Tone Realization in Sung Mandarin

Murray Schellenberg

1 Department of Linguistics, University of British Columbia, Canada
Mhschellenberg@gmail.com

Abstract
Drawing on previous work which showed that singers in Cantonese transfer contour tone information over to singing, this paper examines whether singers in Mandarin exhibit similar kinds of transfer. Analysis of contour slope, duration and fundamental frequency suggest that Mandarin singers do not transfer these elements of tone to their singing.

Index Terms: tone, singing, Mandarin, contour, duration

1. Introduction

Modern Mandarin music reflects very little of the speech melody [2], [4] although there is a tendency to do so in positions of metrical prominence (both musical and verbal) [9]. Cantonese music, conversely, mirrors Cantonese speech melodies very closely with regard to level but omits contours in its composition [2], [11].

Recent experimental research [7] supports earlier observations [2] that singers in Cantonese produce the missing lexical tone contours when singing. To the author’s knowledge, there has been no systematic study as to whether Mandarin singers include comparable contours. This paper provides the results of a pilot study extending the experimental analysis used in the Cantonese study [7] to Mandarin. The analysis will also be extended to examine whether Mandarin singers maintain the durational correlates of tone [3], [12] and also whether there are differences in fundamental frequency associated with sung tones.

2. Methodology

Mandarin has four tones: high level (5-5); rising (3-5); falling-rising (2-1-4); and falling (5-1) [5]. The numbers in parentheses following the descriptions represent equally spaced F0 levels from 1 (the lowest) to 5 (the highest) and can be used to give a good estimation of the tone contour. Additionally, the tones are usually numbered from 1 to 4 in the order given for easy reference: Table 1 lays out that numbering system and provides examples.

2.1. Subjects

Subjects were 8 native speakers of Mandarin (5F, 3M; mean age = 22, s.d. = 3.46) who were all residents of Vancouver. They were all fluent in English, as well. Five subjects were choral singers; 1 was not a choral singer but had taken singing lessons for 5 years; 1 subject had neither lessons nor choral experience but reported singing on her own for at least one hour per week and 1 subject failed to provide information about his singing experience.

2.2. Stimuli

The target stimuli for this study were a minimal set of shi [shi] on all four tones, given in Table 1; in each case, the target syllable is the second member of a compound word where the first syllable carries tone 2. These syllables were included in the 4 stanzas of a specially written poem which was then set to music. The poem was written by Chenhao Chiu and Yuan Lu, both trained linguists and both native speakers of Mandarin. The Chinese lyrics can be found in Figure 1; the English translation of the lyrics is as follows:

As I was walking down the road
I couldn’t see what was around me
I was so tired
I opened my eyes and saw the city

As I was walking through the city
I couldn’t see what was around me
I was so alone
I opened my eyes and saw the teacher

As I was walking with him
I couldn’t see what was around me
I miss her so much
I opened my eyes and saw my love history

As I was walking through the past
I started to see what was around me
I am so afraid
I closed my eyes but saw only my time

The target syllables appear as the second half of the final word of each stanza: city, teacher, love history, time (see Table 1). Modern Mandarin words are usually bisyllabic so it was not possible to create a completely identical carrier phrase that would also work in the context of a song but the word choice was controlled as closely as possible. In all cases, the preceding syllable has tone two and the sense of the final line of each stanza, the carrier sentence, was kept fairly constant.

The music was composed by the author, an amateur musician with training in composition. The song was written in the style of shidai qu (時代曲), also known as guoyu laoge (國語老歌) – Mandarin popular songs produced in Shanghai in the 1930s and ’40s. The author is very familiar with this style of music which still well known. It is strongly formulaic in structure; chapter 5 of Chen [4] provides a very clear articulation of the formula and was used as the template and guide for the composition of the music. The completed song was played to four native speakers of Mandarin and was judged to be comparable to the original style. The score is given in Figure 1.

<table>
<thead>
<tr>
<th>TONE</th>
<th>CHAR</th>
<th>GLOSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>high level</td>
<td>吾師 ‘teacher’</td>
</tr>
<tr>
<td>2</td>
<td>rising</td>
<td>曾時 ‘time’</td>
</tr>
<tr>
<td>3</td>
<td>fall-rise</td>
<td>情史 ‘love history’</td>
</tr>
<tr>
<td>4</td>
<td>falling</td>
<td>城市 ‘city’</td>
</tr>
</tbody>
</table>

Table 1: Target words (second syllable only)
2.3. Procedure

A mock-karaoke system was set up using E-prime [8]. Two lines of the song were displayed at a time on the computer screen while a recorded musical track played the melody with an accompaniment over computer speakers placed just behind the monitor. As each character on the screen was to be sung, it changed colour indicating to the subject when to sing that word. The next two lines of the song appeared just prior to their occurrence in the musical track. Instructions were printed in Chinese on the computer screen prior to the recording session and a pause screen was shown at the end of each repetition of the song. Each repetition was started by the subject at his/her discretion.

Subjects were first presented with a printed version of the song (identical to Figure 1) and listened to the musical track while watching the words on the computer screen. They were allowed to repeat this process as often as they wished until they felt comfortable enough with the song to be able to sing it. They were then recorded singing the song 6 times with breaks in between.

Subjects were recorded at a sampling rate of 44 100 Hz using an AKG C520 head-mounted directional microphone and a Sound Devices USBPre pre-amp. The speakers playing the musical track were placed behind the microphone which either strongly damped the recording of the musical track or eliminated it altogether. Recordings were made in Audacity on a Macintosh Classic notebook and saved as .wav files.

