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

International Symposium on Chinese Spoken Language Processing (ISCSLP 2004)
Hong Kong
December 15-18, 2004

ANALYSIS OF SHANGHAINESE F_0 CONTOURS BASED ON THE COMMAND-RESPONSE MODEL

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

[1]The University of Tokyo, [2]Shanghai Jiaotong University

ABSTRACT

As one of the major Chinese dialects, Shanghainese is well known for its complex tone *sandhi* system. This paper applies the command-response model to represent F_0 contours of Shanghainese speech. Analysis-by-Synthesis is conducted both on carrier sentences with monosyllabic target words and on isolated polysyllabic words, from which a set of appropriate tone command patterns is derived for words of different lengths and different initial citation tones. By incorporating the effects of tone coarticulation, word accentuation and phrase intonation, the model gives high accuracy of approximations to F_0 contours of Shanghainese utterances, and hence provides a more efficient means to quantitatively represent F_0 contours and to describe the tone *sandhi* system of Shanghainese than the traditional 5-level tone code system.

1. INTRODUCTION

An accurate 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 [1], 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 [2] and Cantonese [3]. In this paper we investigate the possibility of its application to Shanghainese, a subset of Wu dialect with a complex tone *sandhi* system.

2. SHANGHAINESE TONE SYSTEM

Wu dialect is the second largest Chinese dialect after Mandarin, spoken by about 77 million people mainly distributed in Zhejiang, Jiangsu and Shanghai. It has a long history of about 2300 years.

Although Wu dialect largely shares the same writing system (monosyllabic characters) with Mandarin, its spoken form is quite distinct. Wu dialect has several unique features. First, it preserves the full set of Middle Chinese (7th~10th century A.D.) voiced initials that does not exist in any other Chinese dialect. Second, it has a complex tone *sandhi* system. The F_0 pattern of a word is completely determined by the lexical tone of one syllable while lexical tones of other syllables are lost.

Shanghainese is a distinct style of Wu dialect spoken by about 15 million people in the city of Shanghai. Since Shanghai is a relatively new city and most early immigrants in Shanghai were from throughout Wu-speaking regions, Shanghainese is an integration of various forms of Wu dialect. Due to the important status of Shanghai, nowadays Shanghainese is becoming the most influential form of Wu dialect.

From the traditional point of view, Shanghainese has five citation tones, as shown in Table 1. They can be grouped into two kinds of tonemes (T1: high falling, and T2~T5: regular) from the phonology point of view, because T2~T5 are regularly determined by syllabic structure (voiced or voiceless initial, with or without glottal stop coda) and have no separate phonological values. Syllables of entering tones (入声, *i.e.* T4 and T5 here) end with a glottal stop /ʔ/, and are comparatively shorter in duration than those of non-entering tones. Traditionally a 5-level tone code system is used to describe Shanghainese tones after Chao [4], though it varies somewhat from one reference to another. In the table the underlined codes denote entering tones.

Table 1: Traditional descriptions of Shanghainese citation tones.

Tone* number	Traditional tone name		Toneme	Initial	5-level code
T1	陰平		High falling	voiceless	53
T2	陰去		Regular	voiceless	34
T3	陽去			voiced	13
T4	入声	陰入		voiceless	<u>55</u>
T5		陽入		voiced	<u>12</u>

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

Table 2: Descriptions of the general forms of tone *sandhi* in Shanghainese polysyllabic words (after [5]).

Initial tone	2-syllabic word	3-syllabic word	4-syllabic word	5-syllabic word
T1 53	55 21	55 33 21	55 33 33 21	55 33 33 33 21
T2 34	33 44	33 55 21	33 55 33 21	33 55 33 33 21
T3 13	22 44	22 55 21	22 55 33 21	22 55 33 33 21
T4 <u>55</u>	<u>33</u> 44	<u>33</u> 55 21	<u>33</u> 55 33 21	<u>33</u> 55 33 33 21
T5 <u>12</u>	<u>11</u> 23	<u>11</u> 22 23	<u>11</u> 22 22 23 <u>22</u> 55 33 21	<u>22</u> 55 33 33 21

Shanghainese is well known for its complex tone *sandhi* system. There are two kinds of tone *sandhi* in Shanghainese: the general form and the special form [5]. Since the special form occurs only in words of specific structures, here we only discuss the general form, which occurs in various word structures and

can be characterized by a set of rules as shown in Table 2. When there is more than one syllable in a word, the tone of the first syllable determines the tonal contour of the entire word [5, 6].

