TONE CONTOUR REALIZATION IN SUNG CANTONESE

Murray Schellenberg

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

ABSTRACT

Twelve native speakers of Cantonese (6M, 6F) were recorded singing a tonal minimal set of words ([si]) in the context of a specially composed children's song. Analysis of slope measurements of the vowels suggests that singers include a rising contour when singing a rising tone. They do not appear to include a falling contour when singing a falling tone. F_0 appears to fulfill multiple demands in singing.

Keywords: singing, Cantonese, tones

1. INTRODUCTION

It has been suggested that singers in tone languages will sometimes add tone information during performance, particularly if there is a mismatch between the sung and spoken melodies. Rycroft [8] comments that in siSwati and Zulu, tonal "on-glides" conditioned in speech by particular consonants can be transferred over to singing. Yung [13] says that in Cantonese Opera (which is highly improvisatory) "syllables with rising and falling linguistic contours are sung either to a single pitch preceded by an ornamental glide or to two- or three-pitch figures which reflect the contour of the linguistic tone" (p. 83). Chao [3] suggests that singers in Mandarin may use grace notes (brief extra-metrical notes) "in order to 'smuggle in' the tone, if not already suggested in the main melody" (p. 57). This paper examines the singer's contributions to the realization of tone during singing in Cantonese.

2. METHODS

2.1. Cantonese

There are six tones in Cantonese (see Table 1): three level tones (tones 1, 3, and 6); two rising tones (tones 2 and 5) and one falling tone (tone 4). The level tones also occur in shortened form in syllables that end with voiceless stops; these are sometimes referred to as leading tones and are occasionally classified as separate tones [12]. This study adopts the six tone classification. The numerical descriptions of the tones in Table 1 follow Chao [2] where each number represents an equally spaced F_0 level from 1 (the lowest) to 5 (the highest).

Wong & Diehl [11] classify the six Cantonese tones into three levels for singing according to the final F_0 level of the tone: high-pitch with tone ending on level 5 (tones 1 and 2), mid-pitch with tone ending on level 3 (tones 3 and 5) and low-pitch with tone ending on level 1 or 2 (tones 4 and 6). This conflates the level tones and the contour tones into three registers which they found are usually reflected in musical composition.

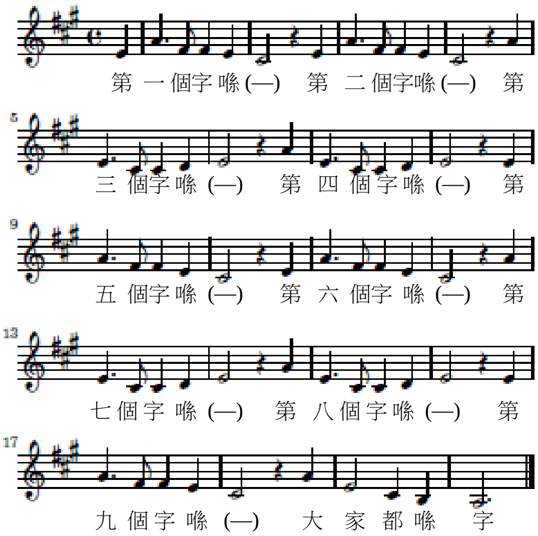
Table 1: The six Cantonese Tones. The numbers in parentheses represent contour descriptions of equally spaced F_0 levels from lowest (1) to highest (5) [2].

TONE	DESCRIPTION	
1	High-level	(5-5)
2	High-rising	(3-5)
3	Mid-level	(3-3)
4	Low-falling	(2-1)
5	Low-rising	(2-3)
6	Low-level	(2-2)

Table 2: Target words.

	TONE	CHAR	GLOSS
si	1	師	'teacher'
si	2	史	'history'
si	3	嗜	'be fond of'
si	4	時	'time'
si	5	市	'market'
si	6	豉	'fermented beans'

Figure 1: Score of the "children's song" used.


