Lexical Tones Learning with Automatic Music Composition System Considering Prosody of Mandarin Chinese

Siwei Qin, Satoru Fukayama, Takuya Nishimoto and Shigeki Sagayama

Graduate School of Information Science and Technology, the University of Tokyo, Japan

{qin, fukayama, nishi, sagayama}@hil.t.u-tokyo.ac.jp

Abstract

Recent research has found that there is an overlap in the processing of music and speech in certain aspects. This research focuses on the relationship between the pitch of tones in language and the melody of songs. We present an automatic music composition system based on the prosody rules of Mandarin and we hypothesize that songs generated with our proposed system can help non-native Mandarin speakers to learn the tones of Mandarin Chinese more easily. To verify this hypothesis, twelve non-Chinese speakers from Japan were asked to identify and pronounce the Mandarin sentence they heard in the experiments with three different learning methods. The result shows that participants got higher accuracies of performances in tone3 with the teaching method of "speech + music" and the teaching method of "music only" is not more effective than "speech only" in some particular tones.

Index Terms: automatic music composition, pitch, prosody, Mandarin, tone learning

1. Introduction

Since Mandarin Chinese is a tonal language, the learning of tones in Mandarin is difficult for non-tonal language speakers [1]. Even for Japanese learners, whose language is defined as a pitch-accented language, tones in Mandarin are the most difficult element and prone to be forgotten in learning Mandarin. This is because in Japanese pitch changes only between syllables, while in Mandarin, pitch changes within the syllables [2].

The way of teaching and acquisition of tones is simple. A survey by Guan showed that all the learners acquired tones from their teachers instead of from textbooks [1]. He mentioned that one of the problems of teaching tones is the weakness of teaching methods.

The only way for the learners to acquire tones is to imitate their teachers, but it is a difficult task for those who are not familiar with the pitch variation in language. Music, which is also represented by the variations of pitches, seems to be a significant aid in learning tones.

Music is suggested to have some relationship with speech [3]. Both musicologists and linguists have realized considerable correlations between the two, a famous example of which is Beethoven's String Quartet No. 16, which is said to have been composed by considering the prosody of German sentences "Muss es sein? Es muss sein". *Yukiko* used songs as teaching aids in the Japanese language classroom [4]. *Chao Y.R.*, the Chinese linguist who invented the method of registering tones in 5 degrees [5], also found some forms of melodies that conform to linguistic tones of the lyrics in Kunqu (a traditional kind of Chinese opera) [6]. *Yu Jiang* argued that the concept of music can be introduced to help learners understanding how pitch changes in each tone [7].

However no system that uses prosody of Mandarin to compose songs for learning the tones has been attempted, which we hypothesize will be helpful to remember the tones.

Table 1. *Tones in Mandarin of "ma"*

type	Syllable	tone	gloss
Tone 1	ma1	high level	"mother"
Tone 2	ma2	rising	"hemp"
Tone 3	ma3	low-falling-(rising)	"horse"
Tone 4	ma4	falling	"scold"

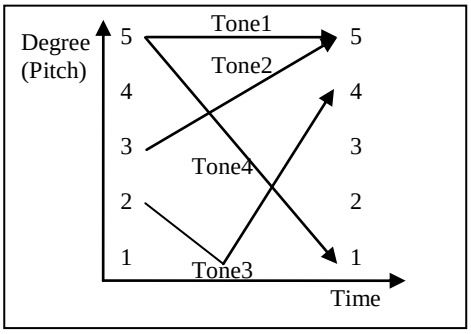

Figure 1: *Chao Y. R.'s 5-degree theory of Mandarin Tone*

2. Automatic Music Composition for Lexical Tone Learning

The objective of the research is to present an easier way to learn tones of Mandarin. Our work is based on the hypothesis that music and melody can be easier to receive and remember by contrast with speech.

We use prosody rules to make our algorithm of song composition. In this section, we discuss the rules of tones used for composition.

2.1. Tones in Mandarin

There are four tones in Mandarin Chinese. They differ from each other by the changes of their pitches.

As shown in Table 1, every syllable in Mandarin can have one of four tones. Every tone can represent different meaning. So if the speaker makes a mistake on tones, his/her message may possibly be misunderstood. *Chao Y. R.* was the first to invent a method of registering tones in 5 degrees (Figure 1) in 1930, which is still widely used for teaching tones.

Among these four tones, Tone 3 often turns to a low-falling shape in connected speech (labeled with "Tone 3-" in this paper), and has a rising tail only in final position of utterances (labeled with "Tone 3*" in this paper) [8].

