

# Automatic Composing Based On Pitch Detection of Chinese Poetry Recitation

Rongfeng Li<sup>1, a</sup> and Xinyun Zhang<sup>2, b</sup>

<sup>1</sup>Beijing University of Posts and Telecommunication, Xitucheng Road 10#, Beijing, China

<sup>2</sup>Peking University, Yiheyuan Road 5#, Beijing, China

<sup>a</sup>lirongfeng@bupt.edu.cn, <sup>b</sup>zhangxinyun@pku.edu.cn

**Keywords:** Computer Music, Digital Audio Signal Processing, Algorithmic Composition

**Abstract:** The balance between randomization and regularization is a key and contradictory issue in the development of automatic composing. Randomized algorithm focuses on creating innovative and novel melodies but will generate unpleasant result. Regularized algorithm uses the existed music and rules to compose pleasant music piece, but the results are similar to the pieces in the training set. In this paper, we set up a new automatic composition model by detecting the pitch of Chinese poetry recitation. First, we utilize the Simplified Inverse Filter Tracking (SIFT) to detect the pitch of the acoustic input signal. Variety of the intonations of the texts of Chinese poetry ensures the innovativeness and randomness of the pitch change. Meanwhile, the congruence between melody and tones makes the change of pitch musically. In order to make the rhythm musical as well, we introduce the database of music scores (in the form of MusicXML) and apply a matching algorithm to determine the duration of each note. The experiment shows that the melody generated by our system is both innovative and in agreement with the musicology in pitch and rhythm.

## 1. Introduction

The balance between control and random is still a key and contradictory issue in the development of automatic composing. In recent times, most automatic composing algorithms are based on certain rules or database of pre-collected music materials which make the result similar to the music pieces in the database. Otherwise, if we overly randomize the composing procedure, the results will become less musical and unpleasant and hard to accept by common audience. The key reason is that, there is no musical information in the process of randomization.

In our work, we turn to non-musical art forms to discover the musical information in them, including poem recitation, painting, dancing, etc. In the 20<sup>th</sup> century, artists tended to be more innovative, and expressionism arose. Art works were trying to exaggerate human intrinsic and emotions rather than represent physical reality. Gaining inspiration from all expressionist art forms such as dancing, painting and other visual art, some music masterpieces were made by music composers.

The middle of the 20th century witnessed the climax of the interaction between composers and painters, when a group New York School [1], including John Cage and Morton Feldman established a profound friendship with a group of painters. Morton Feldman's "Why Patterns" was composed according to the idea of texture derived from Jasper John's crosshatch paintings.

The research of automatic composing using non-musical input has been started in recent years. Satoru Fukayama[2] used the tones of the lyrics to determine the change of the pitch. The tone analysis is based on text; thus it could not obtain the actual pitch from acoustic signal. Robert [3] and Daniel [4] developed this method to transform the pitch curve of the sound from poetry and natural sound into musical notes. Although they both used machine learning model to get the relationship between tones and pitch, their rhythm was not musical.

In this paper, we will introduce an automatic composing system based on pitch detection of the acoustic signal of Chinese poem recitation. In Chinese ancient Tang and Song dynasties, poetry was composed with both texts and melody. Poems could be sung. The texts of the poems must be consistent with their melody. Musical information can be extracted from the texts of Chinese poetry. As for rhythm, we will introduce the database of music scores (in the form of MusicXML) and use a matching algorithm to determine the duration of each note. Chinese poets created poems with regard to their melody, thus the texts of the poems are musical. In addition, the database of music scores is only used to learn the relationship between pitch and rhythm, therefore, our result will not be similar to the music pieces in the database. In this way, we succeed in balancing randomization and regularization.

The rest of this paper is structured as follows. We introduce the automatic composing system based on pitch detection of acoustic signal from Chinese poem recitation in section 2. Section 3 provides the experimental settings and results. Finally, we draw the conclusion and future discussion in section 4.

## 2. Automatic Composing Based On Pitch Detection of Chinese Poetry Recitation

Regarding the congruence between the melody and the tone of the texts, we introduce the automatic composing system based on pitch detection of acoustic signal of Chinese poems. This system consists three principle components: pitch detecting system, pitch matching system and composing system, We will present these three systems in the following paragraphs.

