

EFFECTS OF TONAL CONTEXT AND FOCUS ON CANTONESE F_0

Wentao Gu and Tan Lee

Department of Electronic Engineering, the Chinese University of Hong Kong
 {wtgu, tanlee}@ee.cuhk.edu.hk

ABSTRACT

F_0 variations of the speech of tone languages are not only constrained by individual lexical tones but also affected by various contextual factors. In this work the effects of two factors, i.e. tonal context and emphatic focus, on F_0 contours of Cantonese speech are investigated through a controlled experiment.

For tonal context, all the four combinations of carryover/anticipatory and assimilatory/dissimilatory effects are observed, but with different magnitudes and domains. For focus, the effect covers a wider range and shows a similar nature in pre-focus, on-focus, and post-focus domains, both raising F_0 values and expanding F_0 ranges. The interactions between the two factors are also addressed.

Keywords: Cantonese, F_0 contour, tone, focus.

1. INTRODUCTION

An accurate and linguistically meaningful representation of the voice fundamental frequency (F_0) contours is necessary for the synthesis of high-quality speech for various languages. However, this task is rather difficult, because surface F_0 contours of speech show great variation, which is not random but plays an important role in conveying a variety of linguistic, paralinguistic, and nonlinguistic information. It is especially the case for tone languages, of which F_0 contours of continuous speech are largely constrained by local lexical tones but at the same time deviate from the canonical form of tones in each syllable due to various contextual effects, either local (e.g. by adjacent tones) or global (e.g. by phrase intonation). Hence, for a better understanding we need to study the relationships between surface F_0 contours, lexical tones, and the underlying contextual information.

Among the linguistic factors, tonal context has a direct influence on the realization of F_0 patterns of lexical tones. There are two directions of tonal coarticulation, i.e. carryover and anticipatory effects, which are usually assimilatory and dissimilatory in nature, respectively; carryover effect is usually much stronger than anticipatory effect, e.g., in Mandarin [1].

Among the paralinguistic factors, emphatic focus (henceforth focus) is especially important in speech communication. Xu's study on Mandarin [2] showed

that the effects of focus on F_0 range vary in three distinct domains, viz., on-focus expansion, post-focus suppression (i.e. lowering and compression), and pre-focus intactness; hence, focus also brings about substantial F_0 downtrends. Recently, Jia et al. [3] further clarified that such expansion and suppression of F_0 range result mainly from raising and lowering high pitch targets, while low pitch targets are hardly affected by focus.

In the present study we investigate the effects of these two factors on F_0 contours of Cantonese, a tone language with more lexical tones than Mandarin.

2. CANTONESE TONE AND INTONATION

Cantonese has nine lexical tones, including six non-entering tones and three entering tones – the syllables of entering tones end with an unreleased stop /p/, /t/, or /k/, hence shorter than those of other tones. Since the entering tones show similar F_0 patterns as the non-entering level tones and can be distinguished from the latter by syllable structure, some schemes of transcription (e.g. Jyutping in HK) define only six tones, by merging non-entering and entering ones.

Table 1 gives the descriptions of the six tones in Cantonese. Conventionally, a 5-level tone code system is used for phonetic notation of tones. The levels, from 1 to 5, indicate the relative pitch targets from low to high, respectively. Recently, a fully quantitative scheme based on the command-response model was also proposed [4], characterizing the tones in terms of tone command patterns, i.e. the command polarities (+/0/-/=) in the earlier and later parts of a syllable (here '=' indicates more negative than '-').

The effects of both factors in Cantonese have been studied in previous works. For tonal context, Wong [5] showed similar effects as for Mandarin, viz., carryover assimilation as the primary effect and anticipatory dissimilation as the secondary. For focus, Man [6] claimed that it does not affect tone identities

Table 1: Cantonese tone system.

Tone type	Pitch feature	Tone code	Command pattern
T1	high level	55	++
T2	high rising	35 or 25	-+
T3	mid level	33	00
T4	low falling	21 or 11	==
T5	low rising	13 or 23	-0
T6	low level	22	--

but lengthens syllable duration and expands F_0 range, whereas a recent study based on the command-response model [7] showed that its major effect on F_0 can be characterized by an insertion/increase of phrase command before the on-focus syllable (hence a wide-range increase of F_0 thereafter) but the underlying patterns of tone commands are preserved.