2.2. Melodic notes used to represent tones

Each syllable of Mandarin is suggested to carry two moras except the neutral tone [9]. Hence, two melodic notes can be used to represent a syllable in Mandarin Chinese except Tone 3* which can be represented by three melodic notes.

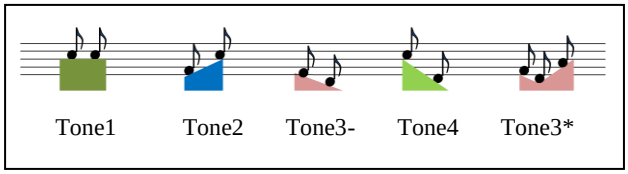

Figure 2: *High-Low constraints of notes representing tones*

The pitch of melody plays a similar role as in speech. Therefore, the pitch contour of speech can be used as a constraint on the relation between two neighboring melody notes when we compose a song. From the Chao Y. R.'s 5-degree theory, we can obtain the High-Low constraints on the notes in one syllable (Figure 2).

Now we have to define the High-Low relation between two neighboring melody notes from different syllables. Turning back to Figure 1, by comparing the starting degree with the ending degree of all tones, we can decide whether the first note of a syllable should be higher or lower than the previous one. For example, since the starting degree of Tone 2 is "3" which is higher than the ending degree of Tone 3- and Tone 4, and lower than that of Tone 1, Tone 2 and Tone 3*, the first melody note should be higher than the last note of Tone 3- and Tone 4, while lower than the ones of Tone 1, Tone 2 and Tone 3*.

2.3. Rhythm in melody

Since Mandarin is a tonal language, the length of the syllable does not affect the distinguishing of two tones. It has been shown that normal length of each syllable in Mandarin is in the range of 200-350 ms and there is no distinct difference between simple finals and compound finals [9]. Therefore, we believe that it is reasonable to set all syllables to the same length. However, for the notes in a single syllable, we do not set them to the same length since a study by Wee [11] showed that the high pitch of Tone 1 and 2 and the low pitch of Tone 3 and 4 undergo phonetic lengthening in Mandarin songs. We set the rhythm of Tone 1, 2, 3- and 4 to a sixteenth note connected with a dotted eight note and Tone 3* by sixteenth–eight–sixteenth note set.

2.4. Melody Composition

In order to aid the composition of a melody, chord progression and accompaniment are also be modeled in the system, and are independent of the tones of lyrics. All the rules of tones, rhythm, chord progression and accompaniment can be seen as a constraint on transition and occurrences of the melody notes. Thus, a song can be composed by finding a melody which optimally satisfies all these limitations.

Melody can be represented as a path, as shown in Figure 3. There are two kinds of constraints: linguistic constraints which ensure that the melody obeys the rules of prosody, and musical constraints which ensure that the melody obeys the music theory. Given the pitch series of the melody as a MIDI note number $X = \{x_0, x_1, \cdots, x_n\}$, the cost for the melody X is calculated as follows:

$$Cost(X) = p_{oc}(x_0)\prod_{t=1}^{n} p_{tr}(x_t \mid x_{t-1})\, p_{oc}(x_t) \qquad (1)$$

where $p_{oc}(x_0)$ is the occurrence probability defined by music constraints, and $p_{tr}(x_t \mid x_{t-1})$ is the transition probability determined by the tone rule. Melody composition can be

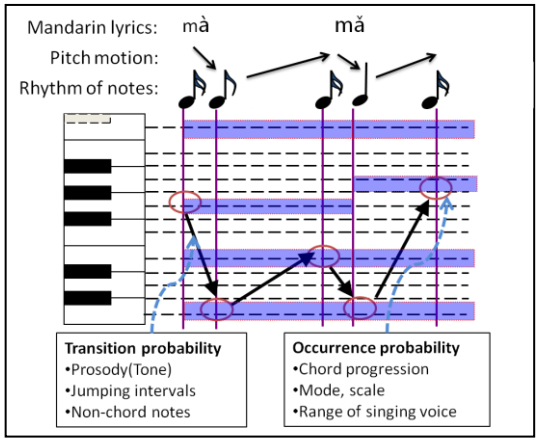

Figure 3: *Model of automatic composition as an optimal-path searching*

formalized by finding the optimal X^* which maximize $\log Cost(X)$:

$$X^* = \arg\max_{x} \log Cost(X). \qquad (2)$$

We can obtain the series of $X^* = \{x_0^*, x_1^*, \cdots, x_n^*\}$ by using dynamic programming.

