ACCEPTABILITY JUDGMENTS OF VARIATIONS IN TONE REPRESENTATION IN CANTONESE SONGS

Esther Y. T. Wong[1], Murray Schellenberg[1] & Bryan Gick[1,2].

University of British Columbia[1]. Haskins Laboratories[2].
esther_wong_yt@hotmail.com, mhschellenberg@gmail.com & gick@mail.ubc.ca

ABSTRACT

Representations of tones in Cantonese songs have been studied extensively from a compositional point of view. In contrast, the perception of lexical tones in sung Cantonese music has received significantly less attention. Existing research has looked at lexical comprehension, but perceptual acceptability has not been examined. This study compares Cantonese native speakers' acceptability judgments for short melodies that contain either matches or mismatches between the musical spoken and sung melodies. Results show that native Cantonese listeners prefer musical contours that match the tonal contour of the spoken tones of the language.

Keywords: Cantonese, tone, music, singing, perception.

1. INTRODUCTION

Cantonese is well known for its relatively rigid correspondence of song melodies to contours that reflect spoken language pitch contours [1, 2, 3, 4, 5]. While spoken Cantonese has a six-tone system (Table 1), songs in Cantonese typically conflate the six tones into three registers based on the final target tone: tones 1 (55) and 2 (35) form one register; tones 3 (33) and 5 (23) form another; and tones 4 (21) and 6 (22) form a third [1].

A song's musical score reflects these registers in the contour of the song [1, 2, 3, 6, 7]. For example, if the lyrics contain a word with either tone 1 (55) or tone 2 (35) followed by a word with tone 3 (33) or tone 5 (23), the melody will most often have a falling contour; the note for the second word will be lower than the note for the first [1, 2, 3]. Furthermore, singers will often add in rising contour information during performance, even when that information is not explicitly included by the composer [2, 5].

While there is known to be a high correspondence between sung and spoken melodies in Cantonese songs, this correspondence is by no means absolute. Studies examining the level of correspondence between Cantonese sung and spoken melodies have found proportions of correspondence ranging from 79.2% in 6 children's songs [8] to 90.7% in 6 popular songs [2] to 91.81% in 4 popular songs [3]. This leaves a small but interesting proportion of non-matching sections of contour. This paper presents a study looking at whether mismatches of this kind are perceived to be aesthetically acceptable to native Cantonese listeners.

		TONE		GLOSS
絲 [si]	1	high level	(55)	'thread'
士 [si]	2	high rising	(35)	'cheese'
肆 [si]	3	mid level	(33)	'to unbridle'
匙 [si]	4	low falling	(21)	'key'
市 [si]	5	low rising	(23)	'market'
豉 [si]	6	low level	(22)	'black bean'

Table 1: The six Cantonese tones showing a minimal set on [si]

2. METHODS

Sixteen Hong Kong-raised native Cantonese speakers, ages 19 to 25, were paid to participate in this study. The stimuli consisted of specially-composed musical couplets disguised as an advertising jingle. These were jointly composed by three native speakers of Cantonese, all with musical training, one of whom had extensive training in composition. Each jingle consisted of 10 syllables divided into two 5-note phrases, with the target (the syllable [si] with tones as in Table 1) appearing as the final syllable of the first phrase. Variants of each jingle were composed where the musical pitch of the target word was shifted without changing the underlying harmonic structure of the music. Therefore, for each jingle, at least one version always followed the proscribed tonal contour of the spoken phrase; the others varied either by opposing the spoken contour direction or were in the correct direction but of differing intervals. These are given in the Appendix. For each melody, one note was deemed to be the "most acceptable" version of the melody by the three speakers involved in the composition. Each melody, therefore, had four versions: one of which followed the "chosen" contour and three of which varied by different degrees. Each stimulus ends with a note from the final chord. It gives us 4 different options, which varies by different contour and interval. Each melody also had one distractor version where the melody deviated at the 10th note instead of the fifth; two further distractor jingles with no tonal matching at all were also created. This produced a total of 32 different stimuli.

A separate background music track was created for each jingle. All versions of each jingle were recorded by a single singer with a Samson C03U mic using Audacity [9] and the corresponding background music was dubbed onto the singing recording with Final Cut Pro X [10]. The experiment was presented as a rating task in OpenSesame 3.2.4 [11] on an iMac 2017 computer through AKG K240 headphones. Participants were seated in a sound-attenuated booth at the University of British Columbia.