The 5-level tone code system shown in Table 2 serves as the first approach to quantifying tones and tone *sandhi*. However, there are some intrinsic limitations of the tone code system, especially when using it for synthesis purpose. First, the five levels are subjective and relative, and in continuous speech the actual F_0 values change with tonal context, word accentuation and phrase intonation. Second, this approach is only semi-quantitative, and cannot characterize the continuous nature of F_0 values. Third, the five levels in Chao's system are perceptually defined, but many researchers interpret them to be the beginning and the end F_0 values within a syllable, and approximate the F_0 contour with straight lines connecting these values. This is a misunderstanding of Chao's system, and the straight lines are too simple to describe the actual F_0 contours accurately.

3. THE COMMAND-RESPONSE MODEL FOR F_0 CONTOUR GENERATION

To overcome the intrinsic limitations of the traditional tone code system, we introduce the command-response model for the generation process of F_0 contours of Shanghainese.

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.

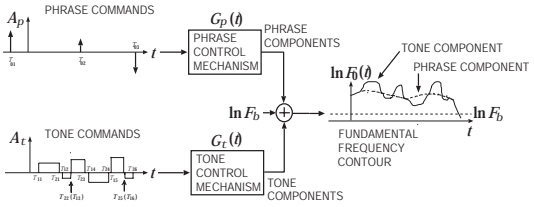

Figure 1: The command-response model for F_0 contour generation with both positive and negative tone commands.

The model can be formulated by the following equations:

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

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

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

The details of the parameters are described in reference [1]. A set of tone command patterns needs to be specified for the model to fit a specific tone language.

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 SHANGHAINESE F_0 CONTOURS

4.1. Speech data

Two sets of speech materials are used. Speech Material A includes carrier sentences with different monosyllabic target words, while Speech Material B includes various isolated polysyllabic words. The materials were recorded by three native speakers of Shanghainese (all from urban districts), one male in his 20s and two females in their 30s.

Speech Material A is used to investigate the command pattern for each tone in its citation form. It consists of carrier sentences "辩个字读__" /ge?5-rhe dzy3 do?5 __/ (This character is pronounced as __), in each of which one of the five monosyllabic target words "啊, 矮, 鞋, 压, and 盒" /(rh)a(?)/ (here the optional /rh/ is a null-initial with voiced sound) respectively carrying one of the five citation tones, is embedded at the underlined position. Since the target word does not combine with any neighboring words into a meaningful compound, the citation form of its F_0 pattern is preserved. Each sentence was uttered 8 times at the speakers' normal speech rate.

Speech Material B is used to investigate the command patterns corresponding to the tone *sandhi* rules shown in Table 2. The bi-syllabic, tri-syllabic and quadri-syllabic words beginning with each citation tone were selected. It has been perceptually confirmed that the F_0 patterns of these words coincide with the general rules shown in Table 2. For quadri-syllabic words beginning with T5, only the words showing the first pattern were selected. In order to investigate whether the command pattern is influenced by intra-word syntactic structure, we selected tri-syllabic words of both (12)3 and 1(23) types, and quadri-syllabic words of (12)(34), 1(234) and (123)4 types simultaneously.

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. Syllable boundaries were labeled by visual inspection of the waveform and the spectrogram.

4.2. Tone command patterns in monosyllabic words

In the first place, F_0 contours of Speech Material A were analyzed by the method of Analysis-by-Synthesis. A sample utterance for each target tone is shown in panels (a)~(e) of Fig. 2 respectively. 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.

