Tone and intonation processing: from ambiguous acoustic signal to linguistic representation

Liu, M.

Chapter 4

Tonal mapping of Xi'an Mandarin and Standard Chinese

Abstract

Tonal information can be a determinant of the phonological similarity or difference between some Chinese dialects and Standard Chinese, yet relatively little empirical research has been conducted on the tonal system of other language varieties in Chinese other than Standard Chinese. Among these dialects, Xi'an Mandarin is particular interesting for its seemingly present one-to-one mapping of tones with Standard Chinese tones. To gain empirical evidence for the mapping, the present study compared the tonal systems of Xi'an Mandarin and Standard Chinese in both tone production and perception. Tones with similar contours from Xi'an Mandarin and Standard Chinese were paired and both tone production and perception experiments were carried out on bi-dialectal speakers of Xi'an Mandarin and Standard Chinese. Acoustic results showed that the F0 difference ranged from no F0 difference (level contour tone pair) through F0 curvature difference (rising contour tone pair) to F0 height difference (falling contour tone pair) and F0 contour difference (low contour tone pair). Except for the falling contour tone pair, all the other tone pairs also exhibited differences in tone duration. The varying acoustic differences in different tone pairs, together with the phonological rule, resulted in varying degrees of tonal similarity in tone perception, but tones with similar contours between the two dialects were basically perceived to be the same. The two experiments together showed the systematic mappings of tones between Xi'an Mandarin and Standard Chinese.

Keywords: Standard Chinese, Xi'an Mandarin, tonal mapping, tone production, tone perception

4.1 Introduction

Chinese is a tonal language where tones are used to distinguish lexical meanings. However, the term "Chinese" refers to a large number of Sinitic language varieties. While numerous studies have been conducted on Standard Chinese (i.e., the official language of China; SC), relatively little attention has been paid to other dialects or language varieties of Chinese. Some of the dialects differ from SC in both segmental and tonal information, whereas others such as dialects within the Mandarin family overlap largely in segmental information with SC. In these latter dialects, tonal information can be important as it determines the phonological similarity or difference between the dialect and SC.

In China, most speakers of SC speak a local dialect (Li & Lee, 2008; Wiener & Ito, 2014). It is therefore of both practical and theoretical value to systematically investigate the tonal similarity or difference between different dialects and SC. Such investigations can be the prerequisite to developing dialect-oriented speech synthesis and speech recognition technology (Czap & Zhao, 2017), to guiding language pedagogy in teaching SC to dialectal speakers (Lam, 2005; Wong & Xiao, 2010), and to addressing issues such as whether the phonological information of one's two or more dialects are stored separately or integrally (Wu, 2015), or how cross-dialect phonological similarity/difference affects lexical access in the minds of bi-dialectal tonal language speakers.

Currently, relatively little empirical research has been conducted on the tonal system of other language varieties except for SC; even less research compared the tonal system of other language varieties with that of SC. As language varieties within the Mandarin family rely largely on tonal information to make distinctions from SC, the present study aimed to empirically compare the tonal systems of two closely related dialects in the Mandarin family, SC and Xi'an Mandarin (XM).

According to Chappell (2001), there are ten major dialect groups in Chinese (but see Yuan, 1989; Li & Thompson, 1981 which argue for a seven major dialect groups). The Mandarin family is the largest Chinese dialect group. It contains a group of Chinese varieties, which are typically spoken in the northern and southwestern China. The most influential language within the Mandarin family is SC. The other dialects within the Mandarin family share a

common logographic writing system with SC and bear high resemblance with SC as to lexical items and syntactic forms. Some dialects such as XM also exhibit large overlap of segmental features with SC. More interestingly, the tones of XM seem to have one-to-one correspondence with those of SC (Li, 2001; Zhang, 2009). This overall correspondence between the two tonal systems is quite unique and makes XM a very compelling case to study.

XM is a Mandarin dialect typically spoken in the urban areas of Xi'an, the capital of Shaanxi Province. It is the representative dialect of the Guanzhong dialect spoken in the Guanzhong area (Li & Stephen, 1987). XM directly originates from the official language in ancient China and has important historical value. As in SC, there are four tonal categories in XM, and they are referred to as T1, T2, T3 and T4. Here, the terms T1-T4 are adopted to suggest that words which share the same tonal categories across the two dialects are etymologically-related translation equivalents in most cases. Across XM and SC, different tones distinguish lexical meanings for syllables with the same segment. For example, the segment *ma* means *mother*, *hemp*, *horse* and *scold*, respectively, when it is combined with the four lexical tones (tonal category: T1, T2, T3, T4). On the 5-point scale notation system (Chao, 1930; 1968), the pitch value of the SC tones has been established as 55 (T1), 35 (T2), 214 (T3) and 51 (T4) respectively. However, there have been discrepancies among researchers regarding the specific pitch value of each XM tone (see Table 1 for a summary of the representative transcription of XM tones).

The majority of the existing studies on XM tones, including the first six studies listed in Table 1, have been based on impressionistic observation. Pitch values of XM tones in these studies could be susceptible to the subjective pitch sensitivity of the researchers. It is therefore not surprising that these studies vary in pitch value for each tonal category. The remaining studies, such as the last three in Table 1, have attempted to study the pitch value of XM tones with more objective acoustic methods. However, these studies either sampled from a very limited number of speakers (e.g., two in Ma (2005); one in Ren (2012)) or lacked control of the lexical properties of the stimuli used (e.g., Zhang & Shi, 2009). It is not known to what degree these results can represent the typical tonal patterns of XM. The present study thus decided to empirically examine

the acoustic properties of XM tones with a larger sample of speakers and stimuli and better control of lexical properties of the stimuli.

Table 1. *Representative transcription of Xi'an Mandarin tones in previous studies.*

Reference	T1	T2	T3	T4
Bai (1954)	21	24	453	45
Luo & Wang (1981)	31	24	42	55
Yuan (1989)	21	24	53	45
Wang (1996)	21	24	53	44
Peking University (2003)	21	24	53	55
Sun (2007)	31	24	53	55
Ma (2005)	21	24	53	44
Zhang & Shi (2009)	31	24	52	55
Ren (2012)	31	24	52	55

Although the specific pitch value of each tonal category in XM varies among previous studies, the basic tonal contour shape tends to be largely consistent across studies. Generally, the four tonal categories of XM possess the tonal contours of low-falling (T1), mid-rising (T2), high-falling (T3) and high-level (T4), respectively. Interestingly, XM tones display almost the same tonal contours as SC tones. In SC, tonal contours of the four tonal categories are described as high-level (T1, 55), mid-rising (T2, 35), low-falling-rising (T3, 214) and high-falling (T4, 51), respectively. As one can see, both tonal systems of SC and XM contain tones of high-level, mid-rising and high-falling tonal contours, and each of these tone pairs of similar contours share similar pitch values across the two tonal systems, though the similar contours do not necessarily represent the same tonal category in the two tonal systems (see Table 2 for details). Moreover, SC has a tone of low-falling-rising tonal contour, whereas XM has a tone of low-falling tonal contour without the rising tail. The former, however, would lose its rising tail when placed before other tones (Dow, 1972; White, 1980) and become similar to the latter. Overall, each XM tone seems to have a corresponding tone in SC with which it shares similar tonal contour and

pitch value, resulting in a very interesting systematic mapping pattern between the tonal systems of XM and SC.

