Proceedings of the 22nd North American Conference on Chinese Linguistics (NACCL-22) & the 18th International Conference on Chinese Linguistics (IACL-18). 2010. Vol 1. Clemens, L.E. & C.-M. L. Liu, eds. Harvard University, Cambridge, MA. 345-362.

The Tonal System of Singapore Mandarin

Leslie Lee
University of California, San Diego

This paper presents an acoustic study of the tonal system of Singapore Mandarin and compares the results with those of Beijing and Taiwan (e.g. Xu 1997, Shih 1988). The results show that Singapore Mandarin differs from the other varieties in a number of ways, particularly in the realisation of Tone 2, which is marked by a low level stretch not found in the other varieties. This system is also influenced by the sociolinguistic variables of gender and language background. Overall, the results support previous studies in showing that phonologically identical tonal systems can exhibit dialectal variation in realisation, while the intra-dialectal variation suggests that tonal realisation can be used for indexical purposes.

1. Introduction

Most acoustic studies of Mandarin tone have focussed on the varieties spoken in Beijing and Taiwan (e.g. Xu 1994, 1997, 1998; Shih 1988; Fon & Chiang 1999 among others), and little attention has been paid to the dialect spoken by the large Chinese population in Singapore. Most Chinese Singaporeans are bilingual in at least two of the four official languages: English and Mandarin. Among the ethnic Chinese (76.8% of the resident population), the predominant home language can either be Mandarin (45.1%), English (23.9%), or other Chinese languages (30.7%) (Leow 2001: viii-ix). Given the sociolinguistic situation of bilingualism, the potential influence of other tonal Chinese languages, and the sheer geographical distance from Taiwan and China, one would expect the realisations of Singapore Mandarin tones to differ from those of Beijing and Taiwan. As published studies of the tones in Singapore Mandarin are wanting in the literature, it is important to document this aspect of the language.[1]

The aim of this paper, then, is precisely to present an acoustic study of the four lexical tones in Singapore Mandarin, and compare the results with the other standard varieties of the language. In addition, this study also takes the sociolinguistic situation in Singapore into consideration by controlling the gender and language background of the speakers, so as to investigate any potential effects of these factors on tonal realisation in Singapore Mandarin (henceforth SM).

[1] I would like to thank Amalia Arvaniti, Sharon Rose, Sarah Creel, Lucien Carroll, Dan Michel, Roger Levy, and members of the audience at IACL18/NACCL22 for useful discussion and feedback.

2. Background

2.1. Acoustic studies

Mandarin tones are typically transcribed using the ‘Chao tone letter system’ (Chao 1930/1980). Following Chao’s (1956, 1968) transcriptions of Peiping (Beijing) Mandarin the four tones are commonly transcribed as shown in (1):

(1) Tonal contrasts in Mandarin

Tone #	Pitch pattern	Chao tone letters	Example	
1	high level	55	*ma1*	‘mother’
2	high rising	35	*ma2*	‘hemp’
3	low falling-rising	214	*ma3*	‘horse’
4	high falling	51	*ma4*	‘to scold’

Shih (1988) presents an acoustic study of Taiwan Mandarin (henceforth TM) tones produced by a female speaker. Using the set of minimal pairs in (1), she plots the tones produced in isolation, reproduced in Figure 1.

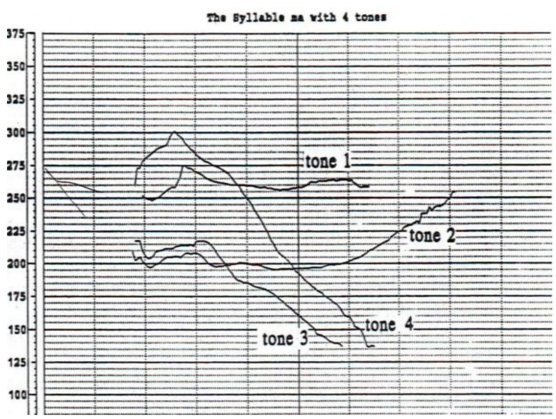


Figure 1: TM tones (Shih 1988: 98)

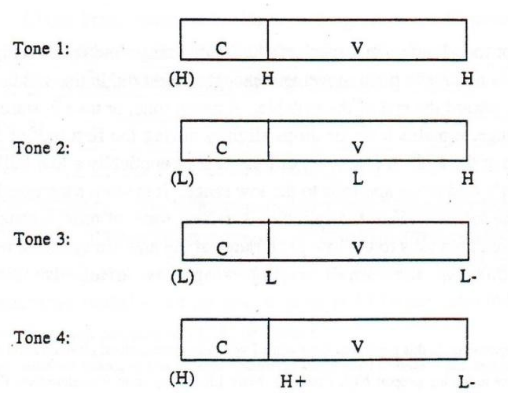


Figure 2: Tonal targets (Shih 1988: 84)

Shih notes that the start and end points of all tones fall on three distinct levels: Tones 1 and 4 both start high, close to where Tones 1 and 2 end; Tones 2 and 3 both begin in the mid range; while Tones 3 and 4 both fall to the low range. Based on these observations, she proposes the relative values and placement of tonal targets for each tone as shown in Figure 2, and suggests that tonal contours are the result of interpolation between targets (H+ represents a pitch level slightly higher than H; L- is lower than L).

Beijing Mandarin (henceforth BM) tones seem to behave differently. Xu (1997) examines BM tones produced by eight male speakers both in isolation and in different tonal contexts. The tones in isolation were elicited using the set of minimal pairs in (1). A program was used to record the F0 values at regular intervals within each segment, and the F0 curves obtained were smoothed using a function incorporated into the program. The mean F0 contours were plotted over normalised time, reproduced in Figure 3.

