHC and ZBR are dipping (Slope 2: N = 205, mean = 5.27, sd = 2.20). Similarly, with only three exceptions, the data points of the falling subtype are positioned on the left of the central vertical axis, indicating that DLL, TJQ and LYH realize L as a falling contour (Slope 2: N = 194, mean = –6.51, sd = 2.99). Finally, Slope 2 of the lengthened falling subtype (panel c) as produced by the four remaining speakers is located around the central vertical axis (Slope 2: N = 266, mean= –0.24, sd = 2.46). By contrast, all three subtypes begin with a falling slope (Slope 1). These data show in more detail that the intra-speaker variation is very much smaller than the inter-speaker variation and that the inter-speaker variation concerns the treatment of L in the second half of its syllable rhyme. Results from a one-way between-subjects ANOVA revealed a significant group effect on Slope 2 ($F(2, 662) = 1062.71$, $p < .001$). Dunnett's T3 post hoc tests revealed that significant differences were found between each pair of the three subtypes.

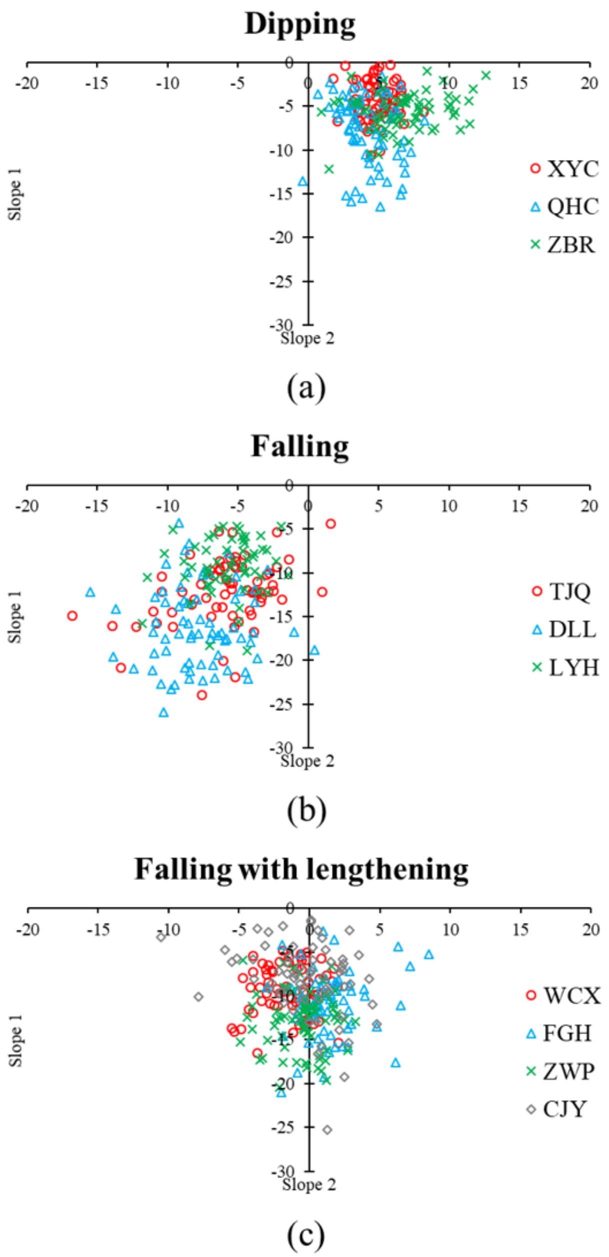


Figure 3.6 Intra-speaker variation in the realization of L. Each realization is represented as a point in the Slope 1/Slope 2 plane. Ten speakers are grouped separately in Panel (a–c), based on the subtypes they were assigned to. Panel (a): XYC (N = 67), QHC (N = 69) and ZBR (N = 69); panel (b): TJQ (N = 67), DLL (N = 70) and LYH (N = 57); panel (c): WCX (N = 62), FGH (N = 67), ZWP (N = 70) and CJY (N = 67).

To see if there was a durational effect of slope type, the Pearson correlation coefficient between rhyme duration and Slope 2 was established. It turned out to be positive and significant (r = .43, N = 665, p < .001), indicating that rising Slope 2s tend to be longer than level and falling ones. A scatterplot summarizes the results (Figure 3.7). This suggests that the longer duration of L and the presence of a flank to the right of the low target go together to a certain extent. This effect is further confirmed by the results from a one-way between-subjects ANOVA, which yielded a significant effect of the three different subtypes on the rhyme durations of syllables (F(2, 662) = 89.14, p < .001; dipping: N = 205, mean = 280.38, sd = 56.41; lengthened falling: N = 266, mean = 247.94, sd = 29,89; falling: N = 194, mean = 217.95, sd = 53.83). Dunnett's T3 post hoc tests revealed significant differences for each pair of subtypes.

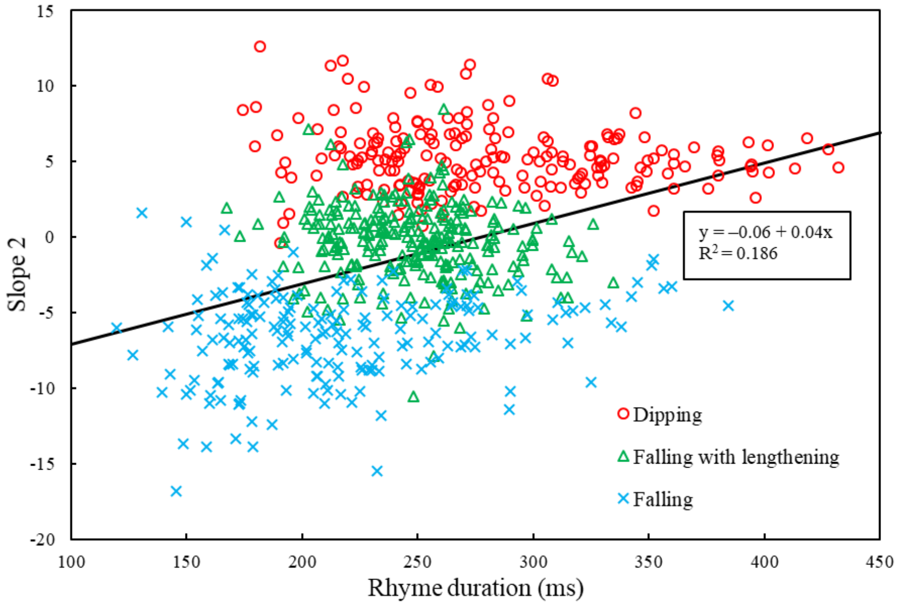


Figure 3.7 Slope 2 as a function of rhyme duration. Subtypes indicated by the colour of the plot symbols (N = 665).

The histogram in Figure 3.8, which bins the lowest values at each of the 21 sampling points for the three subtypes of L, confirms that the location of the f0 minimum within the rhyme reflects the contour differences among the subtypes. The median of the dipping subtype lies at 45% of the rhyme duration, i.e., just before the halfway point. The lengthened falling subtype has a more variable location of the f0 minimum, including cases where it lies right at the end, when the level stretch is in fact slightly descending. The falling subtype has its most frequent location at the rhyme end, as expected. Overall, the location of the f0 minimum varies from the 25% position within the rhyme to the rhyme offset.

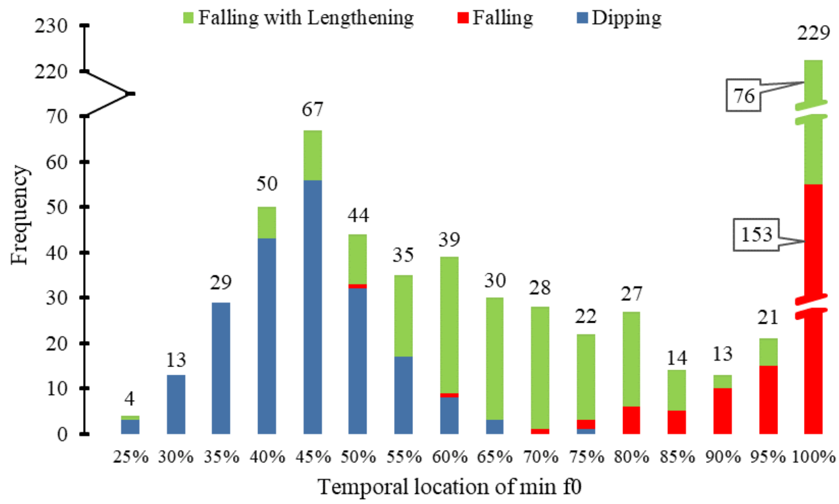


Figure 3.8 Frequency distribution of the f0 minima of the three subtypes of L over 16 points equally spaced over the last 75% of the rhyme, represented as percentages of the normalized rhyme duration on the x-axis. Red = falling subtype; green = lengthened falling subtype; blue = dipping subtype (N = 665).

3.3.5 Creak

Realizations of Kaifeng Mandarin L show varying degrees of creak. Creaky voice is characterized by the 'irregularity of the interval between consecutive glottal pulses,' technically known as 'jitter' (Ladefoged 2003: 172). Figure 3.9 displays the waveforms and spectrograms of realizations with creak obtained from three speakers. Panel (a) shows a realization of L by ZBR in which creak occurs mainly in the central portion of the rhyme. The amplitude of the waveform decreases considerably at the point creak begins. Although there is creak, the intervals between the pulses are regular, and the utterance can still be analyzed in terms of f0. The yellow speckles superimposed on the spectrogram are the pitch track obtained by manually labeling and calculating the intervals of the glottal pulses and smoothed in Praat (Boersma & Weenink 1992–2015). Panel (b) shows a realization of L by FGH with more salient creak. Irregular pulses can be found from the 20% location to the offset of the rhyme. Panel (c) is a realization by LYH. The waveform and spectrogram show that from the onset of the periodicity to the 70% location of the rhyme, intervals between the pulses are brief and regular, while in the last 30% of the rhyme they become longer, such that f0 can no longer be measured.

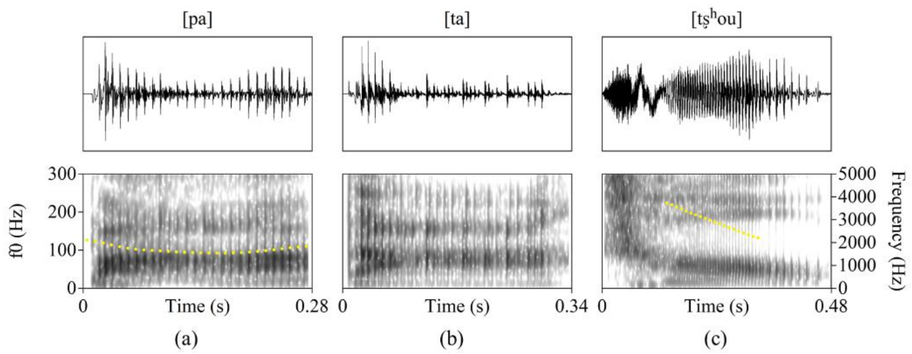


Figure 3.9 Creaky voice in realizations of L. Panel (a) shows the waveform and spectrogram for a syllable [pa] uttered by speaker ZBR, panel (b) for a syllable [ta] produced by speaker FGH, and panel (c) for a syllable [tʂʰou] generated by speaker LYH.

Due to a lack of quantitative measurement of the gradience within creaky phonation, it is difficult to show how creak is distributed across the tone categories. Impressionistically, creak also occurs in LH and HL, though less often, while some realizations reveal a region of low amplitude and breathiness. Only 15 of our discarded tokens show a great disturbance of the periodicity (e.g., panels b–c in Figure 3.9), of which 13 are L and 2 are HL.

3.4 Discussion

3.4.1 The analysis of the low tone as L

Our investigation of the four tone categories of Kaifeng Mandarin has established that:

(1) the rhymes of syllables with L (i.e. Tone 4) are significantly longer than those of the other three tones;

(2) a greater inter-speaker variation is found in the tone shapes of L than those of LH, HL and H;
(3) a smaller intra-speaker variation is found in the case of L than in the case of LH, HL and H; the variation in L can be classified into three subtypes, dipping, falling and falling with lengthening, used by three subgroups of speakers;
(4) the slope of the f0 contours of L in the second half of the rhyme (Slope 2) is positively correlated with rhyme duration: rhymes with descending slopes in the second half are shorter than those with ascending slopes;
(5) over all three L-subtypes, the time stamp of the f0 minimum lies within the last 75% of the duration of the rhyme, with a bias towards 45% in the dipping contour and towards 100% in the falling contour, while f0 minimum varies between 40% and 100% in the lengthened falling contour;
(6) creaky phonation is a frequent feature of L, but it is neither consistent for L nor exclusive to L.

Unlike the interpretation of the rising, falling and high tones as LH, HL and H respectively, the interpretation as L of the dipping, falling and falling+lengthened pronunciations of Tone 4 is not an immediately obvious analysis. Our investigation has revealed that these different forms reflect inter-speaker variation. The dipping contour, [312] in Liu (1997), was used by three of our ten speakers, while contour [31] of Zhang et al. (1993) was used by the other seven, four of whom in fact used a falling contour with a lengthened low target, representable as [311] (see Figure 3.3). Under the incorrect assumption of an HLH tone, which might at first sight be an appropriate representation, the phonetic implementation would have to locate the low target as early as at 25% of the rhyme duration or as late as the rhyme offset, in the latter case without any implementation of the final H-target. Moreover, the brief duration of the initial fall suggests that initial H, too, is a dispensable element. These facts suggest that the most plausible hypothesis is that the f0 variation around the L target does not have a phonological origin. Before supporting this conclusion with phonological arguments, in § 3.4.2 we attempt to explain the phonetic variation as different ways of enhancing the low target.

3.4.2 Explaining the variability in the realization of the isolated L tone

We suggest that the explanation of the variety of f0 contours for phrase-final pronunciations of L lies in the intrinsic lack of salience of low pitch and the resulting need for its phonetic enhancement. An *a priori* expectation about the realization of a low tone is that it should have a low target. However, low targets are not easily lowered, unlike the way high targets can be raised (Pierrehumbert 1980: 68). Moreover, while a high f0 target typically stands out from a lower-pitched context, low pitch may be ambiguous between a tonal target and a low-pitched context, in particular in utterance-final position if utterances are expected to end in low pitch. Three types of enhancement are available to speakers. First, the low target (and the sonorant segment it coincides with) may be lengthened. In our data this is evident in the greater duration of L compared to the other three tone categories, as documented in Figure 3.2. We also found that Ls with an ascending Slope 2 are longer than those with a descending Slope 2, but even rhymes with a descending Slope 2 (230.90 ms, sd = 47.42) are longer than those of HL (209.02 ms, sd = 37.31) and H (203.52 ms, sd = 54.32) tone, while being equivalent to that of LH (229.25 ms, sd = 55.99), which has been shown to be longer than falling shapes for intrinsic reasons (Ohala & Ewan 1973; Xu & Sun 2002). That is, even if the extra duration is explained as a by-product of the flanking movements, part of the lengthening of L is independent of the f0 contour, as observed with reference to Standard Mandarin Tone 3 by Gussenhoven & Zhou (2013). The long duration of Standard Chinese Tone 3 has in fact been shown to be perceptually relevant in distinguishing it from Standard Chinese Tone 2 (Blicher, Diehl & Cohen 1990).

Second, creaky phonation may replace regular phonation at the point where the lowest f0 is expected, the centre of the rhyme for the dipping subtype and the final portion for the falling subtype and clearly serves as an enhancement of low pitch. A perceptual study by Ding, Jokisch & Hoffmann (2010) demonstrated that a creaky or breathy realization of T3 in Standard Chinese improves its recognition, showing that phonation contributes to the perceptual salience of the L-tone. Similar results are found in Yang (2015), who ran a tone identification experiment with stimuli taken from Standard Mandarin f0 continua produced by gradually altering the f0 contour of one natural tone exemplar into one of the other three in the four-tone set, leaving original phonation properties intact. It showed that speakers utilize phonation cues to identify the four tones, particularly T3, even though the creak in the T3 exemplar was not very obvious. It is possible that L-tones may naturally induce non-modal phonation, since LH and HL also have creak. A similar distribution of creak also holds for the tones in Standard Chinese (Belotel-Grenié & Grenié 2004; Kuang 2017, 2018). In fact, a larger and more clearly asymmetrical distribution of creak among different tone categories has been reported for Cantonese, a language with a more complex tone system (six non-checked tones and three checked tones). Based on a similar visual inspection of the disturbance of periodicity depicted in the waveform and spectrogram, Yu & Lam (2014) concluded that creak occurs more frequently in Cantonese T4 ([21/11] in Chao digits) than other tones (24.2% of T4 vs 4.7 % of the corpus). They also reported that Cantonese creaky T4 is identified with a much higher accuracy than non-creaky T4, while listeners also utilize creak in the perceptual separation of T4 and T6 ([22] in Chao digits). This suggests that languages with a greater functional load of f0 in conveying complex tonal contrasts, like Cantonese, tend to rely more on phonation cues in the production and perception of L-tones than languages like Mandarin, where pitch can adequately signal the L.

Third, the slight fall towards the low pitch target in the falling and dipping contours as well as the flanking f0-rise in the dipping contour can be seen as additions that place the low target in relief (Gussenhoven 2007: 256). Although no perception data for the Kaifeng Mandarin tones have been reported, results from studies of Standard Chinese suggest that a low realization lacking f0 movements is more likely to be heard as H than as L (Whalen & Xu 1992; Cao 2010a; 2010b), indicating the lower perceptual salience of a bare, low pitched realization of L. The initial fall is also relevant for the discrimination between L and LH. Perception experiments in Standard Chinese indicate that discrimination of isolated T2 and T3, both of which can be said to have dipping realizations, is cued by the degree of the f0 initial fall, i.e., the frequency change from the onset f0 to the f0 minimum, as well as the timing of the f0 minimum. Specifically, T3 is characterized by a greater f0 initial fall and a later f0 turning point compared with T2 (see Jongman et al. 2006 and references cited there). Thus, a relatively greater f0 initial fall and a later L-alignment may be preferred for a better perceptual separation between a dipping L and a dipping LH.

The pronunciation of words with L under corrective focus suggests that speakers enhance the specific enhancements they use under broad focus, which further strengthens the case for interpreting the subtypes as alternative ways of enhancing L. Thus, the focus for phrase-final L is expressed by lengthening the L-target for speaker ZWP, but by increasing the f0 movement for speaker WC. Specifically, compared to the low-falling realization of L in the broad-focus condition, ZWP lengthens the low f0 level stretch, shown in panels (a–c) in Figure 3.10, in line with his realizations in citation forms (see Figure 3.3), while for WC (panel d–f), additional curvature is created in the focal L, achieved by raising the flanks on both sides of the L-target, resulting in a more dipped realization. This finding is in line with Shih's (1988: 84) informal observation that Northern speakers pronounce Standard Mandarin T3 as falling-rising in isolation or sentence final position, but as low-falling elsewhere, while 'Southern speakers often keep the low-falling pattern even in the final position in casual speech and use the falling-rising pattern only in deliberate, emphatic speech, or in yes-no questions'.

Emphasis in phrase-initial L-tone is generally cued by pitch range expansion of the low tone that raises the beginning of the initial fall, but not of the low target itself, as exemplified in Figure 3.11 for speaker ZWP. It shows pitch contours of L followed by LH, HL and H, with L pronounced in broad focus and corrective focus. In corrective focus, the pitch range is extended at the higher end, while the low target remains unaffected. This is not always true for HL in equivalent circumstances, where the end of the fall may be higher under corrective focus, along with a much higher beginning of the f0 fall.

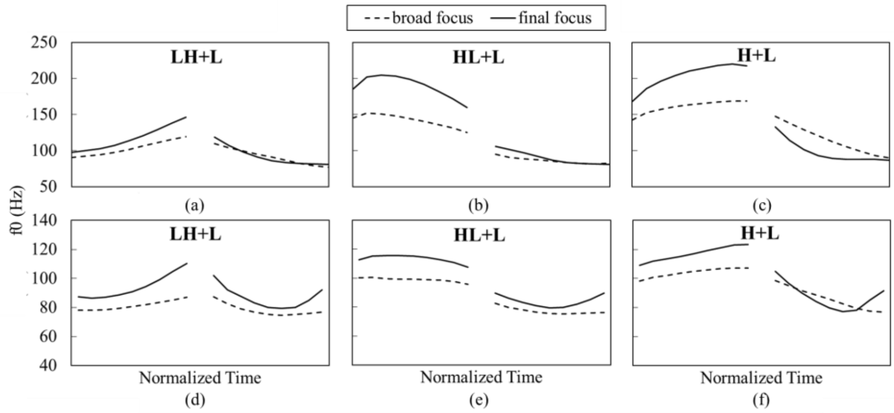


Figure 3.10 Phrase-final L-tones produced in two focus conditions by ZWP (panels a, b, c) and WC (panels d, e, f), with data from Wang (2018). Pitch curves were averaged by ten repetitions per speaker. L-tones preceded by LH, HL and H are presented in separate columns. The vertical axis shows f0 in Hz.

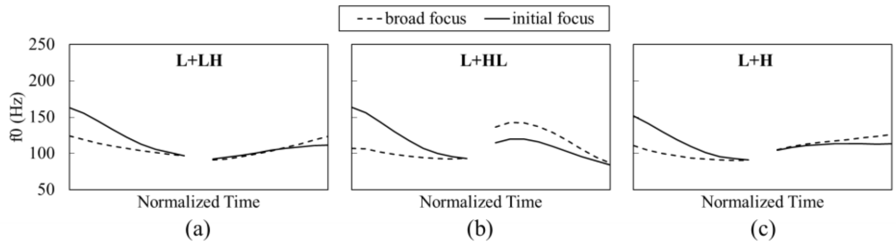


Figure 3.11 Phrase-initial L-tones produced in two focus conditions by ZWP (data from Wang 2018). Also see Figure 3.10.