3. Implementation and Experiments

3.1. Implementation of the composition system

We used Orpheus [12] which is an automatic composition system that we implemented for Japanese lyrics. We changed the interface of it to make it accept Mandarin Chinese. We also changed the rule of processing prosody of it to include the tones of Mandarin. After we get the pinyin with tones from the lyrics, constraint on melody by considering the pitch motion, as we discussed above, can be added to generate to transition probability for automatic composition. The modified Orpheus system accepts a Chinese phrase with 7 or 8 characters and repeats the lyrics four times in an eight-bar song. Since we have not found a singing voice synthesizer of Mandarin, the songs have to be sung by a human. A two-bar example of a song for lyrics "huan1 ying2 ni3- dao4 zhong1 guo2 guan3*" is shown in Figure 4.

3.2. Experiments

3.2.1. Subjects

Twelve Japanese native speakers participated in the experiment. They were all males and ranged from 21 to 26 years of age, with a mean age of 23.7 years and SD of 1.5 years. All the participants reported that they had no previous exposure to Mandarin.

3.2.2. Contents

We prepared six sentences of Mandarin that have some practical significance considering real language education environment. Four sentences consisted of seven characters and the other two consisted of eight characters. Each tone appeared at least once in every sentence. The sum of all tones in all sentences was arranged to be same.

We fed these sentences into our system to compose six songs and we asked the same Mandarin native speaker both to read it clearly in declarative sentence and sing the song strictly

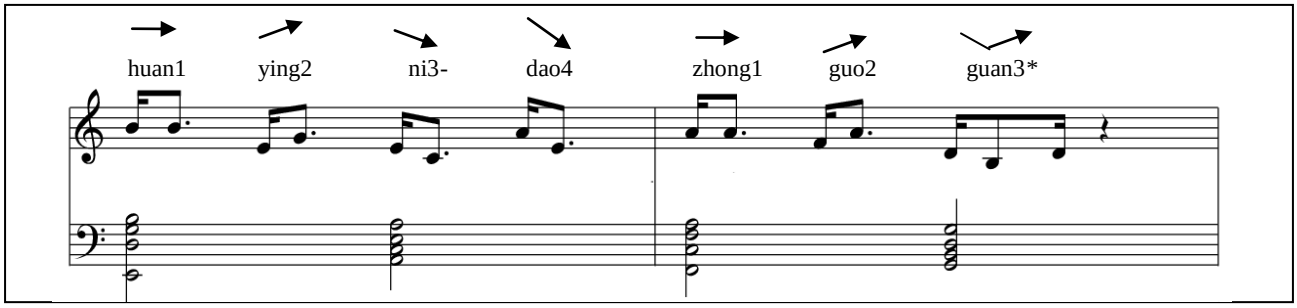

Figure 4: *Example of generated song with the lyrics input of "欢迎你到中国馆 (welcome to China Pavilion)"*

following the melodies composed by our system. Both the reading and singing contents was recorded at 44.1 kHz. The tempo of all the songs was set to 110 beats per minute, so the average duration of each character was 0.545 s. The speech was recorded with the same speed as the song.

3.2.3. Methods

The contents mentioned in section 3.2.2 were played to the participants with three methods listed below:

- "speech only": speech for 8 times
- "music only": music for 8 times
- "speech + music": speech for 4 times and music for 4 times

3.2.4. Procedures

There were two kinds of tasks in this experiment: tone identification and tone reproduction. The same subjects participated in both of them and both experiments took place within one testing session.

3.2.4.1 Before the experiment

Before the start of the experiment, all subjects were given a short tutorial in order to familiarize them with Mandarin tone system. We taught them the different pitch patterns of Mandarin tones and explained how to mark and differentiate them according to the pitch variations. A sound example of the syllable "ma" pronounced in four different tones was played to them. Finally, they learned to pronounce the syllable "ma" in four tones. Any pronunciation mistake would be corrected by the experimenter in the tutorial section. After the short tutorial, they were explained the procedures (which will be introduced in next paragraph) and the tasks and they were asked to join the experiment. They were allowed to ask any questions before the experiment began.

3.2.4.2 In the experiment

Three methods mentioned in section 3.2.3 are matched with three pairs of the sentences mentioned in section 3.2.2 by Latin Square Design [13] to reduce Sequence Error.

