1. Introduction
Mandarin is spoken as a first or second language by well over a billion people across the globe, and it includes many regional varieties, including three national standards: Guoyu ‘national language’ in Taiwan, Putonghua ‘common speech’ in Mainland China, and Huayu ‘Chinese language’ in Singapore (Peng et al. 2005: 230). However, most acoustic studies of tone in Mandarin have focussed on the Beijing and Taiwan varieties of the language (Xu 1994, 1997, 1998; Shih 1988; Fon and Chiang 1999), and Singapore Mandarin has received considerably less attention in comparison to these varieties.

Mandarin is one of the four official languages spoken in Singapore, alongside English, Malay, and Tamil. Malay is, for historical reasons, accorded official status as the national language; however, English is the language of education, work, and administration. Even though English is the medium of instruction in schools, every child, depending on his/her ethnicity, is also required to learn one of the three other official languages in primary (six years) and secondary school (four or five years). Therefore, regardless of the language(s) spoken at home, most Singaporeans are bilingual in at least two of the official languages. Among the ethnic Chinese, who make up 76.8% of the resident population in Singapore, 23.9% use English most frequently at home, 45.1% use mainly Mandarin at home, while 30.7% predominantly use other Chinese languages at home (Leow 2001:viii-ix). These other Chinese languages include Hokkien, Teochew, and Cantonese, reflecting the fact that most of the ethnic Chinese population are descendents of immigrants from the Min and Yue regions of Southern China. Given the sociolinguistic situation of bilingualism, the potential influence of other tonal Chinese languages spoken on the island, 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 Mandarin. As published studies of the tones in Singapore Mandarin are wanting in the literature, it would be interesting to document this aspect of the language, and compare it with the other varieties.

The aim of this paper, then, is to present an acoustic study of the tones in Singapore Mandarin, and compare the results with the other standard varieties of Mandarin, focussing only on the four lexical tones and leaving aside the issue of the ‘neutral tone’ (see Peng et al. 2005 and references therein). 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 the realisation of tone in Singapore Mandarin.

2. Background
2.1. The transcription of Mandarin tones
Researchers working on Mandarin tone typically describe the four tones using what is commonly referred to as the ‘Chao tone letter system’, a five point tonal representation system established by Chao (1930/1980), where 1 represents the lowest pitch value in a speaker’s speaking pitch range and 5 the highest. Tones are transcribed by noting the beginning and ending points, and also the turning point in the case of a circumflex tone. Following Chao’s (1956, 1968) transcription of Peiping (or Beijing) Mandarin, the four tones are commonly represented as shown below (for the sake of consistency, all transcriptions in this paper are provided in pinyin, albeit without the tone diacritics, which have been changed to tone numbers (1, 2, 3, 4) indicated after the syllable):

<table>
<thead>
<tr>
<th>Tone #</th>
<th>Pitch pattern</th>
<th>Chao tone letters</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>high level</td>
<td>55</td>
<td>ma1 ‘mother’</td>
</tr>
<tr>
<td>2</td>
<td>high rising</td>
<td>35</td>
<td>ma2 ‘hemp’</td>
</tr>
<tr>
<td>3</td>
<td>low falling-rising</td>
<td>214</td>
<td>ma3 ‘horse’</td>
</tr>
<tr>
<td>4</td>
<td>high falling-rising</td>
<td>51</td>
<td>ma4 ‘to scold’</td>
</tr>
</tbody>
</table>
Some of the tones display contextual variation when followed by another tone. When followed by another tone in non-final position, Tone 3 is realised as a low falling tone [21] without the final rise, and Tone 4 displays a smaller fall, being realised as [53] rather than [51]. These changes are often referred to as the Tone 3 and Tone 4 reduction respectively. When Tone 3 is followed by another Tone 3, the first Tone 3 in the sequence changes into a Tone 2, reflecting what is commonly known as the Tone 3 sandhi in Mandarin.

Given the wide geographical distribution of Mandarin speakers, one would expect to find variation in the realisation of the four tones. Such variation has indeed been reported in the literature. As first observed by Tai (1978), even the citation form of Tone 3 tends to be produced without the final rise as [21] in Guoyu, rather than as the full dipping tone [214]. Shih (1988) reports the same difference: “Southern [Chinese] speakers often keep the low-falling pattern even in the final position in casual speech, and use the falling-rising pattern only in deliberate, emphatic speech, or in yes-no questions. Northern [Chinese] speakers often use the falling-rising pattern sentence finally in all speech acts” (p. 84). Fon and Chiang (henceforth F&C) (1999) also suggest two differences between Putonghua and Guoyu: while a Tone 3 syllable in its full [214] form is the longest of all citation form types in Putonghua, the corresponding tone in Guoyu is shorter than Tone 2; also, whereas Tone 2 is typically a rising tone in Putonghua, it is realised as a dipping tone in Guoyu, which can be transcribed as [323]. Clearly, the two varieties do not share the same tonal realisations, and differences can be found both in terms of duration and tonal contours.

Even though most researchers of Mandarin tone use the tone letter system developed by Chao, in most cases, it is not clear how the tone letters are assigned. As noted by F&C (1999), most acoustic studies on tone often assume a linear five point scale with four equi-intervals in between: the lowest pitch is subtracted from the highest and the interval is then divided by four. This, F&C argue, does not seem to have been Chao’s intention. In two separate papers, Chao outlines the relationship between the five reference points and the musical scale. In Chao (1956), he picked E, F, bA, bB, and C as the five reference points, while in Chao (1968), F, G, bB, C, and D were chosen instead. These relationships are illustrated in Figure 1.

Figure 1: Chao’s tone letters and the musical scale (F&C 1999: 16)

As F&C note, Chao adjusted the reference points so that the interval between points 1 and 2 was narrowed down to one semitone, while the interval between points 2 and 3 was enlarged to three semitones. This is shown in Figure 2.

---

1 It is not clear how this can be done, since F0 is continuous. It is more likely to have five equi-intervals, such that each of the five reference points in Chao’s system is assigned to an interval.