3.4.3 Phonological arguments for the analysis as L

The phonetic and perceptual data discussed in §3.4.2, argue for an analysis of the low tone as L. There are four further arguments that support this analysis. First, it fits in with a symmetrical four-term set of word-melodies with LH, HL and H. Second, the representation fits a tone sandhi pattern whereby L is inserted between two H word melodies so as to create HL.H, which rule has a parallel in the insertion of H between two L word melodies, leading to LH.L. If our L were to be represented as HLH or ML, the rule would lose its character as an OCP repair, i.e., the separation of identical word melodies consisting of a single tone. Third, if the low tone was analyzed as monosyllabic HLH, it would run counter to the widely reported undershooting of the target of L, quite the reverse of what happens in Kaifeng Mandarin.[23] A fourth argument can be based on the pronunciation of the low tone in context, where it triggers phonetic behaviour that is associated with L-tones in languages more generally. As expected, tones are subject to coarticulatory effects of neighboring tones (cf. Chen 2012). In Kaifeng Mandarin, tonal coarticulation is bidirectional, involving both carry-over and anticipatory effects. The relevant finding is that the anticipatory effect on L is dissimilatory and tends to raise the f0 of a preceding H-tone, as can be seen by comparing the f0 of H, HL and LH before L with that of the same tone melodies before the other three tones, as in Figure 3.12, which displays the f0 contours of the nonce disyllable /mama/ for all 16 tone combinations. In each panel, the first tone is kept constant and the second is varied. As mentioned in our second argument, Kaifeng Mandarin L is replaced with LH before another L, as shown in the thin dashed line in panel (d) of Figure 3.12. This derived LH does not appear to undergo raising due to the following L, but underlying LH as well as HL show quite sizable raising effects before L. In panel (a) of Figure 3.12, the f0 height and f0 offset of LH are raised by the following L compared with LHs before LH, HL and H. Similarly, both f0 peak and f0 offset of HL are raised before L compared with the HLs before LH, HL and H (panel b of Figure 3.12). These raising effects reach back to the initial portions of the rhyme within LH and HL. By contrast, no raising effect is found in the combination of level H with L (panel c of Figure 3.12). The f0 of H before L is only a little higher than that of H before LH, while remaining pretty close to the f0 of H before HL. (Tone sandhi is responsible for replacing H before H with HL, as shown in the thick dashed line in panel c, which we here leave out of consideration.)

This local dissimilatory effect before L-tones has been described as f0 polarization (Hyman & Schuh 1974), anticipatory dissimilation (Gandour, Potisuk & Dechongkit 1994; Xu 1997), H-raising (Connell

[23] Undershooting of non-starred L-tones between H*-tones as a response to time pressure is exemplified by Mexican Spanish, where the low targets move up and down with the targets of the surrounding H*-tones, while also being higher as the H*-tones are closer to each other (Prieto 1998). The distance effect was also reported for Dutch by Ladd & Schepman (2003), while Hanssen (2017: 151) shows quite extreme raising of L in monosyllabic H*LH% in Zeelandic Dutch. See also Arvaniti (2016), which article also provided the model for the title of this paper.

& Ladd 1990), anticipatory raising (Xu 1999) or pre-low raising (Lee, Prom-on & Xu 2017) for a number of languages. For instance, in Standard Chinese, LH and L tend to raise the f0 of the preceding LH, HL and H (Xu 1997). Similarly, in Thai, as reported by Potisuk, Gandour & Harper (1997), high and rising tones are raised by the following low target. In Yoruba, a H-tone before an adjacent L-tone is raised (Connell & Ladd 1990). Also, long-distance raising effects on H induced by nonadjacent L has been reported in Yoruba (Laniran & Clements 1995), while Lee et al. (2017) show that the higher H* associated with the accented mora in Japanese is due to a pre-low raising effect. While Lee et al. (2017) argued that this effect is automatic, there is also speculation suggesting that it may partially be attributed to perceptual motivations. For instance, Gandour et al. (1993) suggest that it is driven by the need to maximize the perceptual distance between contiguous tones. Potisuk et al. (1997: 35) speculate that 'the raising effect would enhance the perceptual separation of tonal categories in the face of a downward trend in f0 throughout a sentence'.

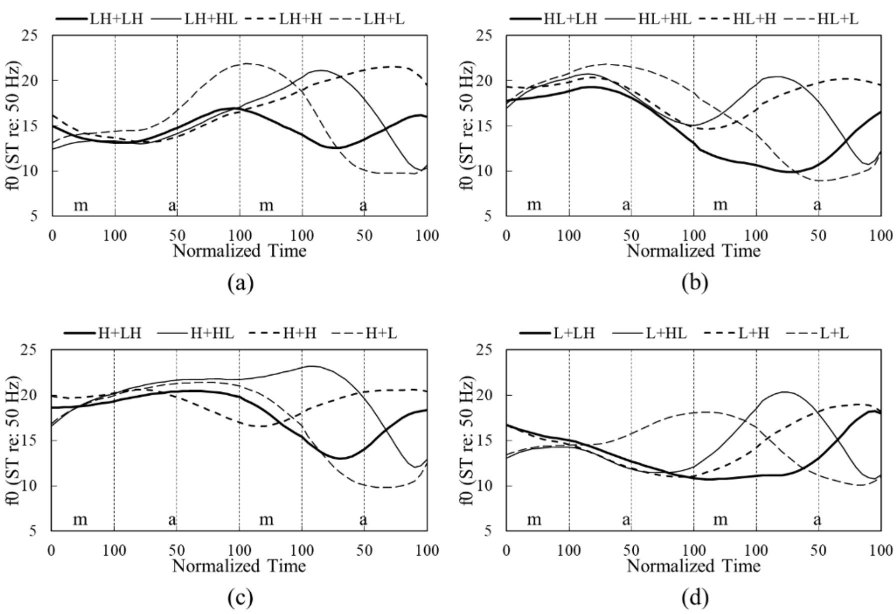


Figure 3.12 f0 contours of /mama/ sequence in Kaifeng Mandarin (data from Wang & Liang 2015). Each f0 curve (ST re: 50Hz) is pooled over 18 stimuli from 3 male speakers (6 repetitions for each speaker). All speakers realize these tone patterns consistently. From panel (a) to panel (d), the first tone is kept intact and the second tone is varied. The duration of the onset nasal /m/ and the vocalic /a/ is normalized to one frame and two frames, respectively.

Just as in the case of the greater salience of the enhancement features under emphasis (§ 3.4.2), the raising effect is even greater when the triggering L-tone is focused, as exemplified in panels (a–f) in Figure 3.10, which compare f0 contours of disyllables with a phrase-final L-tone produced in broad focus as well as corrective focus, by speaker ZWP and WC, respectively. It shows that, for both speakers, focused phrase-final L-tone induces a greater raising effect on the preceding LH, HL and H, which is comparable to that of Tone 3 focus in Standard Mandarin reported in Lee et al. (2016). This suggests that pre-L raising is motivated by enhancement of L, as opposed to an enhancement of the preceding H-tone.

3.5 Conclusion

An acoustic investigation of the four lexical syllabic tone melodies of Kaifeng Mandarin produced by 10 speakers showed that in phrase-final position one of these tones has three different f0 contours depending on the speaker, dipping, falling and falling with lengthening. Fine-grained phonetic data documented the nature and extent of the variation of the four tone melodies. While intra-speaker variation was evident in all four, there was additionally large inter-speaker variation in the realization of one of these, analyzed as L. We have argued that this variation lies behind the divergent phonological characterizations that have been provided for this tone in Zhang et al. (1993) and Liu (1997).

Instead of analyzing these variant pronunciations of the low tone as phonologically different, we argued that they represent alternative ways of enhancing a low tone. In addition to f0 features, the low tone also stands out as being the longest as well as in having creak more often than the other three tones. After indicating that the perception of low tones is inherently more vulnerable than that of high tones or contour tones, we argued that the various enhancements, initial fall, lengthening of the low target or the final rise, serve to increase the salience of the low target. Under corrective focus the enhancements are further enhanced, which strengthens this explanation. The phonological analysis of the low tone as L, instead of ML or HLH, was further defended on the basis of four facts. First, L is the natural complement to a symmetrical four-melody system with H, HL and LH. Second, sequences of L.L are broken up by H, just as sequences of H.H are broken up by L, to produce LH.L and HL.H, respectively, which is another structural argument. Third, the adoption of an HLH melody would uncharacteristically fail to show undershoot of the target of L, which instead remains firmly low in all subtypes. Fourth, L triggers a raising of preceding HL and LH, in line with widely reported phonetic pre-L raising rules.

Thus, the lesson we draw from this investigation is that insightful phonological interpretations of the f0-contours of syllabic tone melodies may well deviate from faithful translations of phonetic forms into elements like Chao digits (e.g., 312) or phonological tones (HLH). In our case, the motivation of the phonological interpretation of the Kaifeng Mandarin low tone as L has led to an increased understanding of the reasons for phonetic forms to deviate from the phonological representation. Of course, while we have demonstrated a persistent presence of similar enhancements of intonational and lexical L in other languages, there is no implication that our three subtypes should make an appearance in L-tone realizations in all languages, any more than that L-tones in all languages will always be enhanced. A final question is why the Kaifeng speech community has not decided on a single way of enhancing their low tone, differently from Northern speakers of Standard Mandarin, where the dipping subtype would appear to be typical (Shih 1988: 84). We believe that the answer may lie in the specific contrasts that the L-tone enters into, as observed by a reviewer. A resolution of this issue will therefore require a perceptual investigation in which the recognizability of different realizations of L is pitted against that of each of the three rival tone categories.

References

Arvaniti, A. (2016). Analytical decisions in intonation research and the role of representations: lessons from Romani. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 1–43. doi: http://dx.doi.org/10.5334/labphon.14

Belotel-Grenié, A. & Grenié, M. (2004). The creaky voice phonation and the organization of Chinese discourse. *International symposium on tonal aspects of languages: With emphasis on tone languages*-2004, Beijing, 5–8.

Blicher, D. L., Diehl, R. & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*, 18, 37–49.

Boersma, P. & Weenink, D. (1992-2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved from http://www.praat.org/

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.

Cao, W. (2010a). 声调感知对比研究：关于平调的报告 [A comparative study on tone perception: Report on level pitch contours]. *世界汉语教学*[Chinese teaching in the world], 24(2), 255–262.

Cao, W. (2010b). 汉语平调的声调感知研究[Level tone perception in Mandarin]. *中国语文*[Studies of the Chinese language], 339, 536–543.

Chao,Y.-R. (1922). Methods of experimental study of tone in Chinese speech. *Kexue* [Science], 7(9), 871–882.

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

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

Chen, Y. (2012). Tonal variation. In A. C. Cohn, C. Fougeron, M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology*, 103–114. Oxford: Oxford University Press.

Connell, B. & Ladd, D. R. (1990). Aspects of pitch realization in Yoruba. *Phonology*, 7(1), 1–29.

Ding, H., Jokisch, O. & Hoffmann, R. (2010). Perception and Production of Mandarin Tones by German Speakers. *Speech Prosody 2010* 100153:1–4. available at http://speechprosody2010.illinois.edu/papers/100153.pdf

Duanmu, S. (2007). *The phonology of Standard Chinese* (2nd edition). Oxford: Oxford University Press.

Gandour, J., Ponglorpisit, S., Dechongkit, S., Khunadorn, F., Boongird, P. & Potisuk, S. (1993). Anticipatory tonal coarticulation in Thai noun compounds after unilateral brain damage. *Brain and Language*, 45, 1–20.

Gandour, J., Potisuk, S. & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, 22(4), 447–492.

Goldsmith, J. A. (1976). *Autosegmental Phonology*. PhD thesis, MIT. Distributed by IULC, 1976, Published by Garland Press, New York, 1979.

Gussenhoven, C. (2007). Intonation. In P. de Lacy. (Ed.), *The Cambridge handbook of phonology*, 253–280. London: Cambridge University Press.

Gussenhoven, C. & Rietveld, T. (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech*, 43(2), 183–203.

Gussenhoven, C. & Zhou, W. (2013). Revisiting pitch slope and height effects on perceived duration. *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 1365–1369.

Hanssen, J. (2017). *Regional variation in the realization of intonation contours in the Netherlands*. Utrecht: LOT.

Hyman, L. M. & Schuh, R. G. (1974). Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry*, 5(1), 81–115.

Huckvale, M. (2014). Prorec: a program for field workers. [Computer program]. Version 1.45, retrieved from http://www.phon.ucl.ac.uk /resource/

Jiang, P. (1999). 汉语诸方言声调分布的优选分析 [Tonal distribution in Chinese dialects: An OT account]. In E. Q. Chen (Ed.), *双语双方言与现代中国* [Bilingualism, diglossia and modern China], 134–152. Beijing: Beijing University of Languages and Culture.

Jongman, A., Wang, Y., Moore, C. B. & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In P. Li, L. H. Tan, E. Bates, O. J. L. Tzeng (Eds.), *The Handbook of East Asian Psycholinguistics Vol. 1: Chinese*, 209–217. Cambridge, UK: Cambridge University Press.

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, 142(3), 1693–1706.

Kuang, J. (2018). The influence of tonal categories and prosodic boundaries on the creakiness in Mandarin. *The Journal of the Acoustical Society of America*, 143(6), EL509–515.

Ladd, D. & Schepman, A. (2003). “Sagging transitions” between high pitch accents in English: experimental evidence. *Journal of Phonetics*, 31(1), 81–112.

Ladefoged, P. (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell.

Laniran, Y. & Clements, G. (1995). A long-distance dependency in Yoruba tone realization. *Proceedings of 13th International Congress of Phonetic Sciences,* Stokholm, Sweden, vol. 2, 734–737.

Leben, W. (1976). The tones of English intonation. *Linguistic Analysis*, 2, 67–107.

Lee, A., Prom-on, S. & Xu, Y. (2017). Pre-low raising in Japanese pitch accent. *Phonetica*, 74(4), 231–246.

Lee, Y-C, Wang, T. & Liberman, M. (2016). Production and perception of Tone 3 focus in Mandarin Chinese. *Frontiers in Psychology*, 7(1058), 1–13. doi: 10.3389/fpsyg.2016.01058.

Lee-Kim, S.-I. (2014). Revisiting Mandarin ‘apical vowels’: An articulatory and acoustic study. *Journal of the International Phonetic Association*, 44(3), 261–282.

Liang, J. & van Heuven, V. (2004). Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language*, 91(3), 282–293. doi: 10.1016/j.bandl.2004.03.006

Liu, D. (1997). 开封方言记略[Kaifeng dialect – a précis]. *方言* [Dialect], 4, 271–285.

Lu, J. (2001). 高/低平调的不对称性和优选论的制约条件 [The asymmetry in H/L and OT constraints]. *Proceedings of the 5th National Conference on Modern Phonetics*, Beijing, China, 243–245.

Ohala, J. & Ewan, W. (1973). Speed of pitch change. *The Journal of the Acoustical Society of America*, 53, 345(A).

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, 34(1), 134–154.

Peters, J. (2007). Bitonal lexical pitch accents in the Limburgian dialect of Borgloon. In T. Riad, C. Gussenhoven (Eds.) *Tones and Tunes. Vol. 1: Typological studies in word and sentence prosody*, 167–198. Berlin & New York: Mouton de Gruyter.

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. PhD thesis, MIT.

Potisuk, S., Gandour, J. & Harper, M. P. (1997). Contextual variations in trisyllabic sequences of Thai tones. *Phonetica*, 54, 22–42.

Prieto, P. (1998). The scaling of the L values in Spanish downstepping contours. *Journal of Phonetics*, 26, 261–282.

Riad, T. (1998). Towards a Scandinavian accent typology. In W. Kehrein, R. Wiese (Eds.) *Phonology and Morphology of the Germanic Languages*, 77–109. Tübingen: Niemeyer.

Rose, P. (1987). Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication*, 6, 343–352.

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

van de Weijer, J. & Sloos, M. (2014). The four tones of Mandarin Chinese: Representation and acquisition. *Linguistics in the Netherlands*, 31(1), 180–191. doi: 10.1075/avt.31.13wei

Vigário, M. & Frota, S. (2003). The intonation of Standard and Northern European Portuguese: A comparative and intonational phonology approach. *Journal of Portuguese Linguistics*, 2, 115–137.

Wang, L. (2018). L-tone focus and word-internal PFC in Kaifeng Mandarin. In *Proceedings of the 9th International Conference on Speech Prosody*, 2018, Poznań, Poland, 423–427.

Wang, L. & Liang, J. (2015). An acoustic study of tones in Kaifeng: Citation forms and contextual variations. *Proceedings of the 18th Oriental COCOSDA and Conference on Asian spoken language research and evaluation*, Shanghai, China, 316–320.

Whalen, D. H. & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25–47.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61–83.

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. (2013). ProsodyPro – A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France, 7–10.

Xu, Y. & Sun, X. (2002). The maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, 111(3), 1399–1413.

Yang, R. (2015). The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics*, 43(1), 453–472.

Yu, K. & Lam, H. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, 136(3), 1320–1333.

Zhang, Q., Chen, T. & Cheng, Y. (1993). *河南方言研究* [A study on Henan dialects]. Kaifeng: Henan University Press.

CHAPTER 4

PHONETIC ENHANCEMENT AND PERCEIVED DISTINCTIVENESS OF THE KAIFENG MANDARIN L-TONE[24]

4.1 Introduction

Contrast enhancement refers to the employment of secondary phonetic parameters to facilitate the perception of the primary phonetic feature in the implementation of a phonological contrast (Stevens, Keyser & Kawasaki 1986). For instance, the auditory distinctiveness of intervocalic voiced and voiceless stops as in English *rabid-rapid* is enhanced by the covariation of duration in the preceding vowel and the closure duration of the stop, which is shorter in /b/ than /p/ (Lotto & Holt 2016). Other things being equal, a closure duration is perceived as even shorter when the preceding vowel is longer and vice versa (Kluender, Diehl & Wright 1988). The enhancement is more effective when the distinction occurs domain-finally, where the primary distinctiveness is weak (Stevens et al. 1986).

Contrast enhancement has been argued to be at least in part controlled by the speakers rather than being entirely automatic (Blicher, Diehl & Cohen 1990; Kingston & Diehl 1994, among others). Yuhuan Wu, which has a tone contrast between HL and ML appearing exclusively in syllables with a sonorant onset, realizes ML in a way that runs counter to the prediction from speech ergonomics (Gussenhoven & Wang 2014). That is, for aerodynamic reasons, the f0 after a sonorant onset would be expected to be higher than after a voiced obstruent onset, in which context only ML occurs, but the f0 immediately after ML is *lower* after a sonorant onset than after a voiced obstruent onset. This is

[24] A version of this chapter has been submitted as: Wang, L., van de Ven, M. & Gussenhoven, C: Phonetic enhancement and perceived distinctiveness of Kaifeng Mandarin L-tone.

explained by the need for enhancement of the HL-ML contrast after sonorants.

By and large, contrast enhancements have mainly been reported for binary contrasts. In addition to pre-obstruent sonorant duration enhancing [+voice] in obstruents, lip rounding has been identified as an accompanying feature of English and French /ʃ/ so as to enhance its distinctiveness from /s/ (e.g., Flemming 2002; Johnson 2012: 159). Also, lip rounding for back vowels generally makes these more distinct from front vowels than they would otherwise be, while backness enhances dental articulations relative to alveolar ones (Stevens et al. 1986: 429, 433–436). In these cases, it is usually not hard to see that the additional phonetic gestures increase the distinctiveness vis-à-vis the other segment or segments involved in the binary opposition. However, when a feature configuration participates in a set of phonological oppositions, it may be more demanding to understand how a specific phonetic shape of the enhancement increases, or at least does not impair, the distinctiveness between the members in each of the contrast pairings. Multiple tone systems are a case in point. Standard Mandarin provides an illustration of a paradigmatic relation between four tonal melodies, T1–T4, which are primarily implemented by the same articulatory parameter, f0. From a perceptual perspective, the optimal form of each of these should be maximally distinct from each of its three rival melodies. T3, the low tone, for instance, has a dipping f0 shape in phrase-final positions, in which the f0 movement towards and after the inflection point has been argued to enhance the low pitch. Gussenhoven (2007: 263) points out that intonational L*s too tend to be enhanced by introducing an f0 fall from a mid-pitch preceding the low pitch, as in European Portuguese (Vigário & Frota 2003), Stockholm Swedish (Bruce 1977), and the Borgloon variety of Limburgish (Peters 2007). Similarly, flanks (i.e., pitch movements towards and after the low target) are produced around the low target of the focal L* in the Romani interrogative contour L*HL% (Arvaniti 2016: 11). As noted by Shih (1988: 84), Northern speakers of Standard Mandarin pronounce phrasal-final T3 consistently as a falling-rising pitch shape, while Southern speakers of Standard Mandarin often pronounce it as low-falling. Thus, compared to the low-falling contour, the dipping contour in T3 would be expected to somehow be the right enhancement of the phrase-final L for Northern speakers of Standard Mandarin. This functional perspective is the central concern in this article. We investigate it on the basis of the Mandarin dialect of Kaifeng (Henan Province), which like Standard Mandarin contrasts four tones, LH, HL, H and L. However, unlike Northern Standard Mandarin, it has three speaker-specific realizations of its phrase-final L (see Chapter 3). In addition to a dipping pronunciation, a low-falling realization occurs, which comes in two variants, one which falls continuously till the end and one which turns to level pitch at the midpoint. They have been argued to be alternative ways of enhancing a low tone (see Chapter 3). There is no age-grading among the three types of speakers, meaning that the speaker-specific distribution has existed for some time, and no relation with any districts within the city of Kaifeng has been established. The question therefore arises which of these three pronunciations provide the best choice for distinguishing L perceptually from the other three tones, LH, HL and H. We decided to answer this question by pitting two variants of Kaifeng L, the dipping and the straight falling types, against each of the three other tones in a word identification task, using stimuli taken from six f0 continua between each of the two ways of enhancing L and each of the three other tone categories. A plan to include the third variant of L, whereby the fall is followed by a level stretch, was abandoned when its continua with the other three tone categories turned out to be barely different from those with the straight falling type. Chinese lexical tones are not only distinguished by means of f0 (Howie 1976; Gandour 1984; Massaro, Cohen & Tseng 1985), but by other phonetic features too. L differs from the other tones in being longest, while the amplitude/intensity profile and phonation have been shown to vary systematically with tone categories and are perceptually relevant to tone identification (amplitude: Whalen & Xu 1992; duration: Blicher et al. 1990; phonation: Yu & Lam 2014). Blicher et al. (1990) created a T2-T3 continuum in two temporal conditions, 350 ms and 450 ms and found that both native Mandarin speakers and native English speakers produce more T3 responses in the long condition. Intriguingly, the native English speakers were unfamiliar with Mandarin, and the effect of duration on the identification of L must therefore have been due to an enhancement that is inherent in low tone. Blicher et al. (1990) further speculated that the duration-induced enhancement of T3 arises from the effect of a longer and more detectable f0 initial fall, which they provisionally believed to perform the main distinctive function of T3.