Table 2. *Paired tones with similar contours from SC and XM.*

Tone Pair	Standard Chinese (SC)			Xi'an Mandarin (XM)		
	Tonal category	Pitch value	Example	Tonal category	Pitch value	Example
Level contour	SC_T1	55	ma1/妈	XM_T4	55/44/45	ma4/骂
Rising contour	SC_T2	35	ma2/麻	XM_T2	24	ma2/麻
Low contour	SC_T3	214	ma3/马	XM_T1	21/31	ma1/妈
Falling contour	SC_T4	51	ma4/骂	XM_T3	52/53/42	ma3/马

In fact, the mapping of tonal contours between the tonal systems of XM and SC has been proposed in previous studies. Li (2001) introduced the mapping pattern of XM tones and SC tones (similar as in Table 2) and suggested that XM learners of SC utilize their knowledge of XM tones to produce SC tones. Zhang (2009) also claimed the presence of a comparable tonal contour for each XM tone in SC. She further statistically compared the F0 contour of each XM tone with its SC counterpart of a similar tonal contour. The results showed that although the paired tones were similar in tonal contour, there were detailed acoustic differences. Specifically, Xi'an low-falling tone was different from the citation form SC low-falling-rising tone in contour shape; Xi'an rising tone was significantly lower than SC rising tone in F0 height except in the early-middle part; Xi'an falling tone had a similar initial F0 height with but higher final F0 height than the SC falling tone, and Xi'an level tone had overall lower F0 height than SC level tone. Zhang (2009) has made an attempt to reveal the acoustic similarities and differences between the two tonal systems empirically. However, it did not include details on the design of the production experiment, and therefore it is not clear how the tonal patterns were obtained and the comparison did not seem to be made on comparable datasets. The present study thus decided to compare the acoustic properties of XM tones and SC tones in a pairwise fashion with a more balanced and comparable design.

In addition to establishing the acoustic similarities or differences between the paired tones of XM and SC in production, we were also interested to know

whether each tone pair of similar contours from XM and SC is perceived to be similar or different in tone perception by bi-dialectal speakers of XM and SC. The tone production and perception experiments together were expected to confirm the mapping pattern of the two tonal systems. So far, there has not been any perception study comparing XM tones and SC tones. Conventionally, tone discrimination relies on several perceptual cues, among which the most widely adopted and important perceptual cues have proved to be F0 height and F0 contour, according to previous cross-language studies (Gandour, 1983, 1984; Gandour & Harshman, 1978; Francis, Ciocca, Ma, & Fenn, 2008). The relative importance of these two cues, however, varies among listeners of different language background. SC listeners tended to attach more importance to F0 contour than F0 height, whereas Cantonese and English listeners gave more weight to F0 height than F0 contour (Gandour, 1983; 1984). Apart from the F0-related features, other acoustic properties such as duration, amplitude contour and voice quality have also been shown to serve as secondary cues for tone discrimination, especially when the primary F0 information was not available (Liu & Samuel, 2004; Whalen & Xu, 1992; Yang, 2011). Furthermore, phonological rules might play a role in tone discrimination. For example, SC native listeners found it more difficult to discriminate between the rising tone and the low-falling-rising tone than other tone pairs in SC (Huang, 2007), which was partly attributed to the tone sandhi rule that makes the two tones conditioned allophonic tonal variants. Specifically, the low-falling-rising tone would be realized as a rising tone when it precedes another low-falling-rising tone (Duanmu, 2007). The native phonological rule can, sometimes even affect tone discrimination in a non-native language. For example, Cantonese listeners with or without SC experience had difficulty distinguishing between the SC high-level tone (55) and high-falling tone (51) (Hao, 2012; So & Best, 2010). This is because in Cantonese the high-level tone (55) has a free allophonic tonal variant, high-falling tone (53) (Hashimoto, 1972; Yip, 2002), which shows phonetic similarity to SC high-falling tone. In this study, based on the acoustic results in the tone production experiment, we ran a tone perception experiment to see whether each tone pair would be perceived as similar or different by the

bi-dialectal speakers and how the acoustic difference in each tone pair affects tone perception.

To sum up, in the present study, tonal categories with similar contours from XM and SC were paired as in Table 2. Both tone production and perception experiments were carried out to test whether each pair of tones is acoustically and perceptually similar or different. In Experiment 1, we compared the acoustic properties of the paired SC and XM tones produced by a group of highly proficient bi-dialectal speakers of these two dialects and established the acoustic difference of each tone pair. In Experiment 2, we further investigated whether each tone pair would be perceived as similar or different by the bi-dialectal speakers of SC and XM and how the acoustic difference in each tone pair affects tone perception with a five-scale tone judgment task. Results from both experiments were expected to reveal the tonal similarity and confirm the mapping pattern of the two tonal systems.

4.2 Experiment 1

4.2.1 Method

4.2.1.1 Participants

Thirty bi-dialectal speakers of SC and XM (16 males, 14 females) were selected and paid to participate in the experiment. All the selected participants were of high and comparable proficiency in both dialects, judging from their performance on a story reading task and their self-reported language proficiency through an adapted version of the LEAP-Q questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007). They acquired both dialects before the age of 6 and were early XM_SC bi-dialectal speakers with the first dialect (D1) being either XM or SC. All of them were born and raised in the urban areas of Xi'an and had no living experience outside of Xi'an. They were all undergraduate or graduate students at local universities. Their age ranged from 19 to 28 ($M \pm SD$: 22.5 ± 3.2). None of them had reported any speech or hearing disorders.

4.2.1.2 Stimuli

Thirty monosyllabic minimal tone sets with full sets of all four tones were selected. The four monosyllables within one minimal tone set were distinguished merely by tone, with the segments being identical. An exemplar of a full minimal tone set was *ma1*, *ma2*, *ma3* and *ma4*. The complete list can be found in Table C1 (see Appendix C). The monosyllables were selected on the condition that no pronunciation difference existed for the segment of each syllable between SC and XM, to avoid any potential effect of segmental pronunciation difference on tones. The monosyllabic items are frequent monosyllabic words with more than 4,500 occurrences in a corpus of 193 million words (Da, 2004). Within each minimal tone set, the monosyllabic words have comparable word frequencies. In total, 120 monosyllabic words (30 Syllables × 4 Tones) were selected. Some disyllabic words were added as fillers.