Since the two factors have close interactions, in this work we investigate their effects jointly, through a deliberate control of the domain of narrow focus.

3. SPEECH DATA

In our controlled experiment, the speech material shares a fixed carrier frame: “Sin1 gong2 _ _ ni1 deoi3 zi6.” (First say the pair of characters _ _), in which a disyllabic target /jau wai/ is embedded at the mid of the sentence, carrying each of the six lexical tones in each target syllable and hence 36 tonal combinations altogether. The disyllabic targets are nonsense words sharing a fixed segmental composition, such that the segmental effect can be minimized. Also, both /jau/ and /wai/ give a voiced initial so as to ensure a continuous F_0 curve throughout the disyllabic interval.

Each sentence was uttered in three different versions: without any narrow focus, with a focus on /jau/, and with a focus on /wai/, respectively. For the latter two versions, a prompt question was provided before recording, i.e., “Sin1 gong2 *jiu* wai ni1 deoi3 zi6 me1?” for a focus on /jau/, and “Sin1 gong2 jau *wui* ni1 deoi3 zi6 me1?” for a focus on /wai/.

The informant is a male native Cantonese speaker who is a professional announcer in Hong Kong. Each utterance was recorded with four repetitions at the normal speech rate of the informant. Hence there are altogether 432 utterances.

The F_0 values extracted by the modified autocorrelation analysis were smoothed and interpolated (for voiceless intervals) to obtain continuous F_0 contours, and syllable segmentation was done manually to ensure a high accuracy. Ignoring durational differences in the present study (in fact, syllables tend to be lengthened by focus [8]), an average time-normalized F_0 contour was obtained by extracting a 10-point (equally spaced) sequence of F_0 values in each syllable and then averaging over the four repetitions of utterances.

4. RESULTS

Figures 1 and 2 (cf. figure file for the full figures) show the time-normalized F_0 contours grouped by the tone in the right syllable /wai/ and in the left syllable /jau/, respectively. Namely, in each panel of Fig. 1 the tone of /wai/ is fixed but /jau/ varies from T1 to

T6, while in each panel of Fig. 2 the tone of /jau/ is fixed but /wai/ varies from T1 to T6. The three columns of panels in each figure correspond to the three versions of focus, respectively. In each panel, to distinguish the varied tones in one of the target syllables, six different colors are used to plot the respective F_0 contours.

