ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2006
Dresden, Germany
May 2-5, 2006

Realization of Cantonese Rising Tones under Different Speaking Rates

Ying Wai WONG

Department of Linguistics and Modern Languages
Chinese University of Hong Kong
ywwong@cuhk.edu.hk

Abstract

The two Cantonese rising tones, high-rising and low/mid-low rising tones, are found to maintain their distinct *slopes of F_0(fundamental frequency)-rise* and *offset F_0* under different speaking rates. This suggests the two as possible acoustic cues for rising tone discrimination. The rising contours, under whichever speaking rate, reside in area temporally near the syllable offset. Furthermore, through tests with different alignment methods, the rising contours are found to show the most significant overlap when aligning with offset of the host syllable. Finally, discussions on characterization of rising tones within the Target Approximation (TA) model are presented.

1. Introduction

1.1. Cantonese rising tones

Among the six tonal categories in Cantonese, there are two rising tones, namely *high-rising* tone-2 (T_2) and *low/mid-low rising* tone-5 (T_5). In canonical form, they differ by both their onset and offset F_0 (fundamental frequency) according to the impressionistically determined tone-letter labels by Chao [3]: 35 for T_2 and 23 for T_5. Following that, several more proposals were put forward [5, 9], similarly labeling the rising tones with different onset and offset tone-letters. Later on, Bauer [1] made use of acoustic measurement to propose a re-naming of T_2 as 25, limiting the difference between the two rising tones to only the offset F_0 values.

The situation becomes more complex in case of connected speech, as a couple of other factors such as intonation and inter-syllabic influence participate in determining the final F_0 contours. According to previous studies reported for a variety of languages [2, 4, 11], in continuous speech, rising tones surface phonetically as rising contours approximately during the second half of their host syllables.

1.2. Rate effect on speech production

Oral speech is composed of sequential phonetic segments. When speaking rate changes, which occurs typically during conversation, temporal organization of component segments changes accordingly. However, the change is not just simply accomplished by linear scaling, with individual phonetic components varying inverse-proportionally with the speaking rate in the time domain. In Cantonese, for example, Zee [15] showed that for the syllable [ta], the duration reduction is contributed mostly by the vowel in fast speech as 63.40% and 16.13% durational reduction of respectively the vowel and the consonant were reported. Voice onset time (VOT), a factor for consonant voicing distinction, was found to become longer for /p/ when speech slows down [7].

When it comes to phonetic realization of rising tones, we are not sure how it interacts with speaking rate. As previously suggested, rising tones are characterized by a rising portion of F_0 in their phonetic manifestation. To implement rising tones under different speaking rates, there are at least two possible strategies, yet to be determined: (1) Agreeing with the tone-letter labels, onset and offset F_0 points of the rising portions are fixed based on the associated tone (e.g. 25 for T_2), such that shortening and lengthening of the host syllable due to different speaking rates linearly scales up or down the slope of rising; (2) The slope of F_0-rise instead is kept somewhat constant for a given tonal specification, in spite of change of speaking rate. One possible consequence of this strategy is that onset/offset F_0 of the rising portion vary across different speaking rates due to different syllable timing.

This study seeks answer to the above query by investigating the phonetic behavior of the two Cantonese rising tones in continuous speech under different speaking rates, with particular focus on the rising portion of the F_0 contours.

2. Method

2.1. Stimuli

In the study, we take "*ngo5 ji4 gaa1 duk6 __ zi6 bei2 nei5 teng1*" (Now, I read to you the character __) as the test sentence, where the underlined test word is a CV syllable *se* associated with either T_2 or T_5. To facilitate acoustic investigation of the test syllable, it was embedded between oral tract closures induced by a preceding stop /k/ (from the closed syllable *duk6*) and a following affricate consonant /ts/ (from the syllable *zi6*), as they can provide clear acoustic landmark for segmentation. However, according to an informal pre-test recording, the preceding stop /k/ diminished quite often in fast speech. To maintain a clear acoustic boundary for accurate syllable segmentation, a fricative consonant /s/ and thus the syllable *se* was used. The loss of initial portion of F_0 contour during voiceless frication as a result is justified by our initial observation that the rising portion, which is the focus of this study, resides well inside the voiced part (approximately the 2nd half) of the host syllable. Also, based on [13], the overall contour, except a brief F_0-raising at onset of the voiced portion, aligns quite well across syllable types. Thus, observations from our study are expected to apply to other syllable structures as well.

