CONSONANTAL PERTURBATION OF F_0 CONTOURS OF CANTONESE TONES†

Ying Wai Wong[1] *and Yi Xu*[2]

[1] Chinese University of Hong Kong, Hong Kong
[2] University College London, UK
E-mail: ywwong@cuhk.edu.hk, yi.xu@ucl.ac.uk

ABSTRACT

A systematic study of F_0 perturbation by voiceless consonants in Cantonese is carried out. Apart from a voiceless interval introduced, a production asymmetry is found: F_0 contours are raised by prevocalic consonants but lowered by postvocalic consonants at the C-V and V-C transitions. Moreover, initial consonants are found to differ in the duration of the voiceless intervals they introduce. Based on the recent finding that F_0 production is synchronized with the syllable, we demonstrate that such durational differences need to be taken into consideration before accurate measurement of F_0 perturbations can be made.

1. INTRODUCTION

The understanding of consonantal perturbation of fundamental frequency (F_0) was significantly advanced by the finding that the characteristics of the effect varied with the underlying intonation contours [11]. Prior to that there were two opposing views regarding the directionality of consonantal perturbation of F_0: (1) the *rise-fall dichotomy* view claims that F_0 falls after voiceless stops but rises after voiced stops (e.g. [7]); and (2) the *no-rise view* claims that F_0 falls after all stops, with the slope being smaller in case of voiced stops (e.g. [9]). None of the earlier studies, however, have taken the underlying F_0 trajectories around the consonants into consideration, even when continuous F_0 contours were obtained, e.g. [4].

By considering also surrounding F_0 contours in a series of perception tasks, Silverman [11], while showing general support for the no-rise view, demonstrated that in certain intonational environment, e.g. in a steeply rising intonation contour through the stop, F_0 could appear to rise after voiced stops. Silverman thus concluded that only by taking into consideration the surrounding F_0 contours related to intonation can one accurately characterize consonantal perturbation.

Another factor that has yet to be taken into full consideration is the fact that obstruent consonants interrupt the otherwise continuous F_0 contours, and such interruptions are likely to also introduce discrepancies in the F_0 measurements depending on the shape of the F_0 contours being disrupted. Furthermore, because the duration of the interruption varies from consonant to consonant, the accuracy of comparison of F_0 perturbation across consonants may also be affected. These difficulties, however, can be eased by taking into consideration the recent finding that the articulation of tones is highly synchronized with the syllable [13, 16, 17]. Such synchronization may provide a reliable reference for aligning the consonants temporally in order to accurately compare their effects on F_0.

Also, previously reported consonantal perturbation is typically about its effect on the following F_0. Little is yet known about the consonantal influence on the preceding F_0 (but cf. [18]). The present paper reports the findings of a tone production study conducted to examine consonantal perturbation of F_0 in Cantonese, a tone language spoken in Hong Kong and the city of Canton.

2. METHOD

To provide a controlled underlying F_0 contour, target syllable is embedded in the carrier sentence, transcribed in *Jyutping*, *ngo5 heoi3 maai5 maa1* ___ *maa1 faan1 uk1 kei2* (I go to buy and bring back home 'maa1 ___ maa1'). With this carrier, the target syllable is placed sentence-medially and phrase-medially (between *maa1* ___ *maa1*) which would minimize reported intonational effects such as sentence-final and phrase-final lowering [6]. Also, the target syllable is surrounded by syllables *maa* ([ma]) with a sonorant onset consonant so as to both obtain continuous F_0 movements and guarantee accurate segmentation [14].

According to Zee [19], there are 19 consonants and 21 vowels in Cantonese. Among them, 12 segmental compositions are chosen, as listed in Table 1. The chosen syllables cover all possible manners of articulation in Cantonese (except approximant, for which accurate segmentation would be a problem) as listed in the first column. Furthermore, syllables are classified into four groups according to F_0 continuity at (1) syllable onset and (2) syllable offset, as indicated by the "+" and "-" symbols in the second row. For the +/- or -/- configurations (i.e. syllables with voiceless offset), voiceless consonants appearing at the postvocalic position is required. Stops [p], [t], [k] are the only qualified coda consonants in Cantonese. Syllables thus formed are listed separately in the last three rows. To focus on consonantal perturbation, the well-known F_0 influence from vowel height [10] is minimized through the use of the central low vowel [a] as much as possible. Finally, each syllable is associated with different tones, either T_1 (high level), T_3 (mid level) or T_6 (mid-low level), shown as numbers in the accompanying brackets. Some combinations (e.g. *paa* with T_6) are absent due to lack of a corresponding Cantonese morpheme, and only level tones are used here to facilitate comparisons with the entering tones. Altogether, there are 29 target syllables, leading to 29 test sentences.