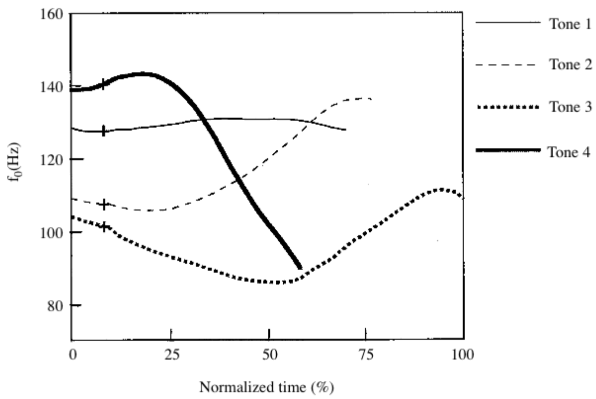


Figure 3: BM tones (Xu 1997: 67)

These data highlight differences between TM and BM. In terms of tonal contours, the final rise on Tone 2 appears to occur later in TM than in BM, while BM Tone 3 has a final rise not found in TM. In terms of duration, while relative durations of BM tones have a descending order of Tones 3>2>1>4, TM tones have the order 2>4>1>3.

In another study, Xu (1998) tested whether the F0 contours of the entire rhyme of a syllable or that of the vowel alone is more consistent across syllable types and speaking rates, and found that regardless of the internal structure of the syllable, the F0 contours for all four tones were consistently aligned with the syllables that carried them. He also investigated the alignment of F0 contours in Tone 2 – Tone 3 sequences, and found that regardless of the internal structure of the syllable, the F0 peak of Tone 2 always occurs close to the offset of the host syllable. He observes that the onset of the F0 rise on Tone 2 always occurs around the centre of the host syllable; that the peak velocity occurs closer to the syllable offset than onset; and that the maximum velocity of the F0 rise does not vary consistently with either syllable duration or structure. These results indicate that the F0 contour of Tone 2 is not spread out evenly over the host syllable, and the entire contour shifts towards the later portion of the syllable when the duration of the host syllable increases. Xu interprets this as evidence that the syllable is the domain of F0 contour alignment in Mandarin and that a contour tone should be treated as a single dynamic target instead of a sequence of static targets. Based on Xu's work on Mandarin tones, Xu & Wang (2001) outline a framework for accounting for surface F0 variations in speech in which there are two types of pitch targets – static and dynamic. "A static pitch target has a register specification, such as [high], [low] or [mid]. A dynamic pitch target has a linear movement specification, such as [rise] or [fall]" (p. 321). On this view, Mandarin has two static pitch targets [high] and [low] and two dynamic targets [rise] and [fall], associated with Tones 1, 3, 2, and 4 respectively. Thus, the contour tones, Tones 2 and 4, are treated as integral units of pitch movement and are not considered to be formed by the concatenation of two static targets, as Shih assumes.

In contrast with Xu and colleagues' assumptions that F0 falls and rises are represented by integral dynamic targets, work on tonal alignment within the autosegmental-metrical framework of intonational phonology (henceforth AM) (Ladd 2008) assumes that speech melodies consist of a series of tones at the phonological level which are phonetically realised as tonal targets and that F0 falls and rises are the result of the transitions between static targets aligned with the segmental string (see Arvaniti 2007 for the comparative evaluation of phonetic models of F0). Work within AM such as Arvaniti et. al. (2006) and Arvaniti & Ladd (2009) thus share with Shih the intuition that complex tonal contours are composed of smaller units aligned with segments.

2.3. Singapore Mandarin

There are few published reports on SM, and to the best of my knowledge, no controlled acoustic survey of tone in SM had been conducted prior to the present study.

Chua (2003) draws a sociolectal distinction between "Singapore Standard Mandarin" (SSM) and "Singaporean Mandarin" (SgM), and notes that speakers code-switch very easily between the two sub-varieties:

> The whole range of the SgM speech continuum provides functional varieties for its speakers. The highest attainable sub-variety a speaker of SgM can use is coupled to his or her educational standard and socio-economic background – but he or she is able to drop quite easily and comfortable into 'lower' sociolects outside the context of his or her own sociolect for functional purposes. (p. 42-43)

He provides the transcriptions of the four tones in the two sub-varieties as shown in (2). Even though Chua makes the distinction between SSM and SgM, he does not provide any useful way of drawing the line between the two sub-varieties; given that both SSM and SgM are parts of a speaker's repertoire and that these varieties lie along a continuum, it is not clear how one can make a reliable distinction between the tones of SSM and SgM. Chua's pitch transcriptions of SgM correspond to Chen's (1983/1993)[2] observations that SM Tones 1, 2, and 4 often appear to be lower than in the other varieties: Tone 1 is often realised as [44], Tone 2 as [24], and Tone 4 as [41] or [42] (Chen 1993: 251). However, these researchers do not specify how the numbers on the tonal scale have been assigned to the tones, and these are likely to be impressionistic judgements.

[2] Chen also reports a fifth tone in Singapore Mandarin, which occurs only in limited distributions. However, she also notes that both Tone 4 and this fifth tone share the same pitch values, and cannot be consistently distinguished. In a cross-generational study, Ho (2000) reports, based on her auditory perception, an inverse relationship between age and frequency of occurrence of the fifth tone. Given that discrimination between the fifth tone and Tone 4 is not reliable even for those speakers who have the distinction, and that the use of the fifth tone is in marked decline, this paper focuses on the four traditional tones of Mandarin.

(2) SSM vs. SgM (Chua 2003: 66)

	SSM	SgM
Tone 1	55	44
Tone 2	35	24
Tone 3	214/211	211
Tone 4	51	41/42

Unlike previous studies, here we recorded speakers who differed in gender and the role of Mandarin in their language repertoire, and subjected the data to acoustic analysis.

3. Methods

3.1. Speakers

16 speakers of SM, 8 male and 8 female, took part in the study. The speakers were ethnic Chinese who, like their parents, had been born and raised in Singapore. All speakers were in their twenties, and were either pursuing or had completed their university education. Within each gender group, the speakers were divided into two categories based on the language used most frequently at home: those whose dominant home language was Mandarin (M_D) and those whose dominant home language was English (E_D). All speakers, M_D and E_D, use some Mandarin on a daily basis and use English at least half the time in their daily lives. The speakers were naïve as to the purposes of the experiment, and none of the speakers had any speech or hearing impairment.

