How to pronounce a low tone: A lesson from Kaifeng Mandarin

Lei Wang

School of Foreign Languages, Tongji University & Centre for Language Studies, Radboud Universiteit Nijmegen
wl1410512@gmail.com

Carlos Gussenhoven

Centre for Language Studies, Radboud Universiteit Nijmegen
c.gussenhoven@let.ru.nl

Jie Liang

School of Foreign Languages, Tongji University
liangjie56@163.net

A multi-speaker acoustic study on citation tones in Kaifeng Mandarin, referred to as LH, HL, H and L, shows that L is realized as three different subtypes by different speakers, i.e. dipping, falling and falling with lengthening, while generally being longer than the other three tones and frequently spoken with creaky voice in part of the vowel. This inter-speaker variation is reflected in the different transcriptions of Kaifeng L that have been given in the literature. We argue that a L-tone is intrinsically less salient than a 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-tones, and thus more likely to be enhanced.

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 four 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

Table 1 Tone scale notations of Kaifeng Mandarin LH, HL, H and L, with their traditional tonal category labels in brackets.

Tonal category	LH (*yinping*)	HL (*yangping*)	H (*shang*)	L (*qu*)
Zhang et al. (1993)	24	41	55	31
Liu (1997)	24	41	55	312

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 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 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, 31/312] (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).

Table 2 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 Mandarin	H	LH	L	HL
	Kaifeng Mandarin	LH	HL	H	L

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*. The sinological literature sub-classifies each tonal category into *yin* and *yang* tones, depending on the voicing status of the onset consonant (Chen 2000) and assigns numerical labels to them. Thus, the four tonal categories in Standard Chinese, *yinping*, *yangping*, *shang* and *qu*, are labelled Tone 1 to Tone 4, respectively. Often, as in the case of Kaifeng Mandarin, the phonetic shapes of the tones used in these etymological classes differ from those in Standard Chinese. For instance, Standard Chinese Tone 4 is a fall from high to low, while Tone 4 in Kaifeng Mandarin is phonologically equivalent to Standard Chinese Tone 3 (see Table 2). While the phonological analysis of [24, 41, 55] as LH, HL and H, respectively, is unproblematic, it is less evident in the case of the proposed L for Kaifeng Tone 4. The analysis is supported by the pitch curve obtained by rudimentary instrumentation by Chao (1922), which resembles the pitch curve of the low tone (Tone 3) in Beijing Mandarin. In addition, 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.

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.

2 Method

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 tonal minimal quadruplets was prepared, shown in Table 3. 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).

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 4 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 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
tʂɻ	知	直	纸	智	tɕy	菊	局	举	锯
tʂʰɻ	尺	迟	齿	翅	tɕʰy	区	渠	娶	去
ʂɻ	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					

Table 4 Ten participants with information of gender, age (years), 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

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 and randomly presented 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 a 10 cm distance from the speaker's mouth on 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.

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
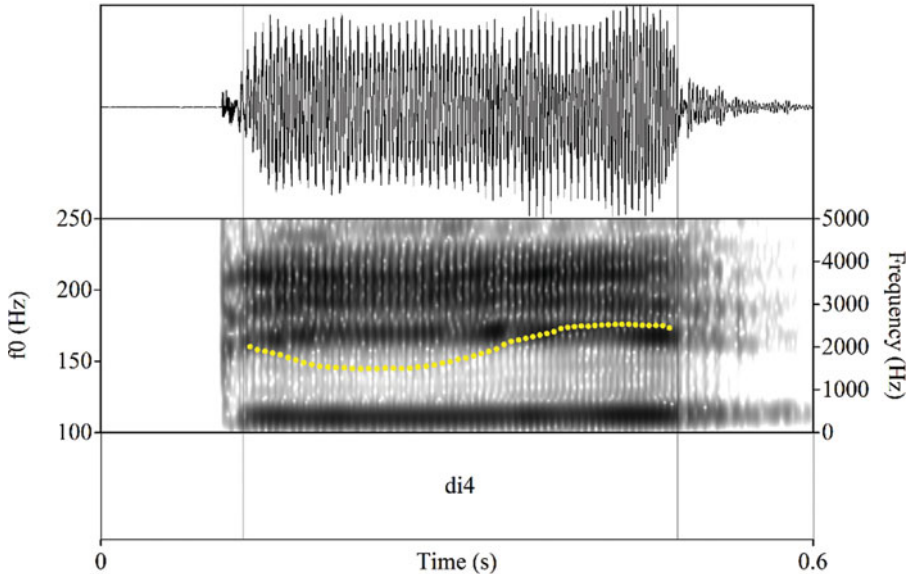
Figure 1 (Colour online) Phonetic annotation of a recorded word in Prorec. The rhyme of each syllable is manually labeled in a Praat textgrid.

manually labeled (see Figure 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 (Xu 1999) and output files with information about time-normalized f0, duration, maximum f0, minimum f0 and f0 velocity were automatically generated. One hundred and forty-nine 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 Section 3.5. From the 2,636 remaining speech files, 21 equidistant f0 measurements (in Hz) were taken from the rhymes.