Since Blicher et al. (1990) only examined the tone pair T2-T3, they left two questions open. First, it remains to be seen whether a reinforcing effect of a longer realization of L is consistently present in its discrimination from the other two non-L tones, T1 and T4. Second, for a language like Kaifeng Mandarin, which has alternative pronunciations of L, it is unclear if syllable duration contributes equally to all the L-tone variants in its discrimination from other tones. For this reason, we decided to add duration as a variable in our experiment, while amplitude and phonation were controlled for.

Our exploratory investigation intends to reveal just how the two types of enhancement of L contribute to its distinctiveness from each of the other three lexical tones, LH, HL and H. The results may help us understand why a relatively close-knit speech community has maintained a situation in which different forms of ‘phonetic conventionalization’ (Dachkovsky 2017) exist, each one apparently used by different speakers.

4.2 Method

4.2.1 Stimuli

Recordings of a H-toned syllable [ma] from one female and one male speaker of Kaifeng Mandarin were used in the present experiment. The reason for choosing a sonorant onset was to eliminate the perturbation effect that might have been caused by an obstruent onset on the following vowel. Another reason for choosing the syllable [ma] is that all four morphemes sharing this syllable, ‘mother’ LH (T1), ‘hemp’ HL (T2), ‘horse’ H (T3) and ‘scold’ L (T4), have high lexical frequencies.[25] Lastly, the choice of an H-toned syllable over an L-toned one is to avoid non-modal phonation (e.g., creak), which has been found to lead to more low tone responses in the identification of tone pairs in both Standard Mandarin (Yang 2015) and Cantonese (Yu & Lam 2014).

---

[25] Though no corpus of Kaifeng Mandarin is available for a direct estimation of the lexical frequency of these morphemes, it can be predicted from their frequency in Standard Mandarin in the sense that these four morphemes are etymologically equivalent to the quadruplet in Standard Mandarin, and most importantly, the same writing system is shared by Chinese speakers (for similar discussions see the Appendix in Zhang & Liu 2011). Indeed, according to Da (2005), all the four morphemes rank top 12% in a frequency list of 12041 characters.

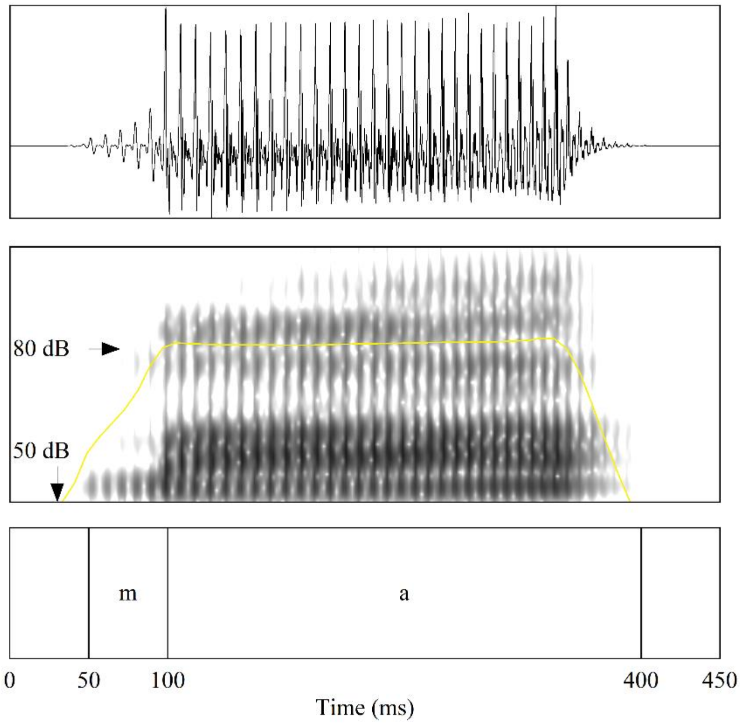


Figure 4.1 The waveform and spectrogram of the short syllable (350 ms) from the male speaker. The manipulated intensity curve is indicated by the yellow line.

A fixed duration of 50 ms for the initial sonorant segment was chosen, while the duration of the remainder was manipulated so as to yield 300 ms and 400 ms, using ‘pitch synchronous overlap add (PSOLA)’ in Praat (Boersma & Weenink 1992–2015). A 50 ms silence was added before and after stimuli in every speech file. Since amplitude/intensity contour alone has been shown to lead to a reasonably successful tone identification (Whalen & Xu 1992), the intensity profile was manipulated using the method described in McCloy (2013: 53–55). Specifically, for a target stimulus, first, the inverted version of the original intensity contour was calculated; second, the original intensity contour was flattened by multiplying it by the inverted one; third, the resulting flattened contour was multiplied by a stylized target contour. For both the short and the long syllable, this target contour increased from 60 dB at the nasal onset to 85 dB at the onset of the rhyme, which value was maintained during the larger portion of the rhyme, and then decreased to 60 dB at the syllable offset. The mean intensity of the syllable was rescaled to 80 dB, shown in Figure 4.1.

A corpus of two repetitions of 35 minimal tonal quadruplets was recorded by a female and two male speakers, A and B. The female speaker and the male speaker A provided the target syllable [ma]. Each tone was sampled by 21 equidistant f0 points in the rhyme. For each speaker, f0 values for each time point within tones were averaged, giving four sets of 21 f0 values per speaker, based on which speaker-dependent $f0_{mean}$ and $f0_{sd}$ were obtained. Raw f0 data, $f0_{Hz}$, were converted into speaker-dependent z-scores based on the formula $f0_{z\text{-}score} = (f0_{Hz} - f0_{mean})/f0_{sd}$ (Rose 1987). Target pitch contours (in z-scores) from four naturally pronounced tones, LH, HL, H and a low-falling L were selected from the female speaker's corpus, while that for the dipping L was selected from the male speaker B's corpus, as displayed in Figure 4.2. The falling L and the dipping L were paired with LH, HL and H respectively and for each of these six tone pairs, eight intermediate f0 versions were created for each duration (see Figure 4.3), giving 240 stimuli (6 tone pairs × 2 duration conditions × 10 pitch steps × 2 voices).

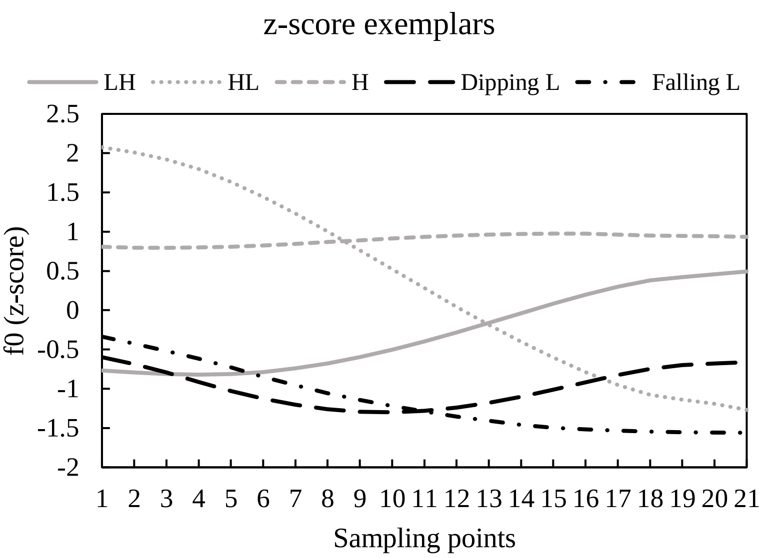


Figure 4.2 Tone exemplars for LH, HL, H and L. L-tone has two exemplars, a dipping one and a falling one. LH, HL, H and Falling L were from a female speaker; Dipping L was from a male speaker (speaker B).

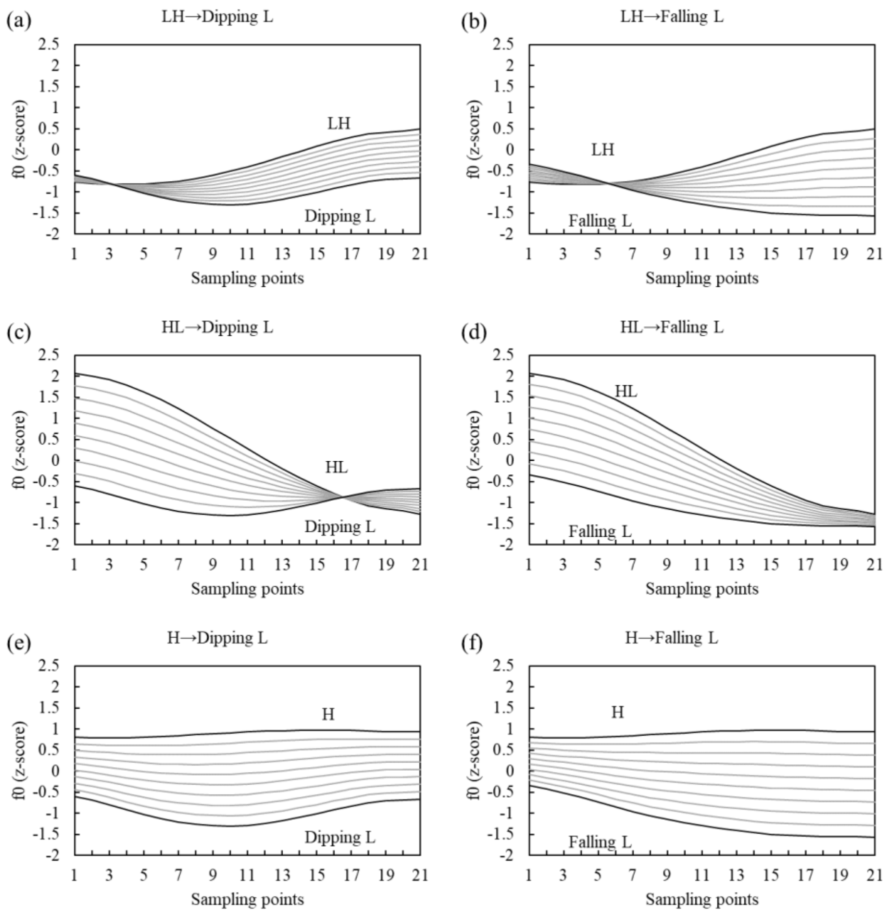


Figure 4.3 f0 contour continua: panel (a) LH-Dipping L, panel (b) LH-Falling L, panel (c) HL-Dipping L, panel (d) HL-Falling L, panel (e) H-Dipping L and panel (f) H-Falling L. There are eight intermediate f0 steps varying incrementally in all 21-sampling points from one continuum endpoint to the other. The continuum endpoints are marked in black and intermediate steps in grey.

Before superimposing the f0 contours on the two target syllables, one for the female speaker and one male speaker A, the z-scores were converted back to Hz based on the speaker-specific $f0_{mean}$ (190 Hz and 123 Hz, respectively) and $f0_{sd}$ (25.63 Hz and 22.58 Hz, respectively) so as to properly situate each tone in their f0 range. With the help of a custom-written Praat script, Hz values for all tone stimuli were evenly distributed over the rhyme portion of the target syllable and linearly interpolated in the Praat manipulation window (PSOLA) so as to match the 300/400 ms rhyme templates. Each stimulus began with the same f0 as the beginning of the rhyme, creating level f0 during the onset [m]. While the syllable is the relevant distributional constituent of the tone melodies, the distinctive f0 shapes in our data began at the beginning of the vowel, suggesting that the rhyme is the tone-bearing constituent (Howie 1974; Hallé 1994; Duanmu 2007).[26]

4.2.2 Participants

Forty native speakers of Kaifeng Mandarin, 19 females and 21 males (age: range = 18–35, mean = 22.47, sd = 5.08), participated in the experiment. Most of them were university students and teachers from Henan University, Kaifeng University and Kaifeng Vocational College of Culture and Art. All of them were born and raised in Kaifeng. None of them had self-reported speaking and listening problems at the time of the experiment. All participants consented to participate in the experiment and were instructed that they could abandon the experiment at any point without giving reasons in accordance with the Radboud University Arts Faculty Protocol. They were paid a small fee for their participation.

4.2.3 Procedure

Participants were randomly and evenly assigned to four groups, who were presented with stimuli created from (1) female voice, all six continua; (2) male voice, all six continua; (3) female voice, continua with dipping L, male voice, continua with falling L; (4) female voice, continua with falling L, male voice, continua with dipping L. Participants were told that they would listen to different stimuli from those presented to other participants and that there was no point in discussing the experiment with them. The experiment was run using ExperimentMFC implemented in Praat, using a binary forced-choice word identification task in which participants had to press a button corresponding to one of two characters presented on a computer screen. The experiment was preceded by a five-trial training session and divided into two 240-item blocks, each encompassing two repetitions of the six ten-step tone pairs under two durational conditions (6 pairs × 10 steps × 2 durations × 2 repetitions); the order of the characters on the

[26] For a treatment of the syllable as the tone-bearing domain, see Xu 1998 and Xu & Wang 2001.

screen was reversed for the second block. Within-block stimuli were randomized. Each stimulus was played twice with an ISI of 500 ms, after a 600-ms 500-Hz beep and a 400-ms pause. The experiment was self-paced. Once the participants chose one of the characters, they could click the ‘next’ button situated in the lower right corner, upon which the next stimulus was presented. They were allowed, but not encouraged, to click a ‘previous’ button situated in the lower left corner to do the previous stimulus again. There was a 5-minute break in between the two blocks. The experiment was conducted in quiet classrooms or offices using two laptops working simultaneously, each equipped with a Sennheiser HD 206 headset, i.e., that two participants were tested at the same time. The participants received verbal notification from the experimenter before the experiment and a written one in the running headline throughout the experiment that they were listening to Kaifeng Mandarin, not Standard Mandarin. The duration of the experiment was about one hour.

4.3 Results

To evaluate the effect of the L-tone variant (dipping L and falling L), rhyme duration (300 and 400 ms), and voice (male and female) on the perception of L in three different phonological tone contrasts (LH-L, HL-L and H-L), the participants’ responses (L or non-L) were analyzed using generalized linear mixed effects models with the logit link function and with random intercepts and slopes (Jaeger 2008) implemented in the package *lme4* (Bates et al. 2015) in R (R Core Team 2016). Separate analyses were performed for each of the three tone contrasts LH-L, HL-L and H-L. Each model was structured with fixed effects of *Step* (1–10), *Variant* (dipping or falling), *Duration* (300 or 400 ms) and *Voice* (female or male) and random effects of *Participants*. The numerical variable *Step* was standardized to a mean of zero and centred in order to avoid multicollinearity (Belsley et al. 1980). Because participants did not judge identical stimuli, no random slopes for *Voice* and *Variant* were included. The fixed and random effects were added to the model in a stepwise manner and the contribution of the effects to the model fit was evaluated using maximum likelihood estimation, i.e., Chi-squared tests (Baayen 2008). Variables (both fixed and random effects) were excluded if they failed to reach significance at the 5% level.

4.3.1 LH-L

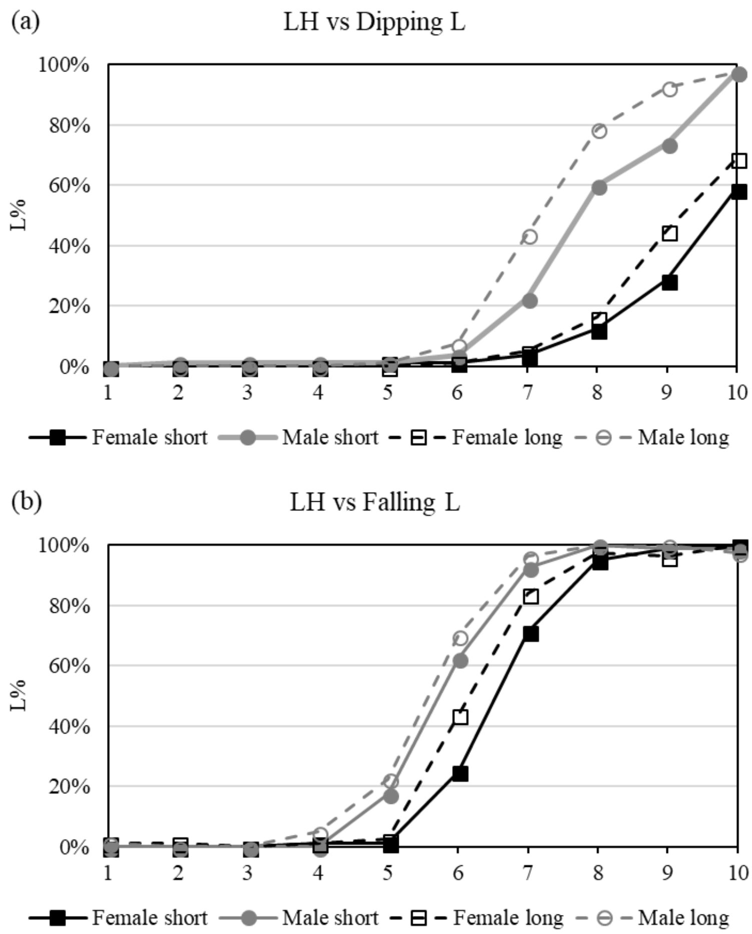


Figure 4.4 L-responses (%) as a function of tone step (1–10) along a continuum from LH to L, with separate curves for long versus short duration and for male versus female base (female = black, male = grey; rhyme duration 300 ms = solid line, rhyme duration 400 ms = dashed line). LH-dipping L and LH-falling L are plotted separately in panels (a) and (b), respectively.

The identification curves for the contrast LH-L under two durational conditions and two voice conditions are shown in Figure 4.4. Through visual inspection, there are three observations. First, across conditions of duration and voice, the falling L-variant (panel b) induces an earlier category boundary and a much larger boundary steepness, compared to the dipping L-variant (panel a). The dipping L-variant reaches maximum identification only at the rightmost step, near-ceiling for the male voice but only around 60% for the female voice. By contrast, the identification curve of the falling L-subtype reaches a ceiling at the eighth step and maintains it thereafter across both voice conditions. Second, lengthening the rhyme increases L-tone responses for both dipping and falling L-variants and it is consistent across two voice conditions (as seen in the dashed line). Third, the male voice yields an earlier and more convincing cross-over boundary than the female voice for both L-tone subtypes.

Table 4.1 Results of the generalized linear mixed effects model for the tone contrast LH-L (300 ms rhyme duration, dipping variant and female voice as baseline).

Predictor: Fixed effects	β	Z
Intercept	–6.6870	–19.001***
Step	4.6065	20.122***
Subtype (Falling)	3.6080	15.895***
Duration (400 ms)	0.7470	6.486***
Voice (Male)	2.9436	13.837***
Step×Subtype(Falling)	1.9635	6.254***
Predictor: Random effects	**Variance explained**	χ^2
Participants	1.5100	-
Slope (Step)	0.3398	7.5308*

* $p < .05$ ** $p < .01$ *** $p < .0001$

The results of the generalized linear mixed effects model are summarized in Table 4.1. First of all, there was a two-way interaction between *Step* and *Variant* ($\chi^2(1) = 41.238$, $p < .001$). As observed above, an increase in *Step* increases L-tone responses more generally across the two variants, but the increase is more dramatic in the falling variant. There was also an effect of *Variant*, indicating that the falling L-variant elicits more L responses than the dipping L-variant. Together, these effects suggest that the perception of the contrast between LH and falling L (panel b) is more categorical than between LH and dipping L (panel a). Second, there was a main effect of *Duration* ($\chi^2(1) = 83.889$, $p < .001$), showing that a longer rhyme duration enhances the perception of L for both dipping and falling L-variants. There was no interaction between *Variant* and *Duration*, indicating that the magnitude of this enhancement is similar across the two variants. There was no interaction between *Step* and *Duration* either, indicating that there is no duration-induced slope difference in the identification curve. Third, there was a main effect of *Voice* ($\chi^2(1) = 268.61$, $p < .001$), showing that the male voice leads to more L-responses for both L-tone variants. The interaction between *Voice* and *Step* was not significant, indicating that there does not appear to be a voice-induced slope difference in the identification curve.