Table 1: Segmental compositions of the target syllables. Numbers in the brackets denote different tones associated with the corresponding syllable.

	F_0 continuity			
	+/+	+/-	-/+	-/-
unasp. stop			*baa* (1, 3, 6)	
asp. stop			*paa* (1, 3)	
nasal	*maa* (1, 3, 6)			
fricative			*sai* (1, 3, 6)	
affricate			*zaa* (1, 3, 6)	
lateral	*lau* (1, 3, 6)			
-[*p*]		*laap* (6)		*daap* (1, 3, 6)
-[*t*]		*laat* (6)		*taat* (1, 3, 6)
-[*k*]		*laak* (6)		*baak* (1, 3, 6)

Six male subjects (including the first author), all native Cantonese speakers, participated in the study. They produced 10 repetitions of each of the 29 sentences in random order. A native Cantonese speaker (the first author) monitored the recording sessions and asked the subjects to reproduce an utterance if it was judged to be incorrect. Recording was done in a quiet room, with electret condenser microphones, at a sampling rate of 22.1 kHz.

3. DATA EXTRACTION

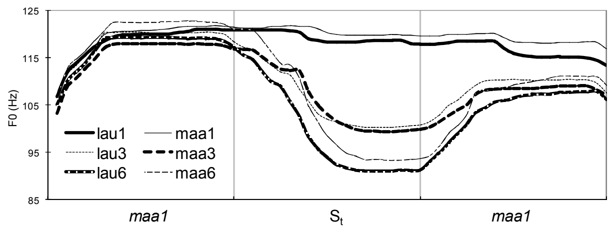
Individual vocal periods were obtained automatically by PRAAT, followed by manual rectification, to compute time-normalized F_0 contours with 20 samples per syllable at 5% temporal steps. The F_0 contours were then trimmed [16] to remove sharp spikes, resulting in the time-normalized F_0 curves presented below.

4. RESULTS

Acoustically, F_0 contours of different tones show appreciable contrasts only at two positions: the target and the subsequent syllables (i.e. ___ *maa1*). To enable closer examination, only the target syllable and two adjacent syllables (i.e. *maa1* ___ *maa1*) are plotted in the following figures. To facilitate discussion, these three syllables are referred to as S_p, S_t and S_f.

To first establish a reference, F_0 contours of target syllables with continuous voicing (i.e. *maa* and *lau*), which are known to generate the least perturbations, are plotted in Fig. 1. Each contour in the figure is obtained by taking the average of 60 contours, 10 from each of the 6 subjects. The three panels show respectively the time-normalized F_0 contours associated with S_p, S_t and S_f. *x*-coordinate of each F_0 point is determined by its normalized time within the host syllable, e.g. the 5th F_0 point in the S_t panel represents mean F_0 value obtained over the 20% - 25% portion of the target syllable. In other words, F_0 contours of different tones are aligned with the syllable. What can be seen clearly there is the effect of tone on F_0. Basically, S_p stays the same across different tones of S_t, surfacing as a gliding movement from around 105Hz (due to a preceding mid-low rising T_5) to around 120Hz followed by a level region; Continuing from there, F_0 contours of S_t diverge to approximate three level tones. The static targets [17] of the three tones become apparent in the second half of S_t. Due to carryover effect from S_t, the first half of S_f forms transition to the high level (T_1) target associated with S_f.

Figure 1: F_0 contours for target syllables *maa* and *lau*, associated with the three level tones.



On the whole, F_0 movements of *lau* match that from a previous study of contextual tonal variations in Cantonese [13]. Moreover, no systematic differences in F_0 can be observed between the two target syllables *maa* and *lau*.