2.5 f0 normalization

In order to eliminate speaker f0 range differences (10.39, sd = 1.22 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 Results

3.1 Rhyme duration

In terms of rhyme duration, Tone L is the longest (249.19 ms, sd = 52.56), followed by LH (229.25 ms, sd = 55.99), while HL and H are the shortest (209.02 ms, sd = 37.31 and 203.52 ms, sd = 54.32, respectively). A one-way between-subjects ANOVA yielded a
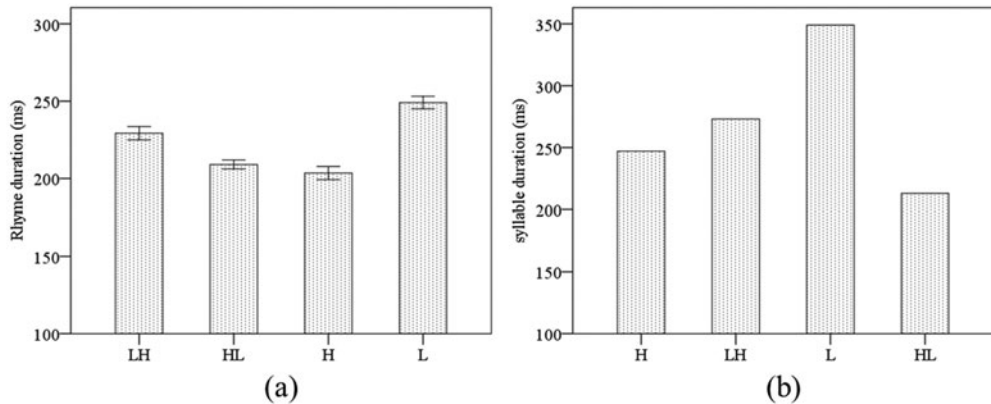
Figure 2 Durations of four tone categories ordered according to their etymological class in Kaifeng Mandarin (panel (a)) and Standard Chinese (panel (b), data from Xu 1997). The data in panel (a) are from ten speakers ($N = 2636$; LH = 663, HL = 652, H = 656, L = 665). Horizontal bars stand for ± 2 standard errors.

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 that between HL and H.

Figure 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, Tone 3, 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 Tone 4, the etymological equivalent of Kaifeng Mandarin L, is in fact the shortest of the four tones in that dialect, and is phonologically HL.

3.2 f0 contours

The mean f0 contours of the four tones are shown in panels (a)–(d) in Figure 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 (0–100). Figure 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 33.18% 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 an 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. Section 3.3 presents the way in which the variation in L deviates from that in the other tone categories.

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 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,
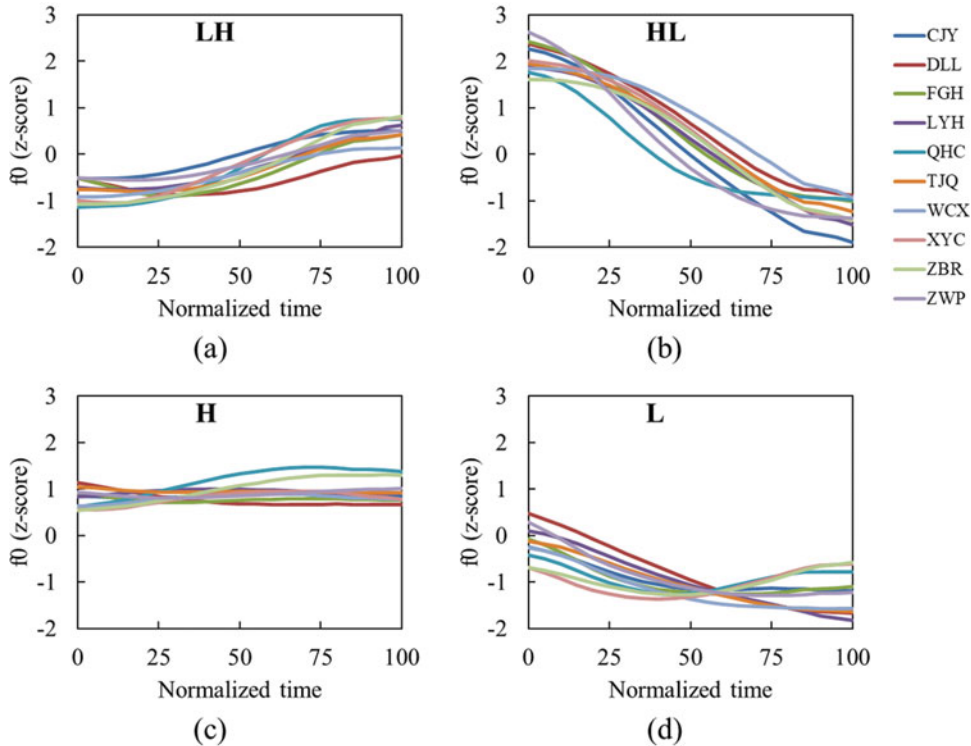
Figure 3 (Colour online) Tonal contour from ten speakers (averaged across tokens). The four tones are grouped in separate panels ((a)–(d)).

