Emotional Intonation Modeling: a Cross-Language Study on Chinese and Japanese

Ai-Jun Li*, Yuan Jia*, Qiang Fang* and Jian-Wu Dang†
* Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China.
E-mail: liaj@cass.org.cn. Tel: +86-10-65237408
† Tianjin University, Tianjin, China;
†Japan Advanced Institute of Science and Technology, Japan
E-mail: jdang@jaist.ac.jp

Abstract—**This study attempts to apply PENTA model to simulate the emotional intonations of two typologically distinct languages, the tone language of Mandarin Chinese and the pitch accent language of Japanese. First, the overall F0 features of the emotional intonations of 4 speakers were analyzed and contrasted across seven emotions and across two languages. And then the performances of the qTA model for simulating each language were numerically evaluated and compared within and across the two languages. The results showed that F0 features have bigger distinctions across the two languages than within them. The qTA model can efficiently encode emotional or pragmatic information for both Chinese and Japanese.**

I. INTRODUCTION

As far as the present study is concerned, we mainly focus on emotional prosody, or communicative prosody, the term proposed by Sagisaka [14,15], which is more related to emotion/affect simulation or synthesis. A number of models have been proposed for intonation/prosody generation, such as Fujisaki Model [2], SFC [1], STEM-ML model [4], the PENTA model/qTA [12-13, 16-18] and Daniel Hirst's Momel and Intsint algorithms [3] for the automatic analysis of speech prosody. Xu [19] gave a critical review of methodology and progress in speech prosody. Apart from Fujisaki's model, most of these models have not been applied to the simulation of emotional speech.

PENTA/qTA model (Parallel Encoding and Target Approximation model), proposed by Xu, is based on the assumption that the surface prosody represents both the articulatory mechanisms and communicative functions, while the communicative functions control F0 contours via specific (and parallel) encoding schemes. [133, 184] It has been widely evaluated and applied in several languages [11] by checking the encoding schemes including the neutral tone, focus, sentence types (interrogation / statement), the coding and speaker styles (read/spontaneous) and the target implementation domain (syllable/word) [11, 13, 16].

While, this study extends the domain of the PENTA model from neutral intonation to emotional intonation in two typologically different languages, the tone language of Mandarin Chinese and the pitch accent language of Japanese.

First, the overall F0 features of the emotional intonations of four speakers are analyzed and contrasted across seven emotions and across two languages. Then the performances of the qTA model in simulating each language are compared objectively or numerically within each language for two speakers and between the two languages as well. Finally, the performances of qTA in simulating these two emotional intonations will be compared with those of the neutral intonations for other languages including Mandarin, English, Brazilian Portuguese and German.

II. THE PENTA AND THE qTA MODELS

The Parallel Encoding and Target Approximation (PENTA) model was proposed by Xu based on the assumption that the surface prosody represent both the articulatory mechanisms and communicative functions, the communicative functions control F0 contours via specific (and parallel) encoding schemes [13]

Xu and Prom-on made a general review on the present prosodic models and described the principle of PENTA and the quantitative target approximation (qTA) model in [12,13] as shown in Fig.1 and Fig.2.

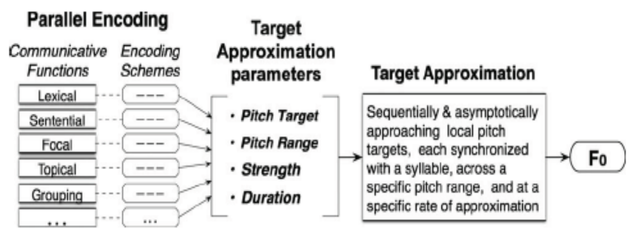

Fig.1 A sketch of the PENTA model from[12,13]

qTA has been tested numerically to simulate tone, lexical stress and focus in Mandarin and English with an automatic analysis-by-synthesis procedure [8]. The core of qTA is the target approximation mechanism. In qTA, a pitch target is a forcing function representing the joint muscular force of the laryngeal muscles that control vocal fold tension. It is represented by a simple linear equation, as in (1)

$$x(t)=mt+b \quad (1)$$

where m and b denote the slope and height of the pitch target, respectively.

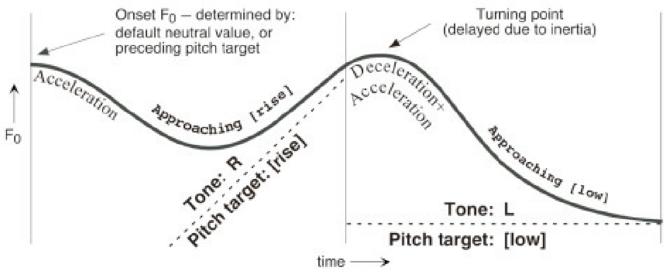

Fig.2. Illustration of the TA model.

The vertical lines represent syllable boundaries. The dashed lines represent underlying pitch targets. The thick curve represents the F0 contour that results from asymptotic approximation of the pitch targets [12,13].

The control of the vocal fold tension in qTA is implemented through a third-order critically damped linear system, in which the total response is

$$F0(t)= x(t)+(c1 + c2.t + c3.t^2).e^{-\lambda.t} \qquad (2)$$

where x(t) is the forced response of the system which is the pitch target and the second part is the natural response, whereas the λ parameter specifies the rate of approaching the target line. The coefficients c1, c2 and c3 are related to the initial conditions. Thus, given these initial conditions, three parameters completely specify the modeled F0 contour for a given linguistic segment. Thus, the qTA model generates a left-to-right local implementation of a F0 contour where the absolute F0 values, together with the first and second F0 derivatives in an utterance segment are defined as the last values of the immediately previous utterance segment.

Since the encoding schemes are hypothesized to be language-specific, this study tries to explore the encoding schemes relating emotional intonation for two typologically different languages: Chinese and Japanese.

For emotional evaluation, i.e. Comparisons of qTA generated F0 contours with those of natural speech are showed in terms of RMSE, Correlation, and perceptual judgment of naturalness [12,13]. In the present paper, the perceptual judgment is excluded.

III. CHINESE AND JAPANESE EMOTION DATA

The Chinese data employed in the study were obtained from the emotional speech corpus Emotion CASS [5-7]. The text material includes a set of 111 sentences with various length (1~14 syllables), different types (narrative/ interrogative), syntactic structures and tonal combinations. The disyllabic sentences contain sixteen tonal combinations.

The Japanese emotion data were obtained from our cross-culture study on emotional speech [9]. The text material includes 10 sentences with various sentence length (1~5 syllables), tonal combinations and grammatical structures.

Both Chinese and Japanese texts are expressive neutral. Two Chinese speakers (male SONG & female AN) and two Japanese speakers (male ZD & female YD) were recruited. The Chinese are university students from a professional oral dubbing department. The Japanese are international students from Tokyo studying in Beijing University. They were selected as the speaker because they are skillful in oral expression.

Seven emotions including 'Disgust, Sad, Angry, Happy, Surprise, Fear and Neutral' were recorded in our sound proof room, resulting 111*7*2=1554 Chinese emotional utterances and 10*7*2=140 Japanese emotional utterances.

All the emotional utterances were annotated with syllable boundaries, and F0 data were extracted and manually corrected by using Praat [10]. The annotated data were stored in TextGrid as requested by qTA trainer.