2.4. Analysis

The recordings were segmented and the target words extracted from the main recording using Praat [1]. The pitch contours were examined and any errors corrected by hand. The results were also normalized for duration: F0 values were extracted at eleven equally-spaced intervals. The data were then exported into R [6] for statistical analysis. Raw duration values were also recorded. Mean contours normalized for duration are shown in Figure 2; the first and last points were not included in the analysis to limit influences from the onset consonant and phrase final position. Error bars have been omitted for legibility. The extraordinarily high F0 for point 9 of tone 1 appears to be an anomaly of the data.

3. Results

3.1. Contours

For each individual token, two slope measurements were computed: early slope (from 1 to 5) and mid slope (from 3 to 7). These have been shown to be the areas where singers included contour information when singing; the tone information is included early in the sung note and the last part of the note is sustained at the prescribed F0 [7]. To minimize the effect of vibrato, the slopes were extrapolated; that is, they were all treated as if they were straight lines.

As the distribution was not normal, these results were fitted to a mixed-effects model with subject as a random-effect factor and tone as a fixed-effect factor; to achieve comparisons across all combinations the intercept was set variously as tone 1, 2 and 4. The complete results are given in Table 2. No significant interactions were found.

<table>
<thead>
<tr>
<th>SLOPE OF FIRST HALF (1-5)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>t</td>
<td>p</td>
<td>β</td>
</tr>
<tr>
<td>-0.17</td>
<td>-0.96</td>
<td>0.34</td>
<td>0.06</td>
</tr>
<tr>
<td>0.25</td>
<td>1.23</td>
<td>0.22</td>
<td>0.11</td>
</tr>
<tr>
<td>0.36</td>
<td>1.76</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SLOPE OF MIDDLE (3-7)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>t</td>
<td>p</td>
<td>β</td>
</tr>
<tr>
<td>-0.18</td>
<td>-0.61</td>
<td>0.54</td>
<td>0.11</td>
</tr>
<tr>
<td>0.29</td>
<td>0.97</td>
<td>0.33</td>
<td>0.21</td>
</tr>
<tr>
<td>0.08</td>
<td>0.27</td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Statistical results comparing slopes of tones in two different sections of the contours
3.2. Duration

The mean raw duration scores are given in Table 3. A repeated measures ANOVA was conducted to compare the effect of tone on duration. A significant effect was found (F(3,28) = 4.7861, p = 0.00815). A post hoc Tukey’s HSD test determined that there were significant differences between the durations of tones 1 and 2 (p = 0.0142) and tones 2 and 4 (p = 0.0162); the difference between tones 2 and 3 was approaching significance (p = 0.062) and all other pairings were not significant.

3.3. Fundamental Frequency

Mid-point measures (point 5 in figure 2) of fundamental frequency were analysed. As with the contour measures, the distribution was not normal so the results were fitted to a mixed-effects model with subject as a random-effect factor and tone as a fixed-effect factor; the intercept was set variously as tone 1, 2 and 4. There were no significant interactions found. The results are given in table 4.

4. Discussion

The results from the slope analysis suggest that singers in Mandarin do not include tonal contour information while they are singing. While the casual listener may observe contour changes over the course of a sung syllable in Mandarin it is most likely that these are present for musical rather than linguistic reasons.

The duration results show that tone two is shorter than the other three tones but the setting of this syllable in the music is on a note of shorter duration (bar 32, second note) so this difference is most likely a result of the music and not of language. Studies of the durational differences associated with tone in Mandarin have shown tone 4 to be the shortest and tone 3 to be the longest with tones 1 and 2 falling in between [3], [12]. The fact that there are no significant differences in duration between tones 1, 3 and 4 suggest that the durational distinctions of tone are neutralized in sung Mandarin.

One of the most striking features observed in Figure 2 is that, although all four of the target words were sung on the same musical note, they appear to have different fundamental frequencies. Tones three and four overlap considerably but tone one is consistently lower than the other tones and tone two is consistently higher. What is particularly interesting is that the patterning observed does not match the F0 levels found in spoken Mandarin tones. Nor does it match the levels of the Zhongzhou or “Central” dialect used in Peking Opera

<table>
<thead>
<tr>
<th>tone 1</th>
<th>mean (msec)</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1418.474</td>
<td>263.8451</td>
</tr>
<tr>
<td>2</td>
<td>1080.955</td>
<td>202.5875</td>
</tr>
<tr>
<td>3</td>
<td>1349.826</td>
<td>218.0666</td>
</tr>
<tr>
<td>4</td>
<td>1400.964</td>
<td>262.5157</td>
</tr>
<tr>
<td>all</td>
<td>1313.485</td>
<td>272.961</td>
</tr>
</tbody>
</table>

Table 3: Means of raw duration scores

where tone 1 = 4-4; tone 2 = 1-1; tone 3 = 5-3 or 5-5; and tone 4 = 2-4 or 4-2-4 [2]. These differences, however, proved not to be statistically significant. It must be remembered, however, that the number of subjects is very small (8 subjects) so this pattern may simply be an anomaly of the data. Conversely, a larger number of subjects may show significance. Further investigation is indicated.

5. Conclusions

The results of this preliminary examination of Mandarin singing suggest that singers in Mandarin do not mark tone while singing in contrast with Cantonese singers who add in missing tonal contours [7]. Tone, as part of language, is a learned behaviour which becomes very deeply engraved – perhaps automatic – in the speakers of that language. The fact that there are differences between Mandarin singers and Cantonese singers suggests that this engrained behaviour can be “turned off” and replaced with a different contour; Mandarin singers seem to turn it off in music while Cantonese singers do not. It appears that there is very little of the tone information transferred over to music in Mandarin singing.

6. References


