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

INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

Analysis of F_0 Contours of Cantonese Utterances Based on the Command-Response Model

Wentao Gu[1, 2], *Keikichi Hirose*[1] *and Hiroya Fujisaki*[1]

[1]The University of Tokyo, Japan [2]Shanghai Jiaotong University, China
{wtgu; hirose}@gavo.t.u-tokyo.ac.jp fujisaki@alum.mit.edu

10.21437/Interspeech.2004-295

Abstract

As a major Chinese dialect, Cantonese is well known for its complex tone system. This paper applies the command-response model to represent the F_0 contours of Cantonese speech. Analysis-by-Synthesis is conducted on both utterances of carrier sentences and utterances with less constrained structures, from which a set of appropriate tone command patterns is derived. By intrinsically incorporating the effects of tone coarticulation, word accentuation and phrase intonation, the model provides a high accuracy of approximation to the F_0 contours of Cantonese, and hence serves as a much better method to quantitatively describe the continuous F_0 contours than the traditional tone letter scale notation system. The constraints in timing and amplitude of tone commands are also investigated, which can be used for synthesis of F_0 contours.

1. Introduction

An accurate and quantitative representation of the essential characteristics of the F_0 contours of speech is necessary for both text-to-speech synthesis and automatic speech recognition, especially for tone languages. For this purpose, the command-response model for the process of F_0 contour generation, proposed by Fujisaki and his coworkers, is useful since it can generate very close approximations to observed F_0 contours from a relatively small number of linguistically meaningful parameters. The model has been successfully applied to tone languages including Mandarin and Thai [1]. In this paper we investigate the possibility of its application to Cantonese, a Chinese dialect with a complex tone system.

2. Cantonese tone system

Cantonese is one of the major dialects of Chinese spoken by over 60 million people in Hong Kong, Guangdong and Guangxi provinces of China, and in many overseas Chinese communities. Although Cantonese largely shares the same writing system and the same monosyllabic nature as Mandarin, it is much richer in the number of tone types. It is usually accepted that Cantonese has nine citation tones, which preserve the tonal categories of Middle Chinese (7th~10th century A.D.). Table 1 gives some traditional descriptions of all the nine citation tones.

The syllables of entering tones end with an unreleased stop coda /p/, /t/ or /k/, and are comparatively shorter in duration than those of non-entering tones. Each entering tone has its counterpart of non-entering tone, showing a similar F_0 pattern. T7, T8 and T9 correspond to T1, T3 and T6 respectively. Therefore in some transcription schemes only six tones are distinguished.

Traditionally a 5-scale tone letter notation system is adopted for Cantonese after Chao [2], though it varies somewhat from one reference to another. Such a notation system provides a simplified canonical form for tones in isolated syllables, but the scales are subjective and relative, and the actual F_0 values change with the tonal context, word accentuation and phrase intonation. In most references, the five scales are interpreted as indicating the beginning and end F_0 values in a syllable, and the F_0 contour is approximated by piecewise straight lines. However, this is too simple to describe the actual F_0 contours accurately.

To overcome the intrinsic limitation of the traditional tone scale notation, we introduce the command-response model for the generation process of F_0 contours of Cantonese.

Table 1: Some traditional descriptions of Cantonese tones.

Tone name in Middle Chinese system		Tone * number	Pitch feature	5-scale notation
Non-entering tones	Upper-level	T1	High level	55
	Upper-elevating	T2	Mid rising	35
	Upper-departing	T3	Mid level	33
	Lower-level	T4	Low falling	21
	Lower-elevating	T5	Low rising	13
	Lower-departing	T6	Low level	22
Entering tones	Upper-entering	T7	High level	5
	Middle-entering	T8	Mid level	3
	Lower-entering	T9	Low level	2

* Note: T1~T4 here are different from those of Mandarin.

3. The command-response model for generating F_0 contours of tone languages

Figure 1 shows the diagram of the model for tone languages. It describes F_0 contours in the logarithmic scale as the sum of phrase components, tone components and a baseline level. The phrase commands (impulses) produce phrase components through the phrase control mechanism, giving the global shape of the F_0 contour, while the tone commands (pedestals) of both positive and negative polarities generate tone components through the tone control mechanism, characterizing the local F_0 changes. Both mechanisms are assumed to be critically-damped second-order linear systems.

