Contextual Tonal Variations and Pitch Targets in Cantonese

Ying Wai WONG

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

Abstract

With Cantonese as the target language, this study investigates the phonetic details of contextual tonal variations in disyllabic tonal sequences. It is found that the main source of F_0 (fundamental frequency) contour deviation from the canonical form comes from carryover effect, which is assimilatory in nature. Furthermore, based on the Target Approximation (TA) model, an optimization problem is formulated as an attempt to unveil mathematically pitch targets of the six lexical tones in Cantonese. Finally, implications of our results on tone production and perception are discussed.

1. Introduction

1.1. Contextual tonal variations

F_0 in tone languages serves as an essential cue for identification of individual lexical items. While syllables of different tones in a particular tone language are always associated with quite distinct F_0 contours when produced in isolation, those F_0 contours show much deviation from their canonical forms in continuous speech depending on neighboring tones, as reported for Mandarin [9], Thai [3], Vietnamese [4] and Cantonese [2, 6].

In particular, for Mandarin continuous speech, Xu & Wang [11] proposed the Target Approximation (TA) model, which states that for Mandarin, each syllable is associated with an underlying pitch target, and the surface F_0 curves in turn show speakers' effort to approximate corresponding pitch targets of individual syllables.

1.2. Cantonese tonal inventory

Cantonese has 6 long tones, including 3 level tones (high level T_1, mid level T_3 and mid-low level T_6), two rising tones (high rising T_2 and mid-low rising T_5) and one falling tone (mid-low falling T_4). Traditionally there are 3 more short tones (T_7, T_8 and T_9) which are associated only with syllables closed with /p/, /t/ and /k/. However, as those 3 tones have F_0 contours observed to be abbreviated versions of the 3 level tones (T_1, T_3 and T_6), only 6 basic tones are in Hong Kong Cantonese [1]. For example, *Jyutping*, the romanization system adopted by LSHK (Lingusitic Society of Hong Kong), denotes T_7 to T_9 as T_1, T_3 and T_6 respectively.

Concerning the onset and offset F_0, which are important for contextual tonal variation discussion, T_2, T_4, T_5 and T_6 have relatively low onset F_0; T_1 and T_2 have relative high offset F_0. A typical trace of F_0 contours for the Cantonese lexical tones is shown in Fig. 1 (from Peng and Wang [7]). The heavy overlap between T_7 to T_9 and their long counterparts are quite obvious.

In the following sections, we will present a study of contextual tonal variations in Cantonese, followed by a mathematical computation of pitch targets [11], and finally some discussions.

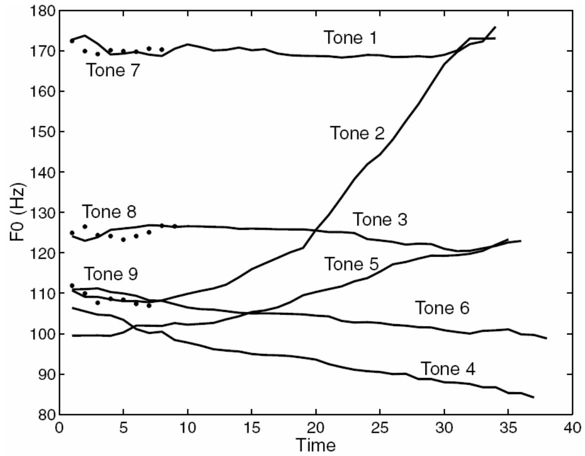


Figure 1: *F_0 contours of Cantonese lexical tones by a male speaker (from Peng & Wang [7]).*

2. Method

2.1. Material

To study contextual variations of tones, previous studies like [2, 3] opted for natural lexical items, inevitably introducing F_0 perturbation from intrinsic pitch across consonant and vowel types [5]. To minimize this undesirable effect, nonsense disyllabic sequences are used in this study. As our focus is on the interactional effect on F_0 contours by adjacent tonal pairs, we use syllables which can show continuous F_0 contours throughout the whole duration. As a result, oral stops and fricatives, which are all voiceless in Cantonese, are ruled out as the candidate. Furthermore, as there is a widely attested free variation between velar nasal stop and null onset consonant, as well as a trend to substitute /l/ for /n/, we finally come out with *lau* (in Jyutping) as the tone-carrying syllable. The disyllabic sequence *lau-lau* is embedded in a carrier sentence "*keoi5 heoi3* ___ *leoi5 hang4*" (He/she goes to ___ for a trip) for tone production task. Exhausting all the six lexical tones in Cantonese, 36 combinations are obtained.