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 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 Section 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
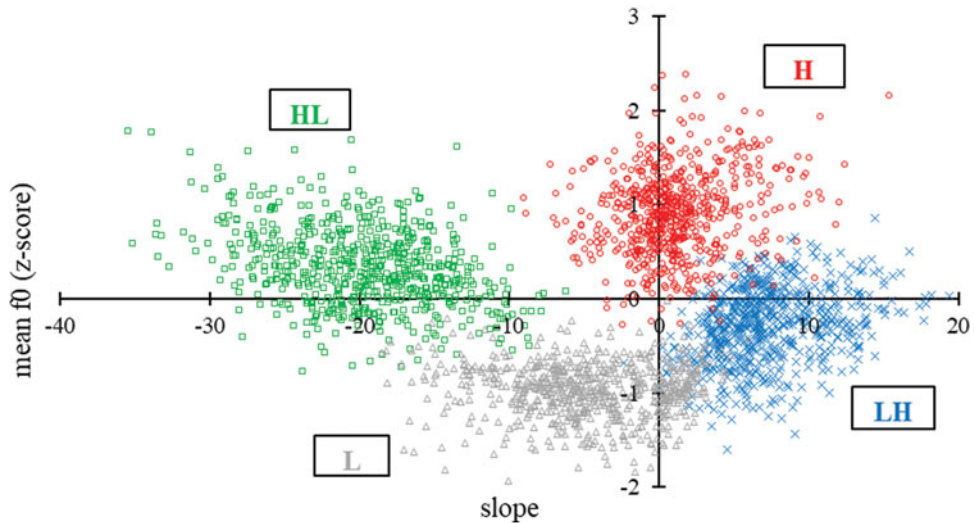
Figure 4 (Colour online) Scatterplot for the citation tones in Kaifeng Mandarin with the horizontal axis showing the slope and the vertical axis showing the mean f0. Ten speakers; $N = 2636$, LH = 663, HL = 652, H = 656, L = 665.

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 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). 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.

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
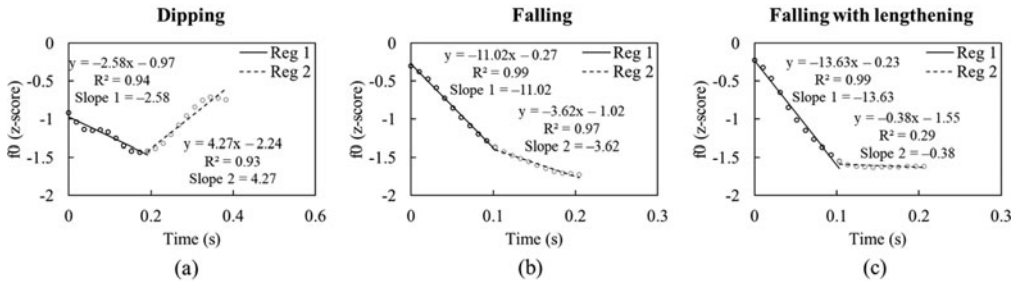
Figure 5 Examples of the dipping (a), falling (b) and falling + lengthening (c) subtypes of L, with two linear regression lines superimposed on the f0 sampling points. Slope 1 (solid line) and Slope 2 (dashed line) are the coefficients of the first and last 11 sampling points with their corresponding real time points as the independent variable.

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 5 exemplify how these data characterize the dipping, falling and falling + lengthening subtypes on the basis of selected tokens.

Figure 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), on the basis of the tone shapes in Figure 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 considerably 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.