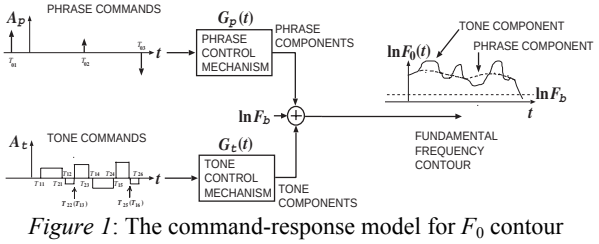


Figure 1: The command-response model for F_0 contour generation with both positive and negative tone commands.

The model can be formulated by the following equations:

$$\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i}) + \sum_{j=1}^{J} A_{tj} \{G_t(t - T_{1j}) - G_t(t - T_{2j})\}, \quad (1)$$

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (2)$$

$$G_t(t) = \begin{cases} \min[1 - (1 + \beta t)\exp(-\beta t), \gamma], & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (3)$$

where $G_p(t)$ represents the impulse response function of the phrase control mechanism, and $G_t(t)$ represents the step response function of the tone control mechanism. F_b is a baseline value. The parameters A_{pi} and T_{0i} denote the magnitude and time of the ith phrase command respectively, while A_{tj}, T_{1j} and T_{2j} denote the amplitude, onset time and offset time of the jth tone command respectively. The constants α, β and γ are set at their respective default values 3.0 (1/s), 20.0 (1/s) and 0.9 in the current study, though the values of β and γ can be set for positive and negative tone commands separately.

A set of tone command patterns needs to be specified for the model to fit a specific tone language. Different sets of tone command patterns have been assigned to Mandarin and Thai respectively [1].

This model incorporates the effects of tone coarticulation, word accentuation and phrase intonation simultaneously in an explicit way. Tone coarticulation is automatically taken care of by the transfer characteristics of the tone control mechanism. Word accentuation can be implemented either by magnifying the amplitudes or by lengthening the duration of tone commands. Phrase intonation is explicitly represented by the phrase components.

4. Analysis of Cantonese F_0 contours

4.1. Speech data

Two sets of speech material were used: Speech Material A was designed with exactly the same carrier sentences, while Speech Material B included various meaningful sentences.

Speech Material A consists of carrier sentences "hon3 gin3 '*maa*' faai3 gong2 ceot7 lai4" (Speak it out quickly when you see '*maa*'), in each of which the target syllable *maa*, carrying each of the nine tones, is embedded. Each sentence was uttered seven times by a native male speaker of Cantonese (from Guangzhou) at his normal speech rate. Although in the lexicon there are no characters corresponding to *maa2* and *maa3*, the speaker was trained to utter pseudo words for them.

Speech Material B consists of 20 declarative sentences each with 5~14 syllables. Each sentence was recorded three times at normal speech rate.

The speech signal was digitized at 10 kHz with 16bit precision. The fundamental frequency was extracted by the modified autocorrelation analysis of the LPC residual. The syllable boundaries and rhyme boundaries were labeled by visual inspection of the waveform and spectrogram.

4.2. Tone command patterns

First, the F_0 contours of Speech Material A were approximated by the procedure of Analysis-by-Synthesis. A sample utterance for each target tone is shown in Fig. 2. The crossed symbols indicate the observed F_0 values, while the solid lines, dotted lines and dashed lines indicate the approximated F_0 contours, baseline frequency and the contribution of phrase components, respectively.