Figure 2: Effect of prevocalic voiceless consonants ([p] in *baa*, [p^h] in *paa*, [s] in *sai*, and [ts] in *zaa*) on F_0.

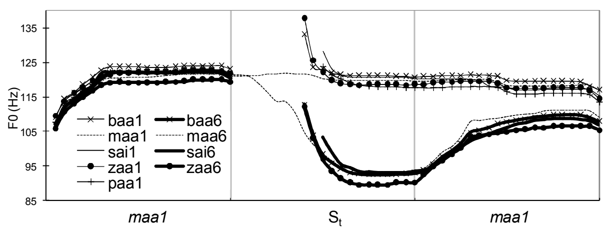
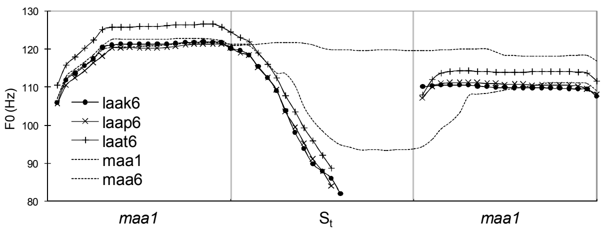


Fig. 2 shows F_0 perturbation by prevocalic consonants in CV syllables with either stop or fricative/affricate onset (i.e. the -/+ configuration). To prevent overcrowding, only F_0 contours for T_1 and T_6 syllables are shown, with F_0 trajectories from *maa* as the reference (thin dash lines). Syllables with the same segmental composition (e.g. *baa1* and *baa6*) are plotted with the same marker while that with the same tone are plotted with the same line thickness.

Briefly speaking, prevocalic voiceless consonants interrupt the otherwise continuous F_0 during the first 40% of total duration of S_t. At the voice onset, F_0 is raised at a similar slope (by visual judgment) across the two tones (syllable durations do not have statistically significant difference across different tones, both for tokens represented in Fig. 2 and for all syllables of the -/+ configuration (including some not shown in Fig. 2, e.g. *baa3*, *paa3*)). Among the consonants, the aspirated stop [p^h] introduces the longest devoiced interval, while the unaspirated stop [p] the shortest. This difference due to onset consonant is statistically significant for T_1 ($F(3, 15) = 13.088$, $p < 0.001$) and T_3 ($F(3, 15) = 20.354$, $p < 0.001$), but marginally for T_6 ($F(2, 10) = 3.759$, $p = 0.061$). A comparison among all the four consonant types reveals that onset F_0 is lowest for the aspirated stop [p^h], but no main effect of *consonant type* is obtained.

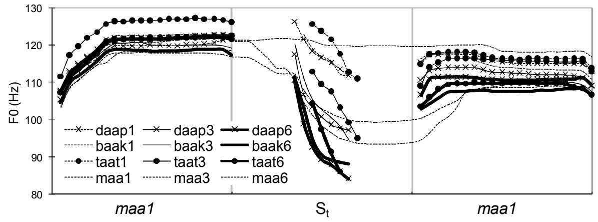
Similar to prevocalic voiceless consonants, about 40% of the total syllable duration extending from the syllable offset is devoiced by postvocalic voiceless consonants [p], [t] or [k], as shown in Fig. 3. However, in contrast to prevocalic ones, voiceless consonants in the coda position lower F_0 relative to the reference *maa* contours. Similar degrees of F_0 lowering are present across F_0 contours, and as a result, by F_0 alone, the three coda stop consonants cannot be reliably distinguished. As only T_6 syllables are available in this +/- configuration set, more contrastive triplets are expected to give finer details on the effect on F_0 resulting from interaction between postvocalic consonants and tone.

Figure 3: Effect of postvocalic voiceless consonants ([p] in *laap*, [t] in *laat*, and [k] in *laak*) on F_0.