Before they listened to a sentence, pronunciation of each syllable without tone information was shown to them in katakana, which is a Japanese syllabary, chosen because it is familiar to the Japanese and so they could concentrate on the tones. To check that they indeed do not know the tones of the sentence, they were first asked to pronounce the sentences according to the katakana shown on the screen. Then they heard the sentence played with a particular method. After that, they were asked to pronounce the sentence after a 3-second direction (1st tone reproduction task, marked as "Repro1" in the figures and tables). They then wrote down the type of Mandarin tones of each syllable in the sentence they listened to on a paper given to them, within a time limit of 30 seconds (tone identification task, marked as "ID" in the figures and tables). No blanks were permitted. In the tone identification task, the accuracy was logged. 30 seconds later, they were asked to pronounce the sentences again (2nd tone reproduction task, marked as "Repro2" in the figures and tables). The time limit at was kept and all the pronunciations of the participants were recorded for calculating the accuracy.

3.2.4.3 After the experiment

After the experiment, all the sound files recorded in the experiment were submitted to PRAAT [14] to analyze the pitch patterns the participants pronounced.

Horizontal pitch patterns were counted as Tone 1; rising pitch patterns were counted as Tone 2; falling-rising pitch patterns were counted as Tone 3*. Since Tone 3- and Tone 4 both show a falling pitch pattern, three Mandarin native speakers were invited to judge the type of tones for falling patterns. They also judged some strange pattern such as "rising-falling" and if there were no more than two same judgments, the strange tone would be counted as "none".

4. Results

The average accuracies of the participants' answers for "speech only", "speech + music" and "music only" in the three tasks are shown in Figure 5. We also calculated the accuracies of the participants' answers for each tone, which are summarized in Table 2.

The data from the experiment was submitted to one-way ANOVA test [15] with learning method as the factor. The analysis showed that there were no significant differences among these three methods ($F_{(2,22)}$=2.30) when using the total accuracies of all tones.

The analysis of data for each tone showed that the accuracy for Tone 2 for "speech + music" was significantly higher than that for both "music only" and "speech + music" in the 1st tone reproduction task (MSe=0.044, $p<0.05$). The accuracy of Tone 2 for "speech only" was significantly higher than that for both "speech + music" and "music only" in the 2nd tone reproduction task (MSe=0.019, $p<0.05$). The accuracy of Tone 3- for "speech only" was significantly higher than that for "music only" (MSe=0.199, $p<0.05$). The accuracy of Tone 3* for "speech + music" was significantly higher than that for "speech only" (MSe=0.216, $p<0.05$), as is shown in Figure 6.

5. Discussions

Despite the fact that the result showed no significant differences among the teaching methods in terms of total accuracy of tone recognition and reproduction in each task, we found significant differences among the methods by analyzing

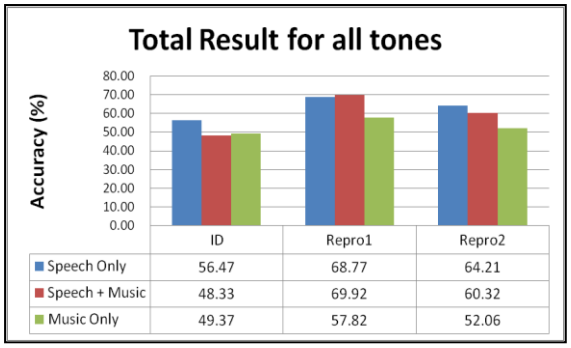

Figure 5: *Accuracies of the participants' performances with the method of "speech only", "speech + music" and "music only" in the three tasks*

Table 2. *Average accuracies of each tone (%)*

Tone	Method	ID	Repro1	Repro2
Tone1	speech only	72.92	75.00	63.19
	speech+music	59.72	56.94	56.25
	music only	65.28	70.14	68.06
Tone2	speech only	36.11	66.67	69.44
	speech+music	34.72	71.53	53.47
	music only	38.19	50.00	45.83
Tone3-	speech only	44.44	54.17	55.56
	speech+music	16.67	61.11	45.83
	music only	13.89	29.17	25.00
Tone4	speech only	67.36	86.81	79.86
	speech+music	66.67	82.64	76.39
	music only	57.64	72.92	59.03
Tone3*	speech only	83.33	41.67	33.33
	speech+music	83.33	91.67	75.00
	music only	83.33	58.33	58.33

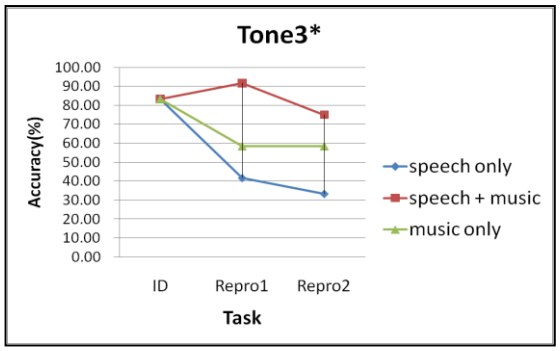

Figure 6: *Accuracies of the participants' performances with three methods in tone3**

the accuracies of each tone separately.