2 Sarah Creel points out that there are discrepancies between Chao’s choices of the five representation points on the musical scale on the one hand (Figure 1), and F&C’s interpretation of the tonal scale-semitone relation (Figure 2) on the other hand. Both figures were reproduced from F&C (1999).
F&C provide formulae for converting acoustic data into each of Chao’s systems:

(2) Formulae for converting acoustic data into Chao tone letters (F&C 1999: 17-18)

\[
N (\text{Semitone}) = 39.86 \times (\log f_i - \log f_{\text{min}}) \quad \text{------------------------ (Function 1)}
\]

\[
\text{Scale} \# = \frac{N}{2} + 1 \quad \text{----------------------------- (Function 2)}
\]

\[
\text{Scale} \# = \frac{N}{2} + 1 \quad \text{(if } N \geq 3) \quad \text{----------------- (Function 3)}
\]

\[
\text{Scale} \# = \text{Scale} 3 \quad \text{(if } 2.5 \leq N < 3) \quad \text{----------------- (Function 4)}
\]

\[
\text{Scale} \# = \text{Scale} 2 \quad \text{(if } 0.5 \leq N < 2.5) \quad \text{----------------- (Function 5)}
\]

\[
\text{Scale} \# = \text{Scale} 1 \quad \text{(if } 0 \leq N < 0.5) \quad \text{----------------- (Function 6)}
\]

For both versions, Function 1 is first employed to transform the acoustic data from Hz to semitones. After the application of Function 1, Function 2 is applied in order to translate the data into the 1956 system. For the 1968 version, Functions 3-6 are applied to the output of Function 1. F&C then tested the two systems against experimental data of Taiwan Mandarin, recorded from a female speaker who read a word list consisting of trisyllabic words of all possible tonal combinations, as well as those syllables in isolation. They measured salient points in each tone: For Tone 1, the beginning and ending points were chosen; For Tones 2 and 3, they measured the initial highest, the medial lowest, and the ending highest points; For Tone 4, the initial highest and final lowest points were measured. The medial of Tone 3 was found to be the lowest of the ten points in the four tones, and it was used as the reference point, \( f_{\text{min}} \), in Function 1. Comparing the values obtained by each set of formulae, they concluded that Chao’s second version (1968) provides more consistent representations of Taiwan Mandarin tones. Regardless of which version of Chao’s system one adopts, the tonal transcriptions of Taiwan Mandarin clearly differ from those proposed by Chao for Beijing Mandarin.

2.2. Acoustic studies

In another study of Taiwan Mandarin tones, Shih (1988) presents an acoustic study of the tones produced by a female speaker both in isolation and in context. Using a set of minimal pairs with the syllable /ma/: ma1 ‘mother’, ma2 ‘hemp’, ma3 ‘horse’, and ma4 ‘to scold’, she plots the tones produced in isolation, reproduced in Figure 3. She describes the four tones as follows:

A high level tone, or tone 1, starts in a speaker’s high pitch range and remains high. As the name implies, there is no drastic pitch movement except a slight dip in the middle of the vowel, and a slight rise toward the end of the syllable. A rising tone, or tone 2, starts at the speaker’s mid pitch range, remains level or drops slightly during the first half of the vowel, and rises up to high at the end. A low tone, or tone 3, is phonetically a low falling tone. It starts at the speaker’s mid range and falls to the low range. It is often accompanied by laryngealization over the second half of the syllable. A falling tone, or tone 4, usually peaks around the vowel onset, then falls to the low pitch range at the end.

(Shih 1988:83)
She notes that the beginning and end points of all tones fall on three distinct levels rather than being scattered across a continuum: tones 1 and 4 both start high, very 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 her observations of the tonal contours, she proposes the summary of the relative values and the placement of tonal targets for each tone shown in Figure 4.

H+ represents a pitch level slightly higher than H; L- is lower than L. Tones 1 and 3 have targets at the C/V boundaries, while Tone 4 has a slightly delayed target. Tone 2 does not have a target until the middle of the vowel, which ensures a relatively late rising contour. Shih includes the tonal targets at the beginning of the consonantal region in parentheses only to account for the tone shapes of the consonant region in isolation and in sentence initial position; she suggests that in non-initial position, the consonantal region is where the tonal transition occurs, and that there is little evidence for the existence of a real target.

The Beijing Mandarin tones seem to behave differently from those of Taiwan Mandarin. Xu (1997) presents data from Beijing Mandarin, where he examines the tones produced by eight male speakers both in isolation and in different tonal contexts. The tones in isolation were elicited using a set of minimal pairs with the syllable /ma/. A program was used to record the F0 values at regular intervals within each segment, and F0 curves obtained were smoothed using a function incorporated into the program. The mean F0 contours were plotted over normalised time, as represented in Figure 5, and he describes the tones as follows:

Tone 1 starts with a high f0 value (near 130Hz) and stays around that level throughout the syllable. Tone 2 starts with a low f0 (near 100Hz), then falls slightly before rising (starting at 20% into the
vowel) throughout the remainder of the syllable. Tone 3 starts with an $f_0$ value slightly lower than the onset of Tone 2, falls to the lowest $f_0$ of all the four tones (about 90Hz) right at the vowel midpoint, then rises sharply to the end of the syllable. Tone 4 starts with the highest $f_0$ value of the four tones (140Hz), continues to rise before reaching the maximum about one fifth of the way into the vowel, then falls sharply to the end of the syllable. (Xu 1997: 67)

Figure 5: Beijing Mandarin tones in isolation (Xu 1997: 67).

The four tones differ in their durations. In descending order, Tone 3 is the longest, followed by Tone 2, Tone 1, and Tone 4.

The data highlight some differences between Taiwan and Beijing Mandarin. In terms of tonal contours, Tones 2 and 3 seem to exhibit different contours in Shih’s and Xu’s studies: the final rise on Tone 2 appears to be later in Taiwan Mandarin than in Beijing Mandarin, while Tone 3 in Beijing Mandarin displays a final rise that does not occur in Taiwan Mandarin. In terms of duration, the two varieties are also different: while relative durations of Beijing Mandarin tones have a descending order of Tone 3 > Tone 2 > Tone 1 > Tone 4, Taiwan Mandarin tones show a descending order of Tone 2 > Tone 4 > Tone 1 > Tone 3.

In another study, Xu (1998) examines the issue of tonal alignment by comparing tone-syllable alignment in syllables with a final nasal to those without a final nasal across three speaking rates. In a first experiment, Xu 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 across speaking rates. For each of his four Beijing Mandarin speakers, Xu computed a mean F0 contour for each labelled segment by taking a fixed number of F0 points at equal time intervals and averaging them across five repetitions. He 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: the F0 contour is distributed between the vowel and the coda when there is a final nasal in the syllable. In another experiment, Xu 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, indicating that syllable structure does not affect F0 contour alignment. He also 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: the low and high points are not distributed evenly throughout the syllable, and the entire contour shifts towards the later portion of the syllable when the duration of the host syllable increases. Xu interprets the results as showing 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 two static targets.