2.2. Procedure

There were altogether 3 male and 1 female native Cantonese speakers participating in our study. Recordings were conducted in a quiet room, with a Sony ECM-MS957 electret condenser microphone. To control the speaking rate, a series of *pace-keeping clips* were presented to the subjects during the course of recording. The pace-keeping clips consisted of

10.21437/SpeechProsody.2006-78

20 “ding” sounds linked up by silence intervals of duration from 1 to 3 seconds, in the step of 0.5 second (i.e. 5 resultant rates). Subjects were instructed to make their best effort to articulate the given test sentences during those silence intervals. Consequently, 20 tokens from one such pace-keeping clip, multiplied by 2 tones and 5 rates presented in random order, 200 utterances were collected per subject.

Recordings were done using the software PRAAT at a sampling rate of 22 kHz. Automatic labeling by PRAAT was followed by manual rectification on missing or wrongly labeled cycles, especially at C-V or V-C boundaries. After discarding 16 wrongly pronounced utterances, we obtained 784 tokens in total for analysis. From each test syllable, 20 samples of mean F_0 were extracted at temporally equally-spaced 5% points for analysis.

3. Results

3.1. Duration

Table 1 shows the obtained averaged syllable length for different subjects. Generally speaking, the pace-keeping clips are successful in roughly controlling the speaking rate of sentence production, with only a few exceptions for Subjects 2 and 4. Due to different speaking style, resulting syllables are observed to have considerably different range of duration across subjects, for instance, Subject 1 produced syllables with the shortest averaged duration while the longest measure is from Subject 2.

Table 1: *Mean duration (in ms) of test syllables recorded with pace-keeping clips of different duration of silence intervals d_s (1.0s to 3.0s).*

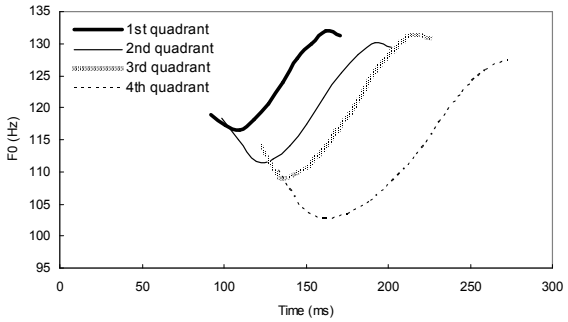
d_s	Subject 1 (M)		Subject 2 (F)		Subject 3 (M)		Subject 4 (M)	
	T_2	T_5	T_2	T_5	T_2	T_5	T_2	T_5
1.0s	180	163	250	294	167	170	240	225
1.5s	191	193	305	322	194	201	253	264
2.0s	229	242	352	364	201	230	278	254
2.5s	230	244	348	348	240	271	302	301
3.0s	288	261	391	390	340	324	336	329

3.2. Realization of rising portion

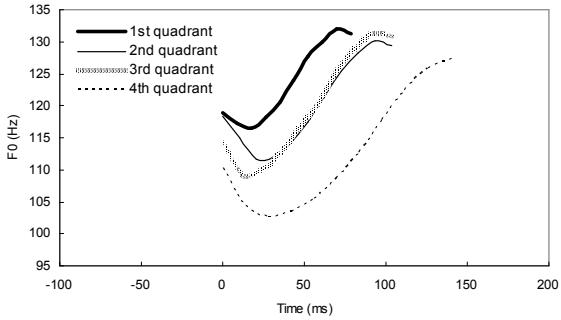
Fig. 1 and 2 show averaged F_0 contours of T_2 and T_5 test syllables produced by Subject 1 with different syllable duration (thus different speaking rates). Judging from Table 1, not all subjects succeeded in producing syllables of duration with a 5-way distinction as the pace-keeping clips indicated. Thus, we finally choose to sort syllables in ascending order of actual syllable duration and divide them into 4 quadrants with balanced number of samples (24 or 25 per group). Excluding frication portion due to the onset consonant /s/, during which F_0 is invisible, F_0 contours are plotted with different alignment respectively in: (a) syllable onset, (b) rhyme onset and (c) syllable offset. Due to space limitation, only data from Subject 1 are presented graphically in this paper, and whatever discussion concerns, we imply similar behavior is exhibited by other subjects as well, unless otherwise specified.