4.3.2 HL-L

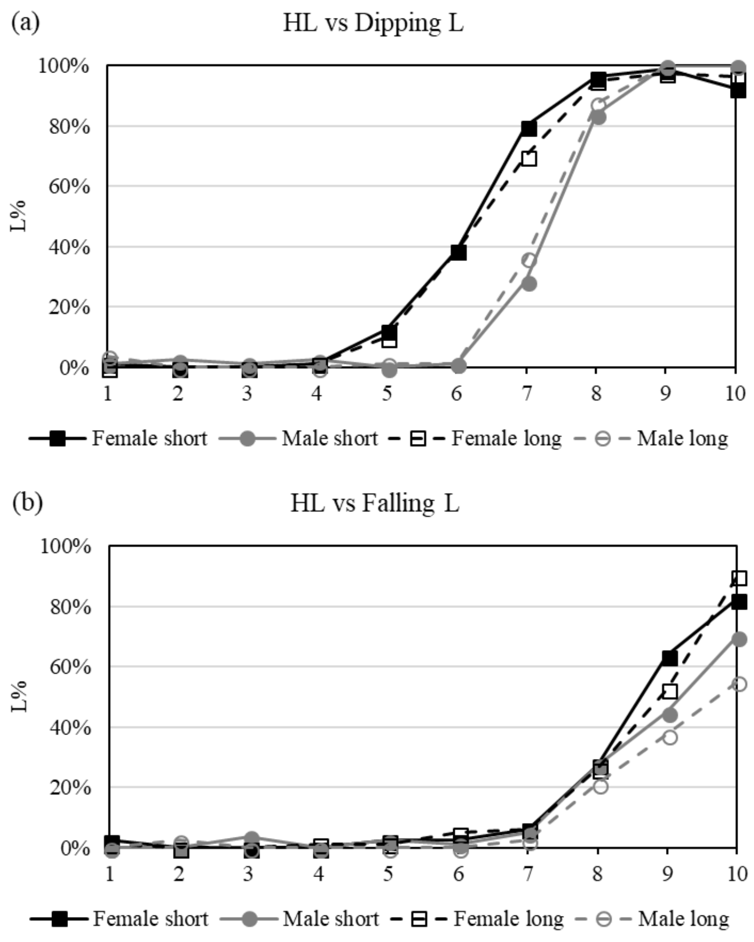


Figure 4.5 L-responses (%) as a function of tone step (1–10) along a continuum from HL to L, with separate curves for long versus short duration and for male versus female base (female = black, male = grey; rhyme duration 300 ms = solid line, rhyme duration 400 ms = dashed line). HL-dipping L and HL-falling L are plotted separately in panels (a) and (b), respectively.

Figure 4.5 displays the identification curves for the tone contrast of HL-L under two durational conditions and two voice conditions. Three observations can be drawn from the visual inspection of Figure 4.5. First, when pooled over the conditions of voice and duration, the dipping variant (panel a) shows an earlier and a steeper category boundary compared with the falling variant (panel b), a pattern running opposite to LH-L (See Figure 4.4). The identification curve of the dipping variant is S-shaped, whereas the L-tone response in the perception of falling variant starts at the seventh step and fails to reach ceiling level even at the terminal step point. Second, syllable duration seems to play a very limited role across both voice conditions. The effect is rather small and neither consistent within variants of L nor within the two voices. Third, across the two L-variants, the female voice appears to lead to more L responses, contrary to the case in LH-L, where the male voice shows more L responses.

Table 4.2 Results of the generalized linear mixed effects model for the tone contrast HL-L (300 ms rhyme duration, dipping variant and female voice as baseline).

Predictor: Fixed effects	β	Z
Intercept	–1.5047	–5.250***
Step	4.7965	14.051***
Subtype (Falling)	–3.1378	–9.270***
Voice (Male)	–2.4157	–6.761***
Step×Subtype (Falling)	–0.5216	–1.332
Subtype (Falling)×Voice (Male)	2.2798	4.210***
Step×Voice (Male)	1.6912	3.684***
Step×Subtype (Falling)×Voice (Male)	–2.5049	–3.982***
Predictor: Random effects	**Variance explained**	χ^2
Participants	1.821	-
Slope (Step)	1.471	94.408***

* $p < .05$ ** $p < .01$ *** $p < .0001$

Results of the generalized linear mixed effect model are summarized in Table 4.2. Most importantly, there was a three-way interaction between *Step*, *Variant* and *Voice* ($\chi^2(1) = 74.47$, $p < .001$). To better spell out this interaction, separate models were constructed for each voice condition. The three-way interaction was captured by a significant two-way interaction between *Step* and *Variant* for male voice ($\chi^2(1) =$

49.605, $p < .001$), which effect was absent for female voice. The presence of the two-way interaction of *Step* and *Variant* indicates that in the case of the male voice the identification curve is steeper in the dipping L-variant than in the falling L-variant. Across both voice conditions, there was also a main effect of *Variant*, indicating that the percentage of L-tone responses was significantly reduced in the falling L-variant, compared with the dipping variant, indicating a much weaker perceived contrast. Second, the main effect of *Duration* was not significant, indicating that the perception of contrast HL and L is insensitive to rhyme duration. Third, separate models for each variant were also created to evaluate the effect of *Voice* independently. The results showed that there was no interaction between *Step* and *Voice* for either the dipping or falling L-variant, indicating that there is no voice-induced slope difference in the identification curves. For the dipping L-variant, there was a main effect of *Voice* ($\chi^2(1) = 14.331$, $p < .001$), indicating that the male voice has more L-tone responses than the female voice in the perception of dipping L-variant. This effect, however, is absent in the falling L-variant, suggesting that there is no voice-induced difference in the perception of falling L-variant.

4.3.3 H-L

The identification curves for the perception of the contrast H and L in the conditions of duration and voice are shown in Figure 4.6. From the visual inspection of Figure 4.6, three observations can be made about the effects of L-variant, duration and voice. First, the identification curves of H-dipping L and H-falling L are both S-shaped across conditions of duration and voice. Second, there is a consistent duration-induced boundary shift in favor of more L-tone responses to longer stimuli across conditions of voice and the L-tone variant. Third, the male voice biases the listeners towards more L-tone responses across both variants of L.

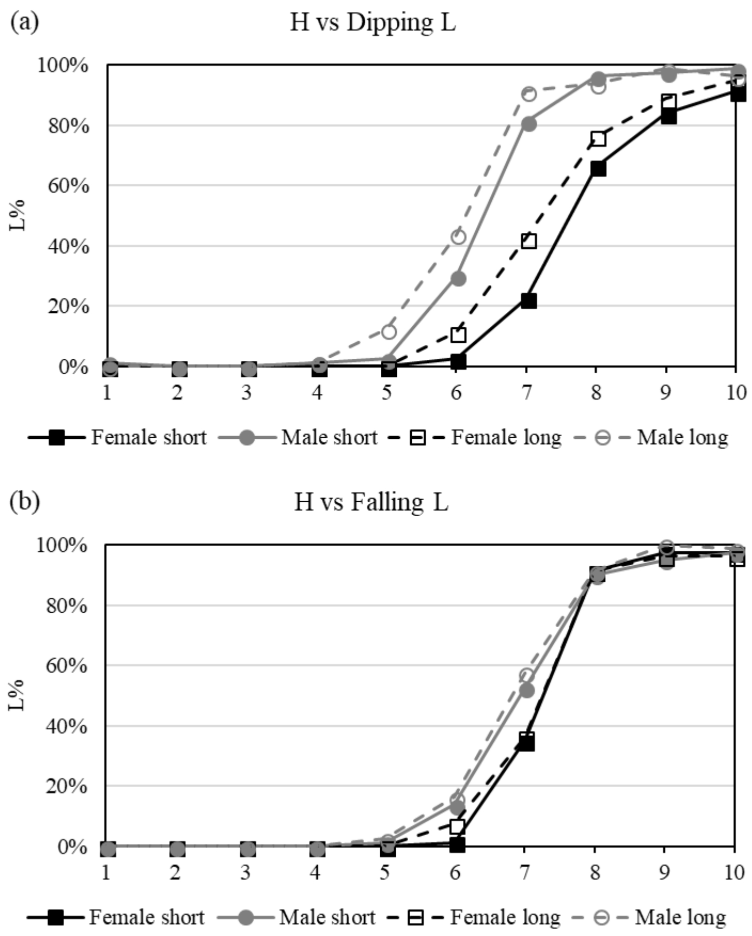


Figure 4.6. L-responses (%) as a function of tone step (1–10) along a continuum from H to L, with separate curves for long versus short duration and for male versus female base (female = black, male = grey; rhyme duration 300 ms = solid line, rhyme duration 400 ms = dashed line). H-dipping L and H-falling L are plotted separately in panels (a) and (b), respectively.

Results of the generalized linear mixed effect model are detailed in Table 4.3. First of all, there was a two-way interaction between *Step* and *Variant* ($\chi^2(1) = 32.789$, $p < .001$), showing that the slope of the identification curve for the falling L-variant is steeper than that of the dipping L-variant. Second, there was a main effect of *Duration* ($\chi^2(1) = 19.401$, $p < .001$), meaning that a longer rhyme duration enhances the perception of L. There was no interaction between *Variant* and *Duration*, indicating that the duration-induced enhancements for the two variants of L are of similar magnitudes. Third, there was an effect of *Voice* ($\chi^2(1) = 110.41$, $p < .001$), indicating that male voice is more likely to be categorized as L than a female voice. There was no interaction between *Voice* and *Variant*, indicating that this voice-induced shift is of a similar magnitude in dipping and falling L-variants.

Table 4.3 Results of the generalized linear mixed effects model for the tone contrast H-L (300 ms rhyme duration, dipping variant and female voice as baseline).

Predictor: Fixed effects	β	Z
Intercept	–3.7682	–14.464***
Step	5.5694	16.004***
Subtype (Falling)	–1.2602	–6.126***
Duration (400 ms)	0.4993	4.434***
Voice (Male)	1.6787	10.127***
Step×Subtype (Falling)	1.7239	5.632***
Predictor: Random effects	**Variance explained**	χ^2
Participants	1.201	-
Slope (Step)	2.638	64.973***

* $p < .05$ ** $p < .01$ *** $p < .0001$

4.4 Discussion

One of the major findings of our perception experiment is that the competition for discriminatory power by the dipping L and falling L ended in a draw. As expected, the perception of both variants of L with H showed an S-shaped identification curve, typical of CP. We take this to be due to the embedding of the low f0 target in a higher f0 context in both the dipping L and falling L, which feature has no counterpart in the f0 shapes for H (cf. Gussenhoven 2007). However, when paired with the dynamic tones HL and LH, dipping L is a better discriminator in the case of HL, because of the different directions of the f0 shapes in the second halves of the contours, but a worse discriminator in the case of LH, when f0 shapes are similar in the two sets of stimuli. Conversely, falling L discriminates better from LH than from HL, for the same reasons. The continua HL-dipping L and LH-falling L show typical S-shaped identification curves; by contrast, continua LH-dipping L and HL-falling L show a response curve showing more non-L responses, while the identification generally failed to reach a ceiling level even at the terminal step point. The interpretation is that listeners disprefer similar f0-shapes for contrasting tones. Since the dipping L scores twice over the falling L through its superior distinctiveness from H and HL and the falling L equally scores twice through its superior distinctiveness from H and LH, the score is tied. In part, this probably explains why Kaifeng speakers are failing to find a uniform way of enhancing their L-tone.

The question arises why speakers of Standard Mandarin agree on a single conventional enhancement form of L in final positions, i.e., dipping (Shih 1988), even if the recruitment of the flanks in the L-tone endangers the perception of LH and hence poses serious problems for both first and second language learners of Mandarin (Li & Thompson 1978; Kiriloff 1969). Most probably, the reason lies in the fact that the T2-T3 distinction has a limited functional load in spontaneous speech, as suggested by three circumstances. First, based on a calculation by Duanmu (2007: 253), T2- and T3-syllables are less frequent than T1- and T4-syllables in a 1255-syllable corpus (T1: 337; T2: 255; T3: 316 and T4: 347). Second, in many Mandarin dialects, T2 and T3, or *yangping-shang* in traditional Chinese tonology, have merged, yielding a three-tone system. In such dialects, the old contrast between *yangping-shang* may surface in phonological processes like tone sandhi and neutral tone. Modern Urumqi Mandarin, for instance, contrasts only three citation tones on the surface, whereby its *yangping* is etymologically equivalent to Standard Mandarin T2 (*yangping*) and T3 (*shang*) (Zhou 1995). Before neutral-tone syllables, Urumqi *yangping* has different reflexes depending on its etymological source (i.e., on whether this is *ping* or *shang* in Middle Chinese). Some Henan Mandarin dialects, such as Luoning, Mianchi and Yima have undergone similar processes (Zhang et al. 1993). These mergers of the *yangping-shang* tone contrast follow from a low functional load of this opposition, as suggested by the inverse relation between the probability of a merger and the functional load of the contrast (Wedel, Kaplan & Jackson 2013). Third, even in those Mandarin dialects including Standard Mandarin, Kaifeng Mandarin, Luoyang Mandarin (He 1984), among many others, in which the tone distinction between *yangping* and *shang* is maintained, tone neutralization between these tones occurs in specific context (Tone 3 sandhi), indicating another case of contrast reduction.

The second finding is that the effect of syllable lengthening enhances the L-tone perception more generally, though not in competition with HL. The durational enhancement was found in both the dipping and falling variants of L. Blicher et al. (1990) hypothesized that the initial fall in Standard Mandarin dipping T3 has the main distinctive function and that lengthening improves its detectability. This is compatible with our analysis that pitch movements towards and after the low target increase its salience. Perceptual discrimination between Standard Mandarin T3 and T2 has been shown to be jointly determined by the timing of the f0 turning point (i.e., the low target) and *Δf0*, i.e., the f0 difference between the f0 onset and the f0 turning point (Shen & Lin 1991; Moore & Jongman 1997). That is, other things being equal, when the f0 turning point of a dipping contour occurs near the f0 onset or if the magnitude of *Δf0* is too small to be perceived, the tone is more likely to be heard as T2 than as T3. Shen & Lin (1991) further showed that even when the turning point occurs near the offset, so that the final rise is no longer salient, the tone is still parsed as T3. This indicates that a later low target is preferred in the identification of L, even when the L target occurs somewhere near the offset in a dipping contour. The reason why lengthening fails to enhance the L when competing with HL is probably that the final L-target in HL is equally enhanced by lengthening. The production experiment in Chapter 3 indicated that the final L-target in HL is lengthened by some speakers (see Figure 3.3). By contrast, T4 in Standard Mandarin does not have this lengthening effect of the final L, even when it is inside the focus (Yuan unpublished). This suggests that the lengthening of the final phonological L in Kaifeng HL is a language-specific feature.

The last finding relates to the effect of voice in the identification of L. Male voice facilitates the identification of L except in competition with HL and this effect is consistent in both variants of L. Because the male voice has the f0 situated in the lower range, it is helpful for a low pitch to contrast with a high pitch. However, in the case of HL which ends up with a low pitch comparable to that of a L-tone (see Figure 4.3), this advantage disappears, as in the case of HL-falling L, or reverses to be a disadvantage, as in the case of HL-dipping L.

4.5 Conclusion

A perception experiment was conducted to examine the effect of two enhancement strategies on the perceived distinctiveness of L-tone with three rival tones LH, HL and H in the Mandarin dialect of Kaifeng. The first type of enhancement relates to two speaker-dependent pitch shapes, referred to as dipping L and falling L, while the second type of enhancement is syllable lengthening. Six sets of f0 continua that vary incrementally from two naturally pronounced L-tone exemplars (dipping and falling Ls) to three non-L exemplars were created. The pitch continua were properly superimposed onto two [ma] syllables, which were naturally pronounced by one male and one female speaker and were manipulated into two different durational conditions. Forty listeners were recruited to perform a forced choice identification task between two Chinese characters that contained either a L-tone or a non-L-tone. The results showed that the dipping L and falling L perform equally well when the rival tone lacks some f0 movement, as in the case of H-L, or when the dynamic movement of the rival tone has the opposite direction to that of the enhanced L, as in the case of LH-falling L and HL-dipping L. However, when the pitch movement in the enhanced L agrees with that of the rival tone, i.e., in the case of LH-dipping L and HL-falling L, stimuli are more likely to be parsed as non-L tones. This balanced discriminability explains the coexistence of both forms of enhancement in the same community, if the assumption is that the best enhancement should have the most discriminative potential. Syllable lengthening additionally yields a more detectable low pitch target, hence serving another form of enhancement. However, the lengthening-induced enhancement fails to come into play when L competes with HL, because the phonological L in HL is equally enhanced by lengthening for language-specific reasons. Finally, a low male voice is advantageous in the perception of L, though only in competition with tone shapes that end in higher pitch.

References

Arvaniti, A. (2016). Analytical decisions in intonation research and the role of representations: lessons from Romani. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 1–43. doi: http://dx.doi.org/10.5334/labphon.14

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01. R package (version 1.1-12) available at http://CRAN.R-project.org/package=lme4

Belsley, D., Kuh, A. & Welsch, R. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley.

Blicher, D., Diehl, R. & Cohen, L. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18, 37–49.

Boersma, P. & Weenink, D. (1992–2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved from http://www.praat.org/

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.

Da, J. (2005). Combined character frequency list of classical and modern Chinese. http://lingua.mtsu.edu/chinesecomputing/statistics/char/list.phpWhich=TO, last consulted 27 August, 2017.

Dachkovsky, S. (2017). Grammaticalization of intonation in Israeli Sign Language: From Information Structure to relative clause relations. PhD dissertation. University of Haifa.

Duanmu, S. (2007). *The phonology of Standard Chinese* (2nd edition). Oxford: Oxford University Press.

Flemming, E. (2002). *Auditory representation in phonology*. New York & London: Routledge.

Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, 12, 235–261.

Gussenhoven, C. (2007). Intonation. In P. de Lacy (Ed.), *The Cambridge handbook of phonology*, 253–280. London: Cambridge University Press.

Gussenhoven, C. & Wang, L. (2014). Yuhuan Wu tone and the role of sonorant onsets. In N. Campbell, D. Gibbon, D. Hirst (Eds.), *Social and linguistic speech prosody. Proceedings of the 7th International Conference on Speech Prosody*, Dublin, Ireland, 619–622.

Hallé, P. (1994). Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese. *Language and Speech*, 37, 281–306.

He, W. (1984). 洛阳方言记略 [Luoyang dialect – A précis]. *方言* [Dialect], 4, 278–299.

Howie, J. M. (1974). On the domain of tone in Mandarin: Some acoustical evidence. *Phonetica*, 30, 129–148.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Jaeger, F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.

Johnson, K. (2012). *Acoustic and Auditory Phonetics,* third edition. Wiley-Blackwell.

Kingston, J. & Diehl, R. (1994). Phonetic knowledge. *Language,* 70(3), 419–454.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, 20, 63–67.

Kluender, K., Diehl, R. & Wright, B. (1988). Vowel length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16, 153–169.

Li, C. & Thompson, S. (1978). The acquisition of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey*, 271–283. New York: Academic Press.

Lotto, A. & Holt, L. (2016). Speech perception: The view from the auditory system. In G. Hickok, S. Small (Eds.), *Neurobiology of Language*, 185–194. Elsevier, Amsterdam.

Massaro, D., Cohen, M. & Tseng, C.-Y. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, 13, 267–289.

McCloy, D. (2013). Prosody, intelligibility and familiarity in speech perception. PhD thesis, University of Washington.

Moore, C. & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, 102(3), 1864–1877.

Peters, J. (2007). Bitonal lexical pitch accents in the Limburgian dialect of Borgloon. In T. Riad, C. Gussenhoven (Eds.), *Tones and Tunes. Vol. 1: Typological studies in word and sentence prosody*, 167–198. Berlin & New York: Mouton de Gruyter.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (version 3.3.0) URL http://www.R-project.org/.

Rose, P. (1987). Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication*, 6, 343–352.

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

Shen, X. & Lin, M. (1991). A perceptual study of Mandarin Tones 2 and 3. *Language and speech*, 34(2), 145–156.

Stevens, K., Keyser, J. & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In J. Perkell, D. H. Klatt (Eds.), *Invariance and Variability in Speech Processes*, 426–449. Hillsdale, NJ: Erlbaum.

Vigário, M. & Frota, S. (2003). The intonation of Standard and Northern European Portuguese: A comparative and intonational phonology approach. *Journal of Portuguese Linguistics*, 2, 115–137.

Wedel, A., Kaplan, A. & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128, 179–186.

Whalen, D. & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25–47.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, 55, 179–337.

Xu, Y. & Wang, Q. (2001). Pitch targets and their realizations: Evidence from Mandarin Chinese. *Speech Communication*, 33, 319–337.

Yang, R. (2015). The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics*, 43(1), 453–472.

Yu, K. & Lam, H. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of Acoustical Society of America*, 136(3), 1320–1333.

Yuan, J. (unpublished). The duration aspect of tones, retrieved from http://www.ling.upenn.edu/~jiahong/publications/a01.pdf

Zhang, Q., Chen, T. & Cheng, Y. (1993). *河南方言研究* [A study of Henan dialects]. Kaifeng: Henan University Press.

Zhang, J. & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, 68, 161–191.

Zhou, L. (1995). *乌鲁木齐方言词典* [A dictionary of Urumqi dialect]. Nanjing: Jiangsu Educational Press.