4.2.1.3 Recording

The recordings took place in a soundproof booth of the behavioral lab at Shaanxi Normal University in Xi'an. Stimuli were randomly presented to the speakers with E-prime 2.0. Each speaker produced all the items in both SC and XM with no repetition in two separate sessions. The order of the sessions was counterbalanced. Half of the speakers did the SC session first and then the XM session, the other half started with the XM session. Each session included one practice block and three experimental blocks. Between each block, there was a 3-minute break. The practice block contained 8 trials, which were not used in the experimental blocks, to familiarize the participants with the specific language mode. An experimental trial started with a 300 ms fixation cross, followed by a 200 ms pause. After that, a stimulus in the form of simplified Chinese character was presented on the screen. Speakers were requested to produce the stimulus in that particular language of the session in a self-paced fashion. They pressed a button to proceed to the next stimulus when finishing producing the current stimulus. The inter-stimulus interval was 500 ms. Instructions were given to the speakers visually on screen in simplified Chinese characters and orally by the experimenter in that particular language before

each session. All the stimuli were recorded at 16-bit resolution and a sampling rate of 44.1 kHz on a laptop via an external digitizer (UA-G1). Altogether, 240 monosyllabic items (30 Syllables × 4 Tones × 2 Languages) were elicited from each of the 30 speakers.

4.2.1.4 Data analysis

The F0 and duration of the speech items were analyzed. All the stimuli were manually annotated in Praat (Boersma & Weenink, 2015). A custom-made script was then used to extract ten equally distanced F0 values from the rhyme part of each time-normalized syllable. Gross errors in F0 extractions were manually corrected afterwards. To eliminate between-speaker acoustic differences, the raw F0 values were transformed to Z-score[4] for each speaker (Rose, 1987), pooling the SC and XM productions.

Statistical analyses of F0 were carried out using the growth curve analysis (Mirman, 2014) with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.1.2 (R Core Team, 2015). The overall F0 curves were modeled with up to second-order orthogonal polynomials, given that the most complex F0 contour in this study has a U-shape curve. Three time terms of the models would be of interest: the intercept, the linear slope and the steepness of the quadratic curvature, which indicate the overall F0 mean, the direction of F0 change such as rising or falling, and the steepness of F0 rising or falling, respectively. If tonal contours under investigation are different, we expect statistical differences in at least one of the three time terms. We built separate models for each pair of tones as listed in Table 2. All the models included random effects of Subjects and Items on all time terms. The fixed effects of Language (XM, SC) on all time terms, as well as random effects of Subjects-by-Language and Items-by-Language on all time terms, were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios.

---

[4] We used *Z*-scores instead of *T*-values to normalize tone values because *T*-values can be easily distorted by extreme values like the maximum F0 or the minimum F0.

Statistical analyses of duration were performed using linear mixed-effects regression models with the package *lme4* (Bates et al., 2015) in R version 3.1.2 (R Core Team, 2015). As for F0, we built separate models for the duration of each pair of tones. All the models first included random intercepts of Subjects and Items. The fixed effect of Language and random slopes of Subjects-by-Language and Items-by-Language were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios.

4.2.2 Results

4.2.2.1 F0

Figure 1 presents the mean F0 (Z-score) contours of the four tones in SC and XM. We report the F0 for each pair of tones of similar tonal contours in what follows.

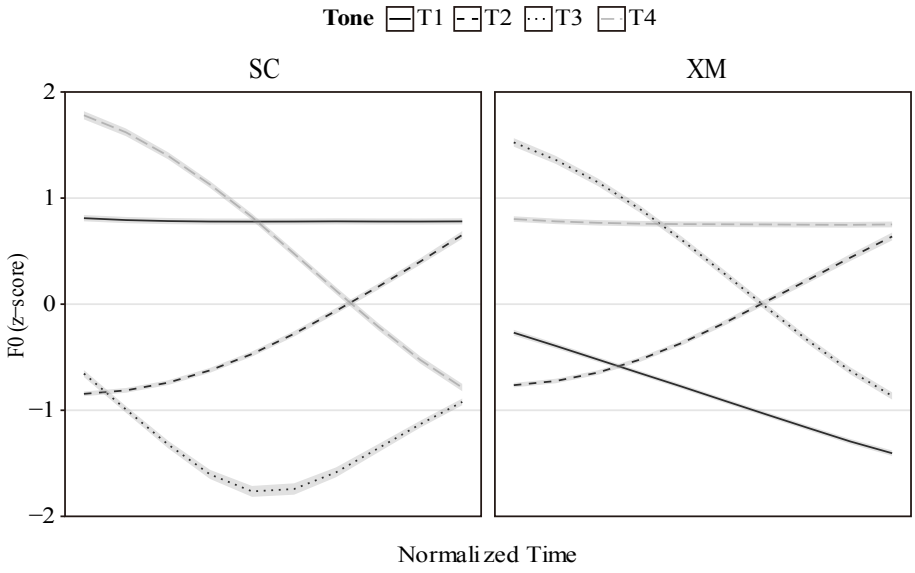

Figure 1. *Mean F0 (Z-score) contours of the four tones in SC and XM. The F0 values of each tone were averaged over 30 speakers and 30 monosyllabic items, with the tone of each item represented by ten equally distanced F0 values taken from the rhyme part of the time-normalized item. Grey areas indicate the 95% confidence interval of the corresponding mean.*

1) Level contour: SC_T1 vs. XM_T4

Results showed that the effect of Language on the intercept did not improve model fit ($\chi^2(1) = 0.03$, $p = 0.87$), nor did the effect of Language on the linear term ($\chi^2(1) = 1.54$, $p = 0.21$) and the effect of Language on the quadratic term ($\chi^2(1) = 0.23$, $p = 0.63$). Overall, it seems that the F0 contours of SC_T1 and XM_T4 did not differ from each other.

2) Rising contour: SC_T2 vs. XM_T2

Results showed that the effect of Language on the intercept did not improve model fit ($\chi^2(1) = 3.17$, $p = 0.07$), nor did the effect of Language on the linear term ($\chi^2(1) = 2.73$, $p = 0.10$). The effect of Language on the quadratic term, however, did improve model fit ($\chi^2(1) = 14.05$, $p < 0.001$). Therefore, the overall F0 mean and the direction of F0 rising did not differ between the two rising tones from SC and XM, whereas their steepness of rising differed, with XM_T2 having a shallower curvature than SC_T2 ($\beta = -0.08$, $t = -4.01$, $p < 0.001$).

3) Low contour: SC_T3 vs. XM_T1

The analyses of SC_T3 and XM_T1 data showed that the effect of Language on the intercept significantly improved model fit ($\chi^2(1) = 58.33$, $p < 0.001$), as well as the effect of Language on the linear term ($\chi^2(1) = 27.89$, $p < 0.001$) and the effect of Language on the quadratic term ($\chi^2(1) = 36.00$, $p < 0.001$). Apparently, SC_T3 was different from XM_T1 in all three time terms. The overall F0 mean of XM_T1 was significantly higher than SC_T3 ($\beta = 0.46$, $t = 9.13$, $p < 0.001$). The direction of F0 change was also different between the two tones ($\beta = -0.96$, $t = -14.06$, $p < 0.001$), with SC_T3 having a falling-rising contour and XA_T1 having a low-falling contour without the rising tail. Moreover, the F0 curvature of XA_T1 was shallower than that of SC_T3 ($\beta = -1.08$, $t = -14.23$, $p < 0.001$).