IV. MODELING THE EMOTIONAL INTONATION USING QTA

The PENTA trainer version 1.4 script [20], run under Praat, was used for modeling the F0 contours. All utterances were segmented into syllables, within which the search for the three qTA parameters, target height and slope, and target strength, took place.

In qTA, two measurements were used to assess the degree of similarity of fit between synthesized and natural F0, RMSE and Pearson's Correlation coefficient. RMSE measures the difference between natural and synthesized F0 contours and Correlation indicates the linear relationship between them. Correlation is used for evaluating F0 contours, not the model parameters. Positive high Correlation indicates consistency between original and synthesized F0 contours not only in height but also in contour shapes. The semitone scale is used for measuring RMSE so as to assess the performance across different speakers.

A. Chinese emotional intonation simulating

The PENTA trainer version 1.4 script, running under Praat, is used for modeling the F0 contours as shown in Fig.3 where the simulated F0 curve of this 'Happy' utterance is well fitted.

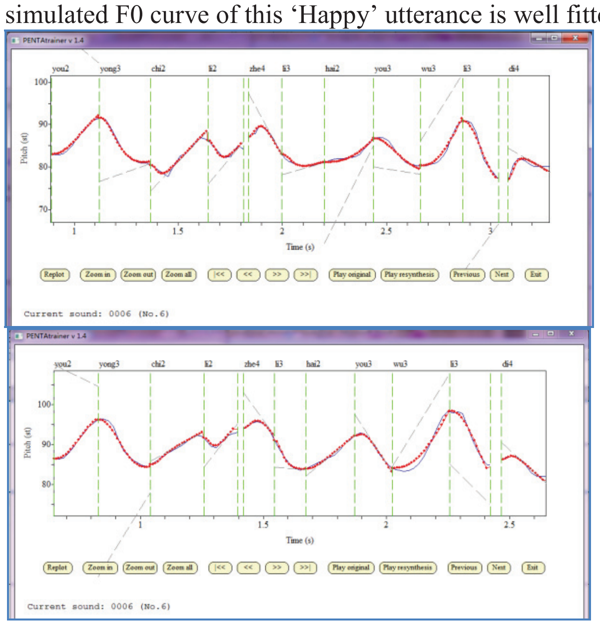

Fig.3 Two synthesized Chinese utterance '游泳池离这里还有 5 里地 (The swimming pool is 5 miles away from here.)' by qTA: the upper panel is for 'Neutral' emotion and the lower panel is for 'Happy' emotion. The blue lines

are the original intonations and the red dotted lines are the simulated intonations.

a. F0 features of Chinese emotional intonation

Figures 4 and 5 depict the mean F0 features of minF0, maxF0 and meanF0 in seven emotions for the two Chinese. The figures show that 'Happy, Surprise and Angry' have higher F0 than other emotions for both speakers.

Comparing the two figures, we see that the variation patterns across seven emotions are consistent except that speaker AN has a lower F0 in 'Sad' intonation than speaker SONG.

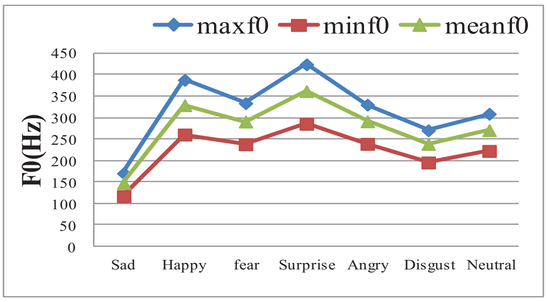

Fig. 4 F0 variations of minF0, maxF0 and meanf0 in seven emotions for Chinese female speaker AN.

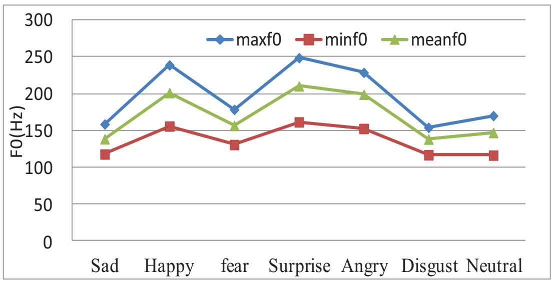

Fig. 5 F0 variations of minF0, maxF0 and meanf0 in seven emotions for Chinese male speaker SONG.

The mean F0 excursion sizes are bigger for 'Happy and Surprise' and smaller for 'Fear and Disgust' as shown in Fig.6. But speaker AN has bigger size for 'Sad' and speaker SONG has smaller size for 'Sad'.

ANOVA analysis on minF0, maxF0 and F0 excursion size indicated that these features are all significantly different across seven emotions for both speakers. (speaker An: maxf0: F(6,3311)=942.30, P=0.0; minf0: F(6,3311)=550.38, P=0.0; meanf0: F(6,3311)=819.37 P-0.0; excursion size: F(6,3311)=13.20, P=0.0). Speaker SONG: maxf0: F(6,3314)=503.05, P=0.0; minf0: F(6,3314)=182.66, P=0.0; meanf0: F(6,3314)=353.04, P=0.0; excursion size: F(6,3314)= 57.91, P=0.0).

The mean F0 excursion sizes are 6.3St and 6.4St for speakers AN and SONG respectively. The F0 features demonstrate that Chinese speakers may adopt different strategies to express emotions.

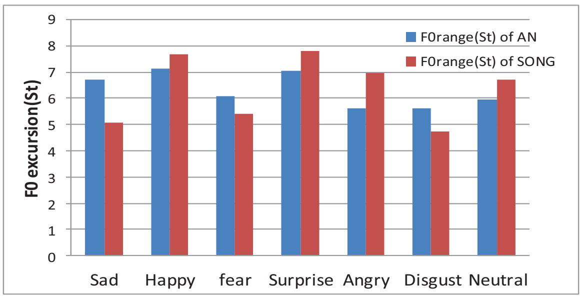

Fig.6 The mean F0 excursion size of the two Chinese speakers in seven emotions.

b. qTA performance of Chinese emotional intonations

How about performance of the qTA to simulate the Chinese intonations in different emotions with so many acoustic deviations? We adopted qTA to simulate 1554 utterances and got the evaluation coefficients of RESE and Correlations. The following figures 7 and 8 show PTA coefficients of mean RMSE and mean Correlations for the two Chinese speakers.

The mean correlations (Standard deviations) are 0.75 (0.02) and 0.76(0.11), the mean RMSEs (Standard deviations) are 0.58St (0.03) and 0.57St(0.09) for speaker AN and speaker SONG respectively.

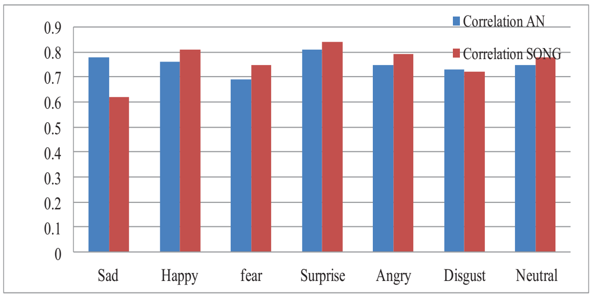

Fig. 7 Mean Correlations of seven emotions for two Chinese speakers.