CHAPTER 5

INHIBITED REALIZATION OF KAIFENG MANDARIN SANDHI TONES[27]

5.1 Introduction

Incomplete neutralization (hereafter IN) refers to the more or less systematic existence of small phonetic differences between words which phonological accounts characterize as having the same pronunciation as a result of the neutralization of an underlying phonological distinction (Mitleb 1981, Port et al. 1984, Port & O'Dell 1986, Port & Crawford 1989). Generally, the phonetic differences, even if they are small, have appeared to be in the direction of what the underlying contrast would lead one to expect. The classic example is the German surface form [bʊnt], which has two underlying sources, one being *bunt* 'colourful', with its neuter form as in *buntes Papier* 'colourful paper', and the other *Bund* 'federation', with its plural form *die Bünde*, whose inflected forms have /t/ and /d/, respectively. Amongst other differences, the closure phase of /t/ in 'federation' has been shown to be shorter and the duration of the sonorant rhyme /ʊn/ longer than in the case of 'colourful'. Many other cases of IN of the voicing contrast in coda obstruents have been reported, e.g., Charles-Luce & Dinnsen (1987) for Catalan, Warner et al. (2004) and Ernestus & Baayen (2006) for Dutch, Dmitrieva et al. (2010) for Russian, in part with additional phonetic features showing up, like longer voice lag and shorter burst duration for the (underlying) voiced plosive.

IN is not restricted to segmental contrasts. It has also been reported for the neutralization of lexical tones, as in the case of Tone 3 sandhi (henceforth T3S) in Mandarin, whereby Tone 3 (L) is replaced by Tone 2 (LH) before another Tone 3, as already reported by Chao (1948). As in the case of coda obstruent devoicing, instrumental studies have

[27] Results of the experiments were presented in the 40th DGfS at the University of Stuttgart. A version of this chapter will be submitted as Wang, L. & Gussenhoven, C: Inhibited realization of Kaifeng Mandarin sandhi tones.

reported small phonetic differences in f0-height and f0-slope between the derived LH and the underlying LH. Specifically, in terms of overall f0-height, the derived LH has been shown to be consistently lower than the underlying LH by a constellation of instrumental studies (Zee 1980; Xu 1993; Peng 2000; however, see Myers & Tsay 2003 for the postulation of complete neutralization of T3S by non-native speakers of Beijing Mandarin). The magnitude of this difference varies among studies, which can be as subtle as 2.3 Hz as reported in Peng (2000) and as large as 17.5 Hz as reported in Zee (1980). In terms of f0-slope, the derived LH has been reported to be flatter than the underlying LH (Zee 1980, with data reanalyzed by Myers & Tsay 2003; Cheng, Chen & Gubian 2013; Yuan & Chen 2014).

Reports on other contrasts suggest that IN is a common phenomenon.[28] By the side of IN in American English Flapping, Turkish g-deletion and English stop intrusion, listed in Dinnsen (1985), Warner et al. (2004) found longer consonant durations for underlying geminates compared to singletons, as in infinitival [pootə(n)], the surface form of both /poot-ən/ *poten* 'plant-INF' and /poot-tən/, from /poot-də-ən/ *pootten* 'plant-PAST-PL'. As observed above, the phonetic realization of underlying contrasts that show IN has generally appeared to be a miniature version of what a full-fledged surface contrast would show, and discussions of the phenomenon have taken this finding as a starting point. One approach among theoretical discussions of the phenomenon involves the introduction of a phonological element that is responsible for the phonetic difference between the neutralized and unchanged forms. Thus, van Oostendorp (2008) proposed that the underlying representation is partially preserved in the output representation. This approach describes IN at the cost of introducing a novel form of representation without laying claims to an explanation of the phenomenon. Nevertheless, a general prediction that can be derived from the phonological nature of the solution is that *the phonetic effects are similar to those produced by phonologically fully distinct representations*. The purpose of this article is to test this prediction of this 'mini-phonology' theory on the basis of

[28] A phenomenon that is often discussed together with IN is near-mergers (Labov 1987; 1994: 310–418). The main difference between IN and near-mergers is that IN involves contextually conditioned neutralizations, while near-mergers are context-free. In this investigation, we limit our discussions on IN.

IN in tone sandhi forms of Kaifeng Mandarin, which has a four-way lexical tone contrast: LH, HL, H and L.

The question we intended to address in our investigation is whether a theory that relies on the underlying representation of a neutralized surface form, like van Oostendorp (2008), is confirmed by the Kaifeng sandhi data. Again, the prediction it makes is that the phonetic features that distinguish sandhi forms from the underived forms correspond to the phonetic facts of what a realization of the underlying representation of the derived form would produce. In Experiment I, we investigate the phonetic realization of derived representations arising through the *insertion of a tone* which occurs in the same context in underived representations. Kaifeng Mandarin has two tone insertion rules, together interpretable as an OCP-repair. In one case, L is inserted to break up a sequence of H-toned syllables, by which H.H changes to HL.H. In the mirror-image case, H is inserted to break up a L.L sequence to create LH.L. Since underlying HL.H and LH.L are generously available, the prediction of the mini-phonology theory is that derived LH (henceforth $LH^{Ø}$) and LH differ in that $LH^{Ø}$ is pronounced somewhat more like an L-tone than is underived LH and that derived HL (henceforth $HL^{Ø}$) and HL differ in that $HL^{Ø}$ is pronounced somewhat more like an H-tone than is underived HL. It may be relevant to realize that while this OCP-repair is part of the synchronic grammar of Kaifeng Mandarin, in general the history of tone-sandhi patterns may be quite complex and the dialect's predecessors in fact had quite different tone systems.

Besides a potential IN effect of tone insertions, a more direct problem for the mini-phonology theory is the existence of multiple underlying forms, all of which may be neutralized with an existing underived representation. Experiment II therefore focuses on a case in which the sandhi forms have a number of different underlying representations, i.e., neutral tones, such that it is impossible for the phonetic facts to be related to any specific one. This situation occurs in the phonology of the neutral tone, since Kaifeng neutral tones are either underlying or derived from up to four different tonal sources.

In addition to syllables with one of the four lexical tone melodies, Kaifeng Mandarin has lexically toneless syllables that are somewhat shorter (underlying 'neutral tone'). Consonants in the onset and pre-nuclear glides are optional. Kaifeng Mandarin has a set of four monophthongs, a set of six diphthongs, and /ə/, which is always followed by coda nasals /n/ or /ŋ/. The four monophthongs are long, unless followed by a coda nasal, while the diphthongs are always long and cannot be followed by a coda nasal. Additionally, the language has a set of retroflex vowels, which occur in underived words as well as in words derived by attaching a diminutive suffix /ɚ/ to nominal hosts, which rhotacizes a vocalic syllable rhyme (see Chapter 2).

The tonal phonology of Kaifeng Mandarin is briefly introduced in § 5.2. Method and results of two production experiments are presented in § 5.3 and § 5.5, respectively. Finally, § 5.4 and § 5.6 discuss the results and conclude the investigation.

5.2 Kaifeng Mandarin tonal phonology[29]

5.2.1 Tone sandhi in disyllables

The tone sandhi rules of Kaifeng Mandarin appear as an OCP repair operation breaking up identical adjacent level tones, causing H.H and L.L to show up as HL.L and LH.L, respectively, with the 'repair' tone added to the left-hand tone melody, as shown in (1), where tones are syllabic melodies. There is thus no ban on the adjacency of identical phonological tones, meaning that HL.L or LH.H are fully grammatical.

(1) a. H-tone sandhi: H→HL/_H

b. L-tone sandhi: L→LH/_L

5.2.2 Neutral tone in disyllables

Kaifeng Mandarin neutral tone provides a case of IN in a situation in which multiple underlying tone contrasts are neutralized. Neutral-tone syllables come from two sources, suffixation and tone deletion. In the first case, underlyingly toneless grammatical morphemes are suffixed to a base syllable, where they remain toneless and have f0-contours that are determined by the base syllable. Typical examples include the possessive marker [leɛ$^{\varnothing}$], nominal suffix [tsɹ̩$^{\varnothing}$] and modal particle [pa$^{\varnothing}$]. In the second case, the underlying tones of a group of lexically

[29] For the pronunciation and phonological representation of the citation tones, see Chapter 3.

determined monosyllabic morphemes are deleted when occurring in specific morphological contexts. Morphemes that lose their tones in this way include post-verbal object pronouns and the reduplicated syllable of nouns and verbs, besides the final syllables in a number of lexicalized disyllables, where tones are syllabic melodies.

(2) Neutral Tone: T(LH, HL, H, L)→Ø/T(LH, HL, H, L)_

Importantly, the pitch shapes of underlying and derived neutral tones are equally conditioned by the preceding full-tone, a mid level after LH, a fall after HL and H, and a low weak rise after L, as shown in Figure 2.8.

5.3 Experiment I: IN and inserted tones

Experiment 1 was designed to compare $LH^{Ø}$ with LH and $HL^{Ø}$ with HL in terms of three acoustic parameters, rhyme duration, f0-height and f0-slope. Ten speakers were recruited to read disyllabic word pairs that were minimally distinguished by the tone of the first syllable. The methodology and the results are detailed below.

5.3.1 Method

5.3.1.1 Design

A total corpus of 256 disyllabic words (76 test words (19 word pairs × 2 sets × 2 members) and 180 fillers), divided evenly into two reading lists A and B, were chosen in the production experiment. The test words include two sets of 19 (putative) homophonous disyllabic word pairs, one with words representing the contrast between underlying HL.H and H.H (Table 5.1) and one with words representing the contrast between underlying LH.L and L.L (Table 5.2). The word pairs [t^hu kai], [mai ma], [fən $tʂ^h$aŋ] and [$tɕ^h$i ma] have been used by Zee (1980) to study T3S in Standard Mandarin. Fillers had multiple underlying tones and most of them had derived neutral tones on the surface in the second syllable. Since reading minimal pairs increases artificial differences between supposedly neutralized items (Port & Crawford 1989), members of minimal pairs were distributed over the two reading lists such that they never appeared in a single list. The shaded words in Table 5.1 and 5.2 were mixed with 90 filler words, which together were randomly inserted in reading list A, while the unshaded words were similarly distributed in reading list B. The onset of the second syllables of selected words were confined to obstruent, nasal or lateral consonants, in an attempt to obtain a clear boundary between the two syllables within a disyllabic word. Inevitably, the segmental makeup and syllable structure varied across word pairs. These within- and between-word pair differences are captured by structuring individual *item* (i.e., word pair) as a random effect in the subsequent mixed-effects models.

Table 5.1. Minimal pairs used for H-tone sandhi. Members within each word pair are underlyingly distinguished from each other in the tones of the first syllable. The tone sequence H.H is positioned in the first column and HL.H in the second column. Shaded words and unshaded ones were presented in two separate reading lists.

IPA	H.H	gloss	HL.H	gloss
xwɤʌ pa	火把	torch	活靶	target
tʰu kai	土改	land reform	涂改	to retouch
mai ma	买马	to buy the horse	埋马	to bury the horse
lau mi	老米	old rice	捞米	to wash rice
fən tʂʰaŋ	粉场	flour factory	坟场	graveyard
fan swɤʌ	反锁	to lock from inside	繁琐	cumbersome
xau mi	好米	good rice	毫米	millimeter
tʂʰau mi	炒米	fried rice	潮米	wet rice
tɕʰi ma	起码	at least	骑马	to ride a horse
tʰu kou	土狗	rural dog	屠狗	to kill a dog
xu tan	虎胆	tiger guts	壶胆	inner surface of a pot
tʂʰaŋ swɤʌ	场所	location	长锁	long lock
mai xau	买好	to have purchased	埋好	to have buried
tɕʰjaŋ tʂɻ	抢纸	to grab paper	墙纸	wallpaper
ma ljan	马脸	horse face	麻脸	pockmarked face
fei ʂou	匪首	mob boss	肥手	fat hand
tɕʰjaŋ ʂou	抢手	popular	强手	master
tsu tʂaŋ	组长	group leader	族长	patriarch
xai tsau	海藻	seaweed	还早	still early

Table 5.2. Minimal pairs used for L-tone sandhi. Word members within each pair are underlyingly distinguished from each other in the tones of the first syllable. The tone sequence L.L is positioned in the first column and LH.L in the second column. Shaded words and unshaded ones were presented in two separate reading lists.

IPA	L.L	gloss	LH.L	gloss
tɕʰi tjan	气垫	air cushion	漆店	paint shop
tʂʰou tɕʰi	臭气	stink	抽气	to remove air
tʂəŋ pjan	政变	coup	争辩	to debate
piŋ pjan	病变	pathological change	兵变	mutiny
ʂaŋ kʰɤʌ	上课	to attend class	商客	investor
ɕin ɻən	信任	trust	新任	newly-appointed
ɕin tɕjau	信教	religious	新教	Protestant
ʂəŋ ɻən	胜任	competent	升任	promotion
kuŋ tɕi	共计	in total	功绩	achievement
pei ʂɻ	被试	participant	笔试	written examination
tsɻ̩ tʂu	自助	self service	资助	funding
ʂɻ tʰai	事态	situation	师太	nun
tɕi ɕy	继续	to continue	积蓄	savings
ʂɻ pjan	事变	event	尸变	reanimation of the dead
kau tʂwaŋ	告状	to sue	膏状	cream
ɕjan lu	线路	line	仙露	delicious wine
ɕin tɕjan	信件	letter	新建	newly constructed
fən pjan	粪便	excrement	分辨	to distinguish
fan lan	泛滥	to flood	翻烂	to read a book thoroughly

5.3.1.2 Participants

Ten native Kaifeng speakers, 5 males and 5 females, were recruited for the experiment. The age and gender distribution of these speakers is well balanced: 27–28 (1 Male, 1 Female), 39–46 (1 Male, 1 Female), 50–60 (3 Males, 3 Females). All speakers were born and raised in the inner city of Kaifeng and they spoke Kaifeng Mandarin in their daily lives. None of them had self-reported speaking or listening problems. They were paid for their participation.

5.3.1.3 Procedure

The production experiment was conducted in a quiet room. The recording task was divided into four sessions. Prior to the recording, participants were given reading list A to familiarize themselves with the stimuli. In the first two sessions, the test words were randomly presented four times and filler words three times. After the first two sessions, participants were then given reading list B. Again, the test words were randomly presented four times and filler words three times in the last two sessions. Individual words were presented one per slide at the centre of the computer screen, using Prorec (Huckvale 2014). During the recording, speakers were instructed to read the words in a natural (i.e., no exaggerated pronunciation, no narrow focus) and consistent manner, and were allowed to repeat the word if they were not satisfied with their pronunciation. After a participant had read a word, the experimenter would click a 'next' button, making the next word available. Consecutive slides had different colored backgrounds to signal the appearance of a new word. For a few noun compounds like 'air cushion' and 'letter', participants were allowed to apply retroflex suffixation to the second syllable to arrive at a more natural and authentic rendering. Between consecutive sessions, participants took a short break. Audio recordings were sampled at a rate of 20,000 Hz, using a Shure SM10A head-mounted dynamic microphone connected to an ASUS laptop via a Shure X2u pre-amplifier. The whole recording task lasted roughly one hour per participant. Post-experiment feedback from the subjects indicated that one careful young female speaker could maximally recognize four minimal pairs in the experiment, while the remaining participants failed to notice any. This indicates that the subjects' awareness of the purpose of the experiment was negligible.

5.3.1.4 Data analysis

The word pairs [tʰu kai], [mai ma], [fən tʂʰaŋ], [tɕʰi ma], [xai tsau] in Table 5.1 and [ʂaŋ kʰɤʌ], [kuŋ tɕi], [ʂɻ tʰai], [tɕi ɕy], [kau tʂwaŋ] in Table 5.2 were removed from the subsequent analysis, either because their pronunciations were inconsistent across the participants or because some participants asked for the meaning of one of the member words in the familiarization process. The remaining stimuli were naturally occurring items as judged by these participants and they had no difficulty in pronouncing any of them. A total of 2240 disyllables (28 word pairs × 2 members per pair × 4 repetitions × 10 speakers) were thus analyzed in Praat (Boersma & Weenink 1992–2015). Syllable rhymes were manually segmented and labeled in Praat textgrids. In word pairs [ɕjan lu] and [fan lan], the nasal coda of the first syllable and the onset of the second syllable form a single sonorant segment (Xu

1989). The rhyme end of the first syllable was located at the midpoint of the geminated segment. Script Prosodypro (Xu 2013) was used in manually checking and correcting each vocal pulse cycle in the waveform and generating rhyme duration and time-normalized f0 within the rhyme. The f0 values were smoothed by using a trimming algorithm (See Appendix 1 in Xu 1999).

The second L-tone syllable in LH.L and L.L from three female speakers were removed from the acoustic analysis due to creak, as was one word pair produced by a male speaker for the same reason. For these discarded tokens, only the first tone was analyzed. For the remaining 4136 syllables, 21 equidistant f0 measurements were extracted from each syllable rhyme. Six tokens either had poor sound quality or a noticeably hyperarticulated pronunciation, reflected in a 40 Hz higher mean f0 than the other three repetitions. Their duration and f0 were replaced by the mean values of the other three repetitions of that word produced by the same speaker. To better reflect pitch perception (Rietveld & Chen 2006), raw f0 values were converted to semitone based on the formula: $f0_{ST}=12*\log_2(f0_{Hz}/50)$.

Since IN is generally very subtle (e.g., Winter & Röttger 2011), the results are heavily dependent on the statistical method used in the analysis. The various statistical analyses employed in earlier investigations to assess the f0-slopes leave room for improvement. For instance, Xu (1993) and Yuan & Chen (2014) analyzed f0 excursion size, whereby Xu used the difference between max f0 and min f0, while Yuan & Chen used the difference between f0-offset and min f0, without actually examining the time-varying tone shapes. Zee (1980) (reanalyzed by Myers & Tsay 2003), Peng (2000) and Myers & Tsay (2003) utilized a repeated-measures ANOVA with sandhi and measurement points as main effects and looked for a *sandhi-by-point* interaction. However, as noted by Myers & Tsay (2008: 61), 'treating f0 measurement point as a simple multi-level factor misses the fact that the [measurement] points are actually ordered, a fact that cannot be captured in ANOVA.' Finally, a functional block-by-block *t*-test used by Cheng et al. (2013) undermines the statistical power on account of the limited amount of data in each block. To address these problems, Growth Curve Analysis (henceforth CGA) (Mirman 2014) implemented in *lme4* package (Bates, Maechler, Bolker & Walker 2015) in R version 3.3.0 (R Core Team 2016) was used to assess the difference in the longitudinal f0-contours of base tones and sandhi tones. The overall f0 contours were modeled with second-order orthogonal polynomials. The time terms in the polynomial function, i.e., the intercept, the linear coefficient and the quadratic coefficient, correspond to the overall f0-mean, the steepness of the overall slope and the sharpness of peaks of the contours, respectively. Using orthogonal polynomials as opposed to natural ones will avoid the collinearity between time terms, hence allowing for independent and more informed parameter estimates of individual time terms. Since items differed in their segmental composition, syllable structure and morphosyntactic structure, the effects of *item* as well as *speakers* were included as random effects on the overall pattern across both sandhi conditions. Similarly, *speaker-by-sandhi* was included as a random factor to assess the speaker-dependent sensitivity towards the fixed effect *sandhi* on all polynomial time terms in the within-subject design. The fixed effect of *sandhi* on all time terms was added in the base model. The contribution to model fit was reported using the likelihood ratio test. Parameter estimates of the *speaker-by-sandhi* random effect on time terms were further used to calculate individual speaker effect sizes.

5.3.2 Results

5.3.2.1 Rhyme duration

Figure 5.1 displays the rhyme duration of disyllables in the sandhi and non-sandhi conditions. HL.H and H.H are displayed in panel (a) and LH.L and L.L in panel (b). Four sets of linear mixed effect models were created, with *sandhi* as fixed effect and *speakers*, *speaker-by-sandhi*, and *items* as random effects, to compare the rhyme durations of the initial sandhi-tone syllable and the final triggering-tone syllable in two sandhi conditions. The results showed that neither the first nor the second syllable shows a significant temporal difference in the rhyme portion under the influence of sandhi for either HL.H vs H.H or LH.L vs L.L (HL.H: $dur_{\sigma1}$ = 143 ms, $dur_{\sigma2}$ = 179 ms; H.H: $dur_{\sigma1}$ = 142 ms, $dur_{\sigma2}$ = 178 ms; LH.L: $dur_{\sigma1}$ = 153 ms, $dur_{\sigma2}$ = 167 ms; L.L: $dur_{\sigma1}$ = 152 ms, $dur_{\sigma2}$ = 167 ms). The statistical results are summarized in Table 5.3.

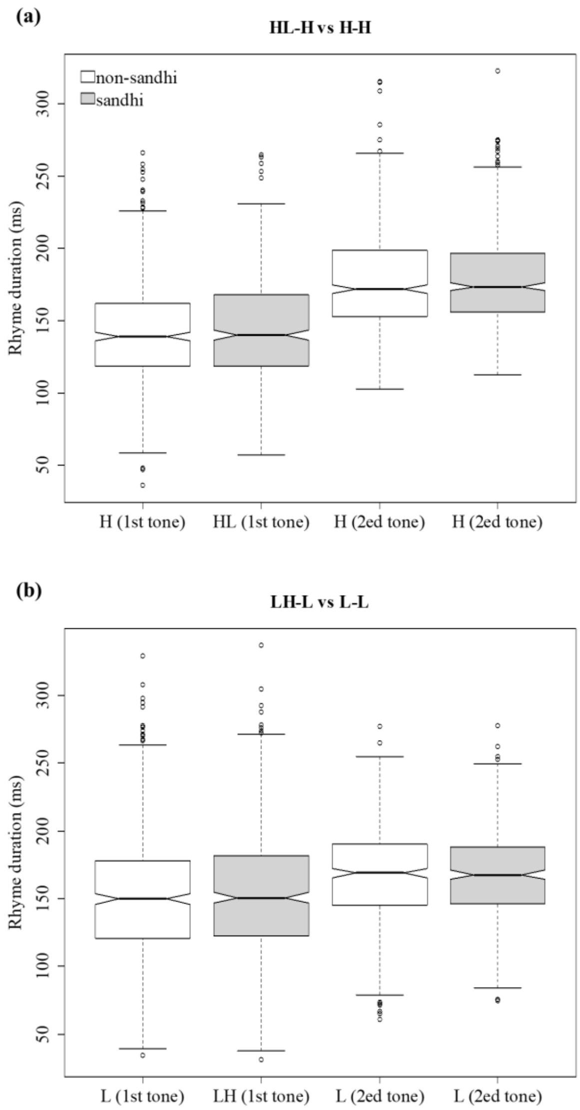


Figure 5.1 Rhyme durations for disyllabic tones in both sandhi and non-sandhi conditions (H.H panel a, white box; HL.L panel a, grey box; L.L panel b, white box; LH.L panel b, grey box). The notches and the bands inside represent the median. The bottom and the top of the box represent the first and third quartiles, the whiskers represent 1.5 IQR (interquartile range), and the small circles represent outliers.