The analysis results show that the tone command patterns for five citation tones in Shanghainese are:

T1: initially positive and later zero or negative

T2: initially zero and later positive

T3: initially negative and later positive
T4: positive
T5: initially negative and later positive

It is noted that the command pattern for T1 (high-falling tone) is somewhat variable. For some speakers, T1 tends to end with a negative command instead of zero (*i.e.*, with a steeper falling) in its isolated form or at phrase ends. On the other hand, all the four regular tones end with a positive command.

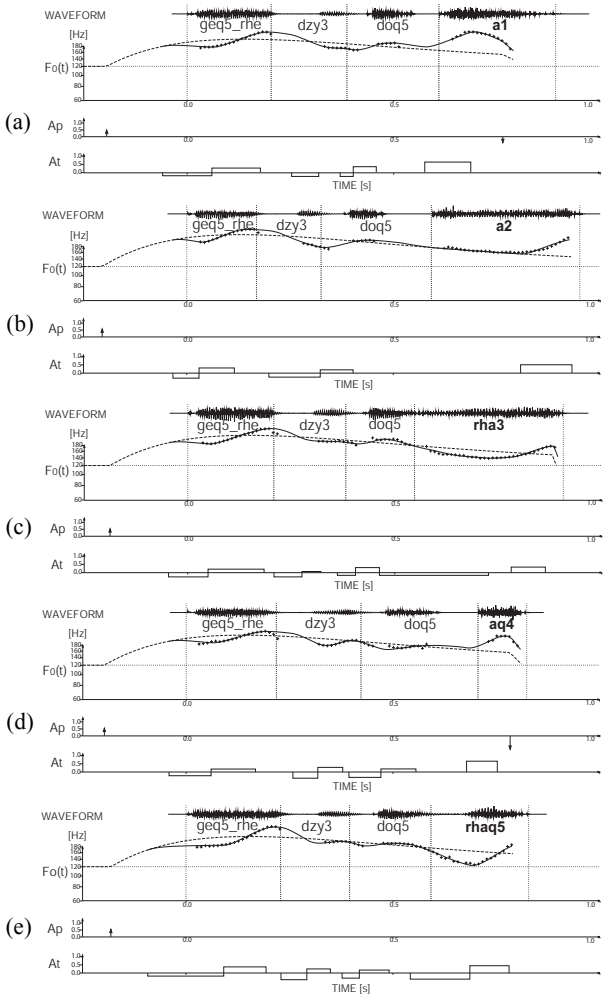

Figure 2: Analysis-by-Synthesis of F_0 contours of the carrier sentences with monosyllabic target words of five tones.

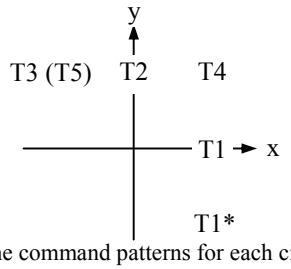

Figure 3: Tone command patterns for each citation tone in Shanghainese monosyllabic words.

The patterns of command polarity for the five citation tones can be represented in a Cartesian coordinate as shown in Fig. 3. The abscissa and the ordinate indicate the polarity of command early in the syllable and late in the syllable respectively. Here T1* denotes a pattern occasionally observed in isolated form or at phrase ends. Also, it is to be noted that T3 (non-entering tone) and T5 (entering tone) occupy the same position, but the duration of T5 is always shorter because voicing is interrupted by the glottal stop coda.

4.3. Tone command patterns in polysyllabic words

Next, F_0 contours of Speech Material B were analyzed by the method of Analysis-by-Synthesis. A sample utterance for each word constitution is shown in Fig. 4. Each row shows the words with the same initial tone, and each column shows the words of the same length. The word texts are given below:

T1:	空调	边浪向	三日两头
T2:	进口	写下来	半夜三更
T3:	牛奶	黄霉天	搞七念三
T4:	发展	吃生活	一塌糊涂
T5:	蹩脚	录音机	复旦大学

The analysis indicates the patterns of tone command polarity for polysyllabic words as shown in Table 3. The symbols +, 0 and – denote positive, zero and negative commands respectively. The vertical lines delimit each syllable. It is shown that tone command patterns change completely with the word length irrespective of the citation tones except for the initial syllable. With this set of tone command pattern definitions for tone *sandhi*, very close F_0 approximations can be achieved as shown in Fig. 4.