It can be observed that each F_0 trace similarly consists of an initial dip, followed by a rising portion, and ends with a final fall. The initial dip and the final fall reflect the F_0 perturbation caused respectively by preceding voiceless fricative /s/ and following voiceless affricate /ts/ [13]. Inspecting medial portion of each visible F_0 trace, rising tones in Cantonese employ the second strategy mentioned in the Introduction Section in that the slope remains largely constant for the same tone, regardless of the speaking rate. Generally speaking, higher speaking rate corresponds to shorter rising portion. Within it, *offset F_0* (of the rising trajectories) is more stable for the same given tone, compared to the onset F_0.

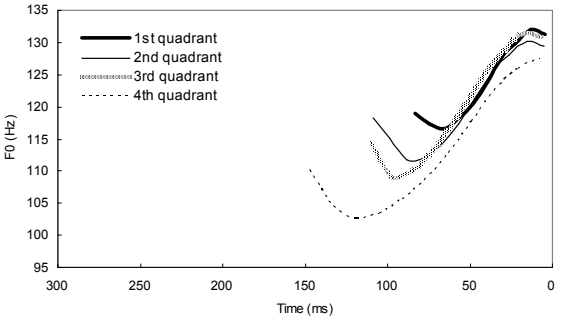
Next, our plot of the rising contours with different aligning methods reveals that for both rising tones, alignment with the syllable offset yields the best match between curves produced with different speaking rates. In each of Fig. 1c and 2c, the 4 F_0 traces have rising portion roughly overlapped along a slant line. It may be concluded that Cantonese rising contours are produced aligned with the syllable offset, regardless of the speaking rates, agreeing with Mandarin results [12] previously reported.



(a) Syllable onset

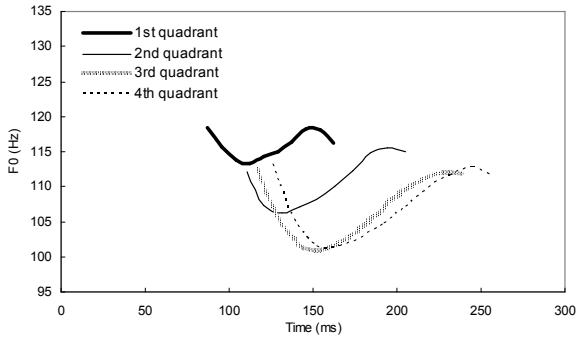


(b) Rhyme onset

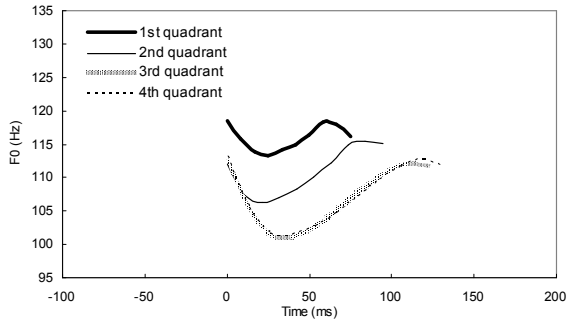


(c) Syllable offset

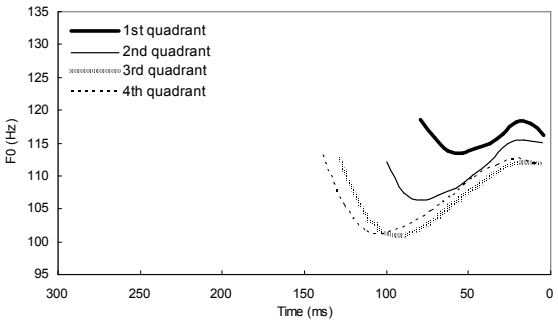
Figure 1: *F_0 traces of T_2 test syllables with different alignment methods.*



(a) Syllable onset



(b) Rhyme onset



(c) Syllable offset

Figure 2: *F_0 traces of T_5 test syllables with different alignment methods.*

To evaluate our previous observations by means of statistical analysis, the two impressionistic measures *slope of F_0-rise* and *offset F_0* have to be extracted acoustically. F_0, as an acoustic correlate of various physiological adjustments (e.g. vocal cord tension) varying in a continuous manner, inevitably exhibits overlaps and transitions between different stages, like the "F_0-perturbation >> rising F_0 >> F_0-perturbation" sequence in our case. To obtain a reasonable estimate of the *slope of F_0-rise*, we resort to extract it by linear regression, excluding those onset and offset portions due to F_0 perturbation from surrounding voiceless regions. For *offset F_0*, it is determined as the maximum points of the rising trajectories near the offset portion of each F_0 contour.