3.2. Materials

In order to establish the canonical form of each of the four SM tones, the four tones were elicited in isolation using the same test words as those used in Shih (1988) and Xu (1997), since using the same test words allows for direct comparison with the results from those studies. In order to elicit the natural reading of these words as one-word utterances, each of these words was preceded by a question that asks how the word is read, and each question-answer pair was printed on a card in Chinese characters as a dialogue. Speakers were only required to read the target word aloud. A sample dialogue is shown below:

(3) Sample dialogue

甲：这个字怎么读？	(A: How is this character read?)
乙：妈	(B: ma)

Each of the four dialogues was printed on five separate cards (4 conditions x 5 cards), and the order of their appearance as randomised.

3.3. Procedure

The recordings took place in quiet rooms either in the speakers' homes, offices, or university that they attended. Prior to the recording, the speakers were given a practice session during which the procedure was explained to them: each of the four dialogues was presented to them on flash cards, and they were instructed to read each word as naturally as possible. During the actual recording, which was monitored, the speakers flipped the flash cards and read at their own pace.

The materials were recorded in Audacity with a sampling rate of 22050Hz and a 16-bit resolution. The recordings were done on a laptop computer using a Beyer dynamic TG-X58 microphone. Three of the five repetitions of each test word were used for measurement; the first and last repetitions were discarded. Measurements from the three tokens were then averaged for each speaker for the purposes of statistical analysis.

3.4. Measurements

The selected tokens were segmented in Praat (Boersma and Weenik 2003) by visual inspection of the spectrograms. The durations of the consonant, vowel, and syllable were then computed from the various time values.

As noted in Section 2, there is no consensus in the literature on how tones should be measured: some researchers measure specific points in the F0 track, while others obtain F0 measurements from regular intervals along the pitch track by automatic means. For this paper, both methods were employed. The first method follows that outlined in Xu (1997), whereby a program was used to record the F0 values at regular intervals within each segment. By plotting the mean F0 contours over normalised time, we then have a picture of the overall configuration of the pitch shapes of the four tones.

In the second method, pitch tracks were obtained for each token using the autocorrelation method in Praat that gives F0 values every 10ms (Boersma 1993), and specific points were measured. In selecting F0 points, care was taken to avoid obvious microprosodic perturbations. F0 measurements were converted from Hertz to Equivalent Rectangular Bandwidth (ERB) using the equation of Hermes and van Gestel (1991: 97):

$$ERB = 16.7 \log(1+ f/165.4), \text{ where } f \text{ is frequency in Hz}$$

This was done because data were collected from male and female speakers, and this scale is standardly used for comparing speech across genders (e.g. Daly and Warren 2001).

For Tone 1, a high level tone, three F0 points were measured: the first F0 point after the onset of /m/ (H1T1), the first F0 point after the vowel onset (H2T1), and the last F0 point before the vowel offset (H3T1) (see Figure 4, panel a)

All tokens for Tone 2 showed an initial fall followed by a low level stretch before the final rise. Four points were measured: the first F0 point after the onset of /m/ (H1T2), the first elbow at which the F0 slope gradually changed from steep to gentle (L1T2), the second elbow at which the F0 slope gradually changed from gentle to steep (L2T2), and the final F0 point before the vowel offset (H2T2) (see Figure 4, panel b).

Tones 3 and 4 were realised as falling tones, with Tone 4 starting at a higher initial F0 than Tone 3. In both cases, some tokens showed a small final rise, and tokens produced by the same speaker could either have the final rise or not. Since the final rise was not found in all tokens, this was not measured, but counts of the number of tokens that exhibited the final rise were kept for both tones and are discussed below. Three F0 measurements were made for Tones 3 and 4: the first F0 point after the onset of /m/ (H1T3/H1T4), the first F0 point after the vowel onset (H2T3/H2T4), and the F0 minima (minT3/minT4 (see Figure 4, panels c & d).

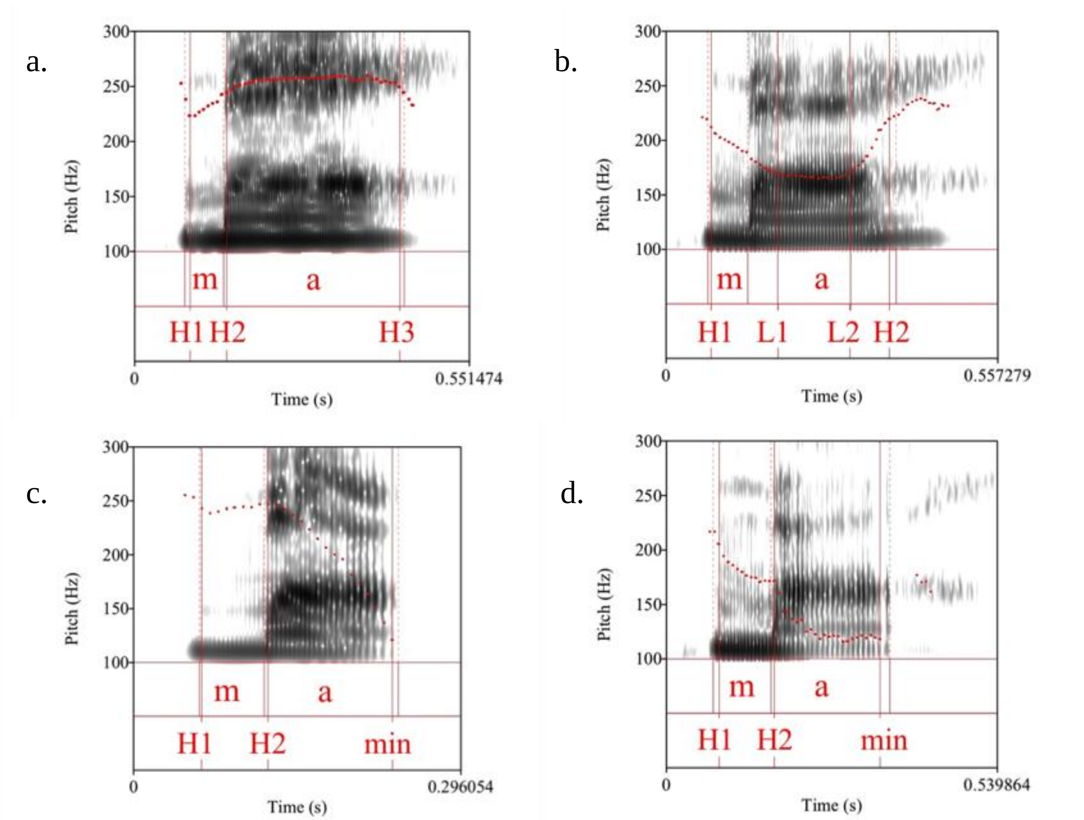


Figure 4: Sample analyses. Clockwise from top left: Tones 1, 2, 3, and 4.