4) Falling contour: SC_T4 vs. XM_T3

The analyses of SC_T4 and XM_T3 showed that there was a significant effect of Language on the intercept ($\chi^2(1) = 9.06$, $p = 0.003$). However, no Language effect on the linear term ($\chi^2(1) = 0.50$, $p = 0.48$) or the quadratic term ($\chi^2(1) =$

0.83, $p = 0.36$) was found. Evidently, the overall F0 mean of XM_T3 was lower than that of SC_T4 ($\beta = -0.19$, $t = -3.08$, $p = 0.002$). The direction of F0 falling and the steepness of F0 falling were not significantly different between the two falling tones.

4.2.2.2 Duration

Figure 2 presents the mean durations of the four tones in SC and XM. The following reports the duration results for each pair of tones of similar tonal contours.

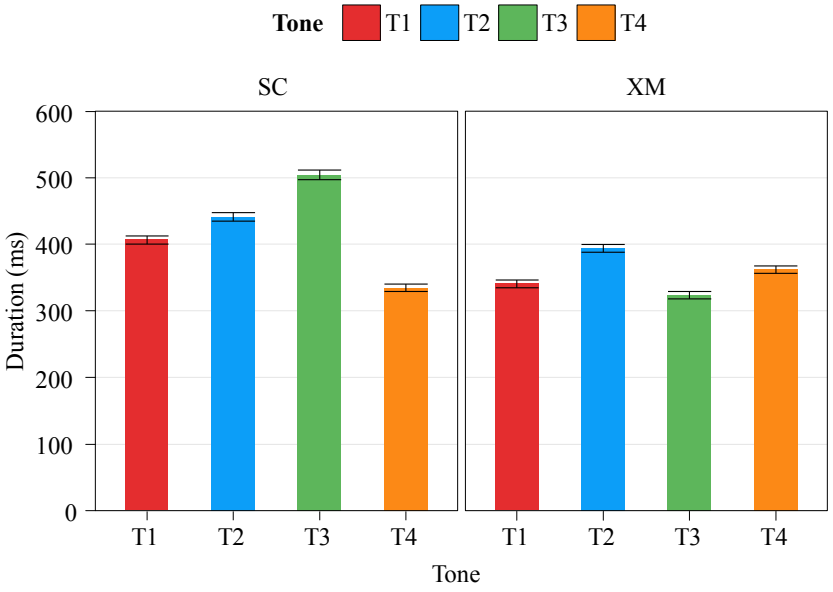

Figure 2. *Mean durations with 95% confidence interval of the four tones in SC and XM.*

1) Level contour: SC_T1 vs. XM_T4

There was a significant main effect of Language ($\chi^2(1) = 10.91$, $p < 0.001$). SC_T1 was significantly longer (45.09 ms) than XM_T4.

2) Rising contour: SC_T2 vs. XM_T2

The effect of Language significantly improved model fit ($\chi^2(1) = 425.36$, $p < 0.001$). SC_T2 was 47.60 ms longer than XM_T2.

3) Low contour: SC_T3 vs. XM_T1

Not surprisingly, a significant main effect of Language was also found for the durations of this tone pair ($\chi^2(1) = 71.33$, $p < 0.001$). SC_T3 was considerably longer than XM_T1, with the duration difference reaching up to 166.29 ms.

4) Falling contour: SC_T4 vs. XM_T3

An investigation of the durations of the tone pair of the falling contour revealed no effect of Language ($\chi^2(1) = 0.64$, $p = 0.42$), indicating that there was no duration difference between SC_T4 and XM_T3.

From the above comparisons of F0 and duration for each pair of tones of similar tonal contours, the acoustic patterns of each tone pair can be summarized as follows. First, the tone pair of level contour did not show any difference in their F0. However, the duration of the tone of level contour in SC was significantly longer than that of its counterpart in XM. Second, the overall F0 mean and the direction of F0 change did not differ between the two tones of rising contour in SC and XM, despite a shallow curvature of the rising F0 contour in XM_T2 relative to SC_T2. In addition, the duration of XM_T2 was considerably shorter than that of SC_T2. Third, the two tones of low contour in SC and XM were significantly different from each other regarding the overall F0 mean, the direction of F0 change and the steepness of F0 change. In fact, their contour shape differed, with the SC tone having a low-falling-rising contour and the XM tone having a low-falling contour without the rising tail. The former also tended to be remarkably longer than the latter. Fourth, having almost parallel F0 contours, the two tones of high-falling contour in SC and XM revealed difference in the overall F0 mean, with an overall higher F0 contour of SC_T4 compared to XM_T3. Their duration nevertheless did not differ.

4.3 Experiment 2

Having established the acoustic difference of each tone pair, the question arises as to whether the acoustic difference in each tone pair is perceivable. In fact, the two tonal systems provide an interesting test case for us to look into the relationship of the production and perception of tones by the bi-dialectal tonal

language speakers. As shown in Experiment 1, the acoustic difference of each tone pair ranged from no F0 difference (level tone pair) through F0 curvature difference (rising tone pair) to F0 height difference (falling tone pair) and F0 contour difference (low tone pair). With this setup of the two tonal systems, we could investigate how different F0 dimensions affect tone perception of the bi-dialectal tone language speakers. In this session, a five-scale tone judgment task was adopted to examine whether the acoustic difference in each tone pair is perceivable. We then compared the tone perception results of different tone pairs to reveal how different F0 dimensions affect tone perception.

4.3.1 Method

4.3.1.1 Participants

Another set of thirty XM_SC bi-dialectal speakers (11 males, 19 females) with high proficiency in both dialects were selected and paid to participate in the experiment. All the selected participants acquired both dialects before the age of 6 and were early XM_SC bi-dialectal speakers with a D1 of either XM or SC. They were born and raised in the urban areas of Xi'an and had no living experience outside of Xi'an. All were undergraduate or graduate students at local universities, between 18 and 29 years old ($M \pm SD$: 21.2 ± 2.6). None of them had reported any speech or hearing disorders. Informed consent was obtained from all the participants before the experiment.

4.3.1.2 Stimuli

The four pairs of tones in Experiment 1 were used for tone judgment. Since the two tones in each tone pair were similar in tonal contour overall, two pairs of tones of distinct tonal contours were added as fillers to avoid potential response strategies. Each tone pair was tested with all the 30 root monosyllables in Experiment 1, resulting in 30 trials. The two speech items in a trial always share the same segment so that participants could focus on the tone judgment. For example, a SC_T1 monosyllable ("妈", ma1, *mother)* was paired with its corresponding XM_T4 monosyllable ("骂", ma4, *scold*).