Stimuli were presented twice over 2 blocks (a total of 64 tokens) and were randomized within each block with a break in between. Participants were told that the stimuli were possible jingles for different prospective products and were asked to judge the acceptability of the jingles based on the predicted publication effects of these jingles for the products. After each jingle, participants were asked to indicate their preference to the jingle on a 5-point Likert scale [12] where 1 indicated that the jingle was "utterly unacceptable" and 5 that the advertisement was "very agreeable".

3. RESULTS

Participants' ratings were collated and are presented graphically in Figure 2 showing mean ratings across all participants, ranked from most acceptable to least acceptable.

The melody variation for each tone is labelled (y-axis) according to tone number and variation. The version deemed best by the composing panel is marked with an asterisk; the remaining three versions are lettered a, b, and c, starting from the lowest pitched note and moving up. Those with the highest ranking by participants are given as black bars, and those ranked lowest are marked by stripes. The distractor jingles are not included in this graph.

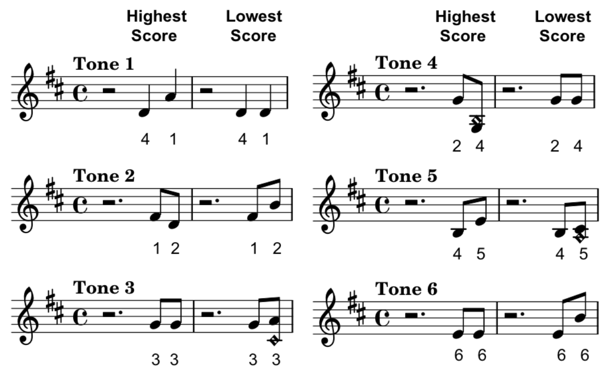

Figure 1: Highest (left) and lowest (right) ranking contours for each tone. Numbers under each note indicate tone of syllable set to that note. Diamond-headed notes mark predicted note, where different from actual results.

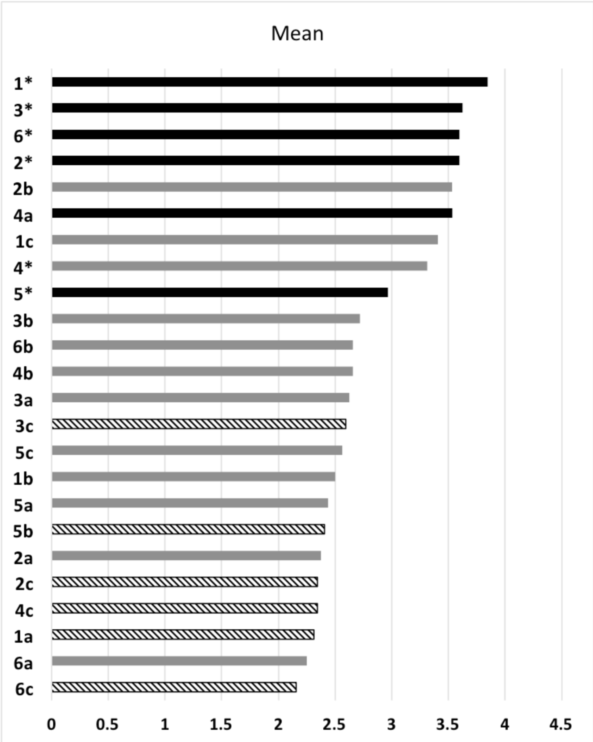

Figure 2: Participants' mean ratings on all stimuli. Black bars indicate highest ranked variant for each tone; striped bars indicate lowest ranked variant.

4. DISCUSSION AND CONCLUSIONS

The version of each jingle judged to be the best by participants corresponded with those thought to be best by the preliminary panel with the exception of tone 4. For this tone, participants preferred the version with the largest falling interval over that predicted by the composition panel.

The contours for tones 3 and 6 both consist of a repeated level tone. In both cases, listeners agreed with the panel that the best representation is to repeat the note (mean ratings of 3.66 ±0.91 and 3.59 ±1.01 respectively). In both cases, the lowest ranked variation was a rising contour (contra the panel's prediction that the large falling interval would be considered even worse for tone 3). The lowest ranked variations (tone 3: 2.59 ±0.98 and tone 6: 2.16 ± 0.77) in both cases involved contours that were rising. This is not only against the tonal contour but also a contour against the natural declination of speech so this may also be having an influence on listeners' preferences.