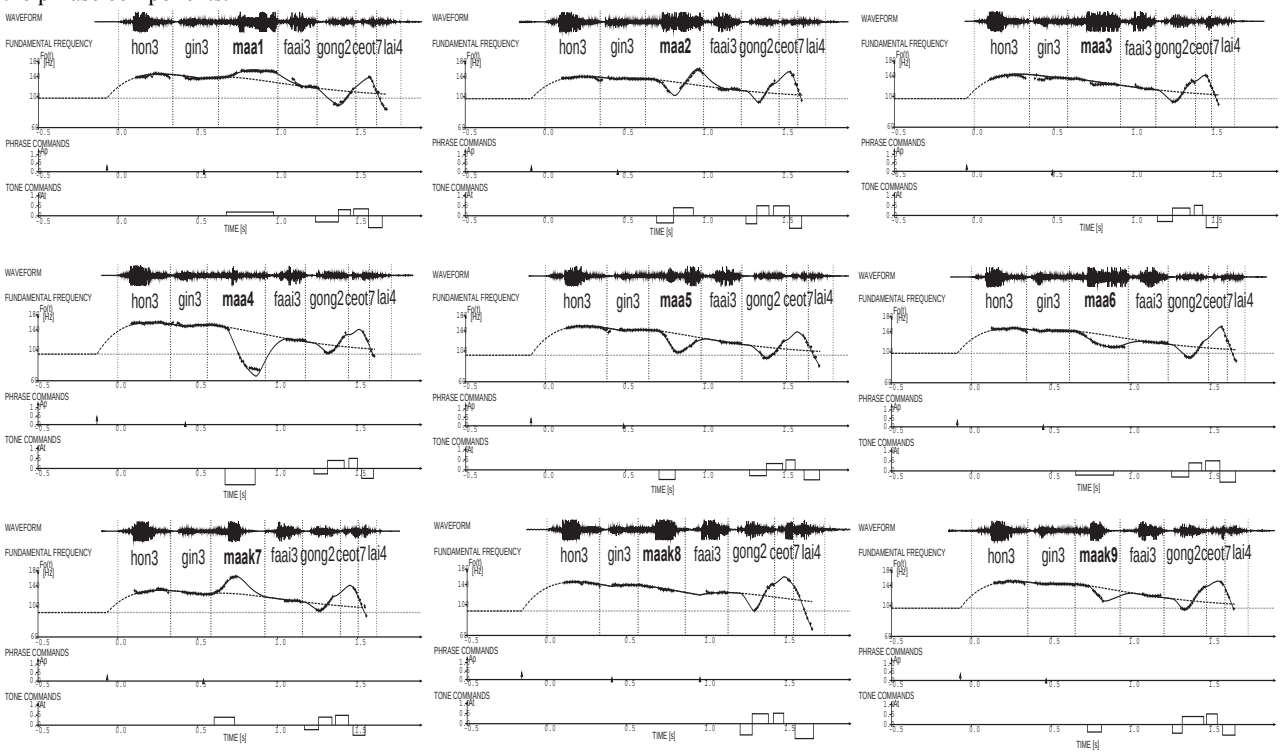


Figure 2: Analysis of F_0 contours for the Cantonese carrier sentences embedded with *maa* of nine tones.

The results in Fig. 2 show that the tone command patterns for Cantonese tones are:

T1: positive
T2: initially negative and later positive
T3: zero
T4: large negative
T5: initially negative and later zero
T6: small negative

We note that T4 gives a command pattern with the same polarity but larger amplitude as compared with T6. As for the three entering tones, the initial command is the same as their respective counterpart of non-entering tones (positive for T7, zero for T8, and negative for T9), while the later command is always extremely negative so that voicing is interrupted.

In the carrier sentence, the two syllables preceding the target syllable and the syllable following the target syllable are all of T3, which has been shown to have no tone command. Therefore, the tone components shown in the target syllables are not affected by immediate tonal context and hence can be considered to indicate their intrinsic command patterns.

With this set of tone command pattern definitions, very close F_0 approximation can be achieved. This was further confirmed by applying Analysis-by-Synthesis to approximate the F_0 contours of Speech Material B. An example is shown in Fig. 3. The sentence means "Among them six persons were dehydrated, so they received intravenous injection".

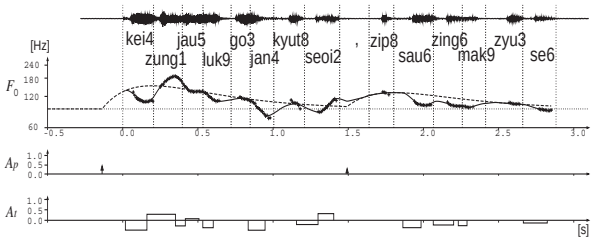


Figure 3: Analysis-by-Synthesis of the F_0 contour of a Cantonese utterance in Speech Material B.

4.3. Model interpretation

From the example utterance shown in Fig. 3, we can observe that the effects of tone coarticulation, word accentuation and phrase intonation have been efficiently embedded in the command-response model.

(1) The T1 syllable *zung1* shows a quite different pattern from its canonical form as given in Table 1. Both the beginning and end sections become substantially lower. This tone coarticulation pattern can be explained by the fact that the negative tone component of the preceding T4 syllable is still continuing, while the negative tone command of the following T5 syllable already occurs before the end of the current T1 syllable.

(2) The second phrase consists of three bi-syllabic words, each including a T6 syllable. The command amplitudes of the three T6 syllables vary with the word position: phrase initial (*sau6*) > phrase mid (*zing6*) > phrase final (*se6*), which may indicate how word accentuation varies with the position in the phrase.