2.2. Recording

4 university graduates (3M1F), who are native Cantonese speakers, participated in the study. Recording sessions were carried out in a quiet room, with a Sony ECM-MS957 electret condenser microphone. In each session, 36 combinations of test sentences were presented to each subject in random order. As all the test sentences contain the disyllabic sequence *lau-*

lau which is non-existent in native Cantonese lexicon, subjects were requested to produce a given sentence with successive repetitions so as to enhance the fluency of sentence production. Recording sessions were monitored by the author and a repetition was requested whenever an utterance was judged incorrect. A total of 5 repetitions for each combination were obtained. The utterances were all recorded in PRAAT, digitized at a sampling rate of 22 kHz for analysis.

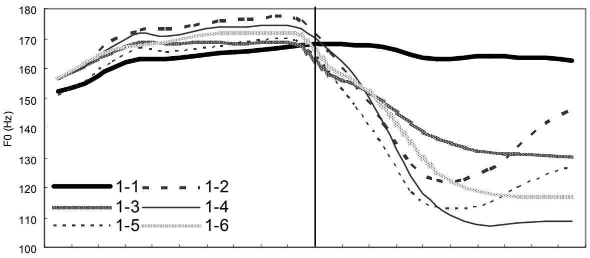
2.3. F_0 extraction

After syllable segmentation, F_0 extraction was carried out in PRAAT, with hand labeling in case of missing vocal pulse cycles by machine labeling, for instance, during creaky voice portions associated with male mid-low falling T_4 in Cantonese. Each syllable was divided into 10 equal parts and average F_0 were obtained for each part. Duration was ignored and all syllables were time-normalized for the current study.

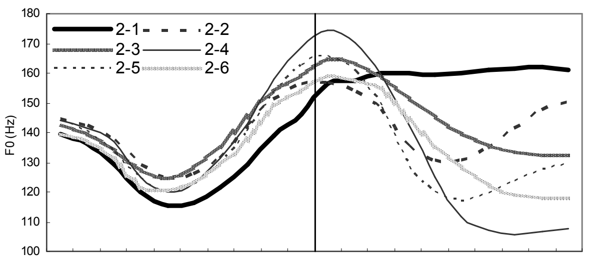
3. Results

3.1. Anticipatory effect

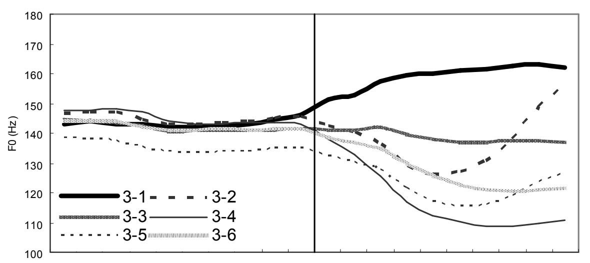
Fig. 2 show the pooled results of F_0 contours of the six tones, each followed by different tones in the testing sentences. To facilitate discussion, we use S_1 and S_2 to denote respectively the two syllables in each graph from now on.