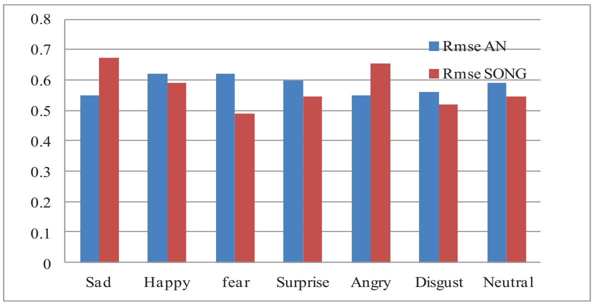

Fig. 8 Mean RMSE of seven emotions for two Chinese speakers.

ANOVA analysis showed that Correlations are significantly different across seven emotions for both Chinese speakers (An: F(6,3311)=9.94, P=0.00<0.05 ; SONG: F(6,3314)=40.10 P=0.00<0.05) ; RMSEs are significantly different across seven emotions for both Chinese speakers (An: F(6,3311)=2.32, P=0.03<0.05 ; SONG: F(6,3314)=13.07 P=0.00<0.05).

TABLE I
ANOVA ANALYSIS FOR RMSE AND CORRELATION BETWEEN TWO CHINESE SPEAKERS

Emotion	P value of RMSE	P value of Corr.
Sad	.000	.000
Happy	**.385**	.005
fear	.000	.000
Surprise	.040	.040
Angry	.000	.008
Disgust	.097	**.476**
Neutral	**.130**	**.123**

The differences of Correlation and RMSE between the two Chinese speakers were statistically analyzed as shown in Table I. The result indicated that 'Disgust and Neutral' have similar 'Correlations' while other emotions have different 'Correlations' between the two speakers; 'Disgust Neutral and Happy' have consistent RMSE while the others have different RMSE between the two speakers.

The results also suggested that qTA performances on 'Happy and Surprise' are higher than 'Neutral' speech, lower for 'Fear and Sad' in some cases. The lower performances may be caused by the 'abnormal' F0 contours such as the trill sound in 'Sad and Fear'.[1]

B. *Japanese emotional intonation simulating*

Based on the F0 data and syllable annotation of the emotional utterances, the simulated emotional intonations were obtained by running Praat script of PENTA trainer on 140 Japanese emotional utterances. Fig. 9 shows two simulated intonation curves which express 'Neutral and Angry' emotions.

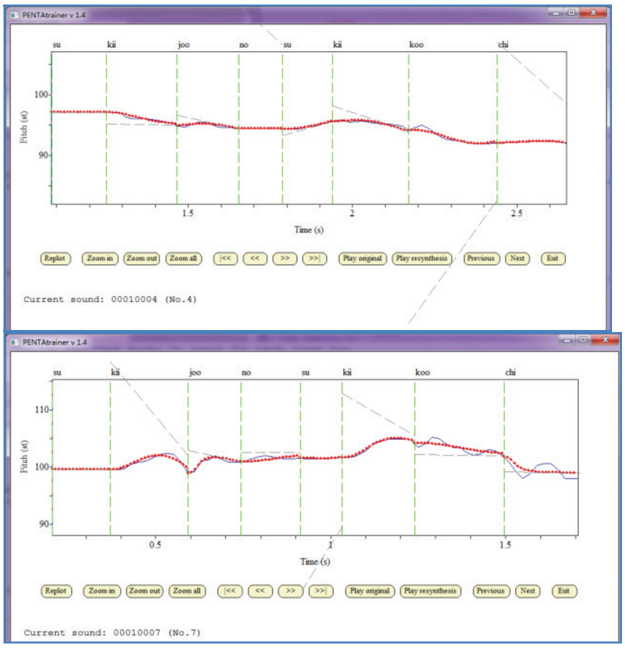

Fig. 9 Two synthesized Japanese utterance 'キー場のスキーコーチ (coach of ski resort)' by qTA: upper panel is for 'Neutral' emotion and lower panel is for 'Angry' emotion. The blue lines are the original intonations and the red dotted lines are the simulated intonations. (Green lines mark the syllable onsets, grey dotted lines are linear target lines)

a. F0 features of Japanese emotional intonation

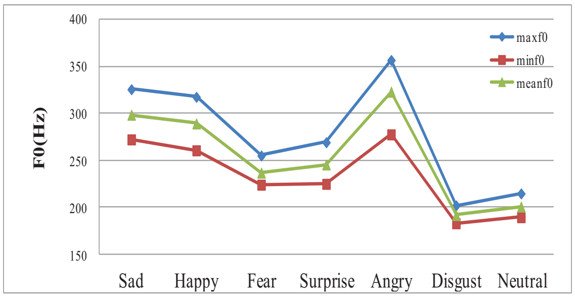

Fig. 10 F0 variations of minF0, maxF0 and meanF0 in seven emotions for Japanese female speaker YD.

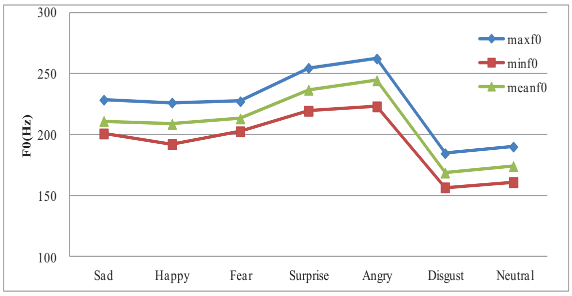

Fig. 11 F0 variations of min F0, max F0 and meanF0 in seven emotions for Japanese male speaker ZD.

F0 features of two Japanese speakers are shown in figures 10 and 11. Based on the ANOVA analysis, it indicated that F0 varies significantly across seven emotions in maxf0, minf0 and meanf0 for both speakers (all P<0.05). For female speaker YD, 'Happy, Angry and Sad' have highest F0 while 'Disgust and Neutral' have lowest F0. For speaker ZD 'Angry and Surprise' have highest F0 while 'Disgust and Neutral' have lowest F0.

For speaker YD (Fig. 12), the mean 'excursion_size' is 2.93St, but the excursion_sizes are significantly different across seven emotions (F(6,382)= 6.883,P=0.0), in which 'Angry, Happy, Surprise and Sad' have wider F0 rang than other emotions. For speaker ZD (Fig. 9), the mean 'excursion_size' is 2.64St, which has no significant difference across seven emotions (F(6,377)= 1.095, P=0.365).

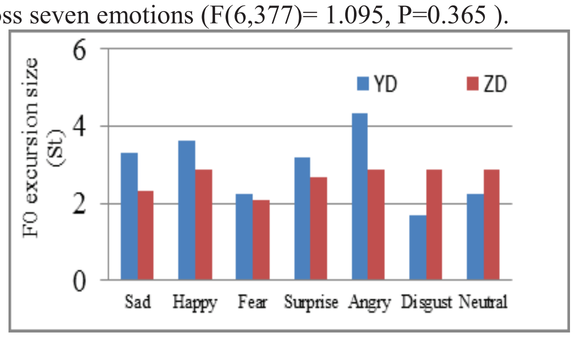

Fig 12: The mean F0 excursion size of the two Japanese speakers in seven emotions.

Comparing the overall F0 varying patterns in Fig. 10 with Fig. 11, we conducted that the two Japanese have less consistent patterns than the two Chinese, especially in 'Fear and Surprise' emotions, which suggests that the two Japanese speakers expressed these two emotions in quite different strategies.

Next section we will focus on how the performance on qTA to simulate the Japanese emotional intonations with so large interpersonal and across emotional deviations.

b. qTA performance of Japanese emotional intonations

The following Figures 13 and 14 show qTA coefficients of mean RMSE and mean Correlation for the two Japanese.