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 7). 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: $N = 266$,
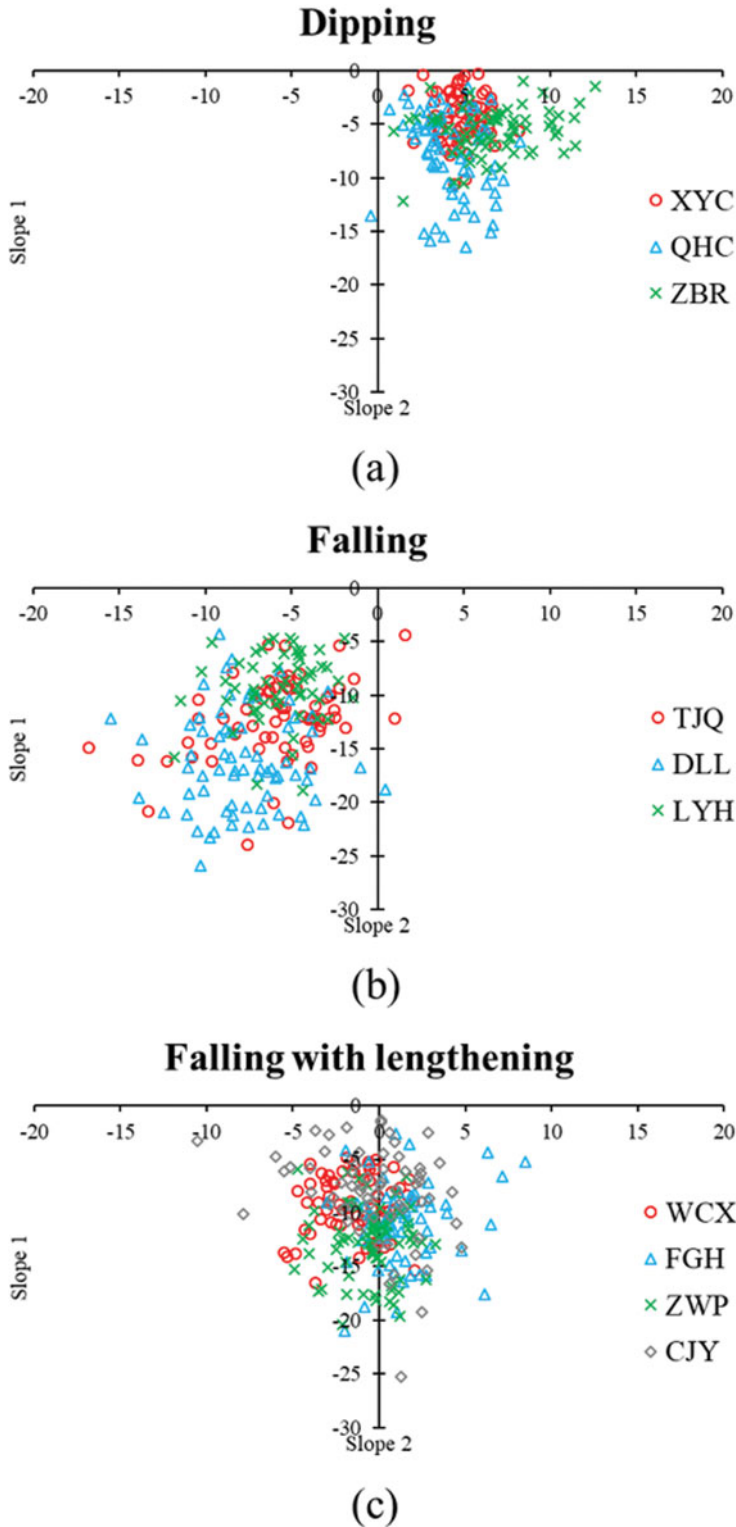
Figure 6 (Colour online) Intra-speaker variation in the realization of L. Each realization is represented as a point in the Slope 1/Slope 2 plane. Ten speakers' data are grouped separately in panels (a)–(c), on the basis of the subtypes the speakers were assigned to. Panel (a): XYC (N = 67), QHC (N = 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).
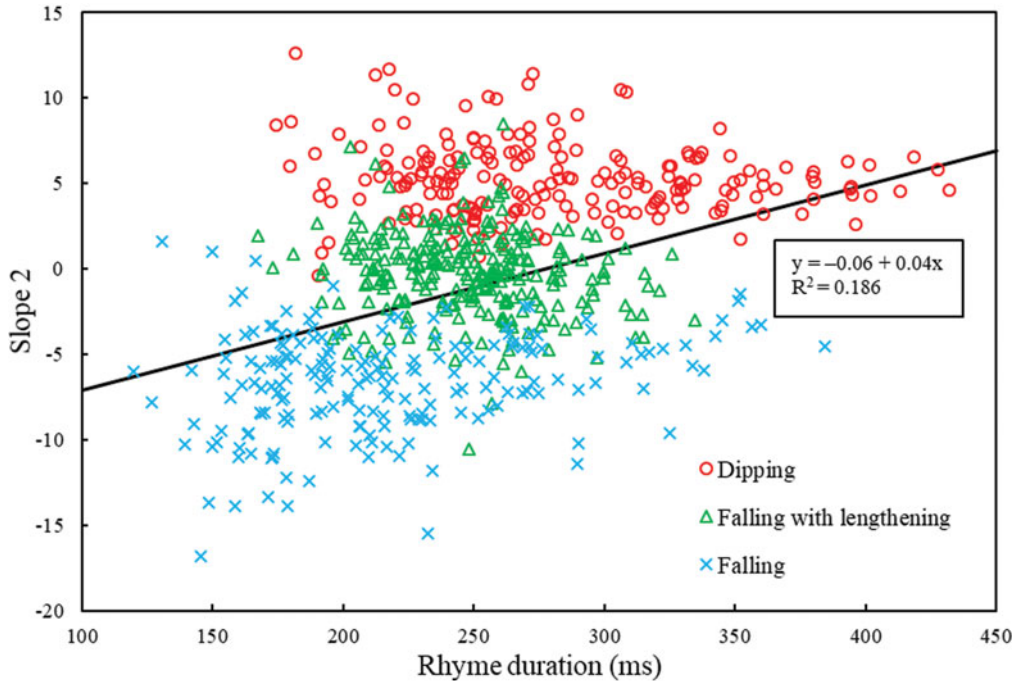
Figure 7 (Colour online) Slope 2 as a function of rhyme duration. Subtypes indicated by the colour of the plot symbols (N = 665).

