

## Based on Gaussian Mixture Model to Explore the General Characteristic and Make Recognition of *XinTianYou*

Juan Li<sup>1</sup>, Kai Wu<sup>2</sup>, Yinrui Wang<sup>2</sup> and Xinyu Yang<sup>2</sup>

<sup>1</sup>Centre for Music Education, Xi'an Jiaotong University, 710049 Xi'an, China

<sup>2</sup>Department of Computer Science & Technology, Xi'an Jiaotong University, 710049 Xi'an, China

**Abstract.** *XinTianYou*, a folk song style from Shannxi province in China, is considered to be a precious traditional culture heritage. Research about *XinTianYou* is important to the overall Chinese folk music theory and is potentially quite useful for the culture preservation and applications. In this paper, we analyze the general characteristics of *XinTianYou* by using the pitch, rhythm features and the combination of these two features. First, we use the Gaussian Mixture Model (GMM) to cluster the *XinTianYou* audio based on pitch and rhythm respectively, and analyze the general characteristics of *XinTianYou* based on the clustering result. Second, we propose an improved Features Relative Contribution Algorithm (CFRCA) to compare the contributions of pitch and rhythm. Third, the probability of a song being *XinTianYou* can be estimated based on the GMM and the cosine similarity distance. The experimental results show that *XinTianYou* has large pitch span and large proportion of high pitch value (about 22%). Regarding the rhythm, we find that *moderato* is dominated while *lento-moderato* keep a similar ratio as *moderato-allegro*. The similarity between pitch features of all *XinTianYou* songs is more significant than rhythm features. Additionally, the average accuracy of *XinTianYou* recognition reaches 92.4% based on our method

### 1 Introduction

*XinTianYou* is an improvisational folk song style which stems from Northern Shannxi, China. It is a unique music form and a treasure of traditional Chinese culture. Exploring the general characteristics of *XinTianYou* can improve the understanding of the aesthetic characteristics and the artistic style of this genre of music, and potentially benefit the cultural preservation.

In this paper we aim to find and understand the general characteristics of *XinTianYou*. We use the Gaussian Mixture Model (GMM) to analyze the *XinTianYou* audio features including pitch and rhythm and do the clustering. Then analyze the general characteristics of *XinTianYou* based on the clustering results and music theory. A new method named Complete Features Relative Contribution Algorithm (CFRCA) is introduced to compare the contribution of pitch and rhythm. Lastly, we use GMM and the cosine similarity distance to recognize *XinTianYou*. Potentially our method can also be applied to other types of folk songs, which enlighten the study of finding the "general characteristics" of different kinds of folk songs and further find the general characteristics or gene of Chinese culture.

There is a variety of related work in Chinese folk music area that has looked at finding general characteristics of *XinTianYou*. In [1] the authors point out

that the tune of *XinTianYou* is generally divided into two types: one has slow speed, long-drawn-out melody and broad range, while the other has a more rigorous structure, graceful and smooth melody. Chen. et al [2] think that *XinTianYou* has free rhythm and unchained melody styles. Zhang. et al [3] points out that the basic structure of *XinTianYou* is stable and clear. The basic structure contains two continuous 4th melodies. *XinTianYou* also features sonorous drawn-out strains and flexible tempo. Liu [4] states that *XinTianYou* has simple tunes but rich mode scale, and "double 4th interval structure" melody is an important feature of *XinTianYou*. These previous studies focus on the music theory. They cannot support their conclusion with precise data evidence and recognize *XinTianYou* automatically because of the lack of quantitative analysis and scientific experiment.

In the area of music information retrieval, quantitative analysis and identification method have been used. But lots of research work focuses on the classification of music genres [5-7], which mainly cares about the accuracy of classification rather than the general characteristics.

In this paper, we apply mathematic methods and modern data analysis techniques to analyze the general characteristics and make recognition of *XinTianYou*

. We first apply GMM to the pitch and rhythm features of the music data, then explore the general characteristics, and use GMM combined with cosine similarity to identify *XinTianYou*. The experiment shows that our results are consistent with music theory and our method can identify *XinTianYou* effectively.

The paper is organized as follows: we describe the use of GMM for audio features in Section 2 then analyze the general characteristics of *XinTianYou* based on pitch and rhythm in Section 3 next we demonstrate the experimental results in Section 4 and finally the conclusions and future work in Section 5.

## 2 The processing of GMM based on features of audio data

Gaussian Mixture model (GMM) [8-9] has been widely used in the acoustic area. According to the statistics theory, using a linear combination of Gaussian probability density functions can approximate arbitrary distribution, thus theoretically GMM could describe the statistical distribution of music elements. We apply GMM to cluster the features of *XinTianYou* audio data. The clustering result can represent overall information of the audio data, and is useful for analyzing its general characteristics and further the reorganization of *XinTianYou*.