For the purposes of statistical analyses, the speaker mean for each point measured was derived by averaging the measurements of that point across the three tokens produced by that speaker. The F0 range of each speaker was calculated by subtracting the highest mean F0 measure from the lowest mean F0 measure for that speaker.

3.5. Statistical analyses

The analysis of duration involved a repeated measures ANOVA with four levels (corresponding to the four tones), conducted with the duration of the syllable as the dependent variable, and gender and language background as categorical variables. In order to investigate the alignment of the pitch elbows in Tone 2 with respect to the segmental material and to each other, several factorial ANOVAs were conducted with gender and language background as categorical variables and the following dependent variables, expressed as percentages of total syllable duration: (i) the duration between the syllable onset and L1T2 (initial fall); (ii) the duration between L1T2 and L2T2 (plateau); and (iii) the duration between L2T2 and the vowel offset (final rise).

For the analysis of F0 range, a factorial ANOVA was conducted with F0 range as the dependent variable, and gender and language background as the categorical variables. To analyse F0 scaling, a repeated measures ANOVA with 13 levels (corresponding to the 13 measured points) was also conducted with F0 as the dependent variable, and gender and language background as categorical variables.

For Tones 3 and 4, two chi-square tests were conducted based on the number of tokens that showed a final rise so as to investigate if there was a correlation between either gender or language background and the occurrence of a final rise on these tones.

4. Results

The plots produced from the method following Xu (1997) are shown in Figures 5 and 6. Figure 5 shows the aggregate mean tone contours for all speakers, while each panel in Figure 6 illustrates the aggregate mean tone contours for each of the four speaker groups. All other results that follow were derived from the second method of measurement.

4.1. Duration

The durations of each of the four tones are illustrated in Figure 7 (all error bars indicate standard deviation). Tone had a significant effect on the duration of the syllable [$F(3,36) = 40.256$, $p<0.001$]. A post-hoc Tukey HSD test indicated that the duration of the syllable bearing Tone 4 was significantly shorter than each of the other tones [$p<0.001$], but the other three tones were not significantly different from one another [$p>0.05$]. Both the effects of language background [$F(1,12) = 3.964$, $p=0.07$] and the interaction of language background and tone [$F(3,36) = 2.622$, $p=0.065$] on syllable duration approached significance. Given the relatively small sample of speakers it is likely that the effects of language background and its interaction with tone on syllable duration would be significant with data from more speakers. There was no significant effect of gender [$F<1$].

4.2. Alignment in Tone 2

There was a significant effect of gender on the duration of the plateau [$F(1,12) = 6.136$, $p<0.030$], as illustrated in Figure 8: male speakers had a longer plateau than female speakers.

Language background had a significant effect on the duration of the final rise [$F(1,12) = 9.192$, $p<0.010$]: the final rise occupied a greater proportion of the Tone 2 syllable for M_D speakers than E_D speakers, as shown in Figure 9. No other main effects or interactions were found.

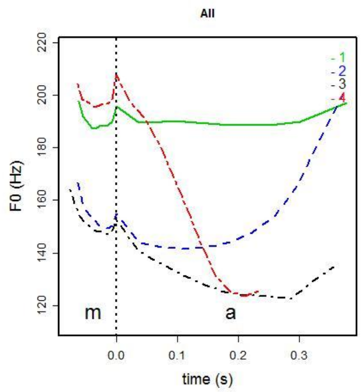
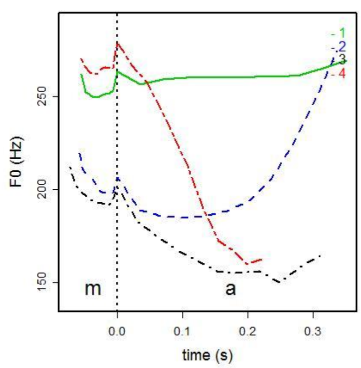
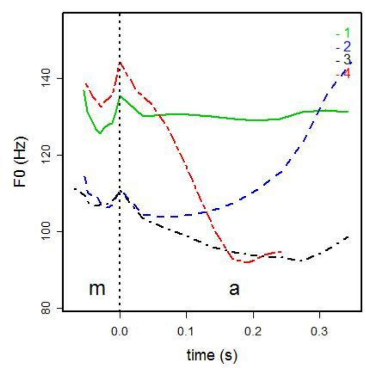


Figure 5: Aggregate mean tone contours for all speakers.

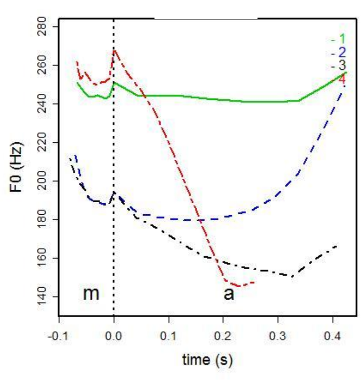
(a) M_D Female



(b) M_D Male



(c) E_D Female



(d) E_D Male

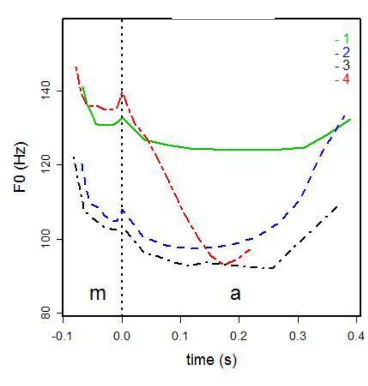


Figure 6: Aggregate mean tone contours for the four speaker groups.

