The Neutralization of Tone-related Duration Differences in Sung Cantonese

Murray Schellenberg

Department of Linguistics, University of British Columbia, Vancouver, Canada

mhschellenberg@gmail.com

Abstract

Cantonese compositional processes reflect register information about lexical tone and Cantonese singers add in contour information about tome while singing. This paper examines whether the singers also carry over durational differences associated with tone into sung Cantonese. 11 native speakers were recorded singing a specially composed song in Cantonese which included a six way minimal set by tone on the syllable [si]. Analysis of the mean durations of these tokens suggest that the durational differences of various tones found in spoken Cantonese are neutralized in sung Cantonese.
Index Terms: Cantonese, singing, tone, duration

1. Introduction

The compositional process of Cantonese music usually includes register information about tones but excludes contour information [1]. Previous work [2] has shown that singers in Cantonese include the missing contour information when singing, even though it is not called for in the music. This paper presents evidence to suggest that durational differences associated with tones is absent during performances of sung Cantonese.

The durational correlates of tone in Cantonese have been well documented [3]. The six non-entering tones of Cantonese are shown in Table 1. The numerical descriptions of the tones in Table 1 follow Chao [4] where each number represents an equally spaced F_0 level from 1 (the lowest) to 5 (the highest). There are 3 level tones (1, 3, and 6); 2 rising tones (2 and 5) and 1 falling tone (which is strongly associated with creaky voice [5]). The six tones cannot be ranked in a definite durational hierarchy but Kong [3] found patterns of duration across speakers: tone 2 (high rising, 3-5) is the longest tone and tones 1(high level, 5-5) and 4 (low falling, 2-1) are the shortest. Of the three level tones, tone 3 (mid-level, 3-3) is the longest. Although duration in music is usually dictated by the song writer/composer, variations of duration are one of the primary markers of "expressiveness" [6]. Performers regularly deviate from the dictated durations of notes for expressive purposes. This paper examines whether such modifications are undertaken for linguistic purposes.

TONE	DESCRIPTION		EXAMPLE
1	High; level	(5-5)	*si* – 'teacher'
2	High; rising	(3-5)	*si* – 'history'
3	Mid; level	(3-3)	*si* – 'to try'
4	Low; falling	(2-1)	*si* – 'time'
5	Low; rising	(2-3)	*si* – 'market'
6	Low; level	(2-2)	*si* – 'yes'

Table 1. *The six Cantonese lexical tones*

2. The experiment

An experiment was conducted using a minimal set of [si] on all six tones; these are given as the examples in Table 1. These six words were embedded in a specially written Cantonese "children's song".

2.1. Stimuli

The music for the song consisted of 2 variations of a melody written by Patrick Wong [1] 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 words all appeared phrase finally before a rest so the opportunity to use duration to mark tone is maximized. The score is given in Figure 1. The lyrics translate as

> *The first word*[i] *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.*

2.2. Subjects

Subjects were 11 native speakers of Cantonese (6M, 5F, mean age = 45.17, sd = 16.17). All were residents of Vancouver, BC (Canada) and all were also fluent in English. Nine 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).

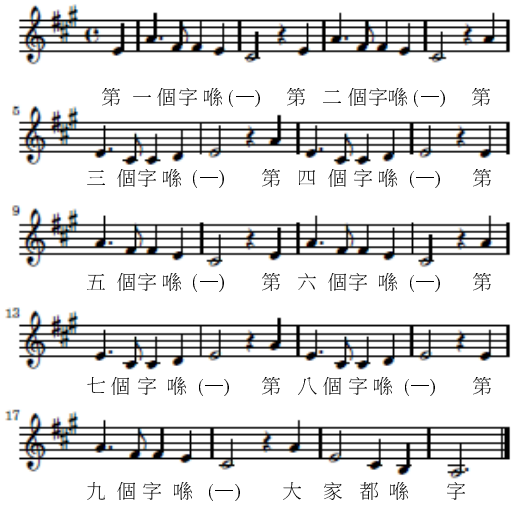

Figure 1: Score of the children's song used

2.3. Procedures

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 [7]. 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 [8]. The duration measures of each occurrence of the target words was extracted by script and exported into R [9]. 361 tokens were analyzed out of a possibility of 396 tokens (6 tokens x 6 tones x 11 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. Figure 2 shows a visual representation of the mean duration measures of each tone across subjects. The mean duration of all the tones was 912.12 ms (sd=101.66). A repeated measures ANOVA was conducted to compare the effect of tone on duration. There was no significant effect ($F(5,5)=0.483$, $p=0.778$). Similarly, when comparing only the three level tones, no significant difference was found ($F(2,2)=0.0713$, $p=0.9318$).

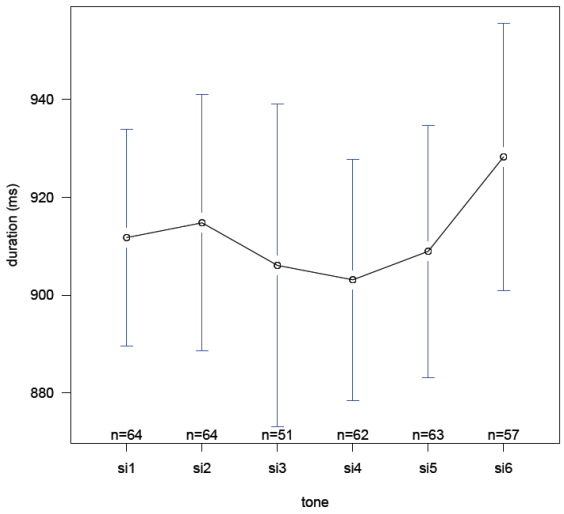

Figure 2: Mean durations of tones on syllable [si]

4. Discussion and conclusions

It appears that singers in Cantonese do not carry the durational differences of spoken tones over to singing. A visual examination of Figure 2 indicates that the patterns found by Kong [3] are not reflected in the singing data. In the singing data, tone 6 (low level, 2-2) is the longest as opposed to tone 2 (high rising, 3-5) in the spoken data. Tone 3 (mid level, 3-3) is the shortest of the level tone whereas in the spoken study it was the longest. Tone 4 (low falling, 2-1) is shortest in both spoken and singing but in the spoken data tone 1 (high level, 5-5) patterned with it whereas in the singing data tone 1 is the third longest. The singing data does not even trend towards the spoken data. The duration differences associated with tone appear to be completely neutralized in singing.

5. References

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

[2] Schellenberg, M., "Tone Contour Realization in Sung Cantonese" *Proceedings of the Seventeenth International Congress of the Phonetic Sciences*. Hong Kong, 1754-1757, 2011.

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

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

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

[6] Kendall, R. and Carterette, E., "The Communication of Musical Expression" *Music Perception* 8(2), 129-164, 1990.

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

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

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

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