**2.1 Pitch Detecting System for Automatic Composing.** Our first job is to detect the pitch of the poem recitation based on the congruence between the tone of the texts and the melody. This congruence is an aesthetic standard in Chinese songs, because each Chinese character has its own tone. The same character with different tones represent different meanings. When the texts and the melody agree with each other, the song can be sung pleasantly and the audience can easily understand the meaning of the lyrics.

In this paper, we utilize Simplified Inverse Filter Tracking (SIFT) to detect the pitch of the input acoustic signal. SIFT algorithm, proposed by Markel [5], is a modern version of Autocorrelation Function (ACF).

In our implementation of the SIFT algorithm, acoustic signal was first band pass filtered using an IIR filter with frequency band of 40~800Hz in order to eliminate the noise from vocal tract.

The key parts of SIFT are linear prediction and inverse filtering. According to the source-filter model, speech signal  $s(n)$  is the convolution of a sound source, such as the vocal cords excitation  $e(n)$ , and a linear acoustic filter, the vocal tract response  $v(n)$ , that is:

$$s(n) = e(n) * v(n) \quad (1)$$

In Z domain, we have:

$$V(Z) = S(Z) / E(Z) \quad (2)$$

$S(Z)$ ,  $E(Z)$  and  $V(Z)$  are the Z transform of  $s(n)$ ,  $e(n)$  and  $v(n)$  respectively. With the above formulation, a four order asynchronous Linear Prediction Analysis is performed, that is:

$$V(Z) = \frac{1}{1 - \sum_{i=1}^P a_i Z^{-i}} \quad (3)$$

Considering (2), we have:

$$S(Z) - S(Z) \sum_{i=1}^P a_i Z^{-i} = E(Z) \quad (4)$$

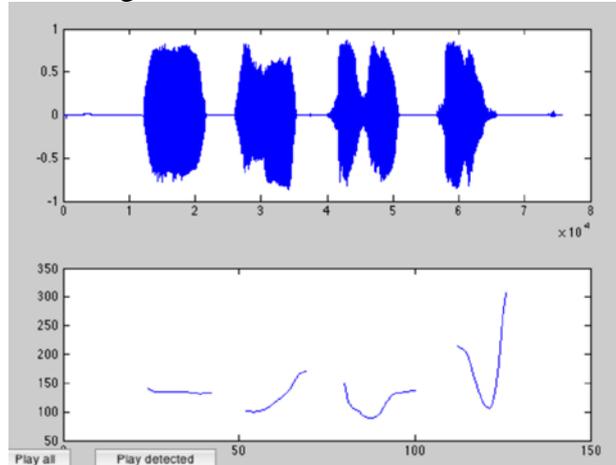
Transforming back to time domain, we have:

$$e(n) = s(n) - \sum_{i=1}^P a_i s(n - i) \quad (5)$$

Then we use the least square estimation rule to get the vocal cords excitation  $e(n)$ .

Finally, the autocorrelation function was searched for peaks to get the pitch curve of the acoustic signal of poem recitation.

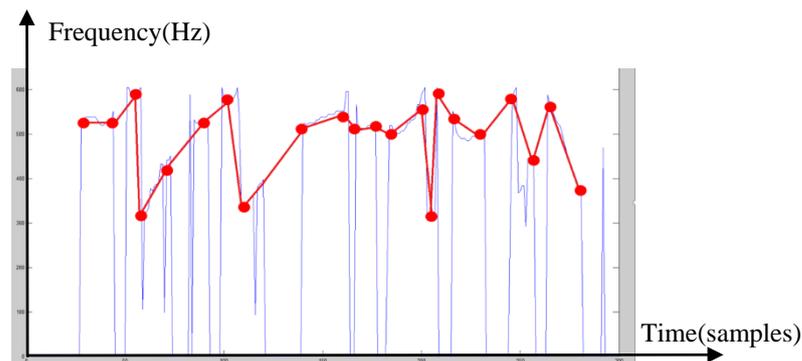
To validate the SIFT algorithm, we read out “阴yin”, “阳yang”, “上 shang”, “去 qu” which is four basic tone in Chinese, and detect the pitch curve. The result is shown in Figure 1 in which the upper part is the acoustic signal in time domain and the down part is the pitch curve. The pitch curve is consistent with the rising and falling in the 4 basic tone.