Based on Xu’s work on Mandarin tones, Xu and Wang (2001) outline a framework for accounting for surface F0 variations in speech. In this framework, 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]” (p321). In this framework, 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. Therefore, the contour tones, Tone 2 and 4, are not considered to be formed by the concatenation of two static targets, as assumed by Shih (1988). Rather, these are treated as realisations of integral units of pitch movement. Although pitch peaks are not included in the framework proper, Xu and Wang claim that it can nonetheless make predictions about the location of F0 peaks: “The occurrence and location of F0 peaks may be predicted by (a) the property of the pitch target, (b) the properties of the adjacent peak targets, and (c) the duration of the host” (p331). Specifically, they suggest that their framework predicts two circumstances under which an F0 peak may be delayed beyond the offset of the host (peak delay): “(a) when the pitch target is [rise] and is followed by a target with a low pitch onset, and (b) when the pitch target is [high] and surrounded by low pitch values and the duration of the host is sufficiently short” (p332). Circumstance (a) would be instantiated by either a Tone 2 – Tone 3 or Tone 2 – Tone 2 sequence in Mandarin, while (b) would be realised by a sequence of Tone 1 – Tone 3, Tone 1- Tone 2, Tone 3 – Tone 1, or Tone 4 – Tone 1. Xu and Wang propose several implementation rules for pitch targets: (i) “[a] pitch target is implemented in synchrony with the host, i.e., starting at its onset and ending at its offset”; (ii) “the approximation of the pitch target is continuous and asymptotic” throughout the duration of the host; and (iii) “[a] falling F0 movement is implemented faster than a rising movement” (p322). An implication of (i) and (ii) is that “When two pitch targets occur next to each other, if the offset of the first one is different from the onset of the second one, the second one will appear as if it has been assimilated or partially assimilated to the first one” (p329). Since the domain of F0 contour alignment in Mandarin is assumed to be the syllable, Xu and Wang’s framework predicts the effects of carryover assimilation reported in Xu (1997). However, given (ii), it is not clear if the anticipatory effects can be accommodated by this framework (see Shih et. al. 2007 and references therein for more evidence of anticipatory effects of tonal coarticulation in Mandarin). In an extension of the framework beyond Mandarin lexical tones, Xu and Xu (2005) argue that Xu and Wang’s model applies also to intonational languages such as English.

In contrast with Xu and colleagues’ assumptions that F0 falls and rises in speech are represented by integral dynamic targets and that the F0 of each syllable is specified, 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 and Arvaniti and Ladd 2009 for the comparative evaluation of phonetic models of F0). Work within AM such as Arvaniti et. al. (2006) and Arvaniti and Ladd (2009) thus share with Shih (1988) the intuition that complex tonal contours are composed of smaller units aligned with segments. However, since AM focuses on the representation of intonation instead of lexical tone, not all syllables are tonally specified, as Shih assumes; in this framework, only a few syllables are tonally specified and the F0 of the rest is derived by interpolation between targets.

In a study of intonation in Greek polar questions, Arvaniti et. al. (2006) present data that strongly suggest a close alignment between tonal targets and the segmental string. By systematically manipulating the position of the focussed word and the position of the last stressed syllable, they showed that the focussed word in a Greek polar question is always associated with a low plateau, and such questions show a final rise-fall movement whose peak is aligned closely with segmental material: when the focussed word is utterance-final, the peak of the rise-fall co-occurred with the utterance-final vowel; when the focussed word occurs earlier in the utterance, the peak co-occurred with the stressed vowel of the last word. However, when tones were crowded by surrounding tones, tonal alignment was adjusted such that a strict alignment relationship cannot be held between targets and segments or syllables: when crowded by other tones, the position of F0 peaks is not consistently aligned either with segments or syllables. In a separate study of Greek wh-questions, Arvaniti and Ladd (2009) present similar evidence of adjustments for tonal crowding that affect the realisation of intonation contours. Their data showed that the exact intonational contour varies according to the length of the question, making it unlikely for the F0 of all
syllables to be specified. Further, the exact alignment of tonal targets depended on the tonal context, suggesting that the phonetic alignment of tones and segmental material can vary within limits. The reported adjustments in tonal alignment suggest effects from both preceding and upcoming tones, and while both the anticipatory and carry-over effects of coarticulation can be easily accounted for in the target and interpolation model of AM, such data cannot be easily accommodated by a model in which syllables are the strict domain of the implementation of dynamic pitch targets.

In a series of experiments on the perception of lexical tone in Thai, Zsiga and Nitisaroj (2007) test 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, which were presented to Thai speakers both as citation forms and in connected speech. In another 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, from a lexical tone language other than Mandarin, 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. However, the duration of the target syllable was held constant in these experiments, and it the role of duration in Thai tone perception, if any, is not clear.

### 2.3. Singapore Mandarin

Unlike Beijing and Taiwan Mandarin, the variety of Mandarin spoken in Singapore has been less carefully studied. There are few published reports on Singapore Mandarin, and to the best of my knowledge, no controlled acoustic survey of tone in Singapore Mandarin had been conducted prior to the present study.

Chua (2003) draws a sociolectal distinction between “Singapore Standard Mandarin” (SSM) and “Singaporean Mandarin” (SgM), and he 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 comfortably into ‘lower’ sociolects outside the context of his or her own sociolect for functional purposes. Many Singaporean Chinese who use the acrolectal sub-variety in lectures, debates and formal discussion use a mesolectal sub-variety in more formal shopping and other transaction situations or in informal talk to strangers but they can also drop comfortably and naturally into the basilectal sub-variety when, for example, talking with friends or colleagues in a coffee shop or in a market place. (Chua 2003: 42-43)

Chua provides the following transcription of the four tones in the two sub-varieties:

<table>
<thead>
<tr>
<th></th>
<th>SSM</th>
<th></th>
<th>SgM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tone 1</strong></td>
<td>55</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td><strong>Tone 2</strong></td>
<td>35</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td><strong>Tone 3</strong></td>
<td>214/211</td>
<td>211</td>
<td></td>
</tr>
<tr>
<td><strong>Tone 4</strong></td>
<td>51</td>
<td>41/42</td>
<td></td>
</tr>
</tbody>
</table>
Chua’s characterisation of the tones in SSM is generally in accordance with the values provided by Chao for Beijing Mandarin, with the exception of Tone 3, which Chua observes can be realised without the final rise even in isolation, as in Taiwan Mandarin. The pitch transcriptions provided for SgM correspond to Chen’s (1983/1993)\(^3\) observations that Tones 1, 2, and 4 often appear to be lower in Singapore Mandarin than 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 did not specify how the numbers on the tonal scale had been assigned to the tones. Chua also notes that the Tone 3 sandhi occurs in both SgM and SSM: when a Tone 3 occurs before another Tone 3 syllable, it changes to a Tone 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.