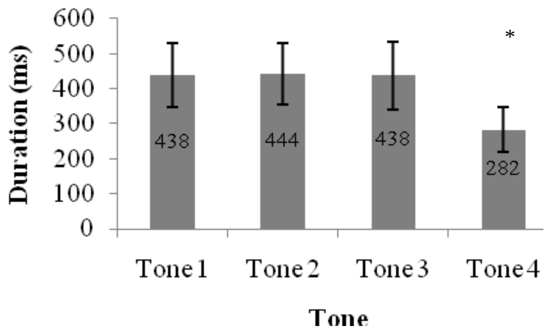


Figure 7: Syllable duration.

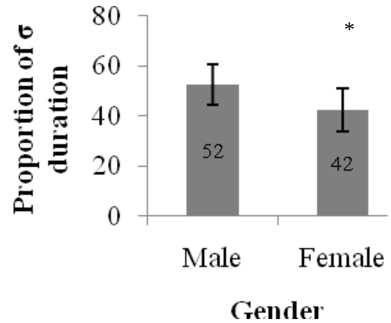


Figure 8: Tone 2 plateau

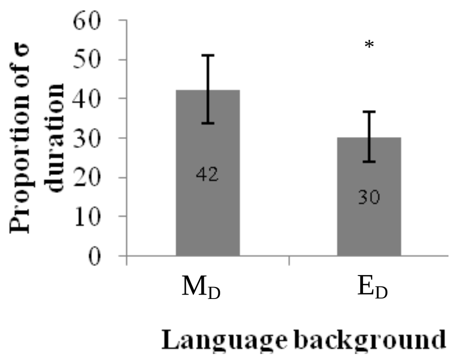


Figure 9: Tone 2 final rise

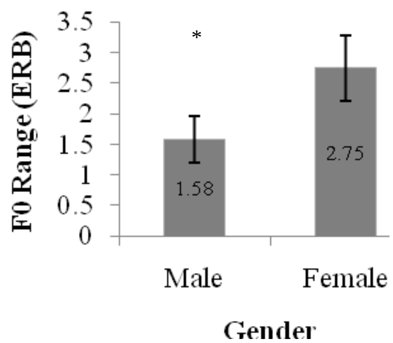


Figure 10: F0 range.

4.3. F0

Unsurprisingly, there was a significant effect of gender on F0 range [$F(1,12) = 23.271$, $p<0.001$]: male speakers used a smaller range than female speakers, as shown in Figure 10. There were no other main effects or interactions on F0 range [$F<1$].

There was a significant effect of F0 points on the scaling of F0 [$F(12,144) = 145.533$, $p<0.001$]. The 13 F0 points can be grouped into four pitch levels, based on statistical significance [Tukey HSD, $p<0.005$], as illustrated in Figure 11. Among the H points, only H1T2, H1T3, and H2T3 are significantly different from the others; H1T1, H2T1, H3T1, H2T2, H1T4, and H2T4 are not significantly different from one another, and as a group, they constitute the highest pitch level among the 13 points. Of the remaining three H points, H1T2 and H1T3 are not significantly different from each other, but both are significantly higher than H2T3 and the L and min points: H1T2 and H1T3 together form the second highest pitch level. H2T3 is not significantly different from L1T2 and L2T2, but these three points are significantly higher than minT3 and minT4, thus forming a third pitch level. At the lowest pitch level, minT3 and minT4 are not statistically different from each other, but are significantly lower than all the other points.

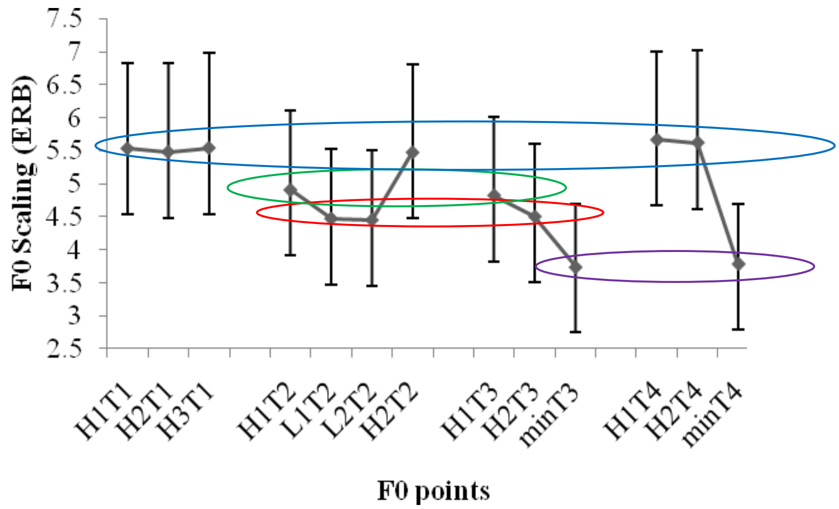


Figure 11: F0 scaling (all speakers).

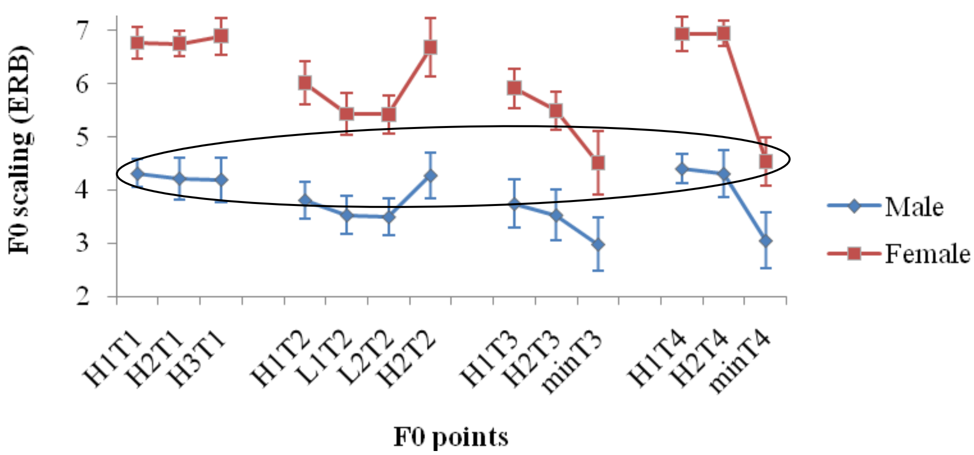


Figure 12: Gender differences in F0 scaling.