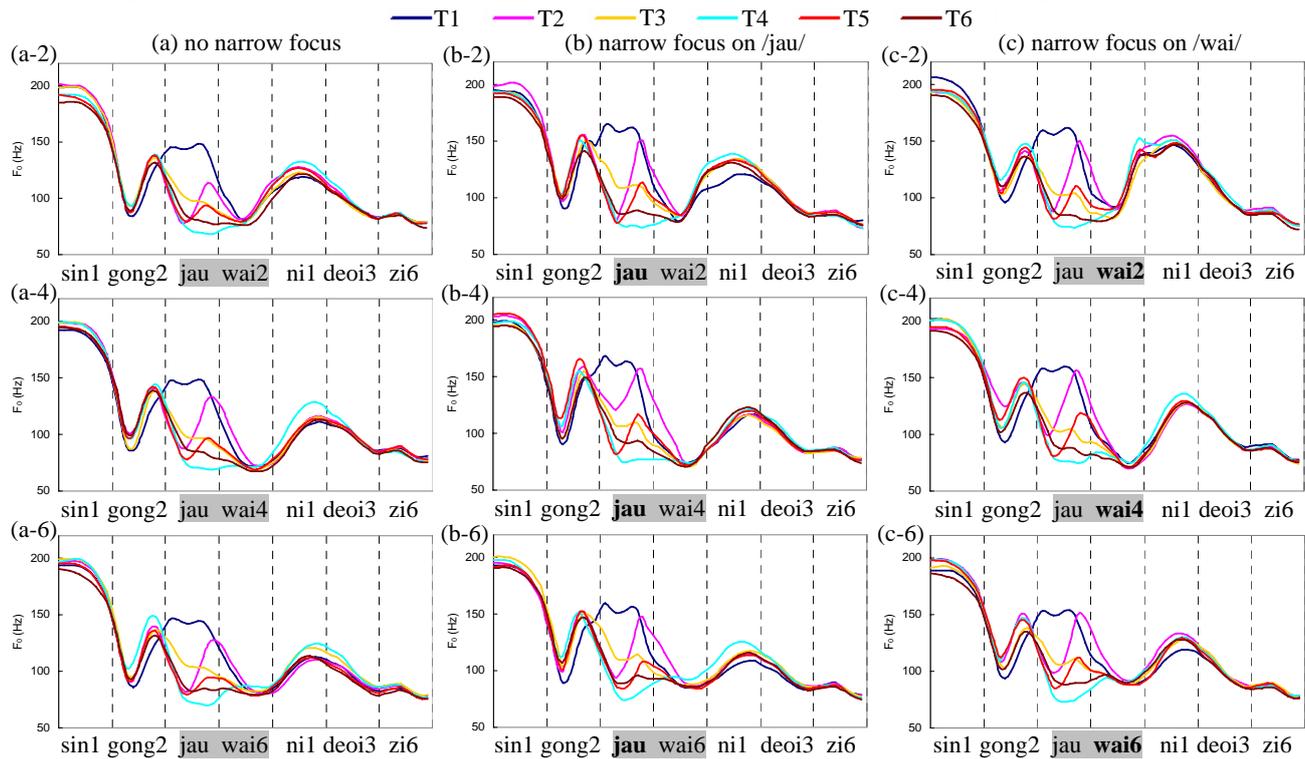
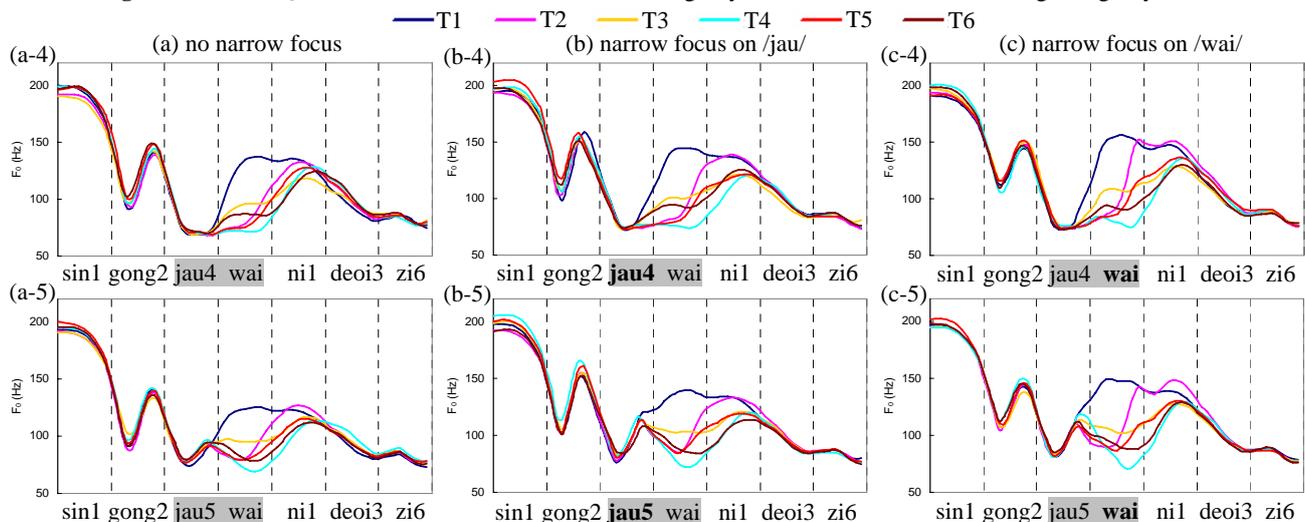
It is straightforward to see that F_0 patterns in each syllable show various deviations from the tones in their canonical form. Also, F_0 values are relatively variable at the beginning of an utterance due to the initial flexibility in speaking (*not* correlated with the target tones), but become rather stable at the end of the utterance even after different target tones (it also implies that the effects of tones are narrow). Due to space limitations, in the following we only describe qualitative results without giving detailed statistics.