Four speakers were recruited to record the stimuli for the perception experiment. They were all university students and in their 20s. Two native speakers (one male, one female) of SC, who were born and raised in Beijing and had no knowledge of any other dialects, recorded the SC monosyllabic sounds. Likewise, two native speakers (one male, one female) of XM, who were born and raised in the urban area of Xi'an and had no living experience outside of Xi'an, recorded the XM monosyllabic sounds. Note that as it is impossible to find monolingual XM speakers nowadays, the two native speakers of XM also speak SC fluently. The recordings took place in Beijing for the Beijing speakers and in Xi'an for the Xi'an speakers. All the speech items were recorded at 16-bit resolution and a sampling rate of 44.1 KHz.

To ensure that the two XM speakers' tone patterns were representative enough of XM, we compared the acoustic properties of their tone patterns with those of the XM tone patterns in Experiment 1 (see Figures 3 and 4) and found no statistical difference in F0 and duration. It was therefore confirmed that the two XM speakers' production of XM tone patterns were representative patterns of XM and suitable for the perception study. We also compared the acoustic properties of the two SC speakers' tone patterns with those of the SC tone patterns in Experiment 1 (see Figures 5 and 6) and did not find statistical difference in F0 and duration either.

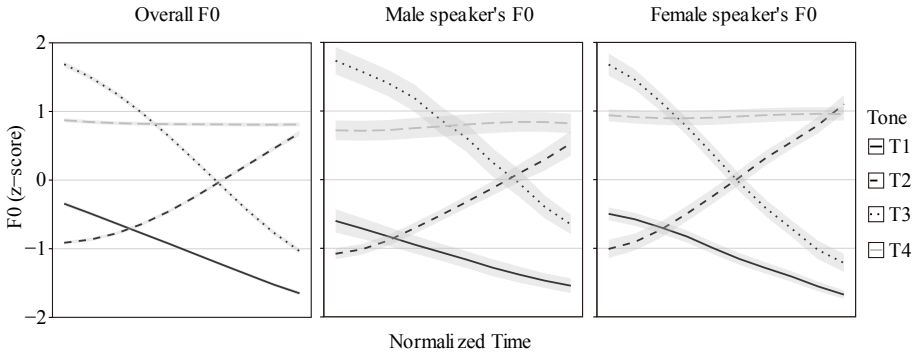

Figure 3. *Mean F0 (Z-score) contours of the four tones in XM by the 30 SC and XM bi-dialectal speakers in Experiment 1 (left panel), the male Xi'an speaker (middle panel) and the female Xi'an speaker (right panel) in Experiment 2. Grey areas indicate 95% confidence interval of the corresponding mean.*

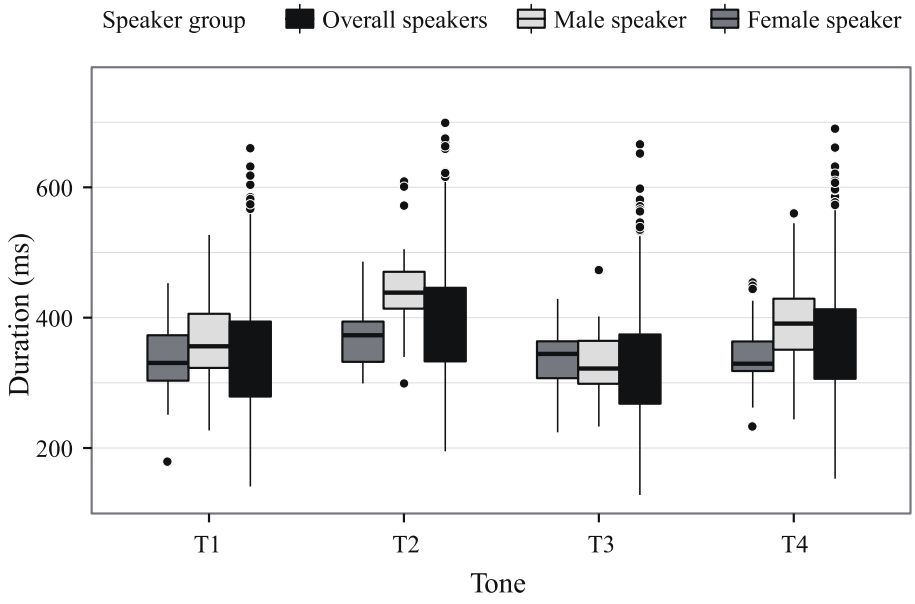

Figure 4. *Mean durations with 95% confidence interval of the four tones in XM by the 30 SC and XM bi-dialectal speakers in Experiment 1 (black boxes), the male Xi'an speaker (light grey boxes) and the female Xi'an speaker (dark grey boxes) in Experiment 2.*

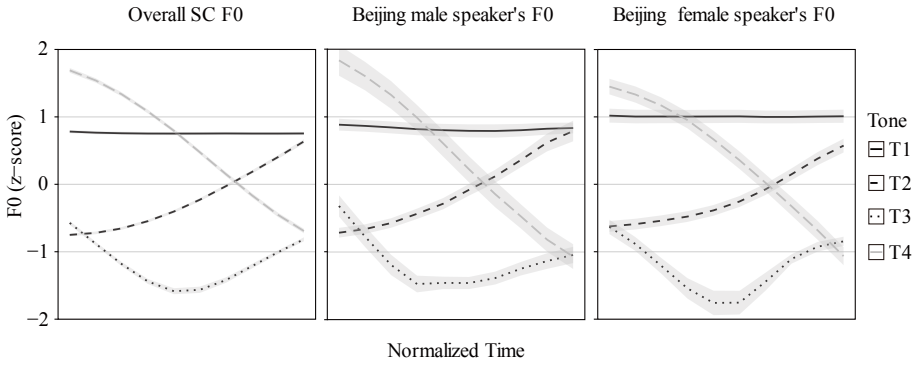

Figure 5. *Mean F0 (Z-score) contours of the four tones in SC by the 30 SC and XM bi-dialectal speakers in Experiment 1 (left panel), the male Beijing speaker (middle panel) and the female Beijing speaker (right panel) in Experiment 2. Grey areas indicate 95% confidence interval of the corresponding mean.*