3. Methods

3.1. Speakers

This study is based on the acoustic analysis of speech materials read aloud by 16 speakers of Singapore Mandarin, 8 male and 8 female. Table 2 summarises the linguistic profiles of the 16 speakers.

<table>
<thead>
<tr>
<th>Gender</th>
<th>L1/L2</th>
<th>Languages and frequency of use in daily life</th>
<th>Languages and frequency of use at home</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>L1</td>
<td>E (50%), M (50%)</td>
<td>M (100%)</td>
</tr>
<tr>
<td>F</td>
<td>L1</td>
<td>E (50%), M (40%), T (10%)</td>
<td>M (90%), T (10%)</td>
</tr>
<tr>
<td>F</td>
<td>L1</td>
<td>E (50%), M (50%)</td>
<td>M (70%), E (30%)</td>
</tr>
<tr>
<td>F</td>
<td>L2</td>
<td>E (70%), M (30%)</td>
<td>E (60%), M (40%)</td>
</tr>
<tr>
<td>F</td>
<td>L2</td>
<td>E (90%), M (10%)</td>
<td>E (90%), M (10%)</td>
</tr>
<tr>
<td>F</td>
<td>L2</td>
<td>E (70%), M (30%)</td>
<td>E (70%), M (30%)</td>
</tr>
<tr>
<td>F</td>
<td>L2</td>
<td>E (80%), M (20%)</td>
<td>E (80%), M (15%), C (5%)</td>
</tr>
<tr>
<td>M</td>
<td>L1</td>
<td>E (60%), M (40%)</td>
<td>M (100%)</td>
</tr>
<tr>
<td>M</td>
<td>L1</td>
<td>E (80%), M (20%)</td>
<td>M (70%), E (30%)</td>
</tr>
<tr>
<td>M</td>
<td>L1</td>
<td>E (50%), M (50%)</td>
<td>M (90%), T (5%), H (5%)</td>
</tr>
<tr>
<td>M</td>
<td>L2</td>
<td>E (90%), M (10%)</td>
<td>E (100%)</td>
</tr>
<tr>
<td>M</td>
<td>L2</td>
<td>E (95%), M (5%)</td>
<td>E (100%)</td>
</tr>
<tr>
<td>M</td>
<td>L2</td>
<td>E (70%), M (30%)</td>
<td>E (80%), M (20%)</td>
</tr>
<tr>
<td>M</td>
<td>L2</td>
<td>E (70%), M (25%), S (5%)</td>
<td>E (100%)</td>
</tr>
</tbody>
</table>

The speakers were ethnic Chinese, and were third generation Singaporeans or beyond (i.e. they and their parents had been born and raised in Singapore). All speakers were in their twenties at the time of recording, 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.

\(^3\) Chen also reports a fifth tone in Singapore Mandarin, which she claims to occur only in limited distributions. However, she also notes that both Tone 4 and the alleged 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 even researchers who claim that this fifth tone exists cannot reliably distinguish between the purported fifth tone and Tone 4, and that the use of this fifth tone, if it exists, seems to be in decline, I will ignore the issue of the fifth tone in this paper.
as reported by the speakers: the first category consisted of speakers whose dominant home language was Mandarin, while the second category consisted of speakers whose dominant home language was English. For the sake of convenience, I shall refer to the first category of speakers as first language (L1) speakers of Mandarin, and the second category of speakers as second language (L2) speakers of Mandarin. All speakers, L1 and L2, 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 reported or was known to have any speech or hearing impairment.

3.2. Materials
In order to establish the canonical forms of each of the four Singapore Mandarin tones, an experiment was conducted to elicit the tones in isolation. The test words were the same as those used in Shih (1988) and Xu (1997) and consisted of the syllable /ma/ with the four Mandarin lexical tones:

<table>
<thead>
<tr>
<th></th>
<th>Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>ma1</td>
<td>‘mother’</td>
</tr>
<tr>
<td>ma2</td>
<td>‘hemp’</td>
</tr>
<tr>
<td>ma3</td>
<td>‘horse’</td>
</tr>
<tr>
<td>ma4</td>
<td>‘scold’</td>
</tr>
</tbody>
</table>

The initial nasal consonant /m/ ensures an unbroken F0 contour throughout each syllable that avoids microprosodic perturbations, and the distinct nasal formant structure allows for easier segmentation of the consonant-vowel boundaries. Using the same test words as those used in previous studies of other dialects also allows for direct comparison with the results from those studies. In order to elicit the 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 flash card in Chinese characters as a dialogue. Speakers were only required to read the target word aloud. A sample dialogue is shown below:

(6) **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 flash cards (4 conditions x 5 cards), and the order of their appearance as randomised.

3.3. Procedure
The recordings were conducted in Singapore over the summer of 2008. 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, with 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 analysis.

3.4. Measurements
The selected tokens were segmented in Praat (Boersma and Weenik 2003) by visual inspection of the spectrograms: the beginning of the /m/ was located at the start of the first formant; the /m/ and /a/ boundary was located at the transition between the nasal and the vowel; while the vowel offset was located at the point where the first and second formants could no longer be identified clearly. 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, and F0 curves obtained were smoothed using a function incorporated into the program. By plotting the mean F0 contours were over normalised time, we then have a picture of the overall configuration of the pitch shapes of the four tones. One possible drawback of this approach is that the F0 measurements obtained might not be actual F0 points corresponding to the output of the pitch extraction algorithm, but the result of interpolation between these points. Furthermore, as Chao notes, tones are never distinguished by the exact shape of the F0 contour:

> practically 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 share 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 Thai (Zsiga and Nitisaroj 2007) that the scaling and alignment of pitch targets provide more consistent cues for tone identification than the overall shape of the tone contour. For these reasons, specific points in the F0 contours are also measured, using the second method.