The 'Correlations' keep consistent across seven emotions for both two speakers, for YD around 0.71 (F(6,382)=0.882, P=0.508); for ZD, around 0.76 (F(6,377)=01.461, P=0.191).

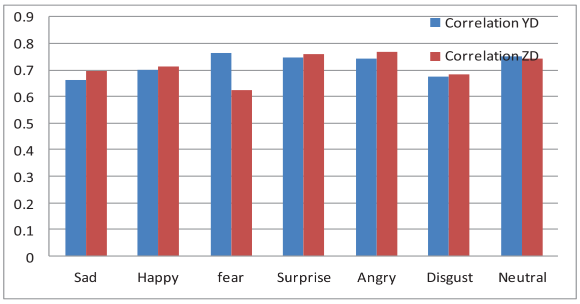

Fig. 13 Mean Correlations of seven emotions for two Japanese speakers.

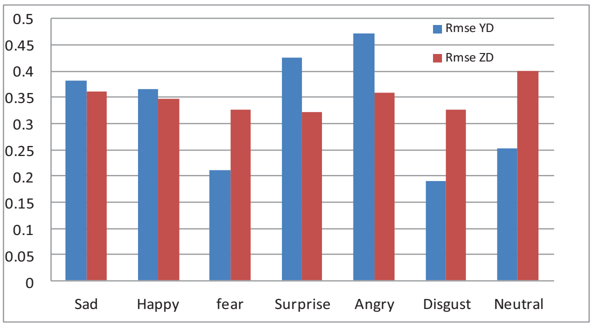

Fig. 14 Mean RMSEs of seven emotions for two Japanese speakers.

For speaker YD, the 'RMSEs'(mean=0.33) are significantly different across seven emotions (F(6,382)= 6.356, P=0.0). 'Disgust Fear and Neutral' have smaller Rmses than the others. While for ZD, the 'RMSEs' (mean=0.35) are not significantly different across seven emotions (F(6,377)= 0.352, P=0.909).

The results suggested that the qTA performance of emotional speech is as good as 'Neutral' speech.

C. Summary on emotional intonations across two languages

a. Overall F0 features

Based on F0 variation patterns for seven emotional states as depicted in Figures 4, 5, 10 and 11, it explicitly showed that Chinese and Japanese express emotions with quite distinct strategies on intonations, but this distinction within the language is smaller than across Languages. For an instance, Chinese always use higher pitch in expressing 'Happy and Surprise' but Japanese don't. Chinese 'Sad' has lower F0 while Japanese don't. Moreover, the mean F0 excursion sizes of the Japanese are almost 3St smaller than those of the Chinese as shown in Fig.4 and Fig. 10, and were summarized in the 1st column in table II.

b. PTA performance

The four speakers have rather consistent mean Correlations ranging from 0.71-0.76 as listed in Table II. Although the mean RMSEs are similar within languages, they are quite different across two languages. The values of Correlation are slightly higher for Chinese but the values of RMSE are smaller for Japanese, so it is hard to reveal which language has better performance than another.

TABLE II.
MEAN PARAMETERS OF THE SEVEN EMOTIONS FOR TWO LANGUAGES

Speakers	F0 Excursion Size(St)	CORR.	RMSE
Chinese AN	6.3	0.75	0.58
Chinese SONG	6.4	0.76	0.57
Japanese YD	2.93	0.71	0.33
Japanese ZD	2.64	0.76	0.35

V. DISCUSSION AND CONCLUSION

The F0 features reveal that the overall variations of intonations are more distinct between Chinese and Japanese than within the two languages. That could be caused by both the typological language distinctions and the cross-cultural differences. In other words, since Chinese is a tonal language, its F0 contour should convey both tone and intonation information, so that the F0 excursion size of Chinese are bigger than that of Japanese. Another reason might be that the Chinese speakers are a professional actor and actress, who vocally expressed the emotions in a quite exaggerated way., while the Japanese speakers expressed the emotions more by using facial actions than vocal actions [8].

TABLE III
CORRELATION AND RMSE COEFFICIENTS IN SIMULATING SEVERAL LANGUAGES.

Language and style	CORR.	RMSE
Read Chinese with tona/focus/position[16]	0.76	2.42
Emotional Chinese (present study)	*0.76*	0.58
Emotional Japanese (present study)	0.74	0.34
Read English(stress/position)[16]	0.77	1.72
Read BP (*Barbosa*) [11]	0.99	1.0
Storytelling BP[11]	0.99	1.2
Read German[11]	0.90	1.4
Storytelling German[11]	0.92	1.3

To check the qTA performances of the simulated emotional intonations for the two languages, the numeric assessment was made on the Correlation and RMSE values, the results indicated that the qTA is efficient for modeling both Japanese and Chinese emotional intonations. Table III summaries these two coefficients when simulating several languages of Chinese [17,18], English[13,16], Japanese, German and Brazilian Portuguese (BP) [11] in different contexts. It reveals that qTA model have quite similar performance in encoding the Chinese and Japanese emotional speech as encoding focus and tone contexts in English and Chinese whose pitch targets are implemented in Syllable domain, while German and BP (basil Portuguese) have higher performance than Chinese, English and Japanese. One of the reasons for this is that tonal targets of German and BP are implemented in word domain in [11].

Until now, we may conclude that PENTA model can used for modeling multi-language intonations including tone language, pitch accent language and stress language.

Besides the numerical assessment, the perceptual assessment has also been made, which indicated that some new encoding scheme has to be adopted for Chinese emotional intonations to express the attitudinal/pragmatic functions as a communicative function.[9]

In previous study on Chinese speech, we know that prosodic features and intonation patterns are quite different between read and spontaneous speech.[21] it seems that the prosodic features of Chinese spontaneous speech are closer to the stress languages as English. So it would be interesting to use PENTA model to typologically check the F0 fitting performance of qTA between spontaneous and read speech with varying target domains across different languages. This is the work we are carrying on.

Acknowledgments

Thanks Dr.Yi XU for providing the script and discussing with the author. This work was supported by the National Basic Research Program (973Program) of China (No. 2013CB329301), NSFC Project with No. 61233009 and 60975081 and CASS innovation project.