Table 5.3 Duration results of the linear mixed effects model

			logLik	deviance	$\chi^2(1)$	*p*
HL.H vs H.H	First tone	intercept	–4941.5	9883.0		
		sandhi: intercept	–4941.3	9882.6	0.447	0.504
	Second tone	Intercept	–5031.3	10063.0		
		sandhi: intercept	–5031.0	10062.0	0.510	0.475
LH.L vs L.L	First tone	Intercept	–5040.6	10081.0		
		sandhi: intercept	–5039.8	10080.0	1.579	0.209
	Second tone	Intercept	–3559.7	7119.5		
		sandhi: intercept	–3559.7	7119.5	0.011	0.917

5.3.2.2 f0 contour

Figure 5.2 shows the time-normalized f0 contours (Hz) of disyllabic tones averaged over tokens from the five male speakers. The f0 contours for the female speakers are shown separately in Figure 5.3. A visual inspection of Figure 5.2 and Figure 5.3 reveals that the neutralization of tones characterized by OCP sandhi rule is phonetically incomplete. There are small acoustic differences in the f0 realizations of the neutralized tones. The phonetic differences occur in both the initial tone and the final tone. Specifically, $HL^{Ø}$ starts with a lower f0-onset and ends up with a higher f0-offset than HL. This pattern is found in the pronunciation of both male and female speakers. The difference in f0-onset is 1.9 Hz for the male speakers and 4.9 Hz for the female speakers. The difference in f0-offset is 1.3 Hz for the male speakers and 0.6 Hz for the female speakers. The f0 of second H in both H.H and HL.H for both gender groups follows the path of the virtual interpolation and starts from the interpolated f0 difference between the f0-offset of the preceding tones, with the former gradually becoming higher in the second half. The f0 difference between them is less than 1 Hz. The phonetic difference between the realizations of L.L and LH.L is largely symmetrical with these data. $LH^{Ø}$ starts with a lower f0-onset for the male speakers (0.6 Hz) and a higher one for the females (2.3 Hz), and ends with a much lower f0-offset for both gender groups (1.7 Hz for males and 6.2 Hz for females). The f0-slope for $LH^{Ø}$ is generally flatter than LH. The f0-differences in the f0-offset extends to the second L-tones. The carry-over from the preceding tones gradually decreases and remains visible till the tone offset (0.7 Hz for males and 0.8 Hz for females).

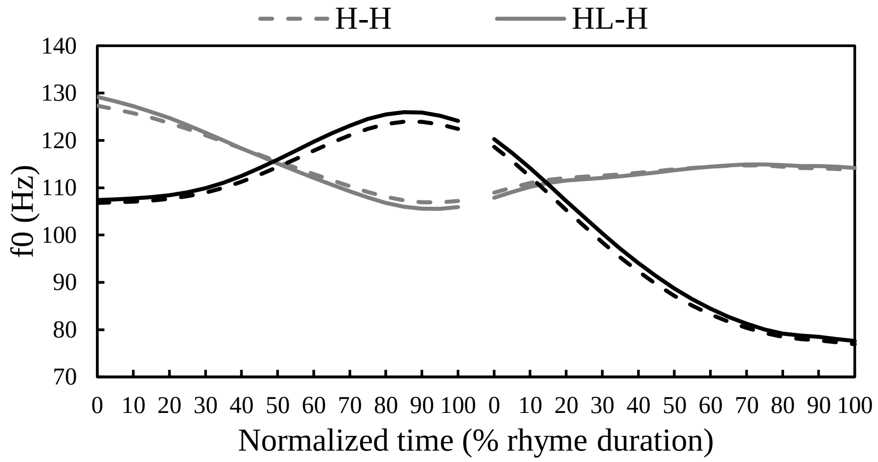


Figure 5.2 Mean f0 contours (in Hz) for disyllabic words averaged over tokens (each contour, n = 280) and five male speakers. The base tones are displayed in solid lines and sandhi tones in dashed lines. Time is expressed as the percentage into the rhyme duration.

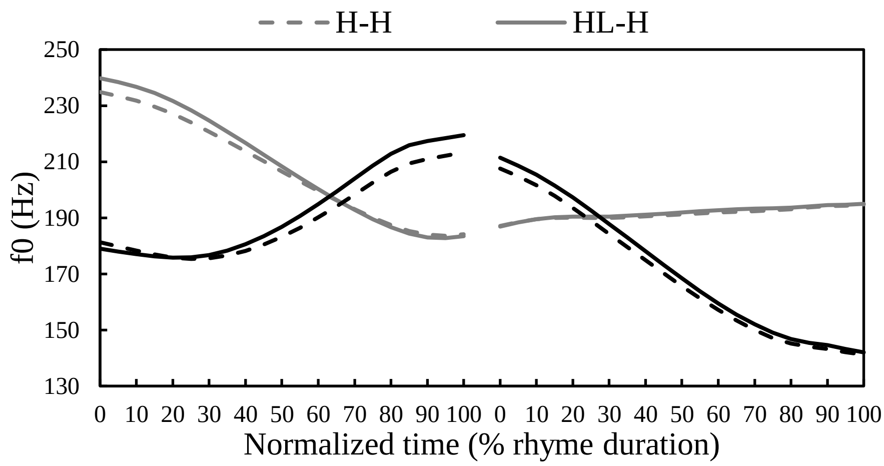


Figure 5.3 Mean f0 contours (in Hz) for disyllabic words averaged over tokens (the second L in LH.L and L.L, n=112, the rest tones, n=280) and five female speakers. The base tones are displayed in solid lines and sandhi tones in dashed lines. Time is expressed as the percentage into the rhyme duration.

GCA analyses on the semitone-transformed f0 contours were conducted. The results are reported using likelihood ratio test; fixed effect parameter estimates and corresponding *p*-values estimated from the normal distribution are summarized in Appendices A–D.

The results by and large confirm the visual observations of the phonetic differences in the base form and the sandhi form. Specifically, for the first tone of H.H and HL.H, results from GCA indicated that the fixed effect of sandhi on the intercept did not significantly improve the model fit ($\chi^2(1) = 0.075$, $p = .784$), indicating that the HL^{Ø} has the same f0-height as HL. The fixed effect of sandhi on the linear term significantly improved the model fit ($\chi^2(1) = 11.098$, $p < .001$). It means that HL^{Ø} has a flatter f0-slope than HL. The fixed effect of sandhi on the quadratic term, however, failed to improve the model fit in a significant level ($\chi^2(1) = 1.187$, $p = .276$), indicating that there is no difference in the degree of peak curvature in the f0 contours between HL^{Ø} and HL. For the second tone, it is shown that the effect of sandhi on the linear term turned out to be marginally significant ($\chi^2(1) = 5.679$, $p = .017$). The presence of the effect of sandhi on the linear term indicates that the second H in HL.H has a steeper f0-slope than that in H.H. The effect of sandhi on the intercept ($\chi^2(1) = 0.000$, $p = .996$) and the quadratic term ($\chi^2(1) = 0.775$, $p = .379$), however, failed to improve the model fit. It indicated that the second tones in two different conditions have the same f0-height and the same degree of centred peak.

As for L.L and LH.L, the intercept of the first tone was significantly influenced by the sandhi manipulation ($\chi^2(1) = 7.630$, $p = .006$), which shows that LH^{Ø} is significantly lower than LH. Additionally, the effect of sandhi on the linear term significantly improved the model fit ($\chi^2(1) = 4.462$, $p = .035$). It means that LH^{Ø} is significantly flatter than LH. The effect of sandhi on the quadratic term, however, failed to improve the model fit ($\chi^2(1) = 2.249$, $p = .133$). It indicates that LH^{Ø} and LH do not differ in the magnitude of the centred peak. The f0-height difference between the second L-tones in L.L and LH.L was captured by a significant improvement of the model fit when the fixed effect of sandhi on the intercept was added in the model ($\chi^2(1) = 9.781$, $p = .002$). By contrast, no significant improvement of model fit was reported, when the fixed effect of sandhi on either the linear ($\chi^2(1) = 0.576$, $p = .448$) or the quadratic terms ($\chi^2(1) = 1.487$, $p = .223$) was included in the model. It shows that the phonetic difference between the two Ls lies solely in the f0-height. That is, the final L in L.L is lower than that in LH.L.

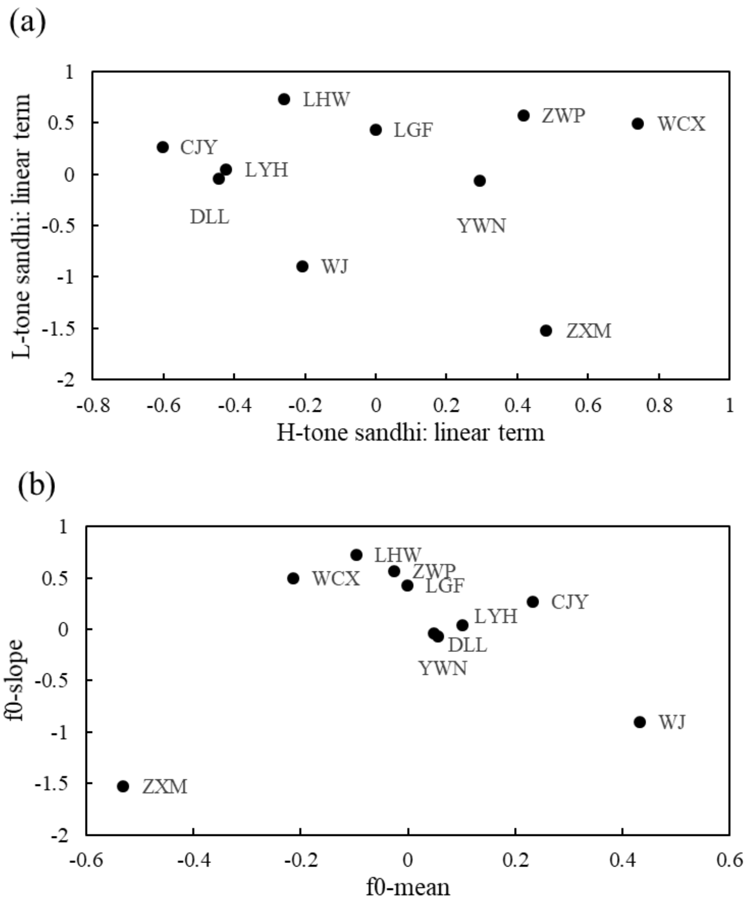


Figure 5.4 Scatterplots of individual effect sizes. Panel (a) summarizes individual effect sizes on the linear term of L-tone sandhi as a function of the effect sizes on the linear term of H-tone sandhi for all speakers. Panel (b) summarizes the individual effect sizes on f0-slope as a function of the effect sizes on f0-mean for L-tone sandhi.

5.3.2.3 Individual variability in sandhi tones

The speaker-dependent sensitivity to sandhi is captured by the speaker-specific effect sizes, obtained by calculating the difference between the random effect parameter estimates of sandhi tones and base tones for each speaker (see Appendix E for the *speaker-by-sandhi* random effect parameter estimates). The random effect parameter estimates (Appendix E) show how individual speakers differ from the overall group mean. In this sense, the individual effect size corresponds to the effect size relative to the group mean effect size. ‘As a result, their individual effect size estimates take into account an estimate of the overall group-level effect size and truly random noise to arrive at a more informed estimate of individual differences' (Mirman 2014: 136). In the present analysis, only the effect sizes for the time terms in the initial tones that showed a significant group effect are calculated, specifically, the linear term of the first tone in H-tone sandhi, and both the intercept and the linear terms of the first tone in L-tone sandhi. Two relevant points are examined.

First, to examine whether participants who show a larger effect size in H-tone sandhi, i.e. a greater difference in f0-slope in $HL^{Ø}$ and HL, also show larger differences between $LH^{Ø}$ and LH, a Pearson's product-moment correlation was calculated between the effect sizes of the linear term in both sandhi conditions (see panel a in Figure 5.4). The result was not significant (r = –0.107, n = 10, p = .768). It indicates that speakers who show more IN in one tone pattern do not necessarily show more IN in the other tone pattern.

Second, to examine whether speakers who show a greater f0-lowering effect for $LH^{Ø}$ relative to LH also show greater slope flattening, a Pearson's product-moment correlation was calculated between the effect sizes of the intercept and the linear term (see panel b in Figure 5.4). The result showed that there is no correlation between the lowered f0-height and the flattened f0-contour (r = 0.177, n = 10, p = .624). It indicates that the magnitude of f0-lowering and slope-flattening in L-tone sandhi is speaker-dependent. Some may show a clear f0-lowering effect but nearly parallel f0 contours, while others may maintain a larger f0-slope difference but situate the f0-height of the sandhi LH at the same level as the non-sandhi LH.

5.4 Interim discussion: Questioning the 'mini-phonology' hypothesis

Experiment I revealed that the insertion of L in H.H (to form HL.H) and of H in L.L (to form LH.L) is acoustically incomplete, because (minute) differences between these sandhi forms and the realizations of the underived LH.L and HL.L in f0-height and f0-slope were found. While differences are both variable, across the two tone insertions and across speakers, as well as small, they are in line with widely attested IN effects, which can be very small, as noted in § 5.1. The overall f0-height difference in Kaifeng Mandarin tone sandhi is comparable to the 2.3 Hz difference reported in Peng (2000) for T3S in Taiwan Mandarin and the 3.5 Hz difference in Xu (1993) for Beijing Mandarin, but smaller than the 17.5 Hz difference in Zee (1980) for Beijing Mandarin or the 28 Hz difference in the f0-offset. The large differences in Zee (1980) may be due to the restricted number of items (five) and speakers (two females) involved in the experiment. Also, the direction of the phonetic difference in L-tone sandhi resembles that of its phonological equivalent, T3S, in Standard Mandarin. Consistently, despite the various magnitudes of the difference, the derived LH is consistently lower than the base LH in T3S (Zee 1980; Xu 1993; Peng 2000).

The comparability of research results is inevitably degraded by the variation in statistical methods. GCA, as adopted in this investigation, is more powerful than any of the measures used in these earlier analyses. Our finding that L-tone sandhi is characterized by both f0-lowering and slope-flattening may well also be valid for Standard Mandarin T3S, for which significant differences in only one of these two measures have been reported (Peng 2000 for only f0-height; Cheng et al. 2013 for only f0-slope), although both have been found in the same data by Zee (1980) and (Xu 1993). Another difference may lie in the inter-subject and inter-dialect variability. Examination of individual cue-weighting effects in L-tone sandhi indicated that the relative weighting between these two cues is speaker-dependent, i.e., that speakers are free to balance one cue against another. Similar speaker-dependent cue-weighting effect is found in Final Devoicing. Piroth & Janker (2004), for instance, found a considerable speaker-dependent variability in the durational cues of German Final Devoicing.

The crucial question to address now is whether the ways in which the sandhi tones are pronounced can be argued to reflect the underlying representation. That is, is $LH^{\emptyset}$ pronounced so as to approximate L in the context before L? Similarly, is $HL^{\emptyset}$ pronounced so as to approximate H in the context before H? There is some ambiguity in the answer. On the one hand, it can be argued that an inserted tone has no corresponding tone in the underlying representation and that there is therefore no basis for comparison. On the other hand, it can be argued that the realizations of the full disyllabic representations $LH^{\emptyset}$.L and $HL^{\emptyset}$.H are more like the expected phonetic reflexes of L.L and H.H than those of LH.L and HL.H, respectively, since L.L would have level low pitch, which is closer to the lowered peak of $LH^{Ø}$.L compared to LH.L, while H.H would have high level pitch, which is closer to the flatter valley of $HL^{Ø}$.H compared to HL.H. In addition, it could be argued that the higher onset of the falling movement in HL.H is to be attributed to the sensitivity of H-raising to the underlying L, an effect that can be argued to be less pronounced if the position of L is Ø in the underlying representation, as it is in the sandhi form $HL^{Ø}$.H.

Equally, however, we might argue that the sandhi forms generally show a small degree of hypoarticulation (Lindblom 1990), an explanation that would readily transfer to other tonal IN data. The flattened f0-slope in L-tone sandhi in Kaifeng Mandarin also holds for T3S in Mandarin. Zee (1980) measured the tones in the onset, dip (if present) and the offset. Based on the result of a repeated-measures ANOVA of the data in Zee (1980) by Myers & Tsay (2003), there was a highly significant interaction between tone class and the three measurement points, indicating that the derived LH is flatter than the original LH. Xu (1993) found a significantly smaller f0 excursion size of derived LH than the original LH. Cheng et al. (2013) compared the velocity profiles of the sandhi LH and the original LH and found that sandhi LH rises less steeply than the original LH in the terminal portion of the tone. Similarly, in a corpus study of T3S, Yuan & Chen (2014) characterized IN in two f0 measurements, the log difference between the offset and the f0-minimum, i.e. the f0 range and the temporal percentage of f0-rise. They showed that sandhi LH has a smaller f0 range and a shorter duration of the f0-rise. In fact, the generally flatter f0-shapes of the Kaifeng Mandarin sandhi tones resembles the '*yi* sandhi' reported in Myers & Tsay (2003), a morphophonemic alternation in Mandarin. The Mandarin morpheme *yi* (一) 'one' has T1 (H) in isolation, T2 (LH) when followed by T4 (HL), and T4 (HL) elsewhere. Myers & Tsay (2003: 46) compared the derived LH, as in yi^{LH} shi^{HL} (一世) 'all one's life' and derived HL, as in yi^{HL} ji^{L} (一己) 'oneself' with tonally homophonous compounds, as in yi^{LH} shi^{HL} (仪式) 'ceremony' and yi^{HL} ji^{L} (异己) 'opponent', respectively, and found that the derived HL and LH are shorter and flatter than their non-derived counterparts. More specifically, derived HL starts with a lower f0-onset and ends with a higher f0-offset and derived LH starts with a higher f0-onset and ends with a lower f0-offset (see Figures 5 and 6 of Myers & Tsay 2003: 53). Importantly, Myers & Tsay concluded that these differences are not attributable to the underlying form of the morpheme. Since the derived tones were shorter, they appealed to a hypothesis that '[the morpheme '一'] was pronounced with less stress than the morphemes it was compared with' (Myer & Tsay 2003: 56). We would like to suggest that the realizations of sandhi tones are neither attributable to their underlying representations nor to prosodic stress as such, but *to a reduction in the confidence with which speakers pronounce them*. That is, a surface form with a synchronic derivational history is treated with a small degree of inhibition, as a result of which phonetic features that cue the phonological elements of identical underived surface forms are less canonically present in the realization of the derived form.

This explanation implies a number of properties. First, IN results from the co-existence of synchronically derived surface forms and otherwise identical underived surface forms. This characterizes 'near-mergers' as a different phenomenon. A near-merger amounts to a phonetic approximation between two phonologically different segments, which resists phonological neutralization in the generation of speakers concerned (Labov 1994). Second, unlike the proposal by van Oostendorp (2008), our proposal characterizes the derived and underived surface phonological representations showing IN as identical, meaning that it is in agreement with circumstantial evidence for neutralization found in spelling errors, rhyme, puns and speaker intuitions. Third, there is no need for the assumption of non-identical exemplar clouds, which must of course be assumed in the case of 'near-mergers' (Yu 2007). This liberates IN from the otherwise inevitable claim that phonetic traces of a pre-neutralization stage are still present long after the neutralization was attested. In the case of German Final Devoicing, this would have to be assumed for some 25 generations, an unlikely time span for such subtle phonetic differences to survive. That assumption is also at odds with the fact that some speakers are apparently deaf to effects of L-insertion, but not to those of H-insertion, or vice versa. Nor would it explain why some speakers select one phonetic feature in signaling the difference and others some other feature.

The rival hypothesis that IN is a phonetic reflection of an underlying phonological contrast would be more directly under threat if it were to be found that IN occurs in cases in which neutralized forms have a number of different underlying representations, such that it is impossible for the phonetic facts to be related to any one specific underlying phonological representation. This case is addressed in Experiment II.

5.5 Experiment II: Multiple underlying forms

Experiment II was designed to compare realizations of neutral tones from different inputs. Specifically, we will make two comparisons. First, realizations of an underlying neutral tone will be compared with realizations of a derived neutral tone. Second, realizations of a derived neutral tone will be compared with realizations of another derived neutral tone which has a different underlying form. In the first case, shown in (3a), the underived neutral tone does not have an underlying representation, while the derived form has LH as its underlying form. If a phonetic difference indicating IN is to be explained in terms of the 'mini-phonology' hypothesis, we should expect the realization of the derived form to show more of a rising pitch shape in the second syllable, compared to a flatter realization of the underived form. Similarly, in the second case, shown in (3b), where the two neutral-toned syllables have different underlying tones, that hypothesis predicts a more rise-like realization in the case of underlying LH and a more fall-like realization in the case of HL. Our 'inhibited pronunciation' hypothesis, by contrast, predicts a slightly hypoarticulated pronunciation of the derived form of (3a) and no difference between the two derived forms of (3b).