In the second method, pitch tracks were obtained for each token using the autocorrelation method in Praat that gives F0 values in Hz every 10ms (Boersma 1993). In order to avoid pitch-halving or pitch-doubling, the pitch settings in Praat were adjusted for each speaker to accommodate his/her pitch range. In selecting F0 points for measurement, 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):

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

This was done because data was collected from both male and female speakers, and this scale is standardly used for the purpose of comparing speech across genders (e.g. Daly and Warren 2001), since the ERB scale gives a very good approximation of the sensation of pitch, as opposed to the physical measurement of F0 in Hz (Hermes and van Gestel 1991).

For Tone 1, which is 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). A sample analysis is shown in Figure 6(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). The elbows were located on the basis of the author’s impressionistic judgement of the point where the transition occurred, and as Del Giudice et al. (2007) have shown, human intuition correlates well with the most robust algorithmic methods for locating pitch elbows. A sample analysis is shown in Figure 6(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. 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). Sample analyses are shown in Figures 6(c) and 6(d).
For each point measured, the speaker mean was derived by averaging the measurements of that point across the three tokens produced by that speaker, for the purposes of statistical analyses. The F0 range of each speaker was calculated by subtracting the highest F0 measure (averaged across the three tokens) from the lowest F0 measure (averaged across the three tokens) for that speaker.

### 3.5. Statistical analyses

The analysis of duration included three repeated measures ANOVAs with four levels (corresponding to the four tones) were conducted, with the duration of /m/, the duration of /a/, and the duration of /ma/ as the respective dependent variables, 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: (i) the absolute duration (ms) between the syllable onset and L1T2 (the initial fall in Tone 2); (ii) the absolute duration (ms) between the vowel onset and L1T2; (iii) the absolute duration (ms) between L1T2 and L2T2 (the low plateau in Tone 2); (iv) the absolute duration (ms) between L2T2 and the vowel offset (the final rise in Tone 2); (v) the duration between the syllable onset and L1T2 (initial fall) expressed as a percentage of total syllable duration; (vi) the duration between L1T2 and L2T2 (low plateau) expressed as a percentage of total syllable duration; and (vii) the duration between L2T2 and the vowel offset (final rise) expressed as a percentage of total syllable duration.

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: H1T1, H2T1, H3T1, H1T2, L1T2, L2T2, H2T2, H1T3, H2T3, minT3, H1T4, H2T4, minT4) 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, with gender and language background as the respective categorical variables. These were to investigate if there was a correlation between either gender or language background and the occurrence of a final rise on these tones.

### 3.6. Chao tone letters
Three methods were used to convert the F0 data, averaged across all speakers, into Chao tone letters. The first two were based on the formulae of Fon and Chiang (1999) given in (2), which convert the acoustic data into Chao’s two systems. As the outputs of the formulae were not whole numbers, the figures were rounded off to the closest whole number (Fon, p.c.). In the third method, the F0 data were first converted into ERB, and the F0 range was then calculated by subtracting the highest F0 measure from the lowest. This range was divided into five equi-intervals, and each tone letter was assigned to each of the intervals, with 1 being assigned to the lowest interval, and 5 the highest. The 13 F0 points were then assigned a tone letter, depending on which interval their values fell within, correct to two decimal places.

4. Results
The plots produced from the method following Xu (1997) are shown in Figures 7 and 8. Figure 7 shows the aggregate mean tone contours for all speakers, while each panel in Figure 8 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.

Figure 7: Aggregate mean tone contours for all speakers.

![Figure 7](image1.png)

Figure 8: Aggregate mean tone contours for the four speaker groups.
(a) L1 Female
(b) L1 Male

![Figure 8](image2.png)
4.1. Duration
There is a significant effect of tone on the durations of the onset, vowel, and syllable. The durations of each of the four tones are illustrated in Figure 9.

Figure 9: Duration of the four tones. Error bars indicate standard deviation of syllable durations.

Tone had a significant effect on the duration of the syllable \( F(3,36) = 40.256, p<0.001 \). The results of a 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 are not significantly different from one another \( [p>0.05] \). There were no significant effects of gender \( [F<1] \) or language background \( [F(1,12) = 3.964, p>0.05] \) on the duration the syllable, and there were no significant interactions (gender x language background \( [F<1] \); tone x gender \( [F<1] \); tone x language background \( [F(3,36) = 2.622, p>0.05] \); tone x gender x language background \( [F<1] \)).

There was also a significant effect of tone on the duration of the onset \( [F(3,36) = 5.151, p<0.005] \). The results of a Tukey HSD test indicated that the duration of the /m/ of Tone 3 was significantly longer than each of the other three tones \( [p<0.018] \), but the other three tones are not significantly different from one another \( [p>0.05] \). There were no significant effects of gender \( [F<1] \) or language background \( [F(1,12) = 1.945, p>0.05] \), and there were no significant interactions \( [F<1] \).

Despite the longer onset in Tone 3, the differences in syllable duration largely reflect the differences in vowel duration, which also showed a significant effect of tone \( [F(3,36) = 47.154, p<0.001] \).
The results of a Tukey HSD test showed that, just as in the case of syllable duration, the vowel of the syllable bearing Tone 4 was significantly shorter than that in each of the other tones \( p<0.001 \), but the other three tones are not significantly different from one another \( p>0.05 \). Thus, syllable duration appears to be affected more by the duration of the vowel than the onset: even though the onset was found to be longer in Tone 3 than the other tones, the effect of this difference on overall syllable duration was cancelled out by the shorter vowel in Tone 4.

There was also an interaction of tone and language background on the duration of the vowel \( F(3,36) = 3.097, p<0.039 \), although language background itself had no significant effect on vowel duration \( F(1,12) = 3.379, p=0.091 \). This interaction is illustrated in Figure 10. Investigation of this interaction by means of a Tukey HSD test showed the following: Within each language background group, the main effect of tone holds: the vowel in Tone 4 is significantly shorter than that in the other tones \( p<0.05 \) for all pairwise comparisons]. For each tone, the lack of a language background effect also holds: the vowel durations between the two language background groups are not significantly different \( p>0.05 \) for all pairwise comparisons]. However, if we compare across language backgrounds and tones, the vowel in Tone 4 of L1 speakers is shorter than the vowels of Tones 1, 2, and 3 of L2 speakers, but the vowel of Tone 4 of L2 speakers is only shorter than the vowel of Tone 3 of L1 speakers \( \text{L1T4 vs. L2T1 } p<0.001; \text{L1T4 vs. L2T2 } p<0.001; \text{L1T4 vs. L2T3 } p<0.003; \text{L2T4 vs. L1T1 } p>0.05; \text{L2T4 vs. L1T2 } p>0.05; \text{L2T4 vs. L1T3 } p<0.048 \).