In case of Tone 3*, the participants got significantly better accuracies in both tasks of tone reproduction for "speech + music", than for "speech only". This finding indicates that melodies generated with a system that considers the tonal contour aided learning of Mandarin Tone 3*. It would suggest that when the participants heard a falling-rising pattern in the melody of the song, they could imitate the variations of the pitch from melody more easily than only from speech.

In cases of Tone 2 and Tone 3-, the accuracies for "speech only" method were higher than the accuracies for "music only". A probable reason might be that it is more difficult to associate musical pitch with pitch accent than to relate them with the hint of speech. Another probable reason is that the melodies did not represent the tones so well since we determined only the pitch motions, but did not control how much a note should be higher or lower from the previous one. It was found in the generated songs that sometimes Tone 3- and Tone 4 were represented by a set of notes with same variation of pitch, which could confuse the participants of the experiment. Hence, the improvement of the composition algorithm to avoid confusion of the tones will be the task for our future work.

Our system currently does not treat neutral tone. Since neutral tones commonly appear in Mandarin, the treatment of this tone will also be included in our future work.

6. Conclusion

This research attempted to design an automatic music composition system that considers the rules of Mandarin tones. We hypothesized that songs generated with this system can aid non-native Mandarin speakers in learning the tones. In a set of tone identification and reproduction experiments using three teaching methods: "speech only", "speech + music" and "music only", Japanese participants got higher accuracies for Tone 3* with "speech + music" method. This suggests that songs generated with our system may help learning of Mandarin Tone 3*. We also found that the "music only" method is not more effective than "speech only" in Tone 2 and Tone 3-. However, we did not find more significant differences among the three teaching methods through current experiment.

This is just a pilot research on the application of automatic music composition on tones learning, we plan to improve our composition algorithm to make the system helpful to learning all Mandarin tones.

7. References

[1] Guan Jian, "Preliminary Exploration on Reformation of tone teaching", Language Teaching and Linguistic Studies, 51-54, 2000-4.

[2] Chen Ziyou, "Ri han tai liu xue sheng han yu sheng diao xi de ji pian wu fen xi yan jiu" (Mistakes and difficulties of the Japanese, Korean, and Thai students study in the tones), Master Thesis of Shaanxi Normal University, 2007.

[3] George List, "The Boundaries of Speech and Song", Ethnomusicology, Vol. 7, pp. 1-16, 1963.

[4] Yukiko S. Jolly, "The Use of Songs in Teaching Foreign Languages", The Modern Language Journal, Vol. 59, No. 1/2, pp. 11-14, 1975.

[5] Chao, Y. R., "A system of tone letters，" Le Maitre Phonetique 45, pp24-27, 1930.

[6] Chao, Y. R., "Tone, intonation, singsong, chanting, recitatives, tonal composition, and atonal composition in Chinese (Mouton, The Hagu", Mouton, The Hague, 1956.

[7] Yu jiang, "A New Teaching Plan for Chinese Tones", Language Teaching and Linguistic Studies, 77-81, 2007-1.

[8] Jialing Wang, Norval Smith, "Studies in Chinese phonology", 82-83, 1997.

[9] Wang Hongjun, "han yu fei xian xing yin xi xue" (Chinese non-linear phonology), 240, 1999.

[10] Feng long, "The Length of Tones in Mandarin", Beijing experimental phonetics, 1985.

[11] Wee, Lian Hee, "Unraveling the Relation between Mandarin Tones and Musical Melody", Journal of Chinese Linguistics, 35.1:128-144, 2007.

[12] Satoru Fukayama, et al. "Orpheus: Automatic Composition System Considering Prosody of Japanese Lyrics," Entertainment Computing - ICEC 2009, pp.309-310, Sep., 2009.

[13] D.C. Montgomery, "Design and Analysis of Experiments. fifth ed.", John Wiley and Sons, pp. 144-150, 1997.

[14] Boersma, P., and Weeknik, D., "Praat: Doing phonetics by computer", http://www.fon.hum.uva.nl/praat/, v5.1.32, 2010.

[15] Satoshi Tanaka, "Practical Psychological Data Analysis",93-116,2006, Online program: http://www.kisnet.or.jp/nappa/software/star/puma/sa.htm