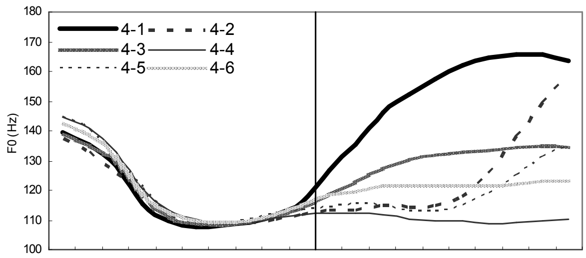
(a) T_1-T_x



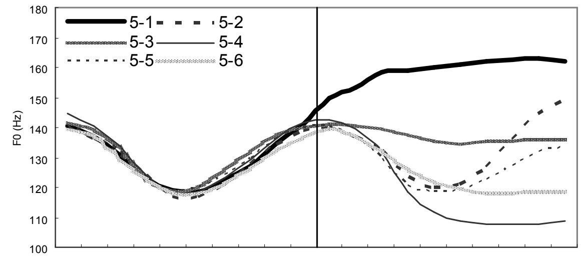
(b) T_2-T_x



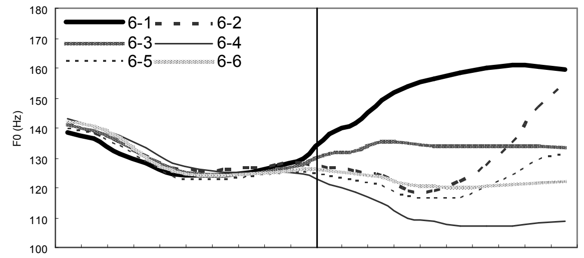
(c) T_3-T_x



(d) T_4-T_x



(e) T_5-T_x



(f) T_6-T_x

Figure 2: *F_0 curves for disyllabic (S_1-S_2) sequences with fixed S_1 (T_1 to T_6).*

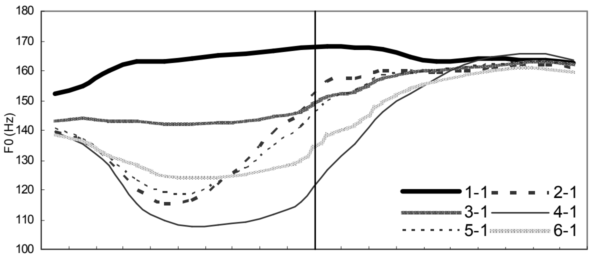
Generally speaking, the anticipatory effect is not large, as observed in the great overlap of F_0 curves of S_1 in Fig. 2c-f. Among the figures, Fig. 2a illustrates most clearly the nature of anticipatory effect. F_0 curves of S_1 are arranged in descending F_0 levels, in the order of $T_2 > T_4 > T_6 > T_3 > T_5 > T_1$ of S_2. S_1 preceding T_2, T_4 or T_6, three tones having relative low onset F_0, show up as having higher F_0 than S_1 preceding tones associated with relative high onset F_0, like T_3 and T_1 in Fig. 2a. For Fig. 2b-f, as there are heavy overlaps as well as crossing points between F_0 curves of S_1, it is less easy to observe the trends. To simplify the picture, we only consider two extremes of relative onset F_0 values by comparing the anticipatory effects due to T_1 (high onset) and T_4 (low onset). We find that the onset F_0 values of S_1 preceding T_4 are always higher than that preceding T_1, while the duration of this difference varies across tones. The difference in F_0 levels are true for the whole duration in case of S_1 being T_1 and T_2 (Fig. 2a-b), while for other tones, around 75% from the onset (Fig. 2c-f), where intersections between S_1-T_1 and S_1-T_4 curves can be observed.

In sum, Fig. 2a-f illustrate the dissimilatory nature of anticipatory effect, agreeing with results previously reported [2, 3, 6, 9].

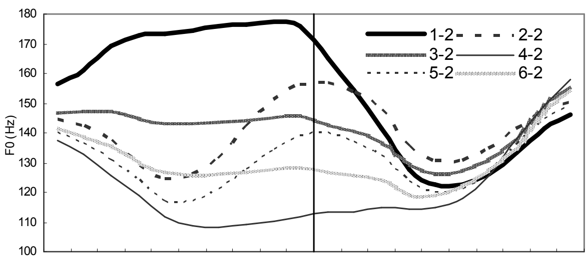
3.2. Carryover effect

Fig. 3 show the pooled results of F_0 curves of the *lau-lau* sequences grouped by S_2. Compared to anticipatory effect, carryover effect is observed more neatly as assimilatory in nature. As an illustration, onset F_0 of S_2 following T_1, T_2 (two tones with relatively high offset F_0) are always raised higher than the case for other preceding tones. Also, T_4, having the lowest offset F_0 in isolation (Fig. 1), always lowers the onset F_0 of S_2 the most. Setting standard deviation of 5Hz as the threshold, we find that carryover effect goes into 50% of S_2. Arranging in ascending order of variation in terms of averaged standard deviation (ASD) values of S_2, we get the sequence: $T_1 < T_3 < T_5 < T_6 < T_2 < T_4$. Given the fact that T_5 and T_6 have very close values of ASD (7.16 Hz VS 7.19 Hz), we find that in Cantonese, level tones are the least susceptible to carryover effect, whereas the falling tone is the most susceptible.