**Figure 10:** Language background differences in vowel duration. Error bars indicate standard deviations.

![Figure 10](image)

This interaction did not carry over to the overall syllable duration, but both the effects of language background \( p=0.07 \) and the interaction of language background and tone \( p=0.065 \) on syllable duration seemed to approach significance levels. Given the relatively small sample of speakers and the low p-values obtained, one might expect the effects of language background and its interaction with tone on syllable duration to appear with data from more speakers, and the current results should be treated with caution.

There were no significant effects of gender \( F<1 \) or language background \( F(1,12) = 3.379, p>0.05 \) on vowel duration, and the other interactions were not significant \( F<1 \).

### 4.2. Alignment in Tone 2

Table 3 provides a summary of the absolute duration measures of alignment in Tone 2. Given the high standard deviations, there appears to be a fair amount of variability in the alignment of the pitch elbows in Tone 2, which is not as strict as the alignment of pitch elbows found in intonation languages such as Greek (Arvaniti et al. 2006, Arvaniti and Ladd 2009).

There was a significant effect of language background on the absolute duration of the low plateau \( F(1,12) = 10.417, p<0.008 \): in absolute terms, the plateau was longer for L2 speakers than L1 speakers, as illustrated in Figure 11(a).
No other main effects or interactions were found in the analysis of absolute duration measures of alignment in Tone 2.

Table 3: Pitch elbow alignment in Tone 2 (absolute duration).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (ms)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable onset to L1T2 (Initial fall)</td>
<td>73.38</td>
<td>41.79</td>
</tr>
<tr>
<td>Vowel onset to L1T2</td>
<td>3.86</td>
<td>39.37</td>
</tr>
<tr>
<td>L1T2 to L2T2 (Low plateau)</td>
<td>211.56</td>
<td>63.53</td>
</tr>
<tr>
<td>L2T2 to vowel offset (Final rise)</td>
<td>158.28</td>
<td>44.91</td>
</tr>
</tbody>
</table>

Figure 11: Error bars indicate standard deviations.
(a) Absolute duration of Tone 2 plateau.
(b) Absolute duration of Tone 2 syllable.
(c) Duration of Tone 2 plateau (% of σ).
(d) Duration of final rise in Tone 2 (% of σ).

Table 4 provides a summary of the analysis of the alignment of the pitch elbows in Tone 2 expressed in proportional terms.
Table 4: Pitch elbow alignment in Tone 2 (proportion of syllable duration).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (% of σ duration)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable onset to L1T2 (Initial fall)</td>
<td>16.31</td>
<td>8.66</td>
</tr>
<tr>
<td>L1T2 to L2T2 (Low plateau)</td>
<td>47.37</td>
<td>9.62</td>
</tr>
<tr>
<td>L2T2 to vowel offset (Final rise)</td>
<td>36.32</td>
<td>9.67</td>
</tr>
</tbody>
</table>

The effect of language background on the absolute duration the low plateau was not reflected when the plateau is expressed as a proportion of the total syllable duration \([F(1,12) = 3.558, p>0.05]\). This suggests that the absolute difference in the duration of the plateau is accompanied by similar differences in absolute syllable duration. In the analysis of duration, language background was not found to significantly affect syllable duration. However, since the effect of language background on syllable duration was close to reaching significance levels \([p=0.070]\), with syllable duration showing a trend of being longer for L2 speakers than L1 speakers, it was felt justified to conduct a separate analysis for Tone 2. In order to explore if the difference between L1 and L2 speakers in the absolute duration of the plateau could in fact be traced to a difference in syllable duration, a factorial ANOVA was conducted for Tone 2, with syllable duration as the dependent variable, and gender and language background as the categorical variables. The results, illustrated in Figure 11(b), indicate a significant effect of language background on the duration of the syllable, with L2 speakers having a longer Tone 2 syllable than L1 speakers \([F(1,12) = 7.797, p<0.017]\).

Although the absolute duration of the plateau is longer for L2 speakers than L1 speakers, this was accompanied by differences in absolute syllable duration in the same direction, explaining why the absolute differences in the plateau duration did not translate into proportional differences. However, since the differences in syllable duration were not statistically significant in the repeated measures ANOVA of duration, these results concerning the differences in absolute duration should be treated with caution, and we should focus instead on the proportional analysis.

Proportionally speaking, there were gender and language background related differences in the alignment of the pitch elbows of 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 11(c). Male speakers had a longer plateau than female speakers, or in other words, male speakers align the two pitch elbows further away from each other than female speakers.

Language background had a significant effect on the duration of the final rise, when expressed as a percentage of the total syllable duration \([F(1,12) = 9.192, p<0.010]\). The final rise occupied a greater proportion of the Tone 2 syllable for L1 speakers than L2 speakers, as shown in Figure 11(d).

No other main effects or interactions were found in the proportional analysis of the alignment of pitch elbows in Tone 2.

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 then female speakers, as illustrated in Figure 12. This result is in line with studies of intonation in languages such as New Zealand English (Daly and Warren 2001) and Dutch (Haan and van Heuven 1999), which report speaker-sex differences in pitch range.
There were no other main effects or interactions on F0 range [F<1].

There was a significant effect of the F0 points on the scaling of F0 [F(12,144) = 145.533, p<0.001]. The results of a Tukey HSD test indicated that the 13 F0 points can be grouped into four pitch levels, based on statistical significance [p<0.005], as illustrated in Figure 13. Among the H points, only H1T2, H1T3, and H2T3 are significantly different from the others, that is, H1T1, H2T1, H3T1, H2T2, H1T4, and H2T4; these other H points are not significantly different from one another, and as a group, they constitute the highest pitch level among the 13 F0 points. Out 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. Therefore, H1T2 and H1T3 together form the second highest pitch level. H2T3 is not significantly different from L1T2 and L2T2, but each of these three points is 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 both are significantly lower than all the other points.