As expected, there was a significant effect of gender on the scaling of F0 [$F(1,12) = 147.842$, $p<0.001$]: each of the F0 points produced by male speakers was significantly lower than the corresponding point produced by female speakers [Tukey HSD, $p<0.05$], as illustrated in Figure 12. There was also a significant interaction of gender and the F0 points on the scaling of F0 [$F(12,144) = 12.356$, $p<0.001$]: minT3 and minT4 of female speakers are not significantly different from all the H points of male speakers, with the exception of H2T3 [Tukey HSD test, $p<0.05$].

Among the female speakers, the relative scaling of the 13 F0 points is similar to that of all the speakers combined, and the 13 points are organised along the four pitch levels in the same way [Tukey HSD, $p<0.05$]. With male speakers however, the four pitch levels are less distinct, as shown in Figure 13. While four pitch levels can still be identified, there is some overlap in the organisation of the 13 points along the four pitch levels. Unlike the pooled data or the female data, H2T1 and H3T1 of male speakers are not significantly different from H1T2. Also, H1T2 and H1T3 do not form a pitch level on their own to the exclusion of the other F0 points: these two points are not significantly different from L1T2, L2T2, and H2T3 [Tukey HSD, $p<0.05$]. That is, those points which were organised into two separate pitch levels (mid-high and mid-low) for females and all the speakers combined are not distinct for male speakers. There were no significant effects of language background on F0 scaling, and there were no significant interactions.

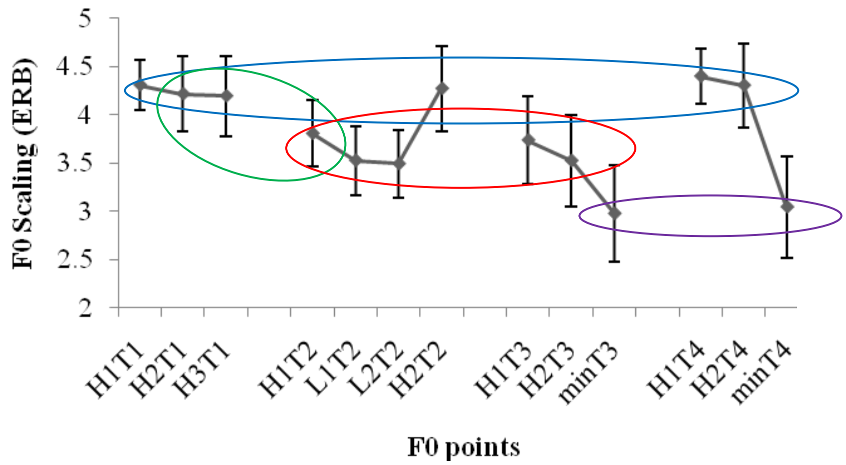


Figure 13: F0 scaling (male speakers).

4.4. Final rise in Tones 3 and 4

Neither gender nor language background had a significant effect on the occurrence of the final rise in Tones 3 and 4 [$p>0.05$]. As both variants could be found in tokens produced by the same speakers, the two variants of these tones appear to be in free variation.

5. Discussion

5.1. Duration

The results show that tone has an effect on syllable duration in SM. The syllable bearing Tone 4 is significantly shorter than each of the other tones, which were not found to be significantly different from one another. Therefore, in terms of syllable duration, SM tones appear to be more similar to BM than TM: like in BM, SM Tone 4 is the shortest among the four tones. However, unlike BM, Tones 1, 2, and 3 of SM are not different in terms of duration. Internally to the variety, E_D speakers showed a trend of producing longer syllables than M_D speakers, i.e. E_D speakers tended to be slower, and perhaps more careful, in their speech. Although not strictly significant, this effect approached levels of significance, and might be detected with data from more speakers.

The lack of temporal difference among the syllables bearing Tones 1, 2, and 3 raises an interesting question: can syllables bearing these tones be reliably distinguished from one another in the absence of F0 cues, such as in whispered speech? In an experiment involving speakers from Northern China, Liu & Samuel (2004) report a correlation between syllable duration and tone perception when F0 information is neutralised. The SM data, however, suggest that syllable duration alone cannot provide the basis for distinguishing Tones 1, 2, and 3. Kong & Zeng (2006) and Chang & Yao (2007) suggest that intensity provides another secondary cue to tone identification in the absence of F0 information. SM seems to provide an ideal testing ground for this hypothesis: since temporal information does not seem to play a role in distinguishing some of the tones, what is the role of intensity in the perception of tones when F0 is neutralised? In addition, although not addressed in the present study, voice quality might also play a role in contributing towards the tonal distinctions.

5.2. F0

As expected, there were gender effects on F0: the F0 range employed by female speakers is both wider and higher than that of male speakers, though in terms of the four levels presented in Table 1 below, the low pitch level of female speakers overlaps with the high and mid-high levels of male speakers. Such gender differences in F0 range are unsurprising, given the physiological differences between males and females that determine their average absolute pitch values (Cruttenden 1997: 3), and the findings here are in line with the results from other languages such as English and Dutch (Daly and Warren 2001; Haan and van Heuven 1999 respectively).

Table 1: Four pitch levels in Singapore Mandarin.

	Pitch level	F0 points
α	Low	minT3, minT4
β	Mid-low	L1T2, L2T2, H2T3
γ	Mid-high	H1T2, H1T3
δ	High	H1T1, H2T1, H3T1, H2T2, H1T4, H2T4

The analysis of the relative scaling of the 13 F0 points reveals that these points can be grouped in four pitch levels, based on statistical difference, as summarised in Table 1. Although the four pitch levels are less distinct for male speakers, this is also accompanied by a narrower F0 range, which compresses the differences between levels. As we saw in Figure 10, male speakers have a compressed F0 range compared to female speakers. This could provide an explanation for the blurred distinctions among the four pitch levels for male speakers: since the F0 range employed by male speakers is narrow, the differences between the levels are smaller, and more data from a larger pool of male speakers would be needed to detect these differences reliably.