4.1. Tonal coarticulation

Carryover effect is distinctly observed in both figures, showing an assimilatory nature in the conjunction of adjacent syllables, as reported in previous works [1, 5]. In terms of the preceding tone, both the onset F_0 of /wai/ in Fig. 1 and the onset F_0 of /ni1/ in Fig. 2 are in the order of T1/T2 > T3/T5 > T6 > T4, which exactly coincides with the order of offset F_0 levels of tones as given in Table 1. In most cases, the domain of this effect ranges from a half to one syllable.

At the same time, a slight but notable nonadjacent dissimilation is also found in carryover effect, not in the onset of the immediately following syllable but after the assimilatory domain. As mainly shown in Fig. 1, compared with other tones, T4 (the tone with the lowest offset F_0) not only lowers the onset F_0 of the following syllable but also often raises F_0 in the post-assimilation domain including the nonadjacent following syllable(s). Hence, the carryover effect shows opposite natures in the two distinct domains, viz., a dominant immediate assimilation, followed by a subtle nonadjacent dissimilation (especially for T4) which was not reported in previous works.

Anticipatory effect, on the other hand, is much smaller than carryover effect, but is still significant. In Fig. 1, the F_0 of /gong2/ tends to be higher before the tone with a lower onset F_0 , hence dissimilatory in nature, as reported in previous works [1, 5]. This is especially clear by comparing the F_0 curves before T1 (the highest tone) and T4 (the lowest tone). In Fig. 2, however, such an effect is shown only in the syllables /wai/ of T2 or T5 (more distinct for T2) but not distinct in those of other tones. It indicates that anticipatory dissimilation is effective mainly when the preceding tone is a *contour* tone (i.e. one of the two rising tones in Cantonese; in fact, T4 is

Figure 1: Pooled F_0 contours with varied tones in the left target syllable but a fixed tone in the right target syllable.**Figure 2:** Pooled F_0 contours with a fixed tone in the left target syllable but varied tones in the right target syllable.

essentially level rather than falling – the frequently observed initial falling is only a transition). This also agrees with the previous observation on Mandarin [1] which however did not receive much attention.

Nevertheless, such an anticipatory dissimilation appears only before a crossing point near the end of the preceding syllable; after this point, *assimilation* is shown instead regardless of the tone identity, though the domain is much shorter than that of the preceding dissimilation (thus almost neglected in previous works). Hence, the anticipatory effect also shows opposite natures in the two distinct domains (both

within the preceding syllable), viz., a very short immediate assimilation, preceded by a relatively longer but rising-tone-specific dissimilation.

By comparison of the three columns of panels in each figure, we can find that all the above-mentioned coarticulatory effects are preserved in the context of focus. The most notable influence of focus is that the domain and magnitude of anticipatory assimilation become larger when either of the two adjacent syllables is on-focus, as shown in Figs. 1(b), 2(b), and 2(c), further indicating that such an effect should not be neglected. The domains of other effects almost

keep intact. Besides, carryover assimilation may also be stronger when the preceding T2 is under focus.

4.2. Focus

By comparison, we further see the influences of focus on F_0 values: (1) almost intact in the utterance-initial syllable /sin1/; (2) raised since the second syllable /gong2/ until the end of the utterance; (3) increase of F_0 in the utterance-final syllable /zi6/ is much smaller than in the earlier part. These indicate that the effect is neither very local nor covering the entire utterance; instead, focus results in a wide-range F_0 increase, which starts slightly before the on-focus syllable and its magnitude decreases gradually over the time.