Tone 2 should in principle behave in the same way as a succession of high level tones [2, 3, 8]. In setting text to music, the rising contour on tone 2 is eliminated and the final pitch target - equivalent to the high level tone 1 - is used as the tone in the song. Based on this, we would expect to see a tone 1 - tone 2 sequence to pattern along with the 3-3 and 6-6 patterns discussed above. However, in this case, the most endorsed variation (3.59 ±0.98) is the rising tone. It should be noted that the level variation (2b) was very close in rating (3.53 ±0.98) and is not significantly different from the level contour. Version c, with an upward contour is least preferred, with a mean rating of 2.34 ±1.00. Interestingly, a drastic melodic drop, as shown in version a (mean rating 2.375), is judged to be equally unpleasant (2.38 ±0.83, not significantly different).

The contours for tones 1 and 5 would also be expected to be viewed similarly as they both consist of a tone 4, the lowest tone, followed by a higher tone (tone 1 is the highest tone and tone 5, the mid-rising tone, should be treated as a mid-level tone). In both cases, the participants' highest ranking choice matches the panel's predicted choice (tone 1: 3.84 ±0.92; tone 5: 2.97 ±0.78). Equally, the least preferred variation was one with a flat musical contour (tone 1: 2.31 ±0.90; tone 5: 2.44 ±0.80). The participants' lowest ranking for tone 5 is not significantly different from the panel's choice.

The contour for tone 4 goes from the highest tone (tone 1) to the lowest tone (tone 4). The musical melody that is ranked highest (3.53 ±0.65) is the one with the largest melodic drop - larger even than the interval predicted by the panel. The lowest ranked was the level contour (2.34 ±0.90). One very interesting observation coming from this set of variations is that the highest and lowest ranking versions are actually the same note, just an octave apart. This suggests that contour is more important to the realization of tone in music than harmonic structure. The exact interval, however, also seems not to be strongly preferred. This can be seen in the results from tone 1 where an equivalent tone separation (lowest to highest) does not show a preference for an octave.

One unusual finding not noted in Figure 2 was the unexpected high ranking of one of the distractors. The two full distractor phrases were created with one native speaker writing two short phrases in the same style as the stimuli and another speaker writing a melody without any reference to the words and the two were simply put together on the assumption that this would produce a randomly mismatched musical phrase. One of these was, by far, the lowest ranked jingle (1.78 ±1.01). However, the other one ended up being surprisingly popular with a mean score of 2.88 ±1.43 (but note the large standard deviation).

The results from the present study support the prediction that native Cantonese listeners will prefer musical contours that match the tonal contour of the spoken tones of the language. The results also suggest that Cantonese listeners engage with the simplified tone system used by composers in ways which match the intended representation of tone.

5. ACKNOWLEDGEMENTS

We would like to thank our composer Jonathan Chan and our singer Ernest Tse for their efforts in producing the stimuli. We would also like to thank all the participants of this experiment.

APPENDIX

5. REFERENCES

[1] Yung, B. (1983). Creative process in Cantonese opera I: The role of linguistic tones. *Ethnomusicology*, *27*, 29-47.

[2] Chan, M.K.M. (1987). Tone and melody interaction in Cantonese and Mandarin songs, *UCLA Working Papers in Phonetics*, *68*, 132-169.

[3] 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.

[4] Zhang, L. (2013). *Intonation effects on Cantonese lexical tones in speaking and singing*. Ph.D. Dissertation. Hong Kong Polytechnical University.

[5] Schellenberg, M., Gick, B. (2018). Microtonal Variation in Sung Cantonese. *Phonetica*, 1-24. Accessed online from https://www.karger.com/ Article/PDF/493755.

[6] Yiu, S. S. Y. (2013). Cantonese tones and musical intervals. In *Proceedings of the International Conference on Phonetics of the Languages in China, ICPLC-2013.*

[7] Yiu, S. S. Y. (2014). Linguistic and musical scaling of Cantonese tones. In *Proceedings of the ICMPC13-APSCOM5 2014 Joint Conference: The 13th International Conference of Music Perception and Cognition and the 5th Conference for the Asian-Pacific Society for Cognitive Sciences of Music.* The College of Music, Yonsei University.

[8] Lau, E. (2010). Tone-Melody Relationship in Cantonese. *Working Papers in Linguistics: University of Hawai'i at Manoa 41*(3), 1-12.

[9] Audacity Team. (2017). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.2.1 retrieved 20 December 2017 from https://audacityteam.org/

[10]Final Cut Pro X. 10.1.2. June 27, 2014. Apple Inc.. USA

[11]Mathôt, S., Schreij, D., Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324.

[12]Boone, H. N., Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, *50*(2), 1-5.