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.

The histogram in Figure 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.

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 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 where 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
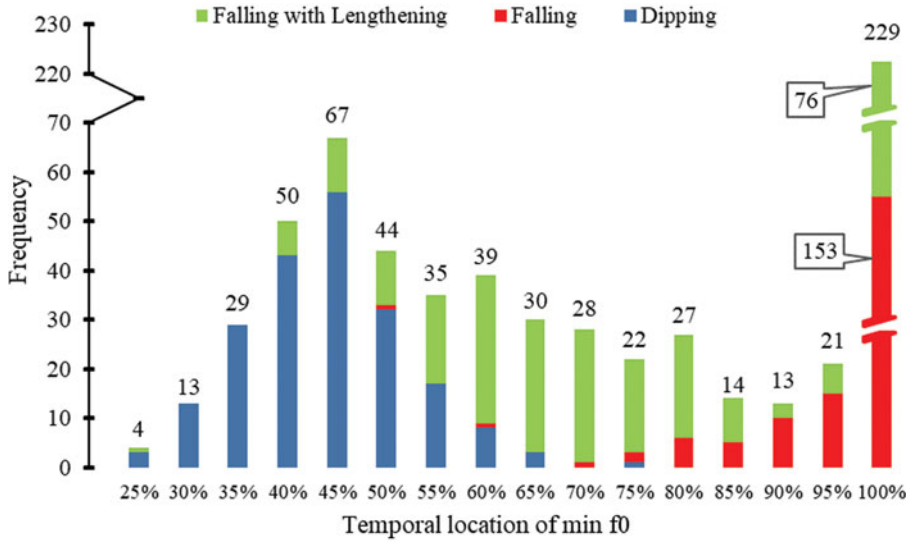
Figure 8 (Colour online) Frequency distribution of the f0 minima of the three subtypes of L over 16 points equally spaced over the last 75% of the rhyme, represented as percentages of the normalized rhyme duration on the x-axis. Red = falling subtype; green = lengthened falling subtype; blue = dipping subtype (N = 665).
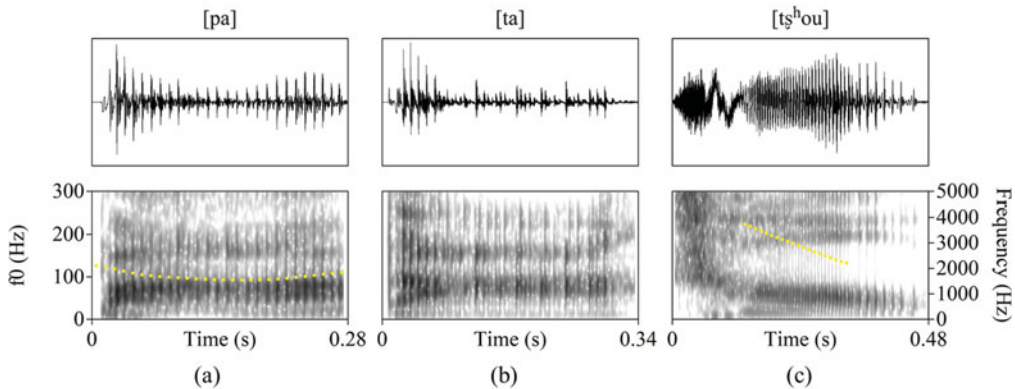

Figure 9 (Colour online) Creaky voice in realizations of L. Panel (a): the waveform and spectrogram for syllable [pa] uttered by speaker ZBR, panel (b): the waveform and spectrogram for syllable [ta] produced by speaker FGH, panel (c): the waveform and spectrogram of syllable [tʂʰou] generated by speaker LYH are provided.

pulses are brief and regular, while in the last 30% of the rhyme they become longer, such that f0 can no longer be measured.