References

[1] G. Bailly, and B. Holm, "SFC: a trainable prosodic mode," *Speech Communication*, vol. 46, pp. 348–364, 2005.
[2] H. Fujisaki, and K. Hirose, "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," *J. Acoust. Soc. Japan*, no. 4, pp. 233-242, May 1984.
[3] D. J. Hirst, and A. D. Cristo, "Intonation Systems,A survey of Twenty Language," *Cambridge University Press*, 1998.
[4] G. Kochanski, and C. Shih, "Prosody modeling with soft templates," *Speech Communication*, vol. 39, pp. 311-352, 2003.
[5] A. J. Li, Q. Fang, and J. W. Dang, "Emotional intonation in a tone language: experimental evidence from Chinese," *ICPhS'2011*.
[6] A. J. Li, Q. Fang, and J. W. Dang, "Emotional Expressiveness of Successive Addition Boundary Tone in Mandarin Chinese," *Speech Prosody, Shanghai, China*, 2012.
[7] A. J. Li, Q. Fang, Y. Jia, and J. W. Dang, "Successive Addition Boundary Tone in Chinese Disgust Intonation," *NACCL24, USA*, June, 2012.
[8] A. J. Li, and J. W. Dang, "A Cross-Cultural Investigation on Emotion Expression under Vocal and Facial Conflict - Also an observation on Emotional McGurk Effect," *International symposium on biomechanical and physiological modeling and speech science, Kanazawa. Japan*, 2009.
[9] A. J. Li, Q. Fang, Y. Jia, and J. W. Dang, "More Targets Simulating Emotional Intonation of Mandarin with PENTA," *ISCSLP'2012*, pp. 271-275, 2012.
[10] P. Boersma, and D. Weenink, "Praat," *an open resource on line: http://www.fon.hum.uva.nl/praat/.*
[11] A. Pl´ınio, Barbosa, H. Mixdorff, and S. *Madureira,* "Applying the quantitative target approximation model (qTA) to German and Brazilian Portuguese," *Interspeech2011, Italy*, pp. 2025-2028, 2011.
[12] S. Prom-on, and Y. Xu, "Articulatory-Functional Modeling of Speech Prosody: A Review Proc." *Interspeech2010, Makuhari*, pp. 46–49, 2010.
[13] S. Prom-on, Y. Xu, and B. Thipakorn, "Modeling tone and intonation in Mandarin and English as a process of target approximation," *J. Acoust. Soc. Am.,* vol. 125, no. 1, pp. 405–424, 2009.
[14] Y. Sagisaka, "Modeling prosody variations for communicative speech and the second language towards trans-disciplinary scientific understanding," *Keynote speech of Speech Prosody*, 2012.
[15] Y. Sagisaka, et al., "Prosody generation for communicative speech synthesis," *SNLP'2005*, pp. 23-28.
[16] Y. Xu, and C. H. Xu, "Phonetic realization of focus in English declarative intonation," *J. Phon.*, vol. 33, pp. 159-197, 2005.
[17] Y. Xu, "Speech melody as articulatorily implemented communicative functions," *Speech Communication*, vol. 46, pp. 220-251, 2005.
[18] Y. Xu, and Q. E. Wang, "Pitch targets and their realization: Evidence from Mandarin Chinese," *Speech Communication*, vol.33, pp. 319–337, 2001.
[19] Y. Xu, "Speech prosody: a methodological review," *Journal of Speech Sciences*, no. 1, pp. 85-115, 2011.
[20] Y. Xu, and S. Prom-on, "PENTA Trainer," *praat,* http://www.phon.ucl.ac.uk/home/yi/PENTAtrainer/.
[21] Y. B. Liu, and A. J. Li, "The difference between read and spontaneous spoken Chinese," *J. of Chinese Information Processing*, vol. 16, no. 1, pp. 13-18, 2002.

[This paper was published in APSIPA,2013]

Acoustic Study of Tonal Patterns of Mayang Dialect in Hunan Province

Xiang Ting
Graduate School of Chinese Academy of Social Sciences
Beijing, China
272367381@qq.com

Jia Yuan
Institute of Linguistic
Chinese Academy of Social Sciences
Beijing, China
Summeryuan_2003@126.com

Abstract—**The present study mainly examines the citation tone and tone sandhi patterns of Mayang dialect through the traditional ways as well as the acoustic experiments. As for the citaion tone, Mayang dialect shows four types, i.e., Yinping, Yangping, Shangsheng and Qusheng. Specific values are: 55, 213, 323, and 44. In regards with the di-syllabic tone patterns, there are basically no tone sandhi of from the two level tones, however, their tonal values change to some extent. Specifically, if the tortuous tone of Mayang dialect locates in the initial of the di-syllabic items, there will be a loss of tail. On the other hand, when the tortuous tone lies in the final position of the di-syllabic items, it shows few changes of tonal values. Further, the tone value of the first Tone2 turns into 23 in tone sandhi patterns of Tone 2. In this case, tone sandhi patterns of Tone3 are different from Mandarin for there are no changes in tonal value. As for neutral-tone syllable, the T-value shape of neutral tone is falling and the duration of the neutral tone shows half length as the citation tone.**

Keywords—Mayang dialect; Southwestern Mandarin; Citation tone; Tone sandhi; Acoustic study

I. INTRODUCTION

Mayang is located in the west of Hunan Province and it belongs to Huaihua City. Mayang dialect is in the border zone of Xiang dialect and Southwestern Mandarin. There are several researches about Mayang dialect, which mainly discuss the dialectal branches Mayang dialect belongs to. Specifically, dialects in Huaihua City are mainly classified into two categories in a general survey of dialects conducted by Hunan Normal University. And, Mayang dialect is regarded as Southwestern Mandarin. But in subsequent studies, there remains controversy over the belongings of Mayang dialect. According to dialectal characteristic, Qian [1] proposed that Mayang dialect belongs to Southwestern Mandarin. However, Huang [2] pointed out that Huaihua dialect is a part of dialect of south Guizhou while Zhao and Yan [3] believed that Mayang dialect belongs to Xiang dialect in dialectal division of Hunan dialect. It is different from them Li [4] considered Mayang dialect as a part of Southwestern Mandarin.

Based on the previous researches, we can find that there shows few researches about the acoustic characteristics of citation tone and tone sandhi of Mayang dialect. Therefore, the present study intends to examine the citation tone and tone sandhi patterns of Mayang dialect mainly through the acoustic experiments. We hope that the present study would provide important evidences for the description of the tonal system of Mayang dialect.

II. METHODOLOGY

A. Subjects

Four speakers were recruited as subjects, i.e., two female and two male. All of them are undergraduate students in the university in Beijing. They were born and grew up in Mayang before they went to the university. During the experiment, each speaker was invited to read a list of 20 mono-syllabic words (5 different word×4 tones), 68 di-syllabic words (17 different tone combinations ×4 words) and 16 neutral-tone di-syllabic items (4 different combinations ×4 words).

B. Materials

There are two criteria for selecting mono-syllabic words: first, since clear boundaries between consonants and vowels are show in stop and affricatives, the study selects three consonants [p], [t], [k] as the initials. Second, the vowels, of [i], [a], [e], [ə] are the classical and standard vowels, therefore, they are selected as the target vowels. The study obtained 20 samples for the analysis of citation tone according to the above criteria. In regard with the di-syllabic constitutions, the study selected four samples for each tonal composition. Finally, the study got 68 samples for acoustic analysis. All the selected constitutions are the commonly used items in Mayang dialect.

C. Recording procedure

The recording software is CUHK-SIAT Recording Tool. All the above tokens were randomly listed in the recording software. Each word would be produced more than once but only one sample is taken for further analysis. The equipment of recording was laptop and head-wear with microphone. Recording was conducted in a quiet room. The source files were obtained from recording software with sample as 44K Hz. All the source files were saved in 'WAV' forms.

D. Data annotation and data collection

All sounds files were first automatically processed by segmentation program to generate both word and segment boundaries. Inaccurate boundaries were modified manually to ensure the accuracy of the data. As for the data analysis, the present study adopts the relatively normalized T-value calculation [5]. It is known that the tone is represented by the F_0 contour. The F_0 data were extracted by Praat script with one syllable being selected ten points to normalize the duration. As for the statistical analysis, the four tones are selected as the independent variable to get the means of the F_0 data. All the F_0 values then were transformed to LOG values. Then the LOG values were calculated by the formula as below:

$$T=[(\lg x-\lg b)/ (\lg a-\lg b)] \times 5 \quad (1)$$

Within the formula, x stands for the target F_0 value of measurement, a means the maximum pitch value and b indicates the minimum value of the pitch range, and T stands for the five-scale value obtained from the formula

III. Analysis the Results

A. Tonal pattern of mono-syllabic item

In this part, the study examines the tonal pattern of Mayang dialect. It further compares the results from the traditional analysis and the acoustic experiment.