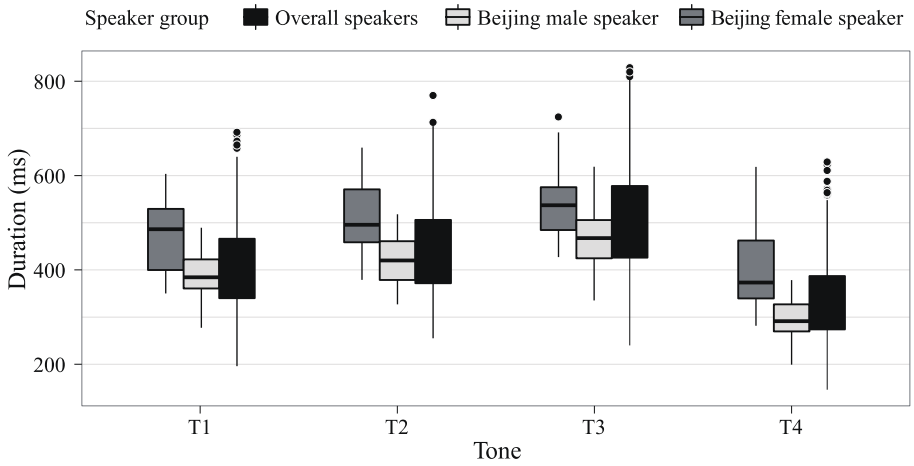

Figure 6. *Mean durations with 95% confidence interval of the four tones in SC by the 30 SC and XM bi-dialectal speakers in Experiment 1 (black boxes), the male Beijing speaker (light grey boxes) and the female Beijing speaker (dark grey boxes) in Experiment 2.*

After normalizing the amplitude of all the speech items in Praat (Boersma & Weenink, 2015), we paired the Beijing female speaker's speech items with the Xi'an female speaker's corresponding speech items according to tone pairs. The same was done for the two male speakers' speech items. Instead of recording all the speech items by a XM_SC bi-dialectal speaker, we recorded the SC speech items by native speakers of SC and the XM speech items by native speakers of XM. This ensured more typical realizations of SC and XM tones. The inclusion of two groups of speakers could avoid potential speaker bias.

4.3.1.3 Procedure

Participants were tested individually in a soundproof booth of the behavioral lab at Shaanxi Normal University in Xi'an. All the trials (30 Syllables × 6 Tone pairs × 2 Speaker groups) were randomly presented to the participants using the E-Prime 2.0 software through headphones at a comfortable listening level.

The experiment included a practice block and four experimental blocks. The practice block contained 6 trials, which were not used in the experimental blocks. Each experimental block contained 90 trials. Between every second block, there was a 3-minute break. An experimental trial started with a 100 ms

warning beep, followed by a 300 ms pause. The first speech item was then presented. After a 600 ms pause, the second speech item was presented. The language order of the two speech items in a trial was counterbalanced for each speaker group of the trials. Half of the trials presented the SC item before its corresponding XM item, while the other half presented the SC item after its corresponding XM item. Participants were requested to judge the similarities of the two tones of the two speech items in a trial on a five-point scale, with 1 indicating "completely different" and 5 indicating "completely the same". Response accuracy rather than speed was stressed. However, if participants did not make any response from the onset of the second stimulus to 2.5 s after the offset of the second stimulus, the program moved on to the next trial automatically with an inter-trial interval of 500 ms. Instructions were given both visually on screen and orally by the experimenter in SC before the experiment. To eliminate any influence of top-down knowledge on tone judgment, we did not mention the source languages of the auditory stimuli to the participants in the instruction.

4.3.1.4 Data analysis

To decide whether each pair of tones was perceived as similar or different, we analyzed the frequency distribution of the responses with chi-square goodness-of-fit test. The observed frequency distribution of the responses was first compared with the expected frequency distribution (null hypothesis: equal proportions) for each tone pair. If the null hypothesis of equal proportions was rejected, the individual response category's contribution to the overall chi-square statistic was determined by calculating the square of the difference between the observed and expected frequencies for a category, divided by the expected frequency for that category. Generally speaking, categories with a larger difference between the observed and expected frequencies make a larger contribution to the overall chi-square statistic. After recognizing the response category that contributed the most to the overall chi-square statistic, we further conducted several pair-wise goodness-of-fit tests to compare this category's frequency with that of the other categories. If all the comparisons are statistically significant (p-value adjusted), the category would be considered as

the best indicator of the similarity/difference of the two tones under investigation.

The second analysis concerned how the varying acoustic differences of different tone pairs affect tone perception of XM_SC bi-dialectal speakers. All the four pairs of tones were merged into one dataset and the tone perception results of different tone pairs were compared. Statistical analyses were carried out with the package *ordinal* (Christensen, 2015) in R version 3.1.2 (R Core Team, 2015). Cumulative link mixed models (CLMMs) were constructed for the dependent variable Response (1, 2, 3, 4, 5) with Tone pair (Level, Rising, Low, Falling), Language order (XM before SC; SC before XM), Speaker group (Female, Male), Listener gender (Female, male) and their interactions as fixed factors, and Subjects and Items as random factors. The fixed factors were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios. Post-hoc pairwise comparisons between different tone pairs were conducted using the *lsmeans* package (Lenth, 2016) with single-step *p*-value adjustment.

4.3.2 Results

4.3.2.1 Level contour: SC_T1 vs. XM_T4

The Chi-square goodness-of-fit test showed that the responses were clearly not equally distributed ($\chi^2(4) = 5634.42$, $p < 0.001$). As can be seen from Table 3, the response category "5" contributed the most to the overall chi-square statistic. Pairwise comparisons showed that the frequency of the response category "5" was significantly higher than that of the other categories (all ps < 0.001), indicating that SC_T1 and XM_T4 were mostly judged as 5, i.e., completely the same.

4.3.2.2 Rising contour: SC_T2 vs. XM_T2

The Chi-square goodness-of-fit test showed that the responses were clearly not equally distributed ($\chi^2(4) = 5677.38$, $p < 0.001$). Again, the response category "5" contributed the most to the overall chi-square statistic (see Table 3). Further pairwise comparisons showed that the frequency of the response category "5"

was significantly higher than that of the other categories (all *p*s < 0.001), indicating that SC_T2 and XM_T2 were mostly judged as 5, i.e., completely the same.

Table 3. *Response counts for each tone pair.*

Tone pair	Measure	Response category				
		1	2	3	4	5
(Level) SC_T1 vs. XM_T4	Observed count	68	22	16	59	**1631**
	Expected count	359.2	359.2	359.2	359.2	359.2
	Contribution to Chi-Sq	236.1	316.5	327.9	250.9	**4503.0**
(Rising) SC_T2 vs. XM_T2	Observed count	55	17	21	68	**1637**
	Expected count	359.6	359.6	359.6	359.6	359.6
	Contribution to Chi-Sq	258.0	326.4	318.8	236.5	**4537.7**
(Low) SC_T3 vs. XM_T1	Observed count	**625**	165	102	149	**754**
	Expected count	359	359	359	359	359
	Contribution to Chi-Sq	**197.1**	104.8	184.0	122.8	**434.6**
(Falling) SC_T4 vs. XM_T3	Observed count	116	26	23	86	**1539**
	Expected count	358	358	358	358	358
	Contribution to Chi-Sq	163.6	307.9	313.5	206.7	**3896.0**

Note. 1 = "completely different"; 5 = "completely the same".