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 greater disturbance of the periodicity (e.g. panels (b) and (c) in Figure 9), of which 13 are L and two are HL.

4 Discussion

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 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). Under the incorrect assumption of a 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 Section 4.2 we attempt to explain the phonetic variation as different ways of enhancing the low target.

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 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 midpoint 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 Tone 3 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 Tone 3, even though the creak in the Tone 3 exemplar was not very obvious. It is possible that L-tones may naturally induce non-modal phonation, as pointed out by a reviewer, since LH and HL also have creak. A similar distribution of creak also holds for the tones in Standard Chinese (Belotel-Grenié & Grenié 2004). 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). On the basis of 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 Tone 4 ([21/11] in Chao digits) than other tones (24.2% of Tone 4 vs. 4.7% of the corpus). They also reported that Cantonese creaky Tone 4 is identified with a much higher accuracy than non-creaky Tone 4, while listeners also utilize creak in the perceptual separation of Tone 4 and Tone 6 ([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; Evans, Yeh & Kulkarni 2018: 525; see also 'post-low bouncing' in Prom-on, Liu & Xu 2012). 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, b), 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 Tone 2 and Tone 3, 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, Tone 3 is characterized by a greater f0 initial fall and a later f0 turning point compared with Tone 2 (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 10, in line with his realizations in citation forms (see Figure 3), while for WC (panels (d)–(f)), additional curvature is created in the focal L, achieved by raising the flanks on both
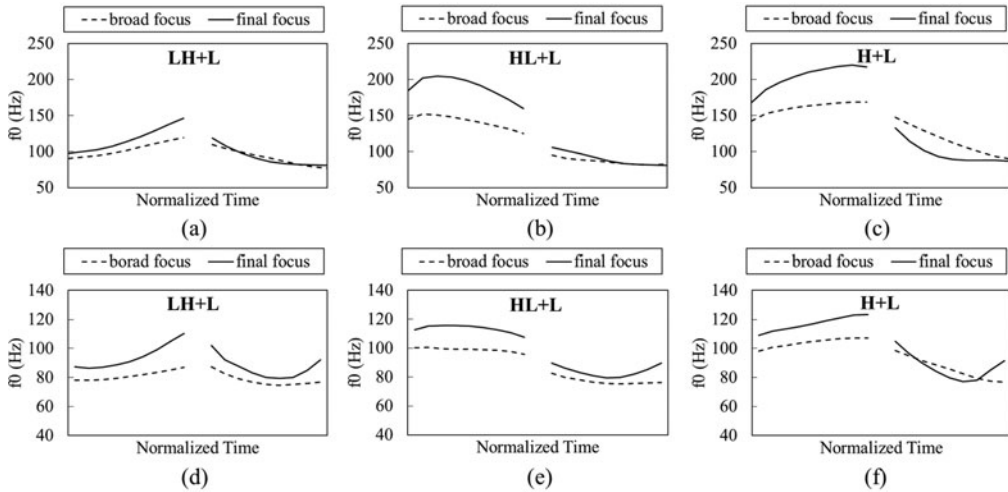
Figure 10 Phrase-final L-tones produced in two focus conditions by ZWP (panels (a)–(c)) and WC (panels (d)–(f)), with data from Wang (2018). Pitch curves were averaged by ten repetitions per speaker. L-tones preceded by LH, HL and H are presented in separate columns. The vertical axis shows f0 in Hz. The duration of tones is normalized.
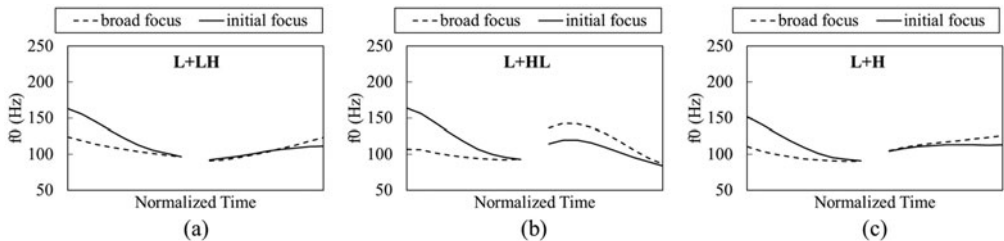

Figure 11 Phrase-initial L-tones produced in two focus conditions by ZWP (data from Wang 2018). Pitch curves were obtained by pooling over ten repetitions. L-tones followed by LH, HL and H are displayed in panels (a)–(c). The vertical axis shows f0 scales in Hz. The duration of tones is normalized.

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 Tone 3 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 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.