(3) a. [kɤʌ$^{\text{LH}}$ pa$^{\text{Ø}}$] 'to cut' vs [kwɤʌ$^{\text{LH}}$ pa$^{\text{LH}\rightarrow\text{Ø}}$] 'scorched rice'

b. [ɕja$^{\text{L}}$ t$^{\text{h}}$jan$^{\text{LH}\rightarrow\text{Ø}}$] 'summer' vs [tɕja$^{\text{L}}$ tɕ$^{\text{h}}$jan$^{\text{HL}\rightarrow\text{Ø}}$] 'price'

5.5.1 Method

One near minimal word pair was chosen for each case, as given in (3a) and (3b), respectively. In (3a), the left-hand word has a modal particle [pa$^{\text{Ø}}$] attached to the verb stem [kɤʌ$^{\text{LH}}$], while the right-hand word is a lexicalized compound. Both words in (3b) are lexicalized compounds. Three repetitions of each word as produced by 13 participants (including the 10 participants in § 5.3.1.2) were recorded under the same conditions as in Experiment I. Scripts for the three speakers who did not participate in Experiment I were shorter, but similar to those used for Experiment I. Acoustic analyses were made in the manner as described in § 5.3.1.4. The pronunciations by two speakers of the words in (3b) had jitter and were not included in the analysis. One male speaker had creak in the first syllable of the word [ɕja^{L} t^{h}jan$^{LH\rightarrow Ø}$] in one token, the f0 of which was replaced by the mean of the other two repetitions.

5.5.2 Results

5.5.2.1 Underlying neutral tone vs derived neutral tone

The f0 contours displayed in Figure 5.5 show that the disyllable with the underlying neutral tone is realized with a wider f0 range than that with the derived neutral tone, ending 2 Hz higher. A second-order GCA analysis on the semitone-transformed f0 contours was constructed, with random effect of *speaker* and *speaker-by-conditio*n and fixed effect of *neutral-tone-type* (underlying/derived neutral tone) on all time terms.[30] It showed that there was a significant effect of *tone type* on the intercept ($\chi^2(1) = 7.811$, $p = .005$), indicating that the underlying neutral tone is higher than the derived one. The fixed effect of *tone-type* on the linear ($\chi^2(1) = 0.214$, $p = .644$) and quadratic terms ($\chi^2(1) = 0.115$, $p = .734$) were not significant, indicating that there is no slope difference between the underlying neutral tone and the derived neutral tone. The LH-tone before the underlying neutral tone has a more dipped realization than that before the derived neutral tone. This was captured by the significant effect of *neutral-tone-type* on the quadratic terms ($\chi^2(1) = 11.095$, $p = .001$). No difference of f0-height between these two LH-tones was reported, as shown by the fact that *neutral-tone-type* did not have a significant effect on the intercept ($\chi^2(1) = 0.6894$, $p = .406$). Nor was there an effect of *neutral-tone-type* on the linear term ($\chi^2(1) = 0.7572$, $p = .384$). The rhyme duration for the underlying neutral tone is 18.7 ms shorter than that of the derived neutral tone ($\chi^2(1) = 14.606$, $p < .001$); no duration difference was found for the first syllables in the two conditions ($\chi^2(1) = 0.661$, $p = .416$), as confirmed by a linear mixed effects model structured with the random effect of *speaker* and *speaker-by-condition* and the fixed effect of *neutral-tone-type* on the intercept

[30] In the comparison of the first LH-tones, the model with maximum random effects failed to converge. Therefore, it was simplified by removing the *speaker-by-conditio*n random effect on the quadratic term.

(see Appendices F–G for the fixed effect parameter estimates and the corresponding *p*-values estimated from the normal distribution).

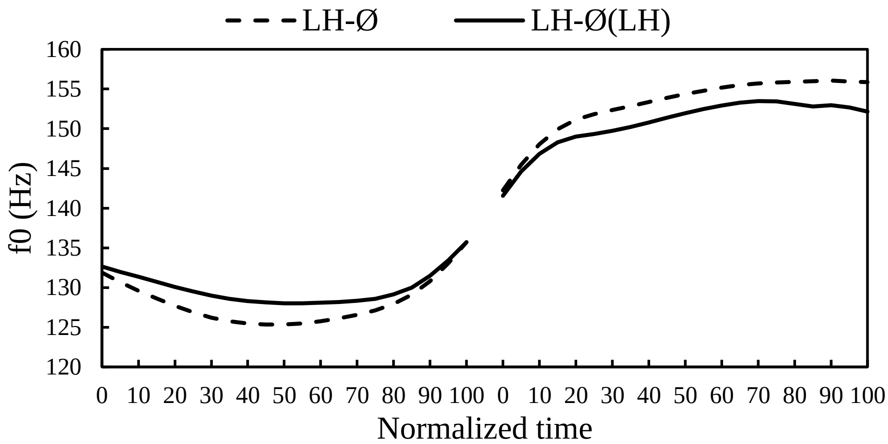


Figure 5.5 Mean f0 contours (in Hz) for disyllabic words averaged over tokens (LH.Ø, n=39; LH.Ø^{LH}, n=39) and 13 speakers. Underlying neutral tone is displayed in dashed line and derived neutral tone in solid line. Time is expressed as the percentage into the rhyme duration.

5.5.2.2 Derived neutral tones with different underlying tones

Figure 5.6 presents f0 contours for the derived neutral tones that have different underlying tones. As shown, the f0 contours of the two derived neutral tones as well as the preceding full L-tones overlap considerably. Results from GCA showed that for the f0 contours of the neutral tones as well as the preceding tones there was no effect of *tone-type* on the any of these time terms (neutral tone, intercept: $\chi^2(1) = 0.0905$, p =.764, linear: $\chi^2(1) = 0.0038$, p =.951, quadratic: $\chi^2(1) = 0.9386$, p = .333; preceding full tone, intercept: $\chi^2(1)$ =0.0138, p = .906, linear: $\chi^2(1)$ = 0.0152, p = .902), quadratic: $\chi^2(1) = 0.6904$, p = .406), indicating complete neutralization between the derived neutral tones (see Appendices H–I for the fixed effect parameter estimates and the corresponding *p*-values estimated from normal distribution).

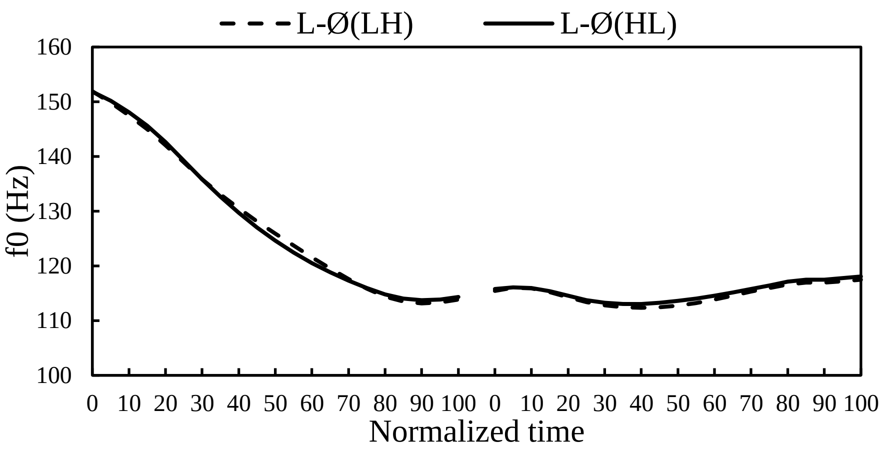


Figure 5.6 Mean f0 contours (in Hz) for disyllabic words averaged over tokens (L.Ø^{LH}, n=33; L.Ø^{HL}, n=33) and 11 speakers. Neutral tone derived from underlying LH is displayed in dashed line and neutral tone derived from underlying HL in solid line. Time is expressed as the percentage into the rhyme duration.

The results of both comparisons, therefore, violate the predictions of the 'mini-phonology' hypothesis and support the inhibition hypothesis. The first comparison made it clear that the f0 contour in the disyllable with an underlying LH in the second syllable is somewhat hypoarticulated relative to the form with the underlying neutral tone. This is true despite the fact that the duration of the underlying form with neutral tone is shorter than that of the derived neutral tone, a reflection of a difference in 'stress' which is inherent in syllables with underlying neutral tone. Our explanation, therefore, is superior to that offered by Myer & Tsay (2003), who held 'stress' responsible for the effects of IN. The second comparison made it clear that the underlying difference between LH and HL is in no way reflected in the neutralized surface forms, which do not show IN, in accordance with our hypothesis that derived forms have inhibited pronunciations relative to underived forms, predicting equal degrees of hypo- or hyperarticulation.

5.6 Conclusion

In this investigation, we have argued against the hypothesis that incomplete neutralization (IN), the existence of small phonetic differences between phonologically identical derived and underived forms, is explained by the phonological representation of the underlying forms. We refer to it as the 'mini-phonology' hypothesis. One explicit proposal is that by van Oostendorp (2008), who introduced a superscript notation specifying the underlying form in the surface form. Instead, we propose that the effects are explained by the speaker's inhibition in pronouncing surface forms that have undergone a derivational change relative to their underlying form. In order to support our 'inhibition' hypothesis, we have presented cases from Kaifeng Mandarin, a language with four lexical syllabic tone melodies (LH, HL, H, L), in which (a) the underlying form has no phonological correspondent for an inserted tone in the surface form (Experiment I); (b) the underlying form has a representation that has no correspondent in the surface form through the deletion of the syllabic tone melody (Experiment II). Our choice of these cases was inspired by the potentially problematic relation between the surface representations and their underlying representations. A strict interpretation of the 'mini-phonology' hypothesis would require the establishment of the phonetic effects of underlyingly empty representations, while surface forms that lack a correspondent of phonological tones in the underlying form are predicted to reflect the underlying representation in phonologically unspecified surface locations.

Experiment I showed that sandhi forms containing an inserted tone, $LH^{\varnothing}.L$ as derived from L.L and $HL^{\varnothing}.H$ as derived from H.H, have different pronunciations from underlying LH.L and HL.H, respectively. An interpretation of this IN effect requires a comparison of the f0 contours of the disyllabic derived and underived forms. The contour for derived $LH^{\varnothing}.L$ was slightly lower and flatter than that of underived LH.L, while that of derived $HL^{\varnothing}.H$ was flatter than that of underived HL.H. These results are readily interpretable as inhibited pronunciations of the derived forms, i.e., slightly hypoarticulated (Lindblom 1990) pronunciations. However, the mini-phonology hypothesis arguably explains this result in that the derived contour can be said to approach the low level and high-level contours that would theoretically be produced by L.L and H.H, respectively. Against this interpretation, it could be argued that the realization of a phonological contrast would be expected to be more consistent than what we found the data to be. Specifically, speakers who showed more IN in one tone pattern do not necessarily show more IN in the other tone pattern, while they also varied in the specific phonetic parameters they used in their IN.

Experiment II attempted to collect less ambiguous data. It revealed IN between an underlyingly toneless syllable and a derived toneless syllable arising through the deletion of LH, i.e., two ‘neutral tones’, both occurring after a LH-tone in disyllabic words. The underived neutral syllable was 2 Hz higher and 18.7 ms shorter than the derived neutral syllable, while the preceding LH before the underived neutral tone had a more dipped realization as compared to the derived neutral tone. These differences are not straightforwardly accounted for by relating the IN to the difference in the underlying representations. Quite apart from the fact that the underlying neutral tone does not have a phonological specification, the mini-phonology hypothesis predicts that the deleted LH in the derived neutral syllable is reflected in the phonetic realization. The prediction must therefore be that there is a more canonical rising contour in the second syllable. However, the opposite is true. The inhibition hypothesis predicts that the derived neutral tone has an inhibited pronunciation of the f0 contour, which prediction is correct, since the f0 range across the disyllable is smaller than in the underived case. What, at first sight, it does not predict is that the duration of the syllable with the derived neutral tone is longer than that of the underived one, since inhibited pronunciations might be associated with shorter durations. However, underlyingly neutral-toned syllables are shorter than toned syllables, as noted in § 5.1, which must be due to their being monomoraic syllables, as opposed to the bimoraic toned syllables. In this interpretation, the neutralization arising from the tone deletion leaves the mora structure intact.

A second data set obtained in Experiment II further undermined the mini-phonology hypothesis. The underlying tones that are deleted in the creation of derived neutral tones include the full set of four tonal melodies of Kaifeng Mandarin. The mini-phonology hypothesis predicts that the f0 contours of derived neutral syllables will show phonetic traces of each of these underlying sources. In reality, while the first data set showed IN in a comparison between a derived neutral tone and an underived one, no IN was observed in any of the comparisons between derived neutral tones. By contrast, the inhibition hypothesis in fact predicts just these results, since an inhibited pronunciation strategy will equally apply to all derived neutral tones.

The assumption that IN does not reflect a difference in phonological surface representations, but arises through the speaker’s inhibition in the pronunciation of surface forms containing phonological adjustments effected in the synchronic grammar suggests that it must be classed with phenomena that can equally be explained by speakers' monitoring and modifying their speech output. Hayes (1994: 64) referred to the mechanism behind one of these as the 'beast' in the speaker, who 'monitors our speech constantly, assessing how much it cares about what we are saying at that instant, and adjusts the pitch range of our voices accordingly'. In addition to linguistic factors, like syntactic phrasing that fails to be reflected in the prosodic hierarchy (the specific phenomenon Hayes was referring to), morphemic differences of the sort investigated by Ingo Plag and colleagues (cf. *tax* vs *tacks*, Plag et al. 2017), and derivation (as hypothesized in this article), the beast must be held responsible for word frequency effects, speaker convergence and divergence (Giles et al. 1991), cultural associations (Kauyumari et al. 2016) as well as any factors that generally relate to the variation between hypo- and hyperarticulation (Lindblom 1990).

In sum, while the mini-phonology hypothesis fails to account for the full set of results of our production experiments, the inhibition hypothesis effortlessly predicts and explains the effects of IN.

References

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01. R package (version 1.1-12) available at http://CRAN.R-project.org/package=lme4

Boersma, P. & Weenink, D. (1992–2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved 24 March 2015 from http://www.praat.org/

Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.

Charles-Luce, J. & Dinnsen, D. (1987). A reanalysis of Catalan devoicing. *Journal of Phonetics*, 15, 187–190.

Cheng, C., Chen, J. & Gubian, M. (2013). Are Mandarin Sandhi Tone 3 and Tone 2 the same or different? The Results of Functional Data Analysis. In *27th Pacific Asia Conference on Language, Information, and Computation*, Taipei, Taiwan, 296–301.

Dinnsen, D. (1985). A re-examination of phonological neutralization. *Journal of Linguistics*, 21, 265–279.

Dinnsen, D. & Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics*, 12, 49–60.

Dmitrieva, O., Jongman, A. & Sereno, J. (2010). Phonological neutralization by native and non-native speakers: The case of Russian final devoicing. *Journal of Phonetics,* 38, 483–492.

Ernestus, M. & Baayen, H. (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In L. Goldstein, D. H. Whalen, C.T. Best (Eds.) *Laboratory Phonology VIII*, 2–49. Berlin:Mouton de Gruyter.

Giles, H., Coupland, J. & Coupland, N. (1991). Accommodation Theory: Communication, Context, and Consequence. In H. Giles, J. Coupland, N. Coupland (Eds.) *Contexts of Accommodation*, 1–68. New York, NY: Cambridge University Press.

Hayes, B. (1994). Gesture in prosody: Comments on the paper by Ladd. In P.A. Keating (Ed.) *Phonological structure and phonetic form: Papers in Laboratory Phonology III*, 64–75. Cambridge University Press.

Huckvale, M. (2014). Prorec: a program for field workers. [Computer program]. Version 1.45, retrieved from http://www.phon.ucl.ac.uk/resource/

Kauyumari, S., Hay, J. & Nilson, E. (2016). Contextual activation of Australia can affect New Zealanders' vowel productions. *Journal of Phonetics*, 48, 76–95.

Labov, W. (1987). The overestimation of functionalism. In R. Dirven, V. Fried (Eds.), *Functionalism in linguistics*, 311–332. Amsterdam: John Benjamins.

Labov, W. (1994). *Principles of linguistic change: Internal factors.* Oxford: Blackwell.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, A. Marchal (Eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Springer.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton: CRC Press.

Mitleb, F. M. (1981). Segmental and Non-segmental Structure in Phonetics: Evidence from Foreign Accent. PhD thesis, Department of Linguistics, Indiana University, Bloomington, Indiana.

Myers, J. & Tsay, J. (2003). Investigating the phonetics of Mandarin tone sandhi. *Taiwan Journal of Linguistics*, 1(1), 29–68.

Myers, J. & Tsay, J. (2008). Neutralization in Taiwan southern min tone sandhi. In Y. E. Hsiao, H.-C. Hsu, L.-H. Wee, D.-A. Ho (Eds.), *Interfaces in Chinese phonology: Festschrift in honor of Matthew Y. Chen on his 70th birthday*, 47–78. Academia Sinica, Taiwan.

Peng, S. H. (2000). Lexical versus “phonological” representations of Mandarin sandhi tones. In M. B. Broe, J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon*, 152–279. Cambridge, England: Cambridge University Press.

Piroth, H. & Janker, P. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics*, 32, 81–109.

Plag, I., Homann, J. & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 53, 181–216.

Port, R. & Crawford, P. (1989). Pragmatic effects on neutralization rules. *Journal of Phonetics*, 16(4), 257–282.

Port, R., Mitleb, F. & O’Dell, M. (1984). Neutralization of obstruent voicing in German is incomplete. In A. Koutsoudas (Ed.), *The Application of Ordering of Grammatical rules*, 55–73. The Hague: Mouton.

Port, R. & O'Dell, M. (1986). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13, 455–471.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (version 3.3.0) URL http://www.R-project.org/.

Rietveld, T. & Chen, A. (2006). How to obtain and process perceptual judgements of intonational meaning. In S. Sudhoff, D. Lenortová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, J. Schieβer (Eds.), *Methods in empirical prosody research*, 283–319. Berlin: Walter de Gruyter.

van Oostendorp, M. (2008). Incomplete devoicing in formal phonology. *Lingua*, 118, 1362–1374.

Warner, N., Jongman, A., Sereno, J. & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, 32, 251–276.

Winter, B. & Röttger, T. (2011). The nature of incomplete neutralization in German: Implications for laboratory phonology. *Grazer Linguistische Studien* 76 (Herbst 2011); S. 55–74.

Xu, Y. (1986). 普通话音联的声学语音学特征[Acoustic-phonetic characteristics of junctures in Mandarin]. *中国语文*[Studies of the Chinese language], 5, 353–360.

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

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, 27, 55–105.

Xu, Y. (2013). ProsodyPro ─ A tool for large-scale systematic prosody analysis. In *proceedings of tools and resources for the analysis of speech prosody* (TRASP 2013), Aix-en-Provence, France, 7–10.

Yu, A. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, 24, 187–214.

Yuan, J. & Chen, Y. (2014). 3rd tone sandhi in Standard Chinese: A corpus approach. *Journal of Chinese Linguistics*, 42, 218–234.

Zee, E. (1980). A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers*, 49, 98–116.

CHAPTER 6

SUMMARY AND CONCLUSION

This dissertation provided a phonetic account of the four syllabic tone melodies, referred to as LH (T1), HL (T2), H (T3) and L (T4), in the Mandarin dialect of Kaifeng, a language which has not been empirically described and analyzed. Following the introductory remarks in Chapter 1, Chapter 2 reanalyzed the synchronic phonology, using a phoneme-based approach, while Chapters 3–5 reported production and perception experiments on the tone structure. The production experiment in Chapter 3 investigated the inter- and intra-speaker variation in the acoustic realizations of L and interpreted the variation as ways of L-tone enhancement. Chapter 4 extended the production experiment in Chapter 3 by conducting a perception experiment in an attempt to see whether the L-tone variations reported in Chapter 3 actually reinforce the perceptual salience of L in lexical tone contrasts. Finally, Chapter 5 reported two cases of incomplete neutralization in the tonal phonology. The present concluding chapter echoes three major objectives of this dissertation, the phonology of Kaifeng Mandarin (§ 6.1), L-tone variation (§ 6.2) and incomplete neutralization of sandhi tones (§ 6.3). Implications for future research and concluding remarks appear in § 6.4 and § 6.5, respectively.

6.1 Synchronous phonology of Kaifeng Mandarin

As noted in Chapter 1, previous descriptions of this Mandarin variety by Chinese dialectologists were syllable-based, as opposed to segment-based. The sound structure has been organized into an initial inventory, a rather complicated inventory of finals (i.e., syllable rhymes) and a tone inventory (see § 1.3). The descriptions were impressionistic and were not supported by empirical data. Chapter 2 aimed to fill this gap. The reanalyzed phonological structure was in terms of consonants, vowels and tones, which has been the standard practice in the Western tradition, and acoustic data were provided to illustrate the sound patterns at issue.

The analysis yielded 22 phonemic consonants, i.e., six plosives /p^h p t^h t k^h k/, four affricates /ts^h ts tʂ^h tʂ/, three nasals /m n ŋ/, four fricatives /f s ʂ x/, one lateral approximant /l/ and four central approximants /ɻ j ɥ w/. The alveolo-palatals [tɕ tɕ^h ɕ] were analyzed as allophones of alveolars /ts ts^h s/ before /i j y ɥ/, i.e., the natural class of high front (semi-) vowels. The onset fricative [ʐ] and apical vowels [ɿ ʅ] were analyzed as allophones of the retroflex approximant /ɻ/.