4.3.2.3 Low contour: SC_T3 vs. XM_T1

The Chi-square goodness-of-fit test showed that the responses were not equally distributed ($\chi^2(4) = 1043.36$, $p < 0.001$). As shown in Table 3, the response category "5" contributed the most to the overall chi-square statistic. However, the response category "1" also made a relatively large contribution to the overall chi-square statistic. Pairwise comparisons showed that the frequencies of the response categories "1" and "5" were significantly higher than those of the rest categories (all *p*s < 0.001). Moreover, the frequency of the response category "5" was higher than that of the response category "1" ($\chi^2(1) = 12.07$, $p = 0.0005$). Overall, participants were more likely to perceive SC_T3 and XM_T1 as the same tone, though they also gave slightly fewer but a comparable number of different responses.

4.3.2.4 Falling contour: SC_T4 vs. XM_T3

The Chi-square goodness-of-fit test showed that the responses were clearly not equally distributed ($\chi^2(4) = 4887.59$, $p < 0.001$). The response category "5" contributed the most to the overall chi-square statistic, as demonstrated in Table 3. Pairwise comparisons showed that the frequency of the response category "5" was significantly higher than that of the other categories (all ps < 0.001), indicating that SC_T4 and XM_T3 were mostly judged as 5, i.e., completely the same.

To summarize, the five-scale tone judgment results showed that the tone pair of level contour (SC_T1 and XM_T4) was mostly judged as the same by the XM_SC bi-dialectal speakers. Similarly, the tone pair of rising contour (SC_T2 and XM_T2) and the tone pair of falling contour (SC_T4 and XM_T3) were mostly judged as the same. Different was the tone pair of low contour (SC_T3 and XM_T1), which elicited a comparable number of same and different responses, though the two were statistically different. It seems that participants had a much harder time discriminating between the two tones of low contour in SC and XM.

4.3.2.5 Comparisons among the four tone pairs

Statistical results for the models of Response showed a significant main effect of Tone pair ($\chi^2(3) = 253.69$, $p < 0.001$), indicating that the rating tendency differed significantly among the four tone pairs (see Table 3). There was also a significant main effect of Language order ($\chi^2(1) = 22.61$, $p < 0.001$) and a significant main effect of Speaker group ($\chi^2(1) = 26.76$, $p < 0.001$). No effect of Listener gender or interaction effect of the above factors was found (all ps > 0.05). Specifically, when a XM tone was presented before its corresponding SC tone, listeners were more likely to rate higher, i.e., more likely to judge the two tones as being more alike ($\beta = 0.35$, $z = 5.08$, $p < .001$) compared to when a SC tone was presented before a XM tone. Likewise, listeners tended to rate higher for the male speaker group' speech than for the female speaker group's speech ($\beta = 0.37$, $z = 5.24$, $p < .001$).

Post-hoc pairwise comparisons showed that the rating tendency of the tone pair of level contour was not significantly different from that of the tone pair of rising contour ($\beta = -0.002$, $z = -0.02$, $p = 1.00$). Both pairs were mostly judged as the same. Their rating tendencies, however, were significantly different from the tone pair of low contour (level vs. low: $\beta = 1.87$, $z = 18.95$, $p < 0.001$; rising vs. low: $\beta = 1.87$, $z = 19.11$, $p < 0.001$) and the tone pair of falling contour (level vs. falling: $\beta = 0.31$, $z = 3.08$, $p = 0.01$; rising vs. falling: $\beta = 0.32$, $z = 3.11$, $p = 0.01$). The rating tendencies of the tone pair of low contour and the tone pair of falling contour also showed significant difference ($\beta = -1.55$, $z = -16.43$, $p < 0.001$). In summary, the rating tendency of the tone pair of low contour was significantly different from that of the other three tone pairs, with the former being judged as either different or the same (there were slightly more "completely the same" responses than "completely different" responses), whereas the latter three tone pairs were mostly judged as the same, though the tone pair of falling contour elicited more different responses than the tone pairs of level contour and rising contour.

4.4 General discussion

The present study investigated the phonological similarity in tones of two closely related Mandarin dialects, SC and XM. Tones with similar contours from SC and XM were paired and their acoustic properties were compared over properly-controlled large samples produced by a group of highly proficient bi-dialectal speakers of XM and SC. F0 results of the four tone pairs ranged from no F0 difference (level contour tone pair) through F0 curvature difference (rising contour tone pair) to F0 height difference (falling contour tone pair) and F0 contour difference (low contour tone pair). Except the falling contour tone pair, all the other tone pairs also exhibited difference in tone duration and the largest duration difference was found in the low contour tone pair. These tone pairs of varying acoustic differences were then presented to the bi-dialectal speakers of XM and SC for tone perception with a five-scale tone judgment task. Results showed that the rating tendency of the tone pair of low contour was significantly different from that of the other three tone pairs, with the former being judged as either different or the same (there were slightly more

"completely the same" responses than "completely different" responses), whereas the latter three tone pairs were mostly judged as the same, though the tone pair of falling contour elicited more different responses than the tone pairs of level contour and rising contour.

With a balanced comparable design, the present production and perception experiments empirically confirmed the systematic tonal mapping pattern between XM and SC proposed in Li (2001) and Zhang (2009). While there were detailed acoustic differences in tone production, tones with similar contours between the two dialects were basically perceived to be the same, resulting in mapped tone pairs of level contour (SC_T1 vs. XM_T4), of rising contour (SC_T2 vs. XM_T2) and of falling contour (SC_T4 vs. XM_T3). Despite having distinct surface tonal contours, the tone pair of low contour (SC_T3 vs. XM_T1) also showed mapping, though to a lesser degree compared to the other three tone pairs.

The mapping pattern of XM tones and SC tones was initially put forward based on the similarity of the tonal contour and pitch value of XM tones and SC tones represented on the 5-point scale notation system (Chao, 1930; 1968). Different from the established pitch value of SC tones, there have been variances as to the specific pitch value of XM tones in previous studies (e.g., Bai, 1954; Luo & Wang, 1981; Ma, 2005; Peking University, 2003; Ren, 2012; Sun, 2007; Wang, 1996; Yuan, 1989; Zhang & Shi, 2009). However, the basic tonal contour shape of each tone was largely consistent across studies, and it has been noted that each XM tone has a mapped tone in SC with which it shares similar tonal contour and pitch value (Li, 2001; Zhang, 2009). Zhang (2009) tested the mapping pattern of the two tonal systems in tone production, but the tonal comparisons were not made on comparable datasets. The present study thus made more of an effort to empirically test the mapping pattern of the two tonal systems in tone production with more balanced comparable design. Tonal comparisons were made on paired tones of similar tonal contours from the two dialects produced by highly proficient bi-dialectal speakers of SC and XM. Our acoustic results showed that except for the tone pair of level contour, all the other tone pairs showed difference in F0. Specifically, the XM rising tone had a shallower rising F0 curvature than the SC rising tone, with an

overall comparable F0 mean. The XM falling tone had an overall lower F0 height than the SC falling tone. The XM low tone, not surprisingly, had a different F0 contour from the SC low-falling-rising tone. Our results of the specific F0 difference for each tone pair, except for the tone pair of low contour, was different from that in Zhang (2009), showing that the manipulation of a comparable design in this study actually resulted in different tonal realizations. It is therefore important to test on comparable datasets in such cross-dialect investigations. Nevertheless, both studies showed an overall compact tonal space of XM tones than SC tones. Apart from F0 difference, we also found duration difference for each tone pair except for the one of falling contour. All XM tones other than the falling tone tended to be shorter than their respective SC counterparts. Overall, there were acoustic differences for each pair of tones with similar contours from SC and XM.