第 一個字 喺(—) 第 二個字喺(—) 第
5
三 個字 喺 (—) 第 四 個字 喺 (—) 第
9
五 個字 喺 (—) 第 六 個字 喺 (—) 第
13
七個字 喺 (—) 第 八個字喺 (—) 第
17
九 個字 喺 (—) 大 家 都 喺 字

2.2. Subjects

Subjects were 12 native speakers of Cantonese (6M, 6F, mean age = 45.17, sd = 16.17). All were residents of Vancouver, BC (Canada) and all were also fluent in English. Ten subjects were choral singers; two of whom had received formal training in singing. The two non-choral singers reported regular singing. The subjects reported an average of 3.3 hours of singing per week (sd = 1.2).

2.3. Stimuli

The target stimuli for this study were a minimal set of [si] on all six tones, given in Table 2.

These six words were embedded in a specially written Cantonese "children's song". The music for the song consisted of 2 variations of a melody written by Patrick Wong [11] to match the spoken contour of the phrase (excluding the target words and the numerals). The two variations were put together in 4 pairs along with a separate concluding phrase to form a short song with the form AABBAABBC. The target word is sung on two different notes: in phrase A the note is E and in phrase B the note is C#. The score is given in Figure 1. The lyrics translate as

> *The first word[1] is (—), the second word is (—), the third word is (—),the fourth word is (—), the fifth word is (—), the sixth word is (—), the seventh word is (—), the eighth word is (—), the ninth word is (—);We are all words.*

Depending on the target tone, the musical melody potentially deviates from the spoken melody. There are two musical contours for the final word: one with a falling interval (phrase A) and one with a rising interval (phrase B). This means that for each target word there is one musical environment that corresponds to its spoken melody and one that contradicts it. There is a similar split for the numerals.

There were 18 words used: the six target words and 12 distracter words. A single repetition of the song used 9 words so two repetitions of the song constituted one block.

2.4. Procedure

Each subject was shown a printed score of the song which also contained a fictitious history of the song in Chinese to provide a plausible explanation for randomization. Subjects were told that in this song the words marked by the dash are traditionally chosen by the singer and that a further challenge is traditionally added by changing the order of the numerals – counting backwards or choosing first the even numbers and then the odd numbers; but that in the study both the word choice and the order of the numerals would be random.

To ensure familiarity with the characters, subjects were also given a list of 36 words in Chinese with English translations which contained the six target words, the 12 distracter words and 18 other words. They were told that the words in the song would be taken from the list. They listened to a recording of the melody as often as they wished until they felt they were comfortable enough with the song to be able to sing it. They were allowed to listen to the recording at any time during the experiment if they felt they needed to be reminded of the melody.

The stimuli were presented using E-prime [9]. Each screen presented a single line of the song, showing the musical score, the words to sing and an arrow to indicate if the first interval was rising or falling (a circle indicated the final phrase). The experiment was timed to present the screens to correspond with a performance tempo of 66 beats per minute. A metronome was used to help the subjects maintain the tempo. Subjects were instructed to push the space bar (which initiated a single repetition of the song) in time with the metronome and count 4 beats before starting to sing. The first screen appeared on the 4th beat. Screens were timed to appear on successive 4th beats until the end of the song at which time a pause screen was presented and the next repetition could begin at the subject's initiation. Just before beginning each song repetition the subject heard the first two notes of the song to give them the starting notes.

There were 2 training blocks and 6 experimental blocks; each subject sang the song 16 times. Subjects were recorded at a sampling rate of 44 100 Hz using an AKG C520 head-mounted microphone and a Sound Devices USBPre pre-amp. Recordings were made in Audacity on a Macintosh Classic notebook and saved as .wav files.