Table 3: Tone command patterns for Shanghainese polysyllabic words showing the general forms of tone *sandhi*.

Initial tone	1-syllabic word	2-syllabic word	3-syllabic word	4-syllabic word
T1	|+0| or |+−|*	|++|−−|	|++|+−|−−|	|++|++|−−|−−|
T2	|0+|	|00|++|	|00|++|−−|	|00|++|+−|−−|
T3	|−+|	|−−|++|	|−−|++|−−|	|−−|++|+−|−−|
T4	|++|	|00|++|	|00|++|−−|	|00|++|++|−−|
T5	|−+|	|−−++|**	|−−|−−|++|	|−−|−−|−−|++|

* Here |+−| only occurs in isolated form or at phrase ends.
** In some cases it can alternatively be |−−|−+|.

Some systematic tendencies can be observed from Table 3. For bi-syllabic words, the command pattern can be roughly regarded as an extension of the pattern for initial citation tone, while with the increase of syllable number, the command pattern gradually changes toward the respective ending pattern: negative for words beginning with T1~T4, and positive for words beginning with T5.

It is also found that different intra-word syntactic structures do not lead to any difference in tone command pattern, as long as the polysyllabic word is regarded as one lexical unit. For example, the three panels of Fig. 5, from left to right, show the analysis of F_0 contours of the following three quadri-syllabic words (all begin with T2) with different structures:

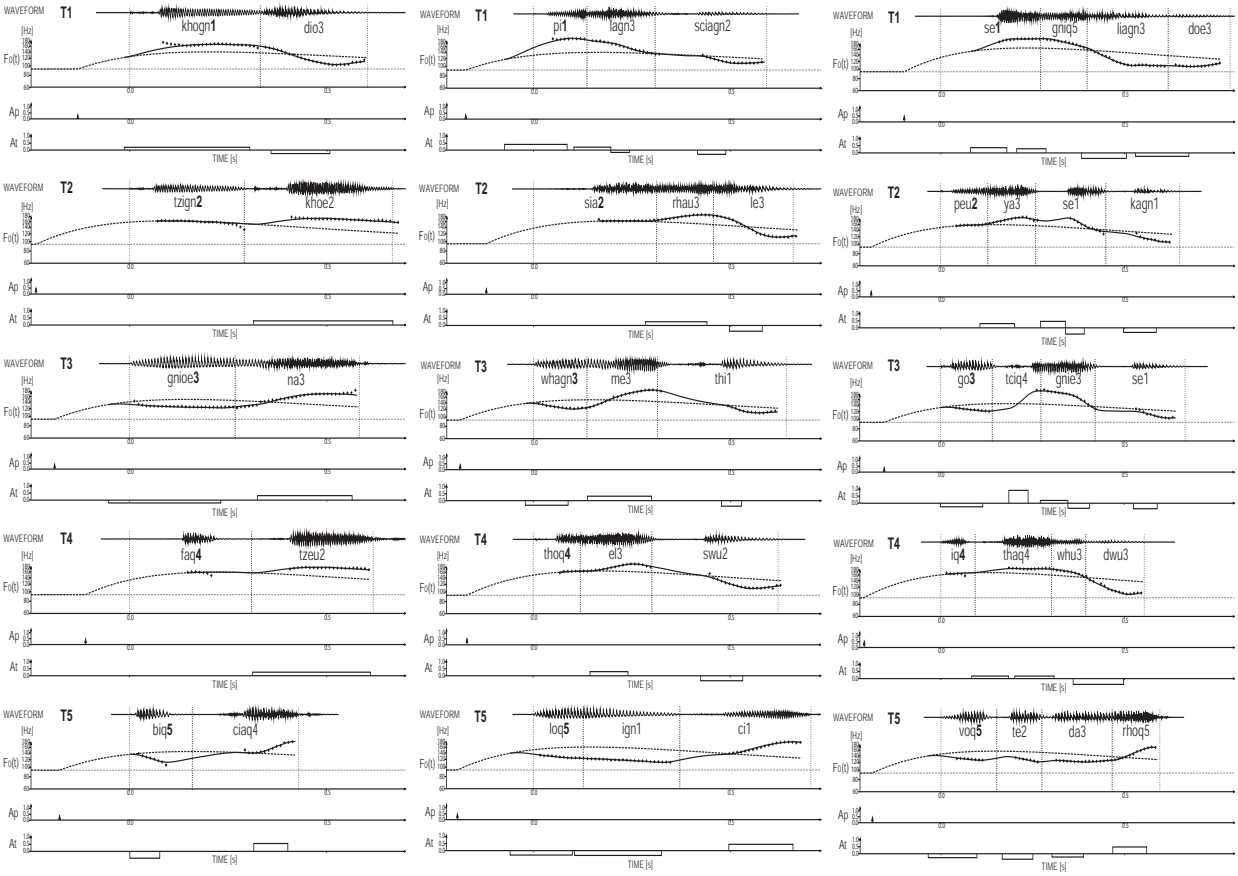

Figure 4: Analysis-by-Synthesis of F_0 contours of Shanghainese words of different lengths and different initial citation tones.

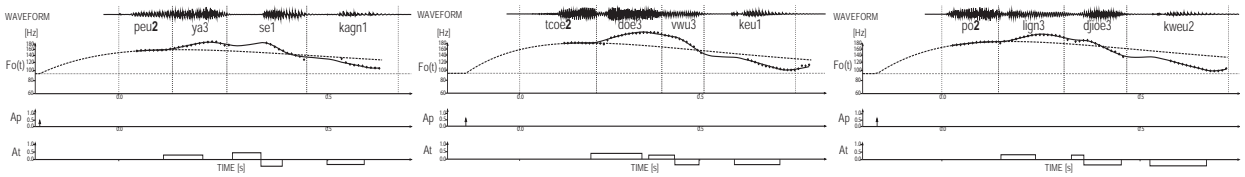

Figure 5: Analysis-by-Synthesis of F_0 contours of Shanghainese quadri-syllabic words with different syntactic structures.

半夜三更	臭豆腐干	保龄球馆
(12)(34)	1(234)	(123)4

No qualitative difference is observed between the three patterns.

5. CONCLUSION

Our experiments have shown that a set of tone command patterns can be defined for Shanghainese tones both in monosyllabic citation form and in polysyllabic connected words. With this set of well-defined command patterns, the command-response model can approximate F_0 contours of Shanghainese utterances with high accuracy. Compared with the traditional 5-level tone code system, this model provides more accurate means to represent F_0 contours and to describe the tone *sandhi* system of Shanghainese.

6. 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] Fujisaki, H., Hallé, P. and Lei, H., "Application of F_0 contour command-response model to Chinese tones," *Reports of Acoust. Soc. Jpn Autumn Meeting*, vol. 1, pp. 197-198, 1987.

[3] Gu, W., Hirose, K., and Fujisaki, H., "Analysis of F_0 contours of Cantonese utterances based on the command-response model," *Proc. ICSLP'04*, Jeju Island, Korea, 2004.

[4] Chao, Y.-R., *Studies of Modern Wu Dialects*, Tsinghua University Press, Beijing, 1928.

[5] Xu, B., and Tang, Z., eds., 上海市区方言志, Shanghai Education Publication House, Shanghai, 1988.

[6] Zee, E., "Tone and syntax in Shanghai dialect," in *From Traditional Phonology to Modern Speech Processing*, Fant, G., Fujisaki, H., Cao, J., and Xu, Y., eds., pp. 507-526, Foreign Language Teaching and Research Press, Beijing, 2004.