With those acoustic measurements, we now conduct two-factor repeated measures ANOVA tests with (1) *syllable length* (i.e. speaking rate), grouped into 4 quadrants, and (2) *tone* (T_2 or T_5) as the independent variables to investigate their effect on *slope of F_0-rise* and *offset F_0* separately. Results indicate that there is no main effect for *syllable length*, $F(3, 9) = 2.44$, $p > 0.05$ (with *slope of F_0-rise* as the dependent variable), $F(3, 9) = 0.48$, $p > 0.05$ (with *offset F_0* as the dependent variable). These lead us to the conclusion that *slope of F_0-rise* and *offset F_0* are largely invariant for a given tone at different speaking rates.

4. Discrimination of rising tones

Comparing Fig. 1 and Fig. 2, *slope of F_0-rise* and *offset F_0* appear to be quite distinct between the two rising tones. The two-factor repeated measures ANOVA tests conducted in previous section also give support by showing a main effect for *tone*, $F(1, 3) = 32.96$, $p < 0.05$ (with *slope of F_0-rise* as the dependent variable), $F(1, 3) = 68.53$, $p < 0.05$ (with *offset F_0* as the dependent variable). This indicates that *slope of F_0-rise* and *offset F_0* are possible acoustic cues for distinguishing between the two Cantonese rising tones in continuous speech, regardless of the speaking rate.

Results for the situation across speakers are shown in Fig. 3, where each data point represents the computed slope of F_0-rise of syllables from tokens within a quadrant, similar to the grouping used in Results Section.

Referring to the figure, for each speaker, the two rising tones reside in distinct regions along the *slope of F_0-rise* axis (*y*-axis), illustrating its aforementioned discriminating role between the two rising tones. More complication is introduced when inter-speaker variations are taken into consideration. Being a female, Subject 2 produced the rising tones with scaled up F_0 range and thus slopes for T_2 and T_5 syllables, compared to other male subjects. As a result, there is an up-shift in the categorial boundary. Worse still, even for Subjects 1, 3 and 4, all being male, behave slightly differently in terms of the slope boundary. This points to the necessity to employ yet other strategies, for instance, speaker normalization [6, 8, 10] in tone perception to compensate those inter-speaker variations. Similar observations can be obtained for *offset F_0*, but with even greater categorial boundary shift across sex, as illustrated in Fig. 4.