3. RESULTS

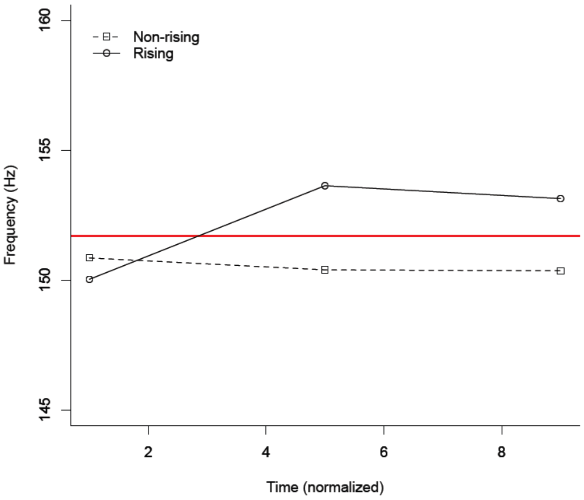
The recordings were segmented and the target words extracted from the main recording using PRAAT [1]. The pitch contours were examined and any errors corrected by hand. For visualization purposes, the pitch contours of the voiced sections were adjusted for target F_0 – most of the subjects sang in the same F_0 range but three of the female subjects sang an octave higher; the results for these three subjects were divided by 2 to place them in the same octave range as the others. The results were also normalized for duration: F_0 values were extracted at eleven equally-spaced intervals. The data were then exported into R [7] for statistical analysis. Raw duration values were also recorded.

A total of 381 tokens were analyzed out of a possibility of 432 tokens (6 tokens x 6 tones x 12 subjects). Most missing tokens were lost due to melody memory errors; singers would forget the melody and either stop completely or relapse to humming to “find” the melody again. Some were also lost to synchronicity errors: subjects would get out of sync with the timing of the computer and lose the general thread of the song.

3.1. Slope of F_0 Contours

Measurements for the higher and lower notes were pooled for each subject. For each individual token, four slope measurements were computed: total slope (from 1 to 9), early slope (from 1 to 5), mid slope (from 3 to 7) and late slope (from 5 to 9); the first and last data points (0 and 11) were not included in the analysis. To minimize the effect of vibrato, all slopes were extrapolated as straight lines. Figure 2 shows the mean extrapolated F_0 tracings across all subjects, normalized for duration; error bars have been excluded for readability. Based on the statistical analysis, the tones have been divided into two groups: rising and non-rising. The target tone (the mean of the two targets) is represented by the solid red line.

Figure 2: Mean F_0 tracings for the vowel portion of [si] for rising tones versus non-rising tones across all subjects, normalized for duration. The solid red line represents the target note.



These results were fitted to a mixed-effects model with subject as a random-effect factor and tone shape (level, falling or rising) as a fixed-effect factor; “level” was set as the intercept. A separate model which also included duration was also fitted. A likelihood ratio model compared the two and duration was found not to contribute significantly to the model; consequently the simpler model was chosen as the final model. Table 3 shows the results for all four slope measurements. For the total slope, rising tones were significantly different from level tones (β = 0.4745, t = 4.628, p >0.0001). The effect is also present in the first half of the contour (β = 0.9158, t = 4.791, p < 0.0001). There is no significant difference between level and rising tones when looking only at the middle of the contour (β = 0.098, t = 0.706, p = 0.4806) nor in the second half (β =0.0349, t = 0.265, p = 0.7910). Falling contours were not significantly different from the level tones across the full length of the syllable (β = 0.0847, t = 0.609, p = 0.5427) but were in the second half of the contour (β = 0.5348, t = 3.212, p = 0.0014) where they exhibited a slight rise. Changing the intercept to “fall” showed that rising tones are significantly different from falling contours (β= 0.3135, t = 2.326, p = 0.0206).

Table 3: Results of mixed-effect model (fixed effects).

SLOPE OF WHOLE CONTOUR (0-10)				
	Estimate	Std Error	t value	p (est)
Intercept (level)	-0.1075	0.1729	-0.622	0.5346
shape-fall	0.1610	0.1298	1.241	0.2154
shape-rise	0.4745	0.1025	4.628	>0.0001*
SLOPE OF FIRST HALF (0-5)				
Intercept (level)	-0.0518	0.2872	-0.180	0.8570
shape-fall	-0.2110	0.2419	-0.872	0.3838
shape-rise	0.9158	0.1912	4.791	>0.0001*
SLOPE OF MIDDLE (3-7)				
Intercept (level)	-0.2389	0.1465	-1.631	0.1038
shape-fall	0.2202	0.1760	1.251	0.2118
shape-rise	0.0847	0.1391	0.609	0.5427
SLOPE OF LAST HALF (5-10)				
Intercept (level)	-0.1622	0.1713	-0.947	0.3444
shape-fall	0.5348	0.1665	3.212	0.0014*
shape-rise	0.0349	0.1315	0.265	0.7910