The acoustically different tone pairs, however, were mostly perceived to be the same, or at least, very similar by the bi-dialectal speakers of SC and XM. In a five-scale tone judgment task, the tone pairs of level contour, rising contour and falling contour were mostly perceived to be completely the same, and the tone pair of low contour was also slightly more likely to be perceived as the same than different. Overall, these tone pairs of similar tonal contours from the two dialects were basically treated as the same during tone perception, despite the presence of acoustic differences. The results of our tone perception experiment confirmed the mapping pattern of XM tones and SC tones proposed in Li (2001) and Zhang (2009), providing new empirical evidence for the mapping of the two tonal systems from a perceptual point of view. Moreover, the mapping pattern seems to be more pronounced in tone perception than in tone production, given that each mapped tone pair was almost perceptually indistinguishable while having acoustically detectable differences.

That the tone pairs of similar tonal contours from XM and SC were basically perceived to be the same by the bi-dialectal speakers does not mean that the participants did not pick up the acoustic differences at all. The mapped tone pairs did vary in the degree to which they were perceived as the same tones. The tone pairs of level contour and rising contour were mostly judged as the

same by the bi-dialectal speakers of XM and SC, followed by the tone pair of falling contour. The tone pair of low contour elicited more different responses relative to the other three tone pairs. As only acoustic information was available to the participants during tone judgment, it is reasonable to assume that the different perceptual results for all the tone pairs resulted from their acoustic differences, in some way. Our acoustic analyses demonstrated that the four tone pairs showed variance in different F0 dimensions, ranging from no F0 difference (level contour tone pair) through F0 curvature difference (rising contour tone pair) to F0 height difference (falling contour tone pair) and F0 contour difference (low contour tone pair). This varying acoustic difference in different F0 dimensions seems to have affected the tone perception results of each tone pair to varying degrees. Compared to the level contour tone pair with no F0 difference, the rising contour tone pair with F0 curvature difference was not perceived any different, seemingly indicating that the bi-dialectal speakers of SC and XM were not sensitive to the F0 curvature difference between the two rising tones. This is not surprising, as F0 curvature has not been identified as a strong perceptual cue for tone discrimination. In contrast, the tone pair of falling contour with F0 height difference was perceived to be less similar than the tone pair of level contour with no F0 difference, suggesting that F0 height difference contributed to the discrimination of the two falling tones. This is consistent with the previous cross-language finding that F0 height is an important perceptual cue for tone discrimination (Gandour, 1983, 1984; Gandour & Harshman, 1978; Francis, Ciocca, Ma, & Fenn, 2008). Lastly, the tone pair of low contour with F0 contour difference was perceived to be much more different than the tone pair of level contour with no F0 difference, as well as than the tone pair of falling contour with F0 height difference. Obviously, F0 contour difference significantly affected the discrimination between the two low tones. Also, the bi-dialectal speakers tended to be more sensitive to the dimension of F0 contour than F0 height in tone discrimination, as has been found by Gandour (1983, 1984) for SC speakers.

Note that although the duration property of each tone pair was maintained in the speech stimuli, participants did not seem to make fully use of it in tone perception, if they used it at all. Duration difference was found in all the tone

pairs except in the falling tone pair. If the participants did use the duration cue for tone perception, with a duration difference of nearly 50 ms, the tone pair of level contour as well as the tone pair of rising contour should have been judged as different tones rather than similar tones. If this is not convincing, a duration difference of about 166 ms in the tone pair of low contour should be certainly salient enough to rule out the possibility that the two tones were judged as similar. However, the pair of low contour tones ended up eliciting even slightly more same responses than different responses. Overall, duration was not adopted as a valid perceptual cue for tone discrimination by the bi-dialectal speakers of SC and XM. They relied primarily on F0 information to make tone judgments.

Acoustic information, especially F0 information, is not the only perceptual cue that listeners employ during tone discrimination. Phonological rules can sometimes play a role in the process, too (Hao, 2012; Huang, 2007; So, & Best, 2010). In this study, the tone pair of low contour (SC_T3 vs. XM_T1) had distinct F0 contours. SC_T3 has a low-falling-rising contour and XM_T1 has a low-falling contour. If participants made tone perceptions purely based on acoustic information, the two low tones would have been judged as different. Instead, the two tones were perceived as either different or similar, with even slightly more same responses than different responses. This could presumably be attributed to the phonological rule of SC_T3. SC_T3 has a low-falling-rising contour when it is in citation form or in the final position of an utterance. When placed before other tones or in a context, it loses its rising tail and becomes a low-falling contour (Dow, 1972; White, 1980), which shows phonetic similarity to the XM low-falling tone. Participants seem to have applied the phonological rule of SC_T3 and used the context form SC_T3 to mediate between the citation form SC_T3 and XM_T1, and therefore classified the citation form SC_T3 and XM_T1 as similar tones. Recall that we did not mention the source languages of the speech stimuli to the participants. It is therefore not clear where and how the context form SC_T3 came into play. There might be two scenarios. One is that XM_T1 here were considered as a representation of the context form SC_T3. Participants then made a comparison between the citation form SC_T3 with the context form SC_T3,

which were judged as similar (citation form SC_T3 vs. (XM_T1 → context form SC_T3)). The other scenario is that when presented with the citation form SC_T3 and XM_T1, participants activated the corresponding context form SC_T3, and they compared the context form SC_T3 with XM_T1, the latter being considered either as a representation of XM_T1 or context form SC_T3, resulting in similar response ((citation form SC_T3 → context form SC_T3) vs. XM_T1). In either scenario, the tone pair of low contour should be judged as similar tones as a result of the phonological rule. In our result, the two low tones from SC and XM elicited a comparable number of same and different responses, suggesting that both the acoustic information and phonological rule played roles in the tone discrimination process, and the two effects seem to counterbalance each other.

4.5 Conclusion

To conclude, the present study investigated the phonological similarity in tones of two closely related Mandarin dialects, SC and XM. Through production and perception experiments, it was established that there is systematic mapping of tones between XM and SC. The degree of the similarity of the mapped tone pair in tone perception was largely dependent on the acoustic phonetic similarity between the tones in tone production, with the phonological rule playing a role in certain circumstance.