According to the field study based on the Character Table of Dialect Investigation [6], we could obtain the citation tonal system of Mayang dialect, as shown in TABLE I.

TABLE I. Tonal value of citation tone of Mayang dialect

Tone	*Tone1*	*Tone2*	*Tone3*	*Tone4*
Tonal value	55	213	312	33

The following Figure1 is the normalized tonal value of the citation tone of Mayang dialect. The Y-axis is the five scale values and the right part of the graph illustrates the contents of the graph. Tone1, Tone2, Tone3 and Tone4 are shown as T1, T2, T3 and T4, respectively.

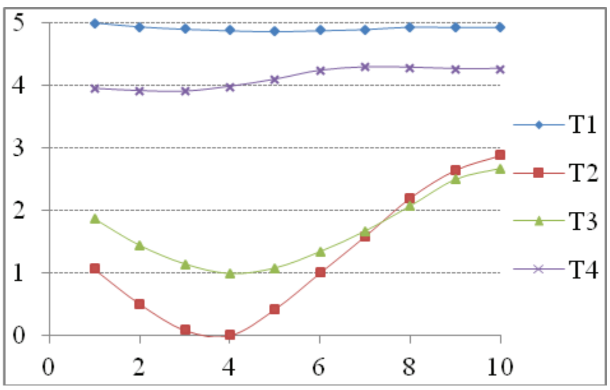

Fig.1 The tonal pattern of mono-syllabic in Mayang dialect

As shown in fig.1, the five-scale for each tone can be seen clearly. For tone1, it performs as a level tone with tonal value as '55', while tonal value of tone4 is '44'. And, both tone2 and tone3 perform as a falling-rising tone. For tone2, the falling part of its contour starts from '1' and drops to 0, and the rising part is from '0' to '3', therefore, we define tonal value as '213' according to the acoustic result. As for tone3, the falling part is from '2' to '1' and the rising part is from '1' to '3', we take its tonal value as '323'.

The following TABLE II presents the results of the field study and the acoustic data through which the study makes a comparison between the two results.

TABLE II. The tonal value of citation tone of Mayang dialect

Tone	*Tone1*	*Tone2*	*Tone3*	*Tone4*
Field study	55	213	312	33
Experimental method	55	213	323	44

In comparison with results of field study, as shown in the TABLE II, the difference lies in following aspects:

a) tonal value of tone3 is '323' in this present study, while as '312' in field study;

b) tone4 performs as '44' in the experiment, while as '33' in the field study.

We can make a conclusion that the acoustic results are similar with that of field study on the tonal pattern. The differences between the field study and the experimental study may lie in the following aspects: i) the differences among subjects, the subjects selected in the study are university students and they were treated Mandarin since primary school. Therefore, their accents may be influenced by the Mandarin; ii) the error from the field study, the tonal value from the field study mainly based on the written record made by one recorder, there may show minor mistakes.

B. Tone sandhi pattern

Field study of tone sandhi of Mayang dialect demonstrated that among the sixteen tonal combinations, Mayang dialect exhibits tone sandhi pattern as described below:

a) There is no tone sandhi when the initial tone is tone1.

b) The tone2 shows as 13 before tone2, and 21 before tone1, tone3, tone4.

e.g. 'parade down street'[yiə$^{213}_{21}$ kai55]
'serious'[yian$^{213}_{13}$ men213]
'running water'[lɪ$^{213}_{21}$ ɕyə312]
'annual interest'[nian$^{213}_{21}$ lɪ33]

c) The tone3 shows as 31 when it locates in the initial of the bi-syllabic items.

e.g. 'navy'[xai$^{312}_{31}$ tɕyen55]
'height above sea level'[xai$^{312}_{31}$ pa213]
'behind schedule'[ʷan$^{312}_{31}$ tian312]
'sea rover'[xai$^{312}_{31}$ tau33]

d) There isn't any change when tone4 locates before tone2, tone3, and tone4. If 'T4+T1' is a V-O construction, tone 4 performs as a neutral tone, otherwise, tone1 performs as a unstressed syllable.

e.g. 'go to the countryside'[xia$^{33}_{1}$ xiaŋ55]
'surgical department'[ʷai33 khuo$^{55}_{4}$]

1) The analysis of experimental data

The following figures *2-5* are tonal patterns of all sixteen tonal combinations, each of them having the same initial tone. These descriptions are listed in the right part of the graphs. The Y-axis illustrates the five-scale values.

a) The tone sandhi pattern of Yinping

Fig.2 shows four combinations which have the same initial tone, i.e., tone1. We can see that Yinping in the initial position of the di-syllable always performs as a high level tone. For the combination of 'T1+T1', both the initial 'T1' and the final 'T1' perform as '44' that is similar with its citation tone '55'. For 'T1+T2', T2 performs as '212'. For T1+T3, T1 still performs as '55'. However, for T3 in the combination of T1+T3, the starting point of T3 is higher than its corresponding citation tone, the ending point is also lower. For 'T1+T4', both of them perform as a level tone, and tone1 performs as '55', tone4 as '33'.

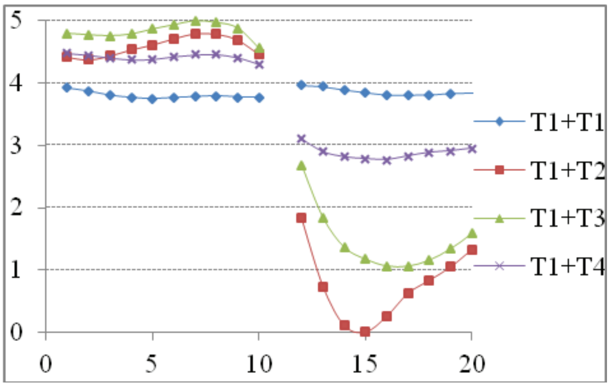

Fig.2 Tone sandhi pattern with the same initial tone1

b) The tone sandhi pattern of Yangping

Fig.3 illustrates four combinations with tone2 as initial tone and the tone1, tone2, tone3 and tone4 are located after tone2. In regard with 'T2+T1', T2 performs as '21' for there are a loss of tail, and T1 performs as '34' or more like a level tone as '44'. For 'T2+T2', the initial T2 still performs as a tortuous tone as its citation form, but the rising part of its contour is more obvious than its falling part, while the second tone2 remains unchanged. Therefore, we can draw a conclusion that if a tone2 is followed by another tone2, it will change to '23'. For 'T2+T3', T2 performs as '21' and T3 still shows as a tortuous tone '323'. As for 'T2+T4', T2 performs as '21' but tone4 goes from a level tone '44' to a rising tone '34'.