Syllables with -/- configuration are plotted in Fig. 4. In general, all F_0 contours are located in the middle of the whole syllable (spanning 30% to 70% of the total duration), appearing falling over time. These surface forms can be viewed as the outcome of both prevocalic F_0-raising and postvocalic F_0-lowering on top of the basic tonal contour. The slope of F_0 contours appears to interact with lexical tone, being steeper for tones at lower F_0 levels (i.e. T_3 and T_6). This visual judgment is verified statistically in a two-factor repeated-measures ANOVA, with *syllable* (different segmental compositions) and *tone* (T_1, T_3 and T_6) as independent factors, and *F_0 slope* (the slope from linear interpolation of the F_0 contours) as the dependent factor. There is a main effect of tone ($F(2, 10) = 10.105$, $p < 0.005$), but not *syllable*, nor interaction of the two. Consistent with the case of -/+ configuration, there is a main effect of *onset consonant* on *duration of prevocalic voiceless interval*, $F(2, 10) = 29.921$, $p < 0.001$.

Figure 4: F_0 contours of syllables with the rhyme vowel surrounded by two voiceless consonants.

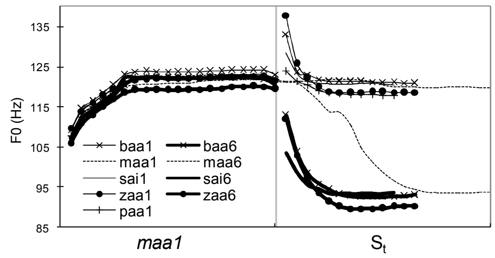


5. GENERAL DISCUSSION

Francis *et al.* [3] studied consonantal perturbation of F_0 due to aspiration in Cantonese and found that F_0 at onset of voicing after unaspirated stops was higher than after aspirated stops (but without statistical significance), which generally agrees with our results (except for T_1 and T_3 in the [t] vs. [t^h] pair where F_0 at voice onset is slightly higher (within 2Hz) for the aspirated [t^h]). A closer examination of Fig. 2 and 4, however, reveals that voice onset F_0 after aspirated stops ([p^h] and [t^h]) are always higher than the corresponding contour for unaspirated ones ([p] and [t]) (though not statistically significant for all pairs). This apparent contrary can be attributed to the alignment method used: In Francis *et al.*'s study, F_0 difference between aspirated vs. unaspirated stops was compared at *onset of voicing*, while comparisons here are based on the assumption of tone-syllable synchronization [17]. Comparing the *baa* ([p]) and *paa* ([p^h]) series in Fig. 2, as well as the *daap* ([t]) and *taat* ([t^h]) series in Fig. 4, we may find that the unaspirated consonants (i.e. [p] and [t]) introduce a significantly shorter prevocalic voiceless interval in the F_0 contours compared to the aspirated ones ($F(1, 5) = 46.548$, $p < 0.005$ for [p] vs. [p^h], $F(1, 5) = 47.595$, $p < 0.005$ for [t] vs. [t^h]) (see Feng [1] for Mandarin data). Consequently, F_0 comparison without considering F_0-syllable alignment can possibly lead to incomplete understanding of consonantal perturbation. An example is given here to illustrate the idea. Prevocalic F_0-raising of 11.42 Hz (for T_1) and 8.12 Hz (for T_6) can be observed in the *baa* ([p]) contour relative to the *maa* ([m]) contour in Fig. 2. If the F_0 contours are aligned with voice onset, as shown in Fig. 5, the differences between voiced and voiceless consonants become 11.67 Hz (for T_1) and -8.18 Hz (for T_6) respectively. Making comparisons at the voice onset leads to an opposite result for T_6 syllables: prevocalic F_0-lowering. Simply put, in addition to surrounding F_0

contours [11], the timing of F_0 contours relative to the syllable is equally important in correct understanding of nature of consonantal perturbations.

Figure 5: F_0 contours from Fig. 2, but aligned at voice onset.