**Figure 1** SIFT Pitch detecting for the four basic Chinese tone.

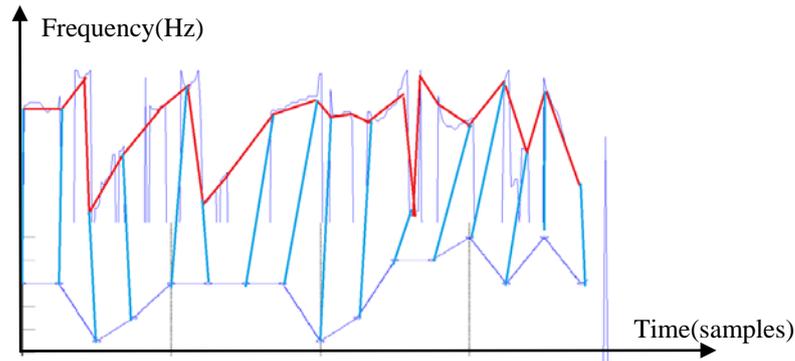
**2.2 Pitch Matching System for Automatic Composing.** Although the tones of the texts of Chinese poems can be transformed into musical pitch curve, the duration of each character is flexible. This makes it difficult to determine the rhythm pattern from the poem recitation. To solve this problem, a matching algorithm is performed to match a proper musical piece from a music score database (in MusicXML form), whose rising and falling are similar to the poem recitation piece.

The pitch curve of poem recitation is somehow continuous and the notes in music score are discrete. In order to perform the matching algorithm, we first discretize the pitch curve of poem recitation by finding the peak point after local smoothing, which is illustrated in Figure 2.



**Figure 2** Peak point of the pitch curve.

Then we perform a Smith-Waterman matching algorithm to determine the similar melody regions between the recitation and the song in the database. Instead of looking at the total sequence, the Smith–Waterman algorithm compares segments of all possible lengths and optimizes the similarity measure. An example of the matching result is illustrated in Figure 3.



**Figure 3** Matching of the rising and falling sequence of poem recitation and music score in the database.

**2.3 Automatic Composing System.** The automatic composing system consists of rhythm assignment and pitch adjustment.

#### **Rhythm Assignment**

We denote the local matching sequences are  $\{p_{i_1}, p_{i_2}, \dots, p_{i_k}\}$  and  $\{m_{j_1}, m_{j_2}, \dots, m_{j_k}\}$ . According to the matching system, we can get the local rising and falling sequence  $\{p_{i_1}, p_{i_2}, \dots, p_{i_k}\}$  and  $\{m_{j_1}, m_{j_2}, \dots, m_{j_k}\}$  which are matching. Then we can get the corresponding notes of  $\{g_{j_1}, g_{j_2}, \dots, g_{j_k}\}$  and look-up the MusicXML fields of “duration” to get duration  $\{d_{j_1}, d_{j_2}, \dots, d_{j_k}\}$  of each note. After that, we assign the duration sequence  $\{d_{j_1}, d_{j_2}, \dots, d_{j_k}\}$  to the corresponding pitch sequence of poem recitation, which is  $\{f_{j_1}, f_{j_2}, \dots, f_{j_k}\}$ . Thus,  $\{f_{j_1}, f_{j_2}, \dots, f_{j_k}\}$  and  $\{d_{j_1}, d_{j_2}, \dots, d_{j_k}\}$  make up a piece of melody.

#### **Pitch Adjustment**

Although the rising and falling of pitch curve is congruence with the music notes, the sequence of poem recitation pitch sequence  $\{f_{j_1}, f_{j_2}, \dots, f_{j_k}\}$  may be unmelodic because rules of pitch changing in poem recitation is not the same as singing. To make the result pleasant, we adjust the pitch interval into harmonious interval, that is major third, minor third, perfect fourth and perfect fifth.