(3) The two T8 syllables, *kyut8* and *zip8*, show quite different F_0 values. With the explicit introduction of phrase components in the model, this can be easily explained by their different positions in the two corresponding phrases.

4.4. Timing and amplitude of tone commands

Like Mandarin, a syllable in Cantonese can be divided into two parts: initial and final. The initial can be a consonant (unvoiced or voiced), a semi-vowel or nil. The final is composed of main vowel(s) and an optional nasal or stop coda. In Cantonese, the nasal /m/ or /ng/ can form a syllable by itself. Such syllabic nasals are also regarded as finals. We define the rhyme, *i.e.* the portion carrying the tones, as the final excluding the stop coda.

Figure 4 shows the timing of tone commands relative to the rhyme timing for all the tones except T3 and T8. The abscissa indicates the rhyme duration, while the ordinate indicates the timing relative to the rhyme onset. The lower and upper groups of points indicate the onsets and offsets of tone commands respectively. For T2, we assume that the onset of the 2nd tone command coincides with the offset of the 1st tone command just for simplicity. The top group of points shown for T2 indicates the offsets of the 2nd tone commands.

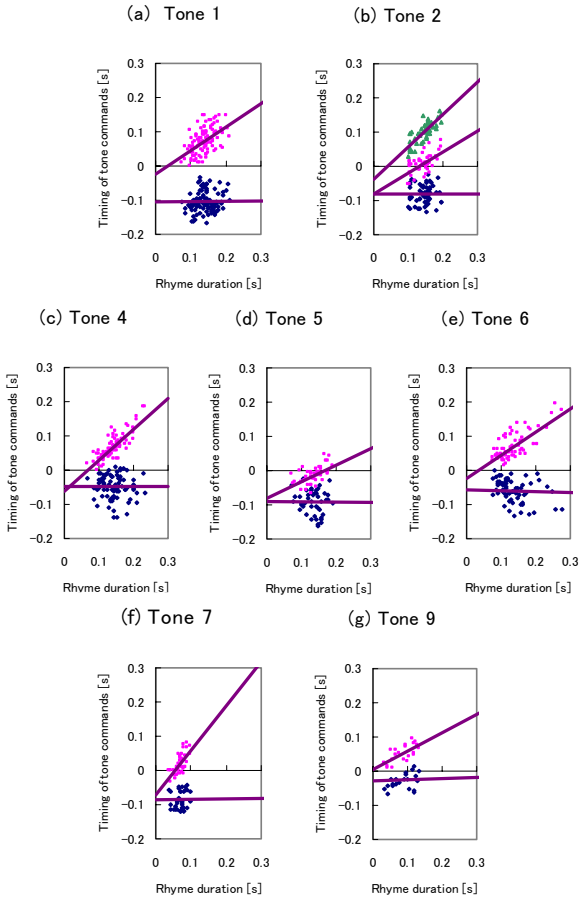


Figure 4: Tone command timing relative to the rhyme.

Some systematic relationships are observed. The onsets of tone commands are found to be concentrated in an interval, *viz.*, 0~100 ms (for T4, T6 and T9) or 50~150 ms (for other tones) prior to the rhyme onset. The offsets of tone commands

can be well approximated by linear regression, indicating a high correlation with the rhyme duration. This finding is similar to those for Mandarin and Thai [3-5]. Such constraints can be used for the synthesis of F_0 contours.

The amplitude of tone commands was also investigated. It is shown to be quite scattered and the correlation with rhyme duration is very low. Such a disperse amplitude, may reflect a continuous strength of word accentuation.

Figure 5 shows the relationship between tone command amplitude and rhyme duration for T4 and T6, both of which correspond to a negative tone command. It is observed that the amplitudes of T4 occupy a more negative region, though there is a big overlapping with that of T6. In fact, the absolute amplitudes for all other tones scarcely exceed 0.6. This confirms the tone command patterns we have defined.

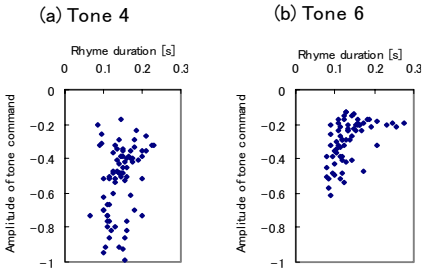


Figure 5: Tone command amplitudes of T4 and T6.