According to the Target Approximation (TA) model [17], a tone is produced by continually approaching its underlying pitch target during the associated syllable, which would result in the best approximation of the target near the end of the syllable, even in the presence of a nasal coda [15]. Perceptually, this would imply more reliable tone information in the later part of a syllable, as has been demonstrated by Wong [12]. These hypothesized targets appear well preserved in Fig. 2, but syllables with +/- configuration (Fig. 3) lose their final portion of F_0 contours due to postvocalic devoicing. Worse still, F_0 of syllables with their rhyme vowel surrounded by two voiceless consonants are perturbed so much that only a short ramp remains visible, as seen in Fig. 4. This may pose a problem for tone identification. Nevertheless, it seems from Fig. 4 that there is still much height difference across these F_0 ramps. A two-factor repeated-measures ANOVA shows a main effect of *tone* ($F(2, 10) = 86.263$, $p < 0.001$), but no effect of *syllable* or *syllable* by *tone* interaction on *average* F_0 (taken as the mean value of points from each F_0 contour at S_t position in Fig. 4). Acoustically, therefore, there are sufficient differences between the three level tones in terms of F_0 height, despite their apparent similarity in contour.

Other than a small number of exceptional instances of tone-sandhi [5], syllables with a voiceless coda consonant [p], [t] or [k] in Cantonese can only be associated with the three level tones. The present results show that this is probably due to the language-specific adjustment to these syllables' insufficiency in bearing non-static tones. In addition, the chance of confusion is further minimized by the rarity of these tones in everyday speech (9.19% in [2] and 9.00% in [8]).

6. CONCLUSIONS

We have found in this study that voiceless consonants in Cantonese introduce F_0-raising and F_0-lowering in prevocalic and postvocalic positions, respectively. Moreover, durational difference of the resulting prevocalic voiceless intervals was observed across onset consonants. Finally, it was demonstrated that the recent finding of tone-syllable alignment is important for the correct understanding of the nature of F_0 perturbations.

7. REFERENCES

[1] Feng, L. 1985. Duration of initials, finals and tones in Beijing dialect. In Lin, T., Wang, L. (eds), *Beijing Yuyin Shiyanlu* [Working Papers in Experimental Phonetics]. Beijing: Beijing University Press, 131-195.

[2] Fok, C.Y.Y. 1974. *A perceptual study of tones in Cantonese*. Center of Asian Studies. University of Hong Kong.

[3] Francis, A.L., Ciocca, V., Wong, V.K.M., Chan, J.K.L. 2006. Is fundamental frequency a cue to aspiration in initial stops? *J. Acoust. Soc. Am.* 120, 2884-2895.

[4] Hombert, J.M. 1978. Consonant types, vowel quality and tone. In: Fromkin, V. (eds), *Tone: A Linguistic Survey*. New York: Academic, 77-111.

[5] Lam, K.P.M. 2002. *A study of intonation in Hong Kong Cantonese*. Ph.D. dissertation. The Chinese University of Hong Kong.

[6] Lee, W.S., Chen, F.X., Luke, K.K., Shen, L.Q. 2002. The prosody of bisyllabic and polysyllabic words in Hong Kong Cantonese. *Proc. SP-2002* Aix-en-Provence, 451-454.

[7] Lehiste, I., Peterson, G.E. 1961. Some basic considerations in the analysis of intonation. *J. Acoust. Soc. Am.* 33, 419-425.

[8] Leung M.T., Law S.P., Fung, S.Y. 2004. Type and token frequencies of phonological units in Hong Kong Cantonese. *Behavior Research Methods, Instruments, & Computers* 36(3), 500-505.

[9] Ohde, R. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *J. Acoust. Soc. Am.* 75, 224-230.

[10] Peterson, G.E., Barney, H.L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175-184.

[11] Silverman, K. 1986. F_0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica* 43, 76-91.

[12] Wong, Y.W. 2006a. Temporal distribution of tonal information in continuous Cantonese speech. *Proc. 2nd International Symposium on Tonal Aspects of Languages* La Rochelle, 125-128.

[13] Wong, Y.W. 2006b. Contextual tonal variations and pitch targets in Cantonese. *Proc. SP-2006* Dresden.

[14] Xu, Y. 1997. Contextual tonal variations in Mandarin. *J. Phonetics* 25, 61-83.

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

[16] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of f_0 contours. *J. Phonetics* 27, 55-105.

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

[18] Xu, Y., Wallace, A. 2004. Multiple effects of consonant manner of articulation and intonation type on F_0 in English. *J. Acoust. Soc. Am.* 115, Pt. 2, 2397.

[19] Zee, E. 1991. Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association* 21(1), 46-48.

† The work was supported in part by a CERG grant, CUHK 1127/04H.