Given the organisation of the 13 F0 points into four distinct pitch levels, the contours of Singapore Mandarin tones can be represented in terms of the relative scaling of salient points along the four pitch levels. In these terms, the four tonal contours of SM can be characterised as follows: Tone 1 starts at a high pitch level and remains at this level throughout the syllable; Tone 2 begins at a mid-high level, falls to a mid-low level stretch, before rising to a high pitch level at the end of the syllable; Tone 3 starts at a mid-high level and falls steadily to the low level; Tone 4 starts and remains high through the onset, and then falls to the low level. Tones 3 and 4 also had variants that exhibited a small final rise, and the two variants appear to be in free variation.

Compared with the other varieties, the contours of Tones 1 and 4 show cross-dialectal consistency. The contour of SM Tone 2 is distinct from that in both BM and TM, and appears to uniquely characterise this variety. Unlike BM Tone 2, which falls slightly before showing an early rise, or TM Tone 2, which remains level or drops slightly during the first half of the vowel before rising towards the end, the contour of SM Tone 2 shows a mid-low level stretch not found in the other varieties. SM Tone 3 is more similar to that of TM than BM, where Tone 3 is realised as a dipping tone instead of ending low.

There are two possible onset pitch levels for the four tones in SM: Tones 1 and 4 start at the high level, while Tones 2 and 3 have mid-high level onsets. Likewise, there are two possible offset pitch levels: while Tones 1 and 2 have high offsets, Tones 3 and 4 have low offsets. The organisation of onset and offset pitch levels into three distinct levels closely reflects Shih's description of the tonal contours in TM (Figure 1). Taken together, such cross-dialectal consistency in F0 scaling at the syllable onsets and offsets points towards the representation of the four Mandarin tones in terms of the concatenation of static tonal targets scaled with respect to one another, whereby the contour of each tone is the result of the interpolation between these static targets. As Chao notes, tones are never distinguished by the exact shape of the F0 contour:

> [P]ractically any tone occurring in any of the Chinese dialects can be represented unambiguously by noting the beginning and ending points, and in the case of a circumflex tone, also the turning point; in other words, the exact shape of the time-pitch curve, so far as I have observed, has never been a necessary distinctive feature, given the starting and ending points, or the turning point, if any, on the five-point scale. (Chao 1968: 25)

This observation is corroborated by evidence from a series of experiments on the perception of lexical tone in Thai (Zsiga and Nitisaroj 2007) testing the hypothesis that the presence and alignment of pitch targets provide more consistent cues to tone identification than the overall shape of the tonal contour. To test the effect of the scaling of pitch targets on tone identification, they manipulated the pitch at the mid and end points of the target syllables, and to test of the effect of the alignment of pitch targets on tone identification, they varied the alignment on F0 peaks on the target syllable, while keeping the overall contour shape constant. Speakers were asked to identify the word that they heard by choosing one of five options that differed only in tone. The results of the F0 scaling experiments showed that the mid and end point values of a pitch contour provide more consistent cues to Thai tone identification than the overall F0 slope. Similarly, the results of the peak alignment experiment indicated that the shape and slope of a tonal contour are not salient in the perception of Thai tones: tonal identification changed when the peak alignment was varied, even if the overall shape and slope remained constant. These results support the hypothesis that contour tones are composed of pitch targets aligned with the segmental material, and provide little evidence in favour of a non-compositional analysis of contour tones.

Furthermore, it is not clear how the variation in the plateau duration or the duration of the final rise in Tone 2 can be modelled in a framework that treats contour tones as integral units of pitch movement. Thus, the SM data appear to be more compatible with the target and interpolation view than the treatment of tonal contours as holistic movements. This is consistent with the findings of Zsiga and Nitisaroj (2007) and there is little reason to treat Tones 2 and 4 as integral units of pitch movement that cannot be accounted for by the concatenation of static targets, as advocated by Xu and colleagues. For further confirmation of this, we await perceptual data of the sort that has been presented for Thai.

5.3. Alignment in Tone 2

As noted earlier, the contour of SM Tone 2 is unlike that of both BM and TM. In particular, it exhibits a low plateau that is flanked by two pitch elbows. Notice that while Tone 2 of Singapore Mandarin is realised as a dipping tone, it is Tone 3 of Beijing Mandarin that has a dipping contour. It would be interesting to investigate if speakers of these varieties would confuse the two tones, and if alignment or pitch level or an interaction between the two would play any role in disambiguating these tones.

The analysis revealed systematic patterns of alignment that relate to gender and language background. Proportionally speaking, male speakers align the two elbows further away from each other than female speakers, while the final rise starts earlier for M_D speakers than E_D speakers. Thus, in addition to the obvious F0 differences between male and female speakers that might be expected on physiological grounds, the proportional distance between the two pitch elbows in Tone 2 could also signal gender identity. Likewise, a proportionally earlier rise on Tone 2 could be indexical of M_D speech, in contrast to the later rise in E_D speech. Given the trend for M_D speakers to have shorter syllables than E_D speakers, the overall perceptual difference of the final rise might be quite large.

5.4. Tonal change

The cross-dialectal differences in the realisation of Tone 2 seem to paint a picture of tonal change in action. In BM, there is an early rise. For TM, there is a mid level stretch before the rise. In SM, we see an initial fall, followed by a low plateau, before the final rise. It appears that the further away from the 'source', the later the rise on Tone 2. The language background distinction in Singapore seems to be moving in the same direction, with the final rise occurring later for E_D speakers than M_D speakers.

Similar observations have been reported for Tone 2 in TM. In a recent study, Sanders (2008) found that older speakers tended to produce a rising tone on Tone 2, while younger speakers tended to use a dipping tone. So we see a cross-dialectal trend for Tone 2 moving away from being a rising tone to becoming a dipping tone.