Contrary to the asymmetric nature found in Mandarin [2] and other languages, the effect of focus in Cantonese shows homogenous nature in the three domains (pre-focus, on-focus, and post-focus), on both height and range of F_0 . First, F_0 values are raised, not evenly but more notably on higher pitch targets – this is in a sense similar to Jia et al.'s finding on Mandarin that focus mostly raises H targets but hardly affects L targets [3]. Second, F_0 ranges are usually expanded, as most clearly shown by the two rising tones (T2 & T5). Besides, although post-focus F_0 downtrend is still observed because the magnitude of post-focus F_0 increment decreases over the time, it is much smaller than that in Mandarin.

Such effects can be interpreted in the framework of the command-response model. The wide-range F_0 increase is explained exactly by an insertion/increase of phrase command before the on-focus syllable [7], while the expansion of F_0 range can be ascribed to lengthened durations of tone commands [8]. Different magnitudes of F_0 increase in high/mid/low pitch targets can be explained by a combined effect of phrase and positive/null/negative tone commands.

5. DISCUSSION AND CONCLUSION

The effects of two factors, i.e. tonal context (linguistic) and focus (paralinguistic), on Cantonese F_0 contours have been investigated in depth through a controlled experiment. It turns out that the effect of focus has a wider domain than that of tones.

For the effect of tonal context, all the four combinations of carryover/anticipatory (direction) and assimilatory/dissimilatory (nature) are observed. Assimilation is ubiquitous in all the tonal contexts, and occurs at the junction of adjacent syllables. It can be deemed a natural result of continuous transition, as also embedded in the framework of the command-response model [4] – the existence of anticipatory assimilation here coincides with the mechanism that a tone command occurs slightly before the rhyme onset.

Apparently, carryover assimilation is dominant, with a much wider domain than anticipatory one.

In contrast, dissimilation in either direction is weak and only observed in particular tonal contexts, with a nonadjacent domain beyond that of immediate assimilation. We conjecture that this effect is associated with a mechanism of maximizing the contrast between adjacent tones with distinct pitch levels, hence it is the most conspicuous on the tones with extreme pitch levels, e.g. T4, which almost touches the lower limit of F_0 variation.

The effect of focus is found to be contrary to that in Mandarin, suggesting that it is rather language-dependent. In Cantonese, the effect shows a similar nature in the three distinct domains, namely, F_0 values are raised and F_0 ranges are expanded.

The interactions between the two factors are also shown. On the one hand, the increase of F_0 caused by focus varies with tonal context – it is larger on higher pitch targets. On the other hand, anticipatory assimilation is enhanced in the context of focus.

By a direct analysis of surface F_0 values in time-normalized F_0 contours, the present study has clarified many effects not revealed in previous works. As an extension to the present study, the model-based analysis of raw F_0 contours will be presented in [8].

6. ACKNOWLEDGEMENTS

This research is partially supported by an Earmarked Research Grant (Ref: CUHK 4227/04E) from the Hong Kong Research Grants Council, and a project grant awarded by SHIAE, CUHK.

7. REFERENCES

- [1] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.
- [2] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of f_0 contours. *J. Phonetics* 27, 55-105.
- [3] Jia, Y., Xiong, Z., Li, A. 2006. Phonetic and phonological analysis of focal accents of disyllabic words in Standard Chinese. *Proc. ISCSLP* Singapore, 43-54.
- [4] Gu, W., Hirose, K., Fujisaki, H. 2007. Analysis of tones in Cantonese speech based on the command-response model. *Phonetica* 64, 29-62.
- [5] Wong, Y.W. 2006. Contextual tonal variations and pitch targets in Cantonese. *Proc. Speech Prosody* Dresden, 317-320.
- [6] Man, C.-H.V. 1999. An acoustic study of the effects of sentential focus on Cantonese tones. M.S. thesis, University of Victoria, Canada.
- [7] Gu, W., Hirose, K., Fujisaki, H. 2006. Modeling the effects of emphasis and question on fundamental frequency contours of Cantonese utterances. *IEEE Trans. Audio, Speech and Language Processing* 14, 1155-1170.
- [8] Gu, W., Lee, T. 2007. Effects of focus on prosody of Cantonese speech – A comparison of surface feature analysis and model-based analysis. *Proc. Paraling'07* Saarbrücken.