### **3. Experimental Results**

In the experiment, we take acoustic signal of reciting Chinese Poems in The Interpretation of Suijin[11] and generate the melody using the automatic composing system introduced in section 2. Then we also use the musical score in [11] to generate the MusicXML database for matching system introduced in section 2.2. [11] includes 96 songs (12617 musical notes in all) from Tang and Song Dynasty. We indicate each song by  $S = \{(f_1, f_2, \dots, f_n), (d_1, d_2, \dots, d_n)\}$ .  $f_i$  and  $d_i$  are the frequency and duration of the  $i^{\text{th}}$  notes in one song. Then the similarity of two songs  $S_1 = \{(f_1^{(1)}, f_2^{(1)}, \dots, f_n^{(1)}), (d_1^{(1)}, d_2^{(1)}, \dots, d_n^{(1)})\}$  and  $S_2 = \{(f_1^{(2)}, f_2^{(2)}, \dots, f_n^{(2)}), (d_1^{(2)}, d_2^{(2)}, \dots, d_n^{(2)})\}$  can be calculated by the following function:

$$\text{sim}(S_1, S_2) = 1 - \frac{\sum (f_i^{(1)} - f_i^{(2)})^2 + \sum (d_i^{(1)} - d_i^{(2)})^2}{\sum f_i^{(1)2} + \sum d_i^{(1)2}} \quad (6)$$

We compare the pitch and duration of each note of the results of automatic composing system and the original score, the similarity is 82.42%. The result shows that the music generated by the automatic composing system is very similar to the original score.

### **4. Conclusion and Future Discussion**

The research on automatic composition has been worked on for many years, but most of them are based on a certain composing mechanisms or machine learning database with some specific characteristics; hence, they can hardly obtain the balance between hearable and random. It is a brand

new research field to compose automatically based on non-musical factors. Though there are a few researchers who have proposed to compose with natural sound or painting, a mature automatic composing system based on non-musical factors has not yet been found. In this paper, we have taken the acoustic input of Chinese poem recitation as a non-musical factor. The users will recite the poem, and the computer will compose based on the recitation. The tunes of acoustic input will ensure the innovativeness of composition, whereas the rhythm based on music pieces in the database will ensure the audibility. We have paid great attention to the important meaning of the matching between the tunes of the texts and the melody. The rising and falling of the tunes are matched with the music pieces in the database. In addition, we adjust the pitch which is detected from the poem reciting, aiming to ensure the audibility of the melody.

Our composing system can be useful for professional composers. Poetry recitation can be transformed into music, with the help of which the composers can study the relationship between poetry and music.

The system will also benefit the amateur composers who may not obtain enough knowledge or skills on music but know well about poetry recitation. Once they recite a poem, a corresponding melody of the poem can be composed automatically, and the melody is well agreed with the texts and can be sung pleasantly.

The evaluation of the composing system is still a tough job because there is no right or wrong in music. In further research, we will develop an online version of composing system in order to collect the feedback of users and evaluate the system by scoring and comments.

## **5. Acknowledgement**

This work is supported by Youth Innovation Project of BUPT under Grant No.2015RC32

## **6. References**

- [1] Steven Johnson, "The New York Schools of Music and the Visual Arts", Routledge, New York, November 7, 2001, ISBN-10: 0415936942
- [2] S. Fukayama, K. Nakatsuma, S. Sako, et al. "Orpheus: Automatic Composition System Considering Prosody of Japanese Lyrics." In: *Entertainment Computing-ICEC 2009*. Springer Berlin Heidelberg, 309-310, 2009.
- [3] Robert Smith, Aaron Dennis, and Dan Ventura. "Automatic composition from non-musical inspiration sources." *Proceedings of the Third International Conference on Computational Creativity, Dublin, Ireland*. 2012.
- [4] Daniel S. Johnson. "Musical Motif Discovery in Non-Musical Media." *All Theses and Dissertations*, Paper 4081. Brigham Young University, 2014.
- [5] Markel, J.D. "The SIFT algorithm for fundamental frequency estimation", *IEEE Transaction on Audio and Electroacoustic*, vol. 20, pp. 367-377, 1972