As expected, there was also 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 test, p<0.05], as illustrated in Figure 14. These differences reflect the physiologically-determined differences in the average absolute pitch values.
used by male and female speakers (Cruttenden 1997: 3). Thus, the F0 range employed by male speakers is both smaller and lower than that of female speakers.

**Figure 14**: Gender differences in F0 scaling. Error bars indicate standard deviations.

There was also a significant effect of the interaction of gender and the F0 points on the scaling of F0 \( F(12,144) = 12.356, p<0.001 \). Comparing the F0 points between the two genders, 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, \( p<0.05 \)]. In other words, in terms of the four pitch levels identified from the pooled data, the lowest pitch level of female speakers overlaps with the two highest levels of male speakers.

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 test, \( p<0.05 \)].

With male speakers however, the four pitch levels are less distinct (Figure 15). The results of a Tukey HSD test \( [p<0.05] \) indicated that while four pitch levels can still be identified, there is some overlap in the organisation of the 13 points along the four pitch levels. Like the pooled data or the female data, the two lowest points of male speakers, minT3 and minT4, remain significantly lower than all the other points, and are not significantly different from each other. H1T1, H2T1, H3T1, H2T2, H1T4, and H2T4 still constitute the highest pitch level, and each of these points is significantly higher than L1T2, L2T2, H1T3, H2T3, minT3, and minT4. Unlike the pooled data or the female data, however, 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. 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.

As shown in Figure 14, 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.

There were no significant effects of language background on F0 scaling \( [F<1] \), and there were no significant interactions (gender x language background \( [F<1] \); language background x F0 points \( F(12,144) = 1, p>0.05 \); gender x language background x F0 points \( F<1 \)).
4.4. Chao tone letters
The results of the three metrics for converting the F0 points to Chao tone letters are summarised in Table 5. The H2 points of Tones 1, 3, and 4, which are the points immediately after the vowel onset, were left out of the conversion, and only the initial and minima/ending points were converted for these tones.

Table 5: Conversion to Chao tone letters. 1 represents the lowest point and 5 the highest on the five point scale.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Chao 1956</th>
<th>Chao 1968</th>
<th>Five equi-linear intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>66</td>
<td>66</td>
<td>55</td>
</tr>
<tr>
<td>T2</td>
<td>4335</td>
<td>4335</td>
<td>3225</td>
</tr>
<tr>
<td>T3</td>
<td>41</td>
<td>41</td>
<td>31</td>
</tr>
<tr>
<td>T4</td>
<td>61</td>
<td>61</td>
<td>51</td>
</tr>
</tbody>
</table>

4.5. 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 Singapore Mandarin. We can see from Figure 7 that Tone 4 appears much shorter than the other three tones, and from Figure 8 we can see that L2 speakers seemed to produce longer syllables than L1 speakers. Indeed, these observations were confirmed by our statistical analyses. 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, the tones of Singapore Mandarin appear to be more similar to Beijing Mandarin (Tone 3 > Tone 2 > Tone 1 > Tone 4) than Taiwan Mandarin (Tone 2 > Tone 4 > Tone 1 > Tone 3): Tone 4 is the shortest among the four tones. However, unlike Beijing Mandarin, Tones 1, 2, and 3 of Singapore Mandarin are not significantly different in terms of duration. Internally to the variety, L2 speakers showed a trend of producing longer syllables than L1 speakers, i.e. L2 speakers tended to be slower, and perhaps more careful, in their speech. Although not strictly significant, the effect of language background on syllable duration seems to be approaching levels of significance, and this effect 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 and Samuel (2004) report a correlation between syllable duration and tone perception when F0 information is neutralised. The data from Singapore Mandarin, however, suggest that syllable duration alone cannot provide the basis for distinguishing Tones 1, 2, and 3, if these tones can indeed be distinguished without F0 cues in the first place. In the case of Tone 3, the longer onset might provide a reliable cue for distinguishing it from the other two tones. However, Tones 1 and 2 do not exhibit any differences in the durations of the onset, vowel, or syllable, and duration cannot possibly provide a cue for the distinction of these two tones in the absence of F0 information. Kong and Zeng (2006) and Chang and Yao (2007) suggest that in addition to temporal information, intensity provides another secondary cue to tone identification in the absence of F0 information. Singapore Mandarin 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? It would thus be interesting to investigate if Tones 1, 2, and 3 can be reliably identified when stripped of their F0 information, and if intensity plays a role in the perception of these tones in Singapore Mandarin.

5.2. F0

As expected, there were gender effects on F0: female speakers employ a wider F0 range and each of the measured F0 points produced by female speakers was higher than the corresponding point produced by male speakers. That is, 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 6, 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, 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).

While the plots obtained from Xu’s method does give us an overview of the picture, mere comparisons of these plots do not give us reliable results. For instance, Figure 8 suggests that male L1 speakers end their Tone 2 higher than their Tone 1. However, our statistical analyses of F0 showed no interactions between gender and language background.

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 6. Although the four pitch levels are less distinct for male speakers, this is accompanied by a narrower F0 range which compresses the differences between levels. It is likely that due to the compressed F0 range of male speakers, the effects are smaller and can only be detected reliably with data from more male speakers.

<table>
<thead>
<tr>
<th>Pitch level</th>
<th>F0 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>α Low</td>
<td>minT3, minT4</td>
</tr>
<tr>
<td>β Mid-low</td>
<td>L1T2, L2T2, H2T3</td>
</tr>
<tr>
<td>γ Mid-high</td>
<td>H1T2, H1T3</td>
</tr>
<tr>
<td>δ High</td>
<td>H1T1, H2T1, H3T1, H2T2, H1T4, H2T4</td>
</tr>
</tbody>
</table>

Table 6: Four pitch levels in Singapore Mandarin.

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 Singapore Mandarin 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 of each tone appear to be in free variation.

The realisations of the four tonal contours show both similarities to and differences from the other varieties. The contours of Tones 1 and 4 show cross-dialectal consistency: in all three varieties, Tone 1 is a high level tone, while Tone 4 is a falling tone that starts at the top of the pitch range and begins falling after the vowel onset to the bottom of the pitch range. The contour of Tone 2 in Singapore Mandarin is distinct from that in both Beijing and Taiwan Mandarin, and appears to uniquely characterise this variety. Unlike Tone 2 of Beijing Mandarin, which falls slightly before showing an early rise, or Tone 2 of Taiwan Mandarin, which remains level or drops slightly during the first half of the vowel before rising towards the end, the contour of Tone 2 in Singapore Mandarin is characterised by an initial dip, followed by a mid-low level stretch, before the final rise. Tone 3 in Singapore Mandarin is more similar to that of Taiwan Mandarin than Beijing Mandarin, where it is realised as a dipping tone and does not end low.