5. Discussion

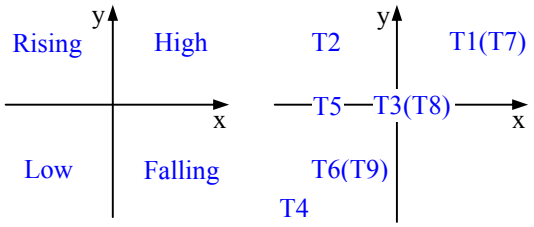
There have already been several works on Cantonese tones in continuous speech. The methods can be either non-parametric or parametric.

The non-parametric method, as given in [6, 7], conducts a statistical analysis directly on F_0 measurements and tries to find out the effect of some specific factors. The parametric method uses a mathematical model to approximate the F_0 contours, and obtains the model's parameters by Analysis-by-Synthesis. Stem-ML [8] uses such a mathematically tractable model. In all these works, however, linear scale is used to represent the F_0 contours.

The use of logarithmic scale adopted in our command-response model is based on the physiological and physical mechanisms of F_0 control, and allows us to generate the F_0 contours from a small number of linguistically meaningful command parameters. All of the findings on Cantonese F_0 contours in [6-8] can be naturally interpreted by the proposed command-response model.

Since the same model has first been applied to Mandarin [3, 4], it is worthwhile to make a comparison between the command patterns for the two dialects. Panel (a) in Fig. 6 shows command patterns for the four lexical tones (High, Rising, Low, and Falling) of Mandarin, while panel (b) shows those for the nine citation tones of Cantonese, where the entering tones are indicated by the parentheses. In both panels, the abscissa indicates the command early in the syllable and the ordinate indicates the command late in the syllable. Besides the lack of an equivalent to the falling tone of Mandarin, Cantonese shows three more non-entering tone command patterns, namely T3, T4 and T5.

In particular, T3 (and the corresponding entering tone T8) has no tone commands. Therefore, accentuation on syllables



(a) Mandarin (b) Cantonese

Figure 6: Tone command patterns of Mandarin and Cantonese.

of T3 (T8) cannot be implemented by the command amplitude. On the other hand, the distinction between T4 and T6 apparently depends on the command amplitude. The acoustic parameters that are used for accentuation in these cases will be elucidated in our future research.

6. Conclusion

Our experiments have shown that with a set of well-defined tone command patterns, the command-response model can be used to approximate F_0 contours of Cantonese with high accuracy. The effects of tone coarticulation, word accentuation and phrase intonation have all been efficiently embedded in the model. Therefore, compared with the traditional 5-scale tone letter notation system, this model provides a much better way to represent the F_0 contours of Cantonese speech quantitatively. The relationship between the tone command timing and rhyme duration shows that certain constraints exist, which can be used in the future work of synthesis of Cantonese F_0 contours.

7. References

[1] Fujisaki, H., "Information, prosody, and modeling – with emphasis on tonal features of speech," *Proc. Speech Prosody 2004*, Nara, Japan, pp. 1-10, 2004.

[2] Chao, Y.-R., *Cantonese Primer*, Harvard University Press, Cambridge, 1947.

[3] Wang, C., Fujisaki, H., Ohno, S. and Kodama, T., "Analysis and synthesis of the four tones in connected speech of the Standard Chinese based on a command-response model," *Proc. Eurospeech'99*, Budapest, Hungary, pp. 1655-1658, 1999.

[4] Wang, C., Fujisaki, H., Tomana, R. and Ohno, S., "Analysis of fundamental frequency contours of Standard Chinese in terms of the command-response model and its application to synthesis by rule of intonation," *Proc. ICSLP 2000*, vol. 3, pp. 326-329, Beijing, China, 2000.

[5] Fujisaki, H., Ohno, S. and Luksaneeyanawin, S., "Analysis and synthesis of F_0 contours of Thai utterances based on the command-response model," *Proc. 15th ICPhS*, Barcelona, Spain, pp. 1129-1132, 2003.

[6] Li, Y., Lee, T. and Qian, Y., "Acoustical F_0 analysis of continuous Cantonese speech," *Proc. ISCSLP'02*, Taipei, pp. 127-130, 2002.

[7] Li, Y., Lee, T. and Qian, Y., "F_0 analysis and modeling for Cantonese text-to-speech," *Proc. Speech Prosody 2004*, Nara, Japan, pp. 467-470, 2004.

[8] Lee, T., Kochanski, G., Shih, C. and Li, Y., "Modeling tones in continuous Cantonese speech," *Proc. ICSLP'02*, Denver, USA, pp. 2401-2404, 2002.