Perhaps what happened in SM is not unlike the great vowel shift in English: When Tone 4 became shorter and distinguishable by duration alone, Tone 3 could lose its final rise and become a falling tone without the risk of creating ambiguity. Since Tone 3 was no longer a dipping tone, speakers then decided to fill that gap by imposing 2 mid-low targets between the start and the end of Tone 2, resulting in a dipping tone. The alignment of these two targets are then manipulated for indexical purposes: the distance between the two pitch elbows indicates gender identity, while the distance between the second pitch elbow and the syllable offset signals the role of Mandarin in a speaker’s repertoire. Or could it be the other way around, such as the tonal push chain suggested by Sanders, where a tone changes when another tone encroaches upon its territory? In order to answer this question, we need cross-generational data from SM.

6. Summary and conclusion

In conclusion, the acoustic study of Singapore Mandarin tone presented here shows some non-trivial differences between the tonal realisations in this variety and those spoken in Beijing and Taiwan. In terms of duration, Singapore Mandarin shows a one-way split between Tone 4 and the other tones not found in the other varieties. With respect to tonal contours, the contour of Tone 3 in Singapore Mandarin is more similar to that of Taiwan Mandarin than Beijing Mandarin, while the contour of Tone 2 is unique to Singapore Mandarin in displaying a low level stretch not found in the other varieties. Internally to the Singapore Mandarin system, there are clear gender-related differences in the scaling and alignment of pitch targets, as well as alignment differences that relate to the speakers’ language backgrounds, suggesting that tonal realisation can be used for indexical purposes. In turn, these types of differences strongly suggest that taking social variation into account, especially in a multilingual context like Singapore, is essential for the adequate description and understanding of grammatical systems, including tone.

References

Arvaniti, Amalia. 2007. On the relationship between phonology and phonetics (or why phonetics is not phonology). *Proceedings of the XVI^th International Congress of Phonetic Sciences*: 19-24.

Arvaniti, Amalia and D. Robert Ladd. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26: 43-74.

Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen. 2006. Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication*, 48: 667-696.

Boersma, Paul. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 17: 97-110.

Boersma, Paul and David Weenik. 2003. Praat 4.19: Doing phonetics by computer. (<http://www.praat.org>).

Chang, Charles and Yao Yao. 2007. Tone production in whispered Mandarin. *Proceedings of the XVI^th International Congress of Phonetic Sciences*: 1085-1088.

Chao, Yuen-Ren. 1930. A system of tone letters. *Le Maitre Phonetique*, 30: 24-27. [Reprinted in *Fangyan* (1980), 2: 81-83.]

Chao, Yuen-Ren. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In Morris Halle (ed.), *For Roman Jakobson*, Mouton: The Hague: 53-59.

Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, Chung Yu. 1983. A fifth tone in the Mandarin spoken in Singapore. *Journal of Chinese Linguistics*, 11.1: 92-119. [Reprinted in Chung Yu Chen (ed.). 1993. *Aspects of Mandarin Chinese*, Singapore: The Commercial Press Ltd.: 249-272.]

Chua, Chee Lay. 2003. The emergence of Singapore Mandarin: A case study of language contact. PhD dissertation, University of Wisconsin-Madison.

Cruttenden, Alan. 1997. *Intonation* (Second edition). Cambridge: Cambridge University Press.

Daly, Nicola and Paul Warren. 2001. Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics*, 5: 85-96.

Del Giudice, Alex, Ryan Shosted, Katherine Davidson, Mohammad Salihie and Amalia Arvaniti. 2007. Comparing methods for locating pitch "elbows". *Proceedings of the XVI^th International Congress of Phonetic Sciences*: 1117-1120.

Fon, Janice and Wen-Yu Chiang. 1999. What does Chao have to say about tones? – a case study of Taiwan Mandarin. *Journal of Chinese Linguistics*, 27: 15-37.

Haan, Judith and Vincent J. van Heuven. 1999. Male vs. female pitch range in Dutch questions. *Proceedings of the XIII^th International Congress of Phonetic Sciences*: 1581-1584.

Hermes, Dik and Joost van Gestel. 1991. The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, 90: 97-102.

Ho, Lui Choon. 2000. 声调的不稳定性在新加坡华语里不同年龄层的表现 [A survey of tonal instability in Singapore Mandarin over a spectrum of ages]. ms. National University of Singapore.

Kong, Ying-Yee and Fan-Gang Zeng. 2006. Temporal and spectral cues in Mandarin tone recognition. *Journal of the Acoustical Society of America*, 120: 2830-2840.

Ladd, D. Robert. 2008. *Intonational Phonology*. (Second edition). Cambridge: Cambridge University Press.

Leow, Bee Geok. 2001. *Census population 2000 statistical: release 2: education, language, and religion*. Singapore: Department of Statistics, Ministry of Trade and Industry.

Liu, Siyun and Arthur G. Samuel. 2004. Perception of Mandarin Lexical Tones when F0 information is Neutralized. *Language and Speech*, 47: 109-138.

Peng, Shu-hui, Marjorie K. M. Chan, Chiu-yu Tseng, Tsan Huang, Ok Joo Lee, and Mary E. Beckman. 2005. Towards a Pan-Mandarin System for Prosodic Transcription. In Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, New York: Oxford University Press: 230-270.

Sanders, Robert. 2008. Tonetic Sound Change in Taiwan Mandarin: The Case of Tone 2 and Tone 3 Citation Contours. *Proceedings of the 20th North American Conference on Chinese Linguistics*: 87-107.

Shih, Chilin. 1988. Tone and Intonation in Mandarin. In Nick Clements (ed.), *Working Papers of the Cornell Phonetics Laboratory*, 3, CLC Publications: 83-109.

Shih, Chilin, Greg Kochanski and Su-Youn Yoon. 2007. The missing link between articulatory gestures and sentence planning. *Proceedings of the XVIth International Congress of Phonetic Sciences*: 35-38.

Xu, Yi. 1994. Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, 95: 2240-2253.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25: 61-83.

Xu, Yi. 1998. Consistency of Tone-Syllable Alignment across Different Syllable Structures and Speaking Rates. *Phonetica,* 55: 179-203.

Xu, Yi and Q. Emily Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33: 319-337.

Xu, Yi and Ching X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33: 159-197.

Zsiga, Elizabeth and Rattima Nitisaroj. 2007. Tone Features, Tone Perception, and Peak Alignment in Thai. *Language and Speech*, 50: 343-383.