There are two possible onset pitch levels for the four tones in Singapore Mandarin: 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 onset and offset pitch levels of the tones in Singapore Mandarin closely reflect Shih’s description of the tonal contours in Taiwan Mandarin (Figure 3). 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. Furthermore, it is not clear how the variation in the plateau duration or the duration of the final rise in Tone 2 can be accommodated in a framework that treats contour tones as integral units of pitch movement, unless arbitrary changes are made to the strengths of the rises and falls. Thus, the data from Singapore Mandarin appear to be more compatible with the target and interpolation view than the position that tonal contours are holistic movements. This view is consistent with the findings of Zsiga and Nitisaroj (2007) from Thai 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 from Thai.

5.3. **Alignment in Tone 2**

As noted earlier, the contour of Tone 2 in Singapore Mandarin is unlike that of both Beijing and Taiwan Mandarin. In particular, it exhibits a low plateau that is flanked by two pitch elbows. The analysis of the alignment of the two pitch elbows showed high standard deviations, reflecting a fair amount of variability in how speakers aligned these pitch elbows. Therefore, the alignment of pitch elbows in Singapore Mandarin appears to be less strict than that found in intonation languages such as Greek. This could simply be due to the uniqueness of the Tone 2 contour in comparison to the other tones: since it is the only tone that is realised as a dipping tone, variation in the alignment of the pitch elbows should be easily tolerated without creating any confusion. 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 one variety would confuse the two tones, and if alignment would play any role in disambiguating these tones.

Even though the alignment of pitch elbows to the segmental material does not appear to be as strict as in intonation languages, there are still systematic patterns of alignment that relate to gender and language background. Notice that while Figure 8 hints at these patterns, the differences that we have found cannot be confirmed by mere observation. Proportionally speaking, male speakers align the two elbows further away from each other than female speakers, while the final rise starts earlier for L1 speakers than L2 speakers. Thus, in addition to the obvious F0 differences between male and female speakers that might be expected on purely physiological grounds, the proportional distance between the two pitch elbows in Tone 2 could also be a signal of gender identity. Likewise, a proportionally earlier rise on Tone 2 could be indexical of L1 speech, in contrast to the later rise in L2 speech. Given the trend
for L1 speakers to have shorter syllables than L2 speakers, the overall perceptual difference of the final rise might be quite large. These issues await proper socio-phonetic investigation.

5.4. Chao tone letters
The two algorithms proposed by Fon and Chiang (1999) for converting the acoustic data into the two versions of Chao tone letters gave the same results: Tone 1 is transcribed as [66], Tone 2 as [4335], Tone 3 as [41], and Tone 4 as [61]. The data from Singapore Mandarin seem to be the exact opposite of Fon and Chiang’s study of Taiwan Mandarin: unlike Taiwan Mandarin, which can be transcribed using only the first four of Chao’s five tone letters, Singapore Mandarin, according to the algorithms proposed by Fon and Chiang, requires a sixth. Clearly, the tones of Singapore Mandarin are organised along a different scale than either of the two systems proposed by Chao, and the tones of Singapore Mandarin are expressed using a wider F0 range than those of Beijing and Taiwan Mandarin. Therefore, neither of the two algorithms provided by Fon and Chiang is viable for transcribing Singapore Mandarin tones using the five letter system of Chao’s.

On the other hand, the data can be converted into Chao tone letters by dividing the F0 range into five equi-intervals, and assigning each tone letter to an interval. Using this method of transcription, the tones of Singapore Mandarin can be given the following transcriptions: Tone 1 is transcribed as [55], Tone 2 as [325]4, Tone 3 as [31], and Tone 4 as [51]. These transcriptions also reflect the fact that even though Chao’s system consists of five levels, only four are relevant for Singapore Mandarin tones: the tone letter [4] does not occur in any of the transcriptions. However, since the first and fifth points of the five point scale are bound by the F0 range of the tonal system, this method does not allow for cross-dialectal comparison of pitch heights, and care should be exercised when comparing cross-dialectal transcriptions of tone. In other words, this transcription method allows one to represent the relative levels of the tonal targets within one variety, and the comparison, across varieties, of how these tonal targets relate to one another within a variety, but does not provide a basis for comparing tonal heights across varieties, unless the same F0 range is used for transcription. Therefore, while Tones 1 and 4 share the transcription values, and thus relative contours, provided by Chao for Beijing Mandarin and Chua for Standard Singapore Mandarin, Tones 2 and 3 are different. Unlike the [35] transcription for Tone 2 provided by Chao and Chua, which suggests a rising tone, the current data show that Singapore Mandarin is a dipping tone which should be transcribed as [325]. As for Tone 3, the data show that Tone 3 in Singapore Mandarin exhibits a steeper fall than in Beijing Mandarin or Chua’s characterisation of Standard Singapore Mandarin: unlike Chao and Chua’s transcriptions, where Tone 3 begins with a [2], Tone 3 in Singapore Mandarin starts with a [3].

6. Summary and conclusion
In conclusion, the acoustic study of tone in Singapore Mandarin presented here shows that despite some similarities to the other varieties, there are some non-trivial differences between the tonal realisations in this variety and the varieties 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, Tones 1 and 4 show cross-dialectal consistencies; the contour of Tone 3 in Singapore Mandarin is more similar to that of Taiwan Mandarin than Beijing Mandarin, which could be due to the influence of Southern Chinese languages spoken on the two islands; while the contour of Tone 2 is unique to Singapore Mandarin. Internally to the Singapore Mandarin system, there are clear gender-related differences in the scaling and alignment of pitch targets, as well as some differences that relate to the speakers’ language backgrounds. While this study focussed on the acoustic properties of Singapore Mandarin tones in isolation, further studies on the realisation of the tones in continuous speech and the perception of these tones are needed for a complete understanding of the tonal system of Singapore Mandarin.

4 Following Chao, a maximum of three letters is used for transcribing each tone. Since the two pitch elbows of Tone 2 have the same pitch value, only one of them is transcribed.
References


Peng, Shu-hui, Marjorie K. M. Chan, Chiu-yu Tseng, Tsan Huang, Ok Joo Lee, and Mary E.