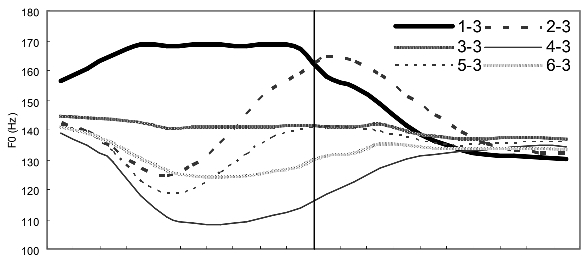
As presented, carryover effect is much larger than anticipatory effect. As a result, for any given tone, the most consistent part of the F_0 contour is in its offset portion, and this led to the Target Approximation (TA) model [11]. According to this model, F_0 contour is the surface manifestation of asymptotic approximation of a series of pitch targets, each associated with a corresponding syllable. From Fig. 3, those hypothesized pitch targets show up as the converging contours in S_2.



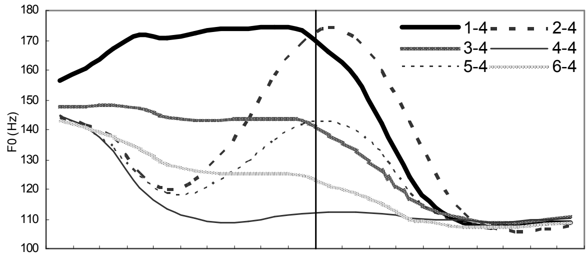
(a) T_x-T_1



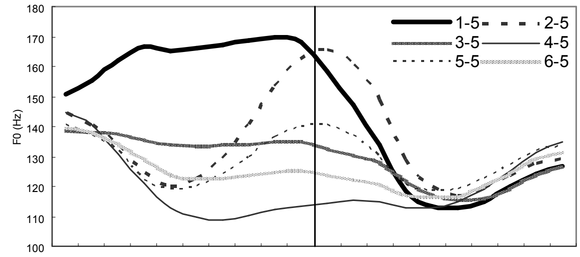
(b) T_x-T_2



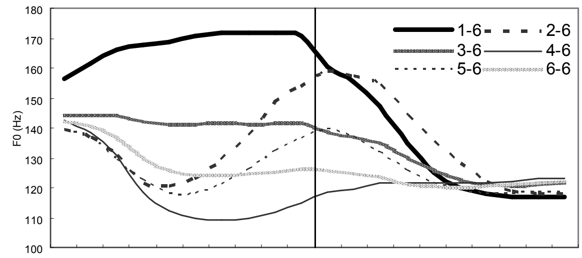
(c) T_x-T_3



(d) T_x-T_4



(e) T_x-T_5



(f) T_x-T_6

Figure 3: *F_0 curves for disyllabic (S_1-S_2) sequences with fixed S_2 (T_1 to T_6).*

By visual inspection, F_0 curves of the three level tones T_1, T_3 and T_6 converge to lines parallel to the *x*-axis, while for the two rising tones T_2 and T_5, F_0 curves converge to two lines with different positive slopes, with T_2 being steeper. Surprisingly, for the canonically falling T_4 (Fig. 3d), F_0 curves converge to again a level line, similar to those three level tones. In other words, according to the TA model, T_4 is observed to have a level pitch target, instead of a falling one as in its isolated form.

4. Mathematical computation of pitch targets

Based on the TA model, we try to use regression analysis technique to work out numerically pitch targets for the six Cantonese tones. Assuming one pitch target per tone, the asymptotic approximation [11] of F_0 curve function $P(t)$ can be expressed mathematically [8] as

$$T(t) = mt + c \quad (1)$$

$$P(t) = \beta \exp(-\lambda t) + T(t) = \beta \exp(-\lambda t) + mt + c \quad (2)$$

where β and λ are rate parameters for the approximation action and $T(t) = mt + c$ is the underlying pitch target.

For a transition from T_i to T_j, $(i, j = 1, 2, \ldots, 6)$, let $F_{ij}(k)$ be the mean F_0 of the k^{th} $(k = 1, 2, \ldots, 10)$ sub-part of S_2 obtained for Fig. 3. Applying (2), the F_0 values predicted by the TA model at the corresponding points $P_{ij}(k)$ can be expressed as

$$P_{ij}(k) = \beta_{ij} \exp(-\lambda_{ij} k) + m_j k + c_j \tag{3}$$

where β_{ij}, λ_{ij} are rate parameters for the transition from T_i to T_j while m_j, c_j are parameters of the underlying pitch target of T_j. To estimate the pitch targets, for each tone T_j, we apply least-square fitting algorithm to minimize the error term

$$E_j = \sum_{i=1}^{6} \sum_{k=3}^{10} \left| P_{ij}(k) - F_{ij}(k) \right|^2 \tag{4}$$