4.3 Phonological arguments for the analysis as L

The phonetic and perceptual data discussed in Section 4.2 argue for an analysis of the low tone as L. There are four phonological arguments that support this analysis. First, it fits in with
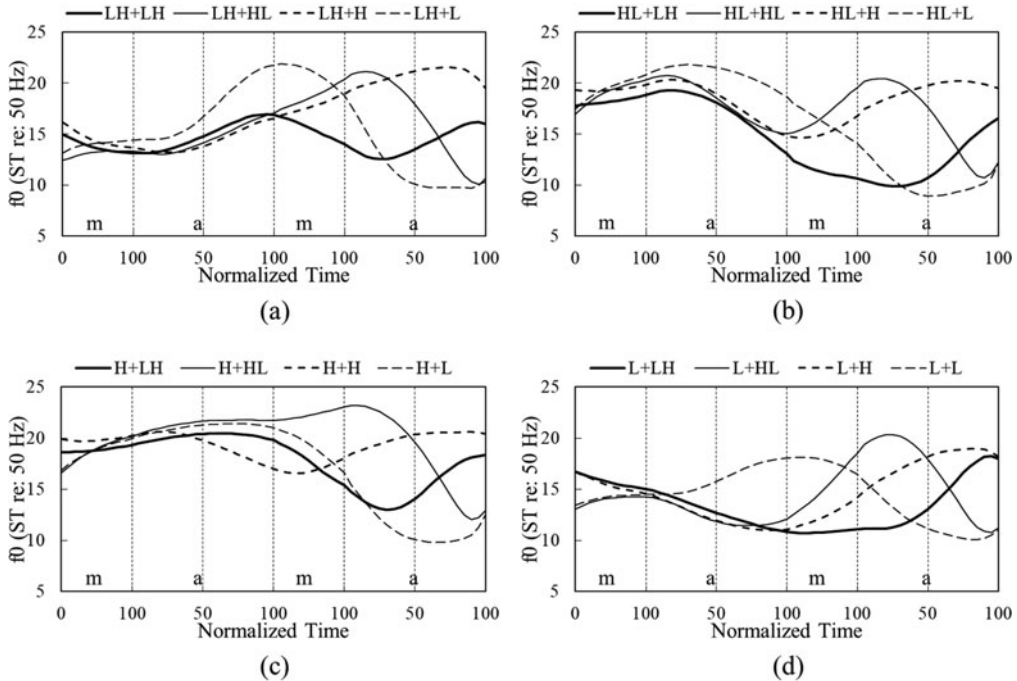
Figure 12 f0 contours of /mama/ sequence in Kaifeng Mandarin (data from Wang & Liang 2015). Each f0 curve (ST re: 50 Hz) is pooled over 18 stimuli from three male speakers (six repetitions for each speaker). All speakers realize these tone patterns consistently. From panel (a) to panel (d), the first 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.

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.[1] 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 neighbouring tones (see 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 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

---

[1] 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), whose article provided the model for the title of this paper.

argument, Kaifeng Mandarin L is replaced with LH before another L, as shown in the thin dashed line in panel (d) of Figure 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 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 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 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 & 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'.

Just as in the case of the greater salience of the enhancement features under emphasis (Section 4.2), the raising effect is even greater when the triggering L-tone is focused, as exemplified in panels (a)–(f) in Figure 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, correctively 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, Wang & Liberman (2016). This suggests that pre-L raising is motivated by enhancement of L, as opposed to an enhancement of the preceding H-tone.

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 have 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 a 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.

Acknowledgments

References

Arvaniti, Amalia. 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.

Belotel-Grenié, Agnès & Michel Grenié. 2004. The creaky voice phonation and the organization of Chinese discourse. *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages–2004*, Bejing, 5–8.

Blicher, Deborah L., Randy L. Diehl & Leslie B. Cohen. 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, Paul & David Weenink. 1992–2015. Praat: Doing phonetics by computer [computer program]. Version 5.4.08, retrieved from http://www.praat.org/ (24 March 2015).

Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.

Cao, Wen. 2010a. Shengdiao ganzhi duibi yanjiu: Guanyu pingdiao de baogao [A compariative study on tone perception: Report on level pitch contours]. *Shijie hanyu jiaoxue* [Chinese teaching in the world] 24(2), 255–262.

Cao, Wen. 2010b. Hanyu pingdiao de shengdiao ganzhi yanjiu [Level tone perception in Mandarin]. *Zhongguo yuwen* [Chinese Linguistics] 339, 536–543.

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

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

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

Chen, Yiya. 2012. Tonal variation. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (ed.), *The Oxford handbook of laboratory phonology*, 103–114. Oxford: Oxford University Press.

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

Ding, Hongwei, Oliver Jokisch & Rüdiger Hoffmann. 2010. Perception and production of Mandarin tones by German speakers. *Speech Prosody* 2010, 100153:1–4. http://speechprosody2010.illinois.edu/papers/100153.pdf (27 October 2016).

Duanmu, San. 2007. *The phonology of Standard Chinese*, 2nd edn. Oxford: Oxford University Press.