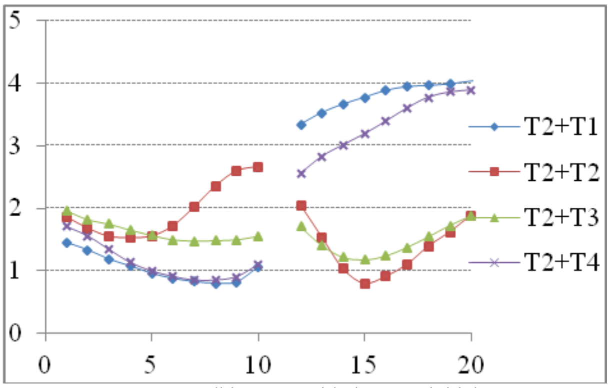

Fig.3 Tone sandhi pattern with the same initial tone2

c) The tone sandhi pattern of Shangsheng

Fig.4 shows four combinations of all which have the same initial tone, i.e., tone 3. The tone3, which is a tortuous tone, performs as a falling tone '32' when it locates in the initial of the bi-syllabic items. For 'T3+T1', tone1 changes into '44' although it remains be a level tone. With regard to T3+T2, T2 shows as a tortuous tone '213'. For 'T3+T3', the initial tone3 changes into '32' while the second tone3 remains unchanged, which is different from tone sandhi pattern of tone3 in Mandarin. For 'T3+T4', tone3 performs as '32' and tone4 changes to a rising tone '34'.

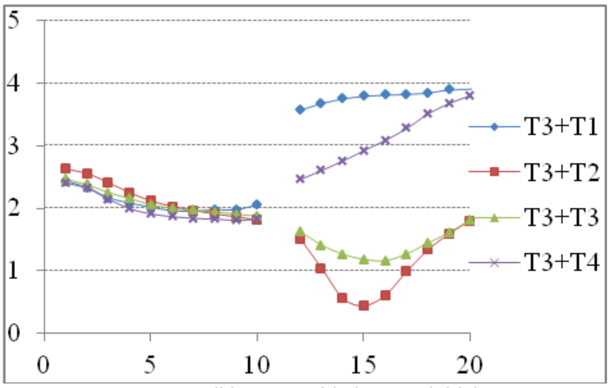

Fig.4 Tone sandhi pattern with the same initial tone3

d) The tone sandhi pattern of Qusheng

Fig.5 shows four combinations with tone4 as the initial tone. Specifically, for'T4+T1', when the di-syllabic constitutions are a V-O construction, the initial tone4 performs as an unstressed syllable and its tonal value is '21' on the one hand. And, the final tone1 performs as a neutral tone'21' on the other hand. As for 'T4+T2', tone4 still performs as a level tone but tonal value changes to '55', at the same time, tone2 changes to '312'. With regard to 'T4+T3', tone4 performs as '44' and tone3 performs as '312'. For T4+T4, the main part of their contour is locating between 3 to 4 in above figure, so we record them as '44+44'.

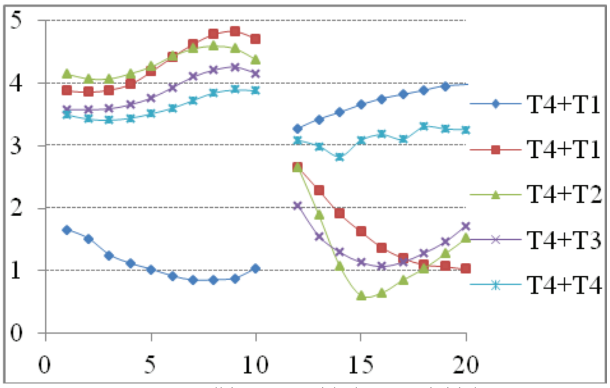

Fig.5 Tone sandhi pattern with the same initial tone4

C. The neutral tone

By the above researches, we know that there is not only unstressed syllable on the initial syllable (e.g. 'cook noodles' [xia33 mian4]) but also on the final syllable (e.g. 'below'[xia1 mian55]) in Mayang dialect. In present study, we mainly discuss the neutral tone on the second syllable.

1) *Field study of neutral tone*

- Before tone1, mid-high, tonal value is '3', e.g. 'things to wear'[ɕyan55 kə3].
- Before tone2, mid-low, tonal value is '2', e.g. 'the family of husband'[pho213 ka2].
- Before tone3, low, tonal value is '1', e.g. 'mine'[ŋo312 kə1].
- Before tone4, high, tonal value is '4', e.g. 'alive'[xuo33 kə4].

2) *The analysis of experiemental data*

a) *The T-value of neutral tone*

Fig.6 illustrates the four combinations of the tone sandhi pattern of neutral tone in Mayang dialect. The contours of the neutral tone are falling. For 'T1+T0', T1 performs as '55' and T0 performs as 31. For T2+T0, T0 performs as 31 and T2 still shows as a tortuous tone. As for T3+T0, T0 performs as 21 and T3 performs as 213.For T4+T0, T4 changes into a rising tone '45', while T0 performs as '41'.

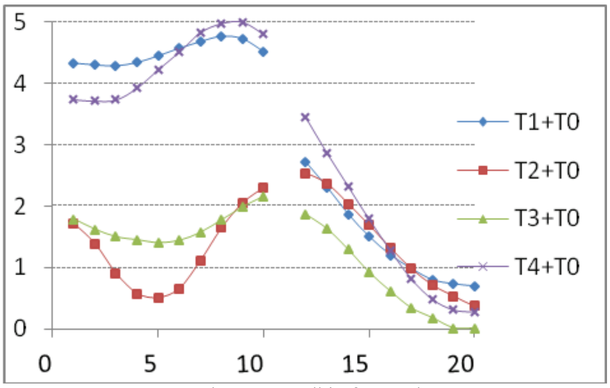

Fig.6 The tone sandhi of neutral tone

b) *The comparison of duration*

Figures 7-10 show the duration of neutral tone as well as the citation tone on the 68 di-syllabic test words. In the figures, the green bars represent the disyllabic words and the red bars represent the combinations including the neutral tone. The durations are divided into two parts, the duration of the initial syllable and the duration of the second syllable. The Y-axis represents the duration. We use the T0 to represent the neutral tone in the following graphs.

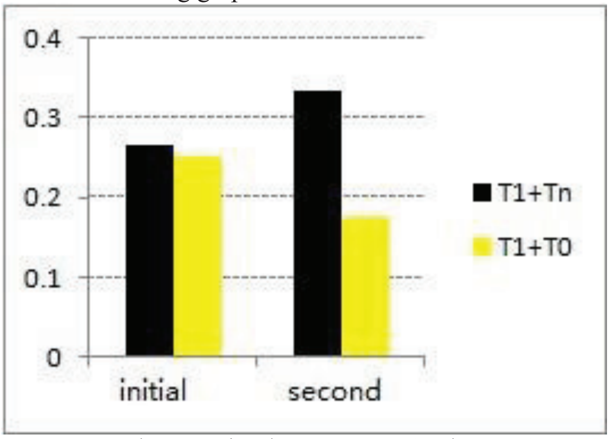

Fig.7 The comparison between 'T1+Tn' and 'T1+T0'

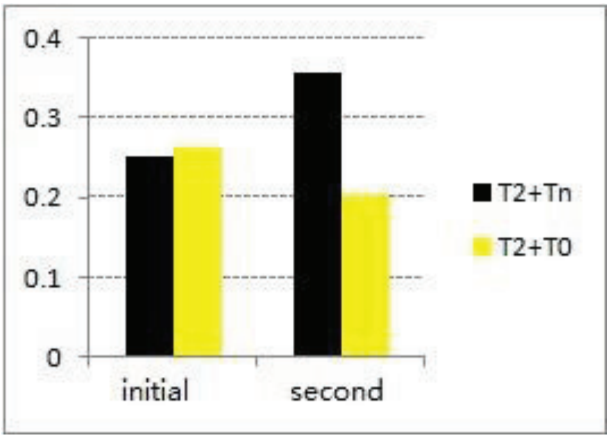

Fig.8 The comparison between 'T2+Tn' and 'T2+T0'