A simpler and clearer vowel system than produced in earlier analyses was proposed. Generally, the analysis of vowels has been very controversial in the Mandarin phonological literature. In the current analysis, first of all, the pre-vocalic semi-vowels, the 'medials' of traditional analyses, were treated as consonantal approximants /j w ɥ/, not the pre-nuclear high vowels /i u y/, excluding GV diphthongs and GVV triphthongs as a result. Next, vowels were grouped into monophthongs /i y u a ə/, diphthongs /ai au ei ou eɛ ɤʌ/ and retroflex vowels /ɯ˞ oɯ˞ ɤʌ˞ aɯ˞ ɚ ɐ˞/. The classification of monophthongs and diphthongs was evidenced by the greater spectral Euclidean distances of the diphthongs than the monophthongs, measured from the [F1, F2] at the 20% of the vowel duration to the [F1, F2] at the 80% of the vowel duration in the F1/F2 acoustic space, as well as by the phonological distribution of the nasal coda which occurs only after monophthongs.

§ 2.5 provided the first comprehensive analysis of the tone structure, encompassing the citation tones, sandhi tones and neutral tones. In disyllables, T3 (H) becomes T2 (HL) before another T3 (H), while T4 (L) becomes T1 (LH) before another T4 (L). The realizations of neutral tones are dependent on the preceding tone, a mid-level after LH, a fall after HL and H, and a low-level (weak rise) after L. The phonological representation of the tones was discussed in Chapter 3. In particular, the representation of T4 as L was defended on four arguments. First, L serves a natural complement of a symmetrical four-tone system with LH, HL and H. Second, it allows for a better generalization of the phonological tone sandhi pattern, H.H→HL.H and L.L→LH.L, as an OCP repair strategy. Third, an incorrect representation of T4 as HLH would predict the undershooting of the L, which is not the case in Kaifeng L. Fourth, L triggers the widely attested pre-L raising effects of preceding HL and LH tones.

6.2 L-tone variation: production and perception

Chapter 3 and Chapter 4 focused on the production and perception of the Kaifeng Mandarin L-tone, which has been transcribed either as a dip (312) or as a fall (31) in the literature. In Chapter 3, ten native speakers were recruited to read 35 sets of tonal minimal quadruplets in an overall randomized order. After extracting, manually correcting and normalizing the speakers' f0 contours, the acoustics brought out L as the most variable tone among the four. The f0 slope of L had the largest inter-speaker variation and the smallest intra-speaker variation. This variation was mainly caused by three speaker-dependent f0 contours in the realizations of L, dipping, falling and falling with lengthening, as evidenced by the visual inspection of tone contours and an analysis of the f0 slope in the second half of the f0 contour. In addition to the f0 feature, L-toned syllables remained the longest among the four and were frequently pronounced with non-modal phonation.

This inter-speaker variation in L explained the different transcriptions of this tone that had appeared in the literature. The variation was interpreted as enhancement of the L-tone, because an L-tone is intrinsically less salient than an H-tone, due to a lack of phonetic space in the low pitch range as well as to a potential ambiguity between contextual low pitch around f0 peaks and low pitch due to L-tone, and thus more likely to be enhanced. The enhancements include the f0 initial fall (in dipping and falling), the final rise (in dipping), the creak as well as the lengthened low f0 stretch and the host syllable. The enhancements were further reinforced in emphatic speech, which strengthened this argument.

The question arises as to which subtype of L better discriminates L from the other three rival tones in the paradigmatic tone system, LH, HL and H. Chapter 4 intended to answer this question by means of a tone identification experiment. Tone exemplars of the dipping L and the falling L were pitted against the exemplars of three rival tones, creating six ten-stepped tone continua. The tone-bearing syllables, pronounced by a male and a female speaker, were manipulated so as to produce two durational conditions in an attempt to investigate the effect of syllable lengthening on the perception of L, while intensity and phonation were controlled for. Tone continua which were properly situated in the speakers' f0 range, were superimposed evenly on the rhyme of the target syllables. Forty Kaifeng native listeners participated in the identification task. They were asked to choose one of the two Chinese characters displayed on the computer screen, one representing a L-tone and the other a non-L tone. To examine the effect of L-tone variant, syllable duration and speaker voice, the resulting categorical data were input into three separate linear mixed effects logistic regression models for three tone contrasts LH-L, HL-L and H-L.

It was revealed that the recognizability of two L-variants is dependent on the contrast L enters into. The dipping L was a good discriminator from HL and H, not from LH, while the falling L was a good discriminator from LH and H, not from HL. When the rival tone lacked f0 movement, as in the contrast L-H, or showed the opposite movement to the enhancement, as in HL-dipping L and LH-falling L, the identification curve was S-shaped, typical of categorical perception. By contrast, when the rival tone had the identical tone shape with the enhancement, i.e., in LH-dipping L and HL-falling L, the L-tone responses were significantly reduced and the bias towards the non-L tones was still present even at the terminal step point of the continuum. Under the functional assumption that the best enhancement form should best separate L from the rival tones, the dipping L and the falling L showed identical discriminative potential. This in part explained the coexistence of both enhancement forms of L in the same speech community.

Syllable lengthening contributes to the recognizability of L more generally. An increased L-tone response was found in the long durational condition. The reason lies in the fact that proportional lengthening yielded a more detectable f0 minimum. For the HL-L contrast, however, lengthening failed to enhance the perception of L, because the phonological L of the HL syllable melody was equally enhanced by lengthening.

Finally, the male voice facilitated the perception of L when the rival tones end in high pitch, because the f0 of the male voice is situated in the lower frequency range. However, this enhancement disappeared in the contrast of HL-L, because HL ends up with a low pitch comparable to that of L.

6.3 Incomplete neutralization of sandhi tones

The main production experiment reported in Chapter 5 concerned the phonological neutralization of an underlying contrast due to the insertion of a tone. Specifically, an L is added to the left-hand H of an H.H sequence and an H to the left-hand L in an L.L sequence, in both cases creating representations that also exist underlyingly, HL.H and LH.L. To investigate the completeness of the neutralization, for each underlying contrast, two sets of word pairs were designed that were minimally distinguished by the tone of the first syllable. Members of the word pair were printed separately on two reading lists, within which the test words were mixed with sufficient amount of fillers. Ten native speakers completed a two-session production experiment, by reading the word lists. Three acoustic parameters, rhyme duration, f0-height, f0-slope were statistically compared. The results showed that the neutralization for both sandhi groups is acoustically incomplete. There was no durational difference between the sandhi tone and non-sandhi tone for both sandhi groups. The incomplete neutralization was robustly cued by the f0 features. Specifically, the sandhi LH and HL were flatter than their non-sandhi counterparts and additionally the sandhi LH was lower than the base LH. There was no correlation in the effect sizes between the f0-flatterning effect in the two sandhi and between the f0-flatterning effect and the f0-lowering effect within L-tone sandhi.

We also compared the acoustic realizations of neutral tones that come from multiple underlying sources. One near-minimal word pair was prepared to examine the neutralization of an underlying neutral tone and a derived neutral tone, both occurring after a LH-tone. The results showed that the underlying neutral tone was realized 2 Hz higher and 18.7 ms shorter than the derived neutral tone, representing incomplete neutralization. However, the small phonetic difference is not attributable to the phonetic reflex of the specific underlying forms, because the underlying neutral-tone syllable does not have an underlying phonological origin. Another near-minimal word pair was created to investigate the neutralization between two derived neutral tones, one from an underlying LH and the other from an underlying HL, both occurring after a lexical L-tone. Results showed that neither the first syllable nor the second showed a difference in the f0 features, representing complete neutralization.

We argue that incomplete neutralization need not reflect a difference in phonological representation, but instead can be understood as a reduction in the confidence with which the derived forms are pronounced, in contrast to the confident pronunciation of the base forms. The main cause of this inhibition lies in the presence of a phonological rule creating the output. We speculate that this inhibition theory of incomplete neutralization has a wider explanatory coverage than explanations based on traces of underlying forms.

6.4 Implications and future research

The findings of this dissertation highlight the importance of detailed phonological and phonetic analyses of Chinese languages. The analysis of vowels has gone some way towards enhancing our understanding of Mandarin phonology. We have provided both phonetic and phonological evidence showing the diphthongal nature of two vowels, /eɛ/ and /ɤʌ/, which have been traditionally treated as monophthongs. In particular, the phonological argument comes from the fact that, unlike monophthongs, they cannot be closed by a nasal, apparently because diphthongs are long, occupying both timing slots in the (bimoraic) rhyme (Duanmu 2007), preventing the addition of a nasal coda. The classification of vowels into monophthongs and diphthongs based on the phonological distribution of the coda nasal might be generalizable to other Mandarin dialects. A third argument in support of the diphthongal nature of /eɛ/ and /ɤʌ/ may come from emphatic speech. That is, the diphthongal target of the vowel under prominence would be expected to be better approximated. Thus, future research is needed to investigate the effect of hyperarticulation on the phonetic implementation of the vowels both within and across Mandarin dialects.

The findings on the production and perception experiments of the citation tones (Chapters 3 and 4) add to a growing body of literature on the inter-individual within-category variation and contribute to the description and theory of prosodic variability. We have shown that the inter-speaker variation of Kaifeng L is reflected in the different transcriptions that have been given in the literature. It suggests that the small sample sizes have been another intrinsic limitation for impressionistic descriptions. The disagreement over the transcriptions of tones that might lead to different phonological generalizations among fieldworkers may have resulted from inherent inter-subject within-category variation. Without a rigorous acoustic analysis that relies on a sufficient number of speakers, the inter-speaker variation would not have been revealed. However, the inter- and intra-speaker variation of tones in this dissertation has been limited to only one dimension, f0 slope. It would be interesting to assess the inter- and intra-speaker variation of tones in other dimensions, such as f0 height, i.e., the vertical dispersion of the f0 contour both within and across speakers, as in the acoustic description of Thai tones in Gandour et al. (1991).

The motivation of the phonological interpretation of Kaifeng Mandarin T4 as L has led to an increased understanding of the reasons for phonetic forms to deviate from the phonological representation. Often, phonological generalizations of tonal processes like tone sandhi have been based on the faithful translations of phonetic forms into elements like Chao digits (e.g., 312) or phonological tones (e.g., HLH) (Bao 1999; Chen 2000; Yip 2002). This approach, however, is sensitive to the choice between various fieldwork descriptions and would not be directly applicable if more than one form within the same category were to be found. The first case may easily lead to drastically different phonological generalizations, as shown in the disyllabic tones in Tianjin Mandarin (see § 1.4), while in the latter case, as shown in the Kaifeng Mandarin L-tone, neither HLH (312) nor ML (31) is the correct phonological representation of Kaifeng T4. Instead, we have shown that the initial f0 fall and the final f0 rise in Kaifeng Mandarin L do not have a phonological origin, casting doubt on the adoption of substance-based representations of syllabic tone melodies in the phonological analysis of tonal alternations. A reasonable approach to examine the phonological relevance of elements like the initial fall and the final rise around the L-target could be to conduct detailed phonetic experiments to examine the variation of f0 contours conditioned by various factors, linguistic or non-linguistic, such as prosodic context, focus, speech tempo, syllable structure (Xu 1998) as well as the variation of f0 contours in different speakers. On the basis of such data, a motivated phonological analysis could then be presented.

The acoustic description on disyllabic tones contributes to the prosodic typology of Chinese languages. We have provided the acoustic data and the phonological generalization of sandhi tones and the acoustic description of neutral tones in Kaifeng Mandarin. Phonologically, Kaifeng Mandarin tone sandhi presents a case of high-ranking markedness ban on adjacent level word melodies, quite different from Mandarin dialects like Standard Mandarin and Tianjin Mandarin (Yip 2002: 179–181).[31] Also, neutral-tone syllables in Kaifeng Mandarin exhibit drastic deviations from other Mandarin dialects. The rhyme duration of the neutral-tone syllables is longer than the preceding full-tone syllables and comparable to the full-tone syllables occurring in the same prosodic position. However, the illustration is only based on a limited amount of data. Future studies need to recruit more speakers, so that a more convincing conclusion can be reached. Furthermore, the realizations of multiple neutral-tone syllables in Beijing Mandarin (Chen & Xu 2006) and Tianjin Mandarin (Li & Chen 2011) suggest that neutral tone syllables in different Mandarin dialects may share 'a common underlying mechanism of the weak implementation of a mid-low pitch target' (Chen 2015). Thus, future multi-speaker studies are also needed to explore how consecutive neutral tone syllables are acoustically realized.

Finally, the tone sandhi data provide additional cases of incomplete neutralization. To the best our knowledge, our carefully-designed production experiment on sandhi tones presented in this study is the first that presents the small but consistent acoustic differences between supposedly neutralized forms in a Mandarin dialect other than Standard Mandarin. In addition to the productive right-dominant OCP-repair case, we have shown another case of incomplete neutralization in a less productive tonal neutralization process, i.e., neutral tone, which typologically resembles the left-dominant sandhi patterns found in Northern Wu in that the non-initial tone undergoes deletion and receives a tone that is dependent on the initial triggering tone. It remains to be seen what a productive left-dominant tone sandhi pattern in languages like Northern Wu contributes to our understanding of incomplete neutralization. Lastly, perception research would be required to see if the small phonetic differences in Kaifeng Mandarin sandhi tones can play any role in speech perception.

[31] In Standard Mandarin, the OCP is only effective in L.L, while in Tianjin Mandarin, there is a ban in L.L as well as in HL.HL and LH.LH.

6.5 Concluding remarks

This dissertation presented a phonetic study on the four syllabic tone melodies, phonologically represented as LH, HL, H and L, in the Zhongyuan Mandarin dialect of Kaifeng, spoken in the east central Henan Province of China. We have shown that L was realized in three speaker-dependent f0 contours, two of which, the dipping L and the falling L, performed equally well in the discriminative power in lexical tone contrasts between L and the rival tones. The inter-speaker variation of L was interpreted as phonetic enhancement of an L-tone, due to its intrinsic low salience compared with an H-tone, while the equal discriminative function in the dipping L and the falling L in part explained their co-existence in the same speech community. Finally, two cases of incomplete neutralization in the tonal phonology were presented and it was proposed that incomplete neutralization can be accounted for by the speaker's inhibition in the pronunciation of derived forms.

References

Bao, Z. (1999). The structure of tone. Oxford: Oxford University Press.

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

Chen, Y. (2015). Neutral tone. In R. Sybesma (Ed.), *Encyclopedia of Chinese language and linguistics*, http://dx.doi.org/10.1163/2210-7363_ecll_COM_000162 [First pulished online: 2015, consulted online on 09 May 2018]. Leiden: Brill.

Chen, Y. & Xu, Y. (2006). Production of weak elements in speech – evidence from f0 patterns of neutral tone in Standard Chinese. *Phonetica*, 63, 47–75.

Duanmu, S. (2007). *The phonology of Standard Chinese* (2nd edition). Oxford: Oxford University Press.

Gandour, J., Ponglorpisit, S. & Dechongkit, S. (1991). Inter- and intraspeaker variability in fundamental frequency of Thai tones. *Speech Communication*, 10, 355–372.

Li, Q. & Chen, Y. (2011). Neutral tone realization in Tianjin Mandarin. In W.-S. Lee, E. Zee (Eds.), *17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 1214–1217.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, 55, 179–337.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

APPENDIX

A. Fixed effect parameter estimates: first tone H.H vs HL.H (baseline H.H).

	Estimate	SE	*t*	*p*
Intercept	19.390	1.733	11.190	0.000
Linear	–5.838	0.584	–10.004	0.000
Quadratic	0.170	0.205	0.828	0.408
sandhi: Intercept	0.070	0.054	1.293	0.196
sandhi: Linear	–0.771	0.160	–4.824	0.000
sandhi: Quadratic	0.075	0.068	1.103	0.270

B. Fixed effect parameter estimates: second tone H.H vs HL.H (baseline H.H).

	Estimate	SE	*t*	*p*
Intercept	18.535	1.571	11.801	0.000
Linear	0.833	0.239	3.492	0.000
Quadratic	–0.295	0.186	–1.587	0.112
sandhi: Intercept	–0.018	0.066	0.282	0.778
sandhi: Linear	0.194	0.075	2.599	0.009
sandhi: Quadratic	–0.058	0.066	–0.882	0.378

C. Fixed effect parameter estimates: first tone L.L vs LH.L (baseline L.L).

	Estimate	SE	*t*	*p*
Intercept	18.577	1.487	12.496	0.000
Linear	4.793	0.806	5.944	0.000
Quadratic	0.972	0.475	2.048	0.041
sandhi: Intercept	0.246	0.080	3.086	0.002
sandhi: Linear	0.705	0.231	3.058	0.002
sandhi: Quadratic	–0.228	0.144	–1.585	0.113

D. Fixed effect parameter estimates: second tone L.L vs LH.L (baseline L.L).

	Estimate	SE	*t*	*p*
Intercept	13.200	1.944	6.789	0.000
Linear	–10.939	0.956	–11.447	0.000
Quadratic	1.983	0.392	5.057	0.000
sandhi: Intercept	0.269	0.061	4.424	0.000
sandhi: Linear	–0.225	0.164	–1.374	0.170
sandhi: Quadratic	–0.125	0.100	–1.260	0.208

E. Speaker-by-sandhi random effect parameter estimates for H-tone sandhi and L-tone sandhi.

Speaker	Tone	Intercept	Linear	Tone	Intercept	Linear
LHW	H.H	–0.027	–0.104	L.L	–0.078	0.300
LHW	HL.H	0.028	0.155	LH.L	0.019	–0.429
ZWP	H.H	0.003	0.224	L.L	–0.007	0.268
ZWP	HL.H	–0.009	–0.193	LH.L	0.019	–0.300
WJ	H.H	0.033	–0.098	L.L	0.184	–0.483
WJ	HL.H	–0.036	0.109	LH.L	–0.248	0.416
LYH	H.H	–0.031	–0.211	L.L	0.044	–0.037
LYH	HL.H	0.030	0.211	LH.L	–0.058	–0.081
LGF	H.H	0.194	0.019	L.L	0.019	0.270
LGF	HL.H	–0.205	0.020	LH.L	0.020	–0.160
DLL	H.H	–0.001	–0.232	L.L	0.008	–0.098
DLL	HL.H	0.008	0.211	LH.L	–0.041	–0.056
ZXM	H.H	–0.100	0.171	L.L	–0.267	–0.701
ZXM	HL.H	0.106	–0.311	LH.L	0.264	0.824
YWN	H.H	0.052	0.088	L.L	0.027	0.071
YWN	HL.H	–0.049	–0.206	LH.L	–0.028	0.137
CJY	H.H	–0.040	–0.262	L.L	0.118	0.059
CJY	HL.H	0.035	0.341	LH.L	–0.114	–0.207
WCX	H.H	–0.082	0.405	L.L	–0.049	0.352
WCX	HL.H	0.092	–0.336	LH.L	0.165	–0.144

F. Fixed effect parameter estimates: first tone LH.Ø vs LH.$Ø^{LH}$ (baseline LH.$Ø^{LH}$).

	Estimate	SE	*t*	*p*
Intercept	15.517	1.605	9.668	0.000
Linear	0.360	0.419	0.861	0.389
Quadratic	1.037	0.184	5.641	1.689
LH.N(0): Intercept	–0.174	0.143	–1.215	0.224
LH.N(0): Linear	0.196	0.222	0.883	0.377
LH.N(0): Quadratic	0.524	0.157	3.337	0.000

G. Fixed effect parameter estimates: second tone LH.Ø vs LH.$Ø^{LH}$ (baseline LH.$Ø^{LH}$).

	Estimate	SE	*t*	*p*
Intercept	18.076	1.603	11.278	0.000
Linear	1.431	0.342	4.178	0.000
Quadratic	–0.969	0.212	–4.576	0.000
LH.N(0): Intercept	0.338	0.002	3.028	0.002
LH.N(0): Linear	0.123	0.285	0.432	0.667
LH.N(0): Quadratic	–0.054	0.159	–0.340	0.734

H. Fixed effect parameter estimates: first tone L.$Ø^{LH}$ vs L.$Ø^{HL}$ (baseline L.$Ø^{HL}$).

	Estimate	SE	*t*	*p*
Intercept	15.130	1.818	8.320	0.000
Linear	–7.205	1.021	–7.058	0.000
Quadratic	1.392	0.260	5.360	0.000
L.N(LH): Intercept	0.022	0.230	0.094	0.925
L.N(LH): Linear	0.077	0.918	0.084	0.933
L.N(LH): Quadratic	–0.255	0.302	–0.844	0.399

I. Fixed effect parameter estimates: second tone L.$Ø^{LH}$ vs L.$Ø^{HL}$ (baseline L.$Ø^{HL}$).

	Estimate	SE	*t*	*p*
Intercept	13.547	1.622	8.351	0.000
Linear	0.518	0.637	0.813	0.416
Quadratic	0.581	0.231	2.513	0.012
L.N(HL): Intercept	0.045	0.265	0.171	0.865
L.N(HL): Linear	0.068	0.395	0.172	0.863
L.N(HL): Quadratic	0.206	0.212	0.973	0.331

CURRICULUM VITAE

Lei Wang was born in 1987, in Kaifeng, Henan, People's Republic of China. He obtained his Bachelor's degree in English from Zhengzhou University, Zhengzhou, Henan, in 2011. In 2014, he finished a Master's degree program in linguistics at Tongji University, Shanghai. In the same year, he started a PhD degree program at Tongji University. In 2016–2018, he worked as a guest researcher and as a PhD student at Radboud Universiteit, Nijmegen.