4. DISCUSSION

The results suggest that singers make a distinction between rising tone contours and non-rising tone contours when singing in Cantonese. Furthermore, this rising contour appears to be prominent during the first part of the syllable. In spoken Cantonese, the initial portion of the vowel does not appear to play a significant role in distinguishing tones [4] but the greater duration of the sung syllables may possibly contribute to this difference. The mean duration of the voiced portion of the sung syllables was 722.55 ms (sd=110.6); the mean duration of spoken [i] in the same syllabic structure is 252.39 msec (sd=22.55) [5]. Furthermore, singing requires F_0 to fulfill a musical role; the longer time available in singing may contribute to this ability to fulfill dual functions.

The fact that tone 4, the low falling tone, patterns more closely with the level tones is not particularly surprising in Cantonese. Tone 4 is often associated with creaky voice [6, 10]. Vance [10] found that synthesized stimuli with low F_0 and a falling contour were rarely identified as tone 4 by Cantonese listeners. Unfortunately, examinations for creaky voice as an indicator of tone 4 were not possible in this study as most of the female subjects sang the song in the lower range of their voices and exhibited a substantial amount of creak throughout the recordings. The slight rise found in the second half of the falling tone is surprising. It may be that there is a slight (but statistically not significant) decline through the first half and an attempt to return to the target note in the second half. Further examination of this phenomenon is necessary.

5. CONCLUSIONS

The results suggest that rising contour appears to be a component of tone that is transferred over to singing in Cantonese. As register is usually represented musically in Cantonese [11], the addition of a contour by the singer may help to distinguish tones further. It does not appear that contour information is used to correct mismatches between the sung and spoken melody; the rising contours are present even when the song melody matches the spoken melody. If the singers were only correcting mismatched speech and song melodies, there should be a marked difference between tones 2 and 5: there should be a correction only on the higher note for tone 5 and on the lower note for tone 2. Initial visual examination of the data suggests this is not the case; there appears to be no distinction between the two rising tones in singing, but closer examination is necessary.

It is possible that voice quality could be used as a marker of tone 4 (the low falling tone) but the pitch range of the song used for this study precludes examination of that question. There are also indications of F_0 over- and under-shoot that warrant further examination. It is not known whether these rising contours could aid listeners in understanding. These are all questions for further research.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2010. Praat: doing phonetics by computer [Computer program]. Version 5.2.03, retrieved 19 November 2010 from *http://www.praat.org/.*

[2] Chao, Y.R. 1947. *Cantonese Primer*. Cambridge: Cambridge University Press.

[3] Chao, Y.R. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition and atonal composition in Chinese. In Halle, M., Lunt, H.G., McLean, H., van Schooneveld, C.H. (eds.), *For Roman Jakobson: Essays on the Occasion of his Sixtieth Birthday, 11th October 1956.* The Hague: Monton & Co, 52-59.

[4] Khouw, E., Ciocca, V. 2005. Perceptual correlates of Cantonese tone. *Journal of Phonetics* 35, 104-107.

[5] Kong, Q.M. 1987. Influence of tones upon vowel duration in Cantonese. *Language and Speech* 30(4), 387-400.

[6] Lam, H.W., Yu, K.M. 2010. The role of creaky voice quality in Cantonese tonal perception. *Proc of the 159th Meeting Acoust. Soc. Am* Baltimore, MD.

[7] R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www.R-project.org*.

[8] Rycroft, D. 1970. The National Anthem of Swaziland. *African Language Studies* 11, 298-318.

[9] Schneider, W., Eschman, A., Zuccolotto, A. 2002. *E-Prime: User's Guide, version 1.0*. Psychology Software Tools.

[10] Vance, T.J. 1977. Tonal distinction in Cantonese. *Phonetica* 34, 93-107.

[11] Wong, P.M., Diehl, R.L. 2002. How can the lyrics of a song in a tone language be understood? *Psychology of Music* 30(2), 202-209.

[12] Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.

[13] Yung, B. 1989. *Cantonese Opera: Performance as Creative Process*. Cambridge: Cambridge University Press.

[1] In Chinese the cardinal number is used; the literal translation of the first line, for example, is ‘number one word is (—)’.