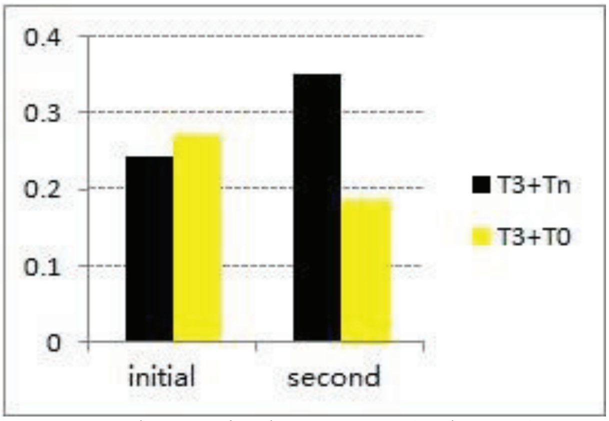

Fig.9 The comparison between 'T3+Tn' and 'T3+T0'

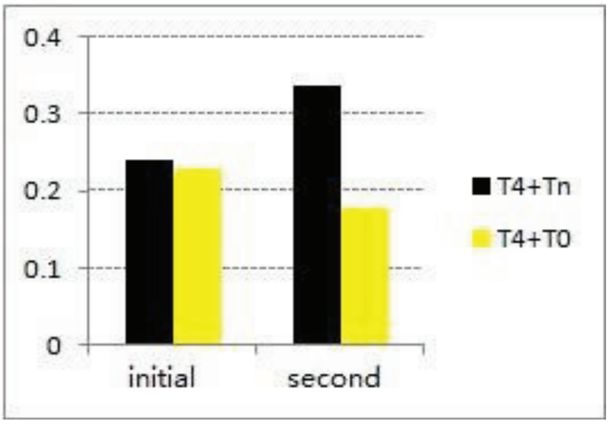

Fig.10 The comparison between 'T4+Tn' and 'T4+T0'

As can be seen in figures 7-10, the neutral tone is always shorter than the preceding citation tone, whether the citation tone is [55], [213], [323] or [44]. The duration of neutral tone ranges from 170-200ms, which is half as much as the duration of second syllable in the di-syllabic combinations without neutral tone. As for the initial syllable, there are little difference between the normal di-syllabic combinations and the neutral tone.

IV. CONLUSION AND DISCUSSION

In the present study, we adopt both the traditional research and acoustic experiments to investigate the citation tone and tone sandhi of Mayang dialect. The purpose of this paper endeavors to use the experimental method of phonetics to summarize the tonal pattern of more dialect of China. Results of citation tone are as following: Yinping (tone1) shows as 55, Yangping (tone2) shows as 213, Shangsheng (tone3) shows as 323, Qusheng (tone4) shows as 33. The conclusion we draw from traditional research is different from that of acoustic experiments. In traditional research, different informants or investigators often lead to different conclusions. But acoustic experiments provide the data more objectively from the sound materials and get an acoustic conclusion. Compared with traditional ways (even though we have to admit that it is an important method), the outcomes from acoustic experiments are more objective, so we chose the latter one as the effective results.

The discussion of tone sandhi of Mayang dialect found more differences according to the results of field study. There are two tortuous tones in Mayang dialect, i.e. tone2'213' and tone3'323'. As can be seen from TABLE III, the tonal value of the first tone 2 turns into 23 in tone sandhi patterns of tone 2, while in tone sandhi pattern of tone3 there is any change except that the first tone3 witness a loss of tail. Therefore, why does tone sandhi occur in tone sandhi of tone2 but not in tone sandhi of tone3? We try to give 2 hypotheses. The first hypothesis argues that the rising part of tone2 is a main character that distinguishes itself from other tortuous tones in Mayang dialect. On the other hand, the main difference between Shangsheng(tone3) and Yangping(tone2) is the falling part of tone3. The second hypothesis supposes that tone sandhi which occurs in tone sandhi of tone3 of Mandarin doesn't occur in tone sandhi of tone3 of Mayang dialect because of the inadequate materials or the limited informants. In this way, more studies are needed to make sure whether there is tone sandhi of tone3 in Mayang dialect.

For the neutral tone in Mayang dialect, it always performs as a 'light' and 'short' tone. The uniqueness of neutral tone lies in the fact that its tonal value varies according to its tonal environment. Besides, the duration of neutral tone is always half of the duration of the second syllable in tone sandhi.

TABLE III. The comparison of tone sandhi

	Field study	*experimental method*
T1+T1	55+55	55+55→44+44
T1+T2	55+213	55+213→55+212
T1+T3	55+312	55+312
T1+T4	55+33	55+44→55+33
T1+T0	55+3	55+31
T2+T1	213+55→21+55	213+55→21+55
T2+T2	213+213→13+213	213+213→23+213
T2+T3	213+312→21+312	213+323→21+323
T2+T4	213+33→21+33	213+44→21+34
T2+T0	213+2	213+31
T3+T1	312+55→31+55	323+55→32+44
T3+T2	312+213→31+213	323+213→32+213
T3+T3	312+312→31+312	323+323→32+312
T3+T4	312+33→31+33	323+44→32+34
T3+T0	312+1	213+21
T4+T1	33+55→1+55	44+55→21+44
	33+55→33+4	44+55→44+31
T4+T2	33+213	44+213→55+323
T4+T3	33+312	33+312
T4+T4	33+33	44+44
T4+T0	33+4	45+41

V. Acknowledgment

This research was supported by Innovation Program of Chinese Academy of Social Sciences "The acquisition and cognition of spontaneous speech".

VI. References

[1] Xia Xianzhong, "The division and characteristics of Huaihua Dialects in Hunan Province", in Journal of Southwest University for Nationalities(Humanities and Social Science), in press.

[2] Huang Xuezhen, "The division of Southwestern Mandarin", in Dialect, in press.

[3] Bao Houxing and Yan Sen, "The division of dialects in Hunan Province", in Dialect, in press.

[4] Li Lan, "Review and subcategorization of the division of Hunan Dialect", in Linguistics Study, in press.

[5] Shi Feng, "The analysis of tonal value of di-syllable of Tianjin Dialect", in Studies in Language and Linguistics, in press.

[6] Chinese Academy of Social Sciences, Character Table of Dialect Investigation, Beijing: The Commercial Press, 2011.

[7] Cao Zhiyun, "A introduction of dialects in west of Hunan", in Linguistic Researches, in press.

[8] Chen Zerong, "Tona sandhi and neutral tone of Southwestern Mandarin", in Journal of Tibet Nationalities Institute(Philosophy and Social Sciences Edition), in press.

[9] Bao Houxing and Chen Hui, "Chinese dialect in Hunan province", in Dialect,in press.

[10] Li Xia, The Speech Research of Southwestern Mandarin, Shanghai, Masteral Thesis of Shanghai Normal University, 2004.

[11] Zhao Binhui, The Study of the Dialect Tonal Experimental Research in the District of Zhangjiajie and Huaihua, Hunan, Masteral Thesis of Shanghai Normal University, 2010.

[12] Australian Academy of the Humanities and Chinese Academy of Social Sciences, Language Atlas of China, Beijing: The commercial Press, 2010.

[This paper was published in O-COCOSDA, 2013]