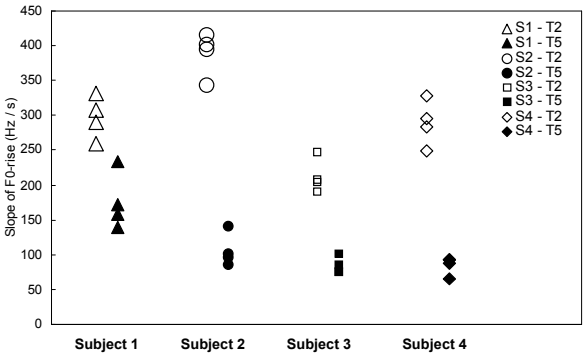


Figure 3: *Slope of F_0-rise from linear regression across speakers.*

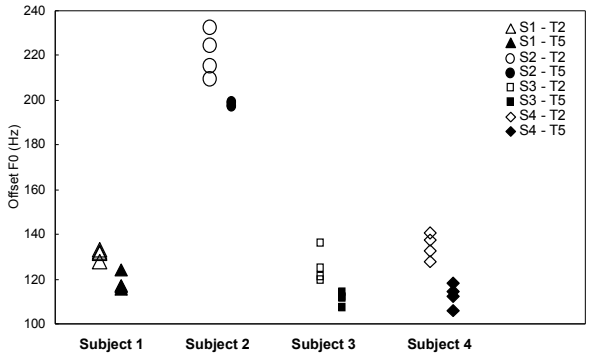


Figure 4: *Offset F_0 across speakers.*

5. Characterization of rising tones

Xu & Wang [14] proposed the *Target Approximation* (TA) model in which surface F_0 contours are argued to be the result of continuous and asymptotic approximation of F_0 to underlying pitch targets. In particular, F_0 traces of rising tones correspond to approximation to *dynamic* targets, i.e. a slant line on a *time-F_0* graph representing F_0 movement. We may reveal some properties of dynamic targets from Fig. 1c and 2c. Within the TA framework, those overlapping rising F_0 portion are interpreted as the surface trajectories in approaching underlying dynamic targets. Up till now, we have shown at least two candidate parameters in characterizing a rising tone: *slope of F_0-rise* and *offset F_0*, which are rather distinctive for a given tone. As a result, a minimal specification of dynamic targets should include at least these two parameters. These concepts are illustrated in Fig. 5, which is a snapshot from Fig. 1c. The estimated dynamic pitch target (thick dotted line), along with other possible outputs (solid grey lines) by varying *slope of F_0-rise* (i.e. the angle at which the slant lines intercept the *x*-axis) and *offset F_0* (i.e. placement of the slant lines along the *y*-axis) are shown. This two-parameter specification of dynamic target is also capable of explaining rising tone contours under different speaking rates, given previously discussed invariance of those two properties in Results Section.

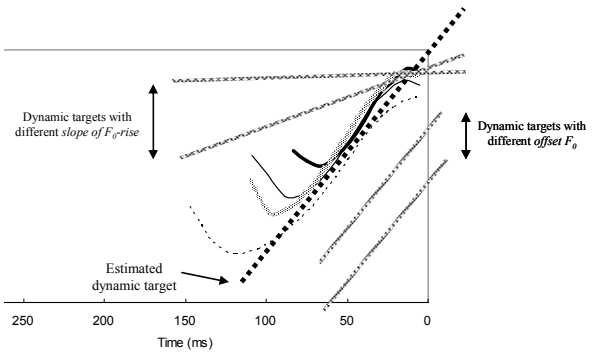


Figure 5: *Dynamic pitch targets by varying the two parameters: "slope of F_0-rise" and "offset F_0".*

6. Conclusions and further work

In this paper, we investigate the interaction of speaking rate and phonetic realization of the two Cantonese rising tones. Among many phonetic features, *slope of F_0-rise* and *offset F_0* are found to be largely invariant under different speaking rates, possibly contributing to tonal identification. Statistical analyses confirm that these two measures can clearly distinguish between the two rising tones within the same speaker. However, queries like whether both of them are employed by our perception system, or which is the primary cue for tone discrimination, still await further experiment. Finally, our data also reveals two important parameters in specifying dynamic targets for Cantonese rising tones within the TA model.

7. Acknowledgement

We are grateful to Yi Xu and Margaret Lei for their suggestions on an earlier version of this paper. Also we thank Thomas Lee, William Wang and Eric Zee for discussions on background issues of the study.

8. References

[1] Bauer, R., 1998. Hong Kong Cantonese tone contours. *Studies in Cantonese Linguistics* 1-33.

[2] Chang, C. Y., 2003. *Intonation in Cantonese*. LINCOM Studies in Asian Linguistics, vol. 49, Munich: Lincom Europa.

[3] Chao, Y. R., 1947. *Cantonese primer*. Greenwood Press: New York.

[4] Gandour, J. T.; Potisuk, S.; Dechongkit, S., 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22, 477-492.

[5] Hashimoto, O.-K. Y., 1972. *Studies in Yue dialects 1: Phonology of Cantonese*. Cambridge University Press.

[6] Leather, J., 1983. Speaker normalization in perception of lexical tone. *Journal of Phonetics* 11, 373–382.

[7] Miller, J. L.; Green, K. P.; Reeves, A., 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43, 106-115.

[8] Moore, C. B.; Jongman, A., 1997. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America* 102, 1864–1877.

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

[10] Wong, P. C. M.; Diehl, R. L., 2003. Perceptual normalization of inter- and intra-talker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research* 46, 413-421.

[11] Xu, Y., 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.

[12] Xu, Y., 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179-203.

[13] Xu, Y.; Wallace, A., 2004. Multiple effects of consonant manner of articulation and intonation type on F0 in English. *Journal of the Acoustical Society of America*, 115, Pt. 2, 2397.

[14] Xu, Y.; Wang, Q. E., 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33, 319-337.

[15] Zee, E., 2002. The effect of speech rate on the temporal organization of syllable production in Cantonese. In *SP-2002*, 723-726.