From Fig. 3, we can observe F_0 peak delay due to overshooting [11] for the two rising tones T_2 and T_5 from S_1, which is not accounted for by (3) at the current stage. To avoid them, we only consider F_0 points from the 3^{rd} to the 10^{th} sub-parts in actual calculation, such that the index of k in (4) counts from 3 to 10. The six pitch targets thus obtained are shown in Fig. 4. Basically, the resulting pitch targets agree with visual inspection in terms of slope. More interestingly, the computed pitch target for T_4 has a slope very close to the three level tones, similar to visual inspection, contrasting with traditional classification of them into separate level and falling tones.

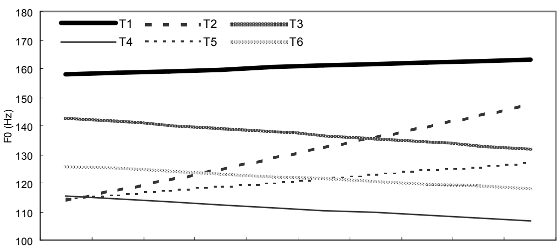


Figure 4: *Pitch targets computed by minimizing (4).*

5. Discussion and future work

The tonal variations investigated in this study are entirely phonetically based, contrasting with phonological variations of tones (tone-sandhi) widely attested in various languages, which may be syntactically or morphologically driven.

While with a lateral consonant, the *lau-lau* sequence in this study can provide continuous F_0 curves, the situation is unknown in case of oral stops and fricatives (all voiceless in Cantonese). First, those consonants leave substantial length of undetectable F_0 regions within the syllables; Second, different consonant types cause different degree of F_0 perturbation [5]. Similarly, entering tones (traditionally T_7, T_8 and T_9) shorten the voiced duration of syllables. Whether the F_0 curves still align with the whole syllable or just the voiced portion is left for further research. Preliminary results on syllable-final nasals [10] indicate it is the former case.

From the observation that the relative degree of variation is much larger for carryover effect than that for anticipatory effect, the later portion of a syllable shows more consistent patterns for the same tone. One possible consequence is relatively heavier reliance on that later portion of syllable in perceiving tonal identities of syllables in continuous speech, though further works have to be carried out to confirm this.

To investigate a single factor: *tone*'s effect on contextual variations, durational differences between tones were removed by time-normalization. This creates possible confusions in Cantonese, where the two rising tones (T_2, T_5) differ in terms of slope of F_0 rise. Xu [10] provided a preliminary solution for Mandarin by discussing the invariance of slope of rise of Mandarin tone-2 under different speaking rate. Parallel work for Cantonese has yet to be done. Furthermore, timing information has to be taken into consideration before the mathematical formulation results can be used in applications like speech recognition and synthesis.

6. Conclusions

With disyllabic tonal sequences, our study shows directional asymmetries in terms of nature and magnitude of contextual tonal variations in Cantonese. Carryover effect is assimilatory whereas it is dissimilatory for anticipatory effect. Furthermore, the former one induces larger deviation of F_0 contours from the canonical form.

In addition, a non-linear optimization problem is formulated based on the TA model [11] to work out the hypothesized pitch targets for the six Cantonese lexical tones.

7. Acknowledgement

The author wishes to thank Yi Xu and Thomas Lee for many helpful comments on a variety of methodological and other related issues, and William Wang and Eric Zee for discussions on background knowledge 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] Gandour, J. T.; Potisuk, S.; Dechongkit, S., 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22, 477-492.

[4] Han, M. S.; Kim, K. O., 1974. Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics* 2, 223-232.

[5] House, A. S.; Fairbanks, G., 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25, 105-113.

[6] Liu, J., 2001. *Tonal behavior in some tone languages*. Ph.D. Dissertation. City University of Hong Kong.

[7] Peng, G.; Wang, W. S. Y., 2005. Tone recognition of continuous Cantonese speech based on support vector machines. *Speech Communication* 45, 49-62.

[8] Sun, X. J., 2001. Predicting underlying pitch targets for intonation modeling. *4th ISCA Tutorial and Research Workshop on Speech Synthesis*.

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

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

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