Evans, Jonathan, Wen-Chi Yeh & Rukmini Kulkarni. 2018. Acoustics of tone in Indian Punjabi. *Transactions of the Philological Society* 116(3), 509–528.

Gandour, Jack, Suvit Ponglorpisit, Sumalee Dechongkit, Fuangfa Khunadorn, Prasert Boongird & Siripong Potisuk. 1993. Anticipatory tonal coarticulation in Thai noun compounds after unilateral brain damage. *Brain and Language* 45, 1–20.

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

Goldsmith, John A. 1976. *Autosegmental Phonology*. Ph.D. dissertation, MIT. [Distributed by IULC, 1976, Published by Garland Press, New York, 1979.]

Gussenhoven, Carlos. 2007. Intonation. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 253–280. London: Cambridge University Press.

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

Gussenhoven, Carlos & Wencui Zhou. 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, Judith. 2017. *Regional variation in the realization of intonation contours in the Netherlands*. Ph.D. dissertation, Radboud University. [Utrecht: LOT]

Huckvale, Mark. 2014. Prorec: A program for field workers [computer program]. Version 1.45, http://www.phon.ucl.ac.uk/resource/ (17 October 2014).

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

Jiang, Ping. 1999. Hanyu zhufangyan shengdiao fenbu de youxuan fenxi [Tonal distribution in Chinese dialects: An OT account]. In En Quan Chen (ed.), *Shuangyu shuangfangyan yu xiandai zhongguo* [Bilingualism, diglossia and modern China], 134–152. Beijing: Beijing University of Languages and Culture.

Jongman, Allard, Yue Wang, Corinne B. Moore & Joan A. Sereno. 2006. Perception and production of Mandarin Chinese tones. In Ping Li, Li Hai Tan, Elizabeth Bates & Ovid J. L. Tzeng (eds.), *The handbook of East Asian psycholinguistics*, vol. 1: *Chinese*, 209–217. Cambridge: Cambridge University Press.

Ladd, D. Robert & Astrid Schepman. 2003. "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics* 31(1), 81–112.

Ladefoged, Peter. 2003. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Oxford: Blackwell.

Laniran, Yetunde O. & George N. Clements. 1995. A long-distance dependency in Yoruba tone realization. *Proceedings of 13th International Congress of Phonetic Sciences* (ICPhS XIII), Stockholm, Sweden, vol. 2, 734–737.

Leben, Wiliam R. 1976. The tones of English intonation. *Linguistic Analysis* 2, 67–107.

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

Lee, Yong-Cheol, Ting Wang & Mark Liberman. 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, Sang-Im. 2014. Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association* 44(3), 261–282.

Liang, Jie & Vincent J. van Heuven. 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.

Liu, Dongbing. 1997. Kaifeng fangyan jilue [A description of Kaifeng dialect]. *Fangyan* [Dialect] 4, 271–285.

Lu, Jilun. 2001. Gao, dipingdiao de buduichenxing he youxuanlun de zhiyuetiaojian [The asymmetry in H/L and OT constraints]. *Proceedings of the 5th National Conference on Modern Phonetics*, Beijing, China, 243–245.

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

Peng, Gang. 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, Jörg. 2007. Bitonal lexical pitch accents in the Limburgian dialect of Borgloon. In Tomas Riad & Carlos Gussenhoven (eds.), *Tones and tunes*, vol. 1: *Typological studies in word and sentence prosody*, 167–198. Berlin & New York: Mouton de Gruyter.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT.

Potisuk, Siripong, Jack Gandour & Mary P. Harper. 1997. Contextual variations in trisyllabic sequences of Thai tones. *Phonetica* 54, 22–42.

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

Prom-on, Santitham, Fang Liu & Yi Xu. 2012. Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling. *The Journal of the Acoustical Society of America* 132(1), 421–432.

Riad, Tomas. 1998. Towards a Scandinavian accent typology. In Wolfgang Kehrein & Richard Wiese (eds.), *Phonology and morphology of the Germanic languages*, 77–109. Tübingen: Niemeyer.

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

Shih, Chi-lin. 1988. Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory* 3, 83–109.

van de Weijer, Jeroen & Marjoleine Sloos. 2014. The four tones of Mandarin Chinese: Representation and acquisition. *Linguistics in the Netherlands* 31(1), 180–191.

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

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

Wang, Lei & Jie Liang. 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, Douglas H. & Yi Xu. 1992. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49, 25–47.

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

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics* 27(1), 55–105.

Xu, Yi. 2013. ProsodyPro: A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France, 7–10.

Xu, Yi & Xuejing Sun. 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, Ruoxiao. 2015. The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics* 43(1), 453–472.

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

Zhang, Qihuan, Tianfu Chen & Yi Cheng. 1993. *Henan fangyan yanjiu* [A study on Henan dialects]. Kaifeng: Henan University Press.