### 2.1 Formatting the title, authors and affiliations

The database we use in this paper contains audio data of 109 *XinTianYou* songs. They are selected from Integration of Chinese folk songs [1], which is one of the most authoritative collections of Chinese folk songs. This collection contains the most rich, comprehensive and authentic Chinese folk song music [10]. After getting audio data for each song, we decompose each audio data to short segments in order to get locally stable signal. To make sure the adjacent segments transfer smoothly, we allow overlap between them. Assume that a piece of audio data has been sampled with frequency F, we can get NL samples. The number of frames Nframes of this data can be computed as follows:

$$N_{frames} = \left\lfloor \frac{N_L - N_M}{N_f - N_M} \right\rfloor \quad (1)$$

In which Nf is frame length - the number of samples in each frame, and NM, is overlap length - the number of samples in the overlapping region. Practically F, Nf and NM are set to 22050Hz, 1024 and 896 respectively.

After the above preprocessing, we compute two types of features: the pitch-domain features LSP (Linear Spectrum Pairs) [11], PF (Pitch Frequency) [12], and time-domain features Tempo [13], FP(Fluctuation Pattern) [14]. We use the software MARSYAS [15] (Music Analysis, Retrieval and Synthesis for Audio Signals) and MIR-TOOLBOX[16] to extract these features.

## 2.2 The process of GMM

GMM Gaussian mixture model uses the weighted linear combination of Gaussian probability density functions to representation the statistical distribution of data. Here we build a GMM for each of the features extracted from our audio data.

Each audio piece includes Nframes frames and for each frame we compute features. Each feature is N-dimensional (N may change for different features). Then each feature vector for each frame can be described as:

$$\mathbf{x}_j = [x_{j,1}, x_{j,2}, \dots, x_{j,N}], j = 1, 2, \dots, N_{frames}$$

### 2.2.1 Dynamic K-means to initialize GMM

When building GMM, we need to find a proper cluster number to balance the accuracy and computational efficiency. Here we apply the dynamic K-Means algorithm to determine the cluster number automatically.

The idea of dynamic K-Means is finding a cluster number which can minimize the following error function:

$$E_K = \sum_{i=1}^K \sum_{x \in L(k)} |x - \mu_k|^2 \quad (2)$$

Where K is clustering number, L(k) is the k-th cluster, x is feature vector,  $\mu_k$  is the mean of cluster L(k). Through dynamic K-Means algorithm, we can determine the cluster number automatically, and calculate the mean, variance and weights of each cluster. Then we initialize the Gaussian mixture model with these parameters. In our experiment, the K for feature LSP, PF is 30 and for feature Tempo, FP is 6, 10 respectively.

### 2.2.2 Estimate parameters of GMM

Assume each feature of each frame data obeys the Gaussian mixture distribution:

$$\mathbf{x}_i \sim \sum_{k=1}^K \pi_k N(\mathbf{x}_i; \mu_k, \Sigma_k) \quad (3)$$

We use the Expectation Maximization (EM) method to find the Maximum Likelihood Estimation (MLE) in order to estimate the model parameters  $\pi_k$ ,  $\mu_k$  and  $\Sigma_k$ . More specifically we find the parameters which make the expectation of the following logarithmic likelihood function to be maximal:

$$\sum_{i=1}^{N_{frames}} \log \left\{ \sum_{k=1}^K \pi_k N(\mathbf{x}_i; \mu_k, \Sigma_k) \right\} \quad (4)$$

The EM algorithm is an iterative algorithm that contains two main steps. First, in the E-step, it tries to “guess” the probability that sample xi is produced by each Gaussian component; second, in the M-step, it updates the parameters of the model based on the guesses.

For each feature we can compute a GMM by using above method. All the GMMs can form a database. The process of building such a database is illustrated in Fig. 1. In the following sections we show how to use this database for the analysis and recognition of *XinTianYou*.



**Table 2.** GMM parameters of LSP (GMM\_LSP).

cluster	$\mu$	$\Sigma$	$\pi$	cluster	$\mu$	$\Sigma$	$\pi$
1	(0.0478,0.0599)	(0.0008,0.0018)	0.08	16	(0.0819,0.0886)	(0.0011,0.0010)	0.02
2	(0.0725,0.0811)	(0.0015,0.0010)	0.01	17	(0.0879,0.0934)	(0.0019,0.0016)	0.04
3	(0.0953,0.1057)	(0.0001,0.0001)	0.01	18	(0.0791,0.0860)	(0.0007,0.0009)	0.03
4	(0.0676,0.0779)	(0.0044,0.0045)	0.06	19	(0.0662,0.0753)	(0.0010,0.0012)	0.07
5	(0.0839,0.0898)	(0.0001,0.0007)	0.02	20	(0.0514,0.0629)	(0.0016,0.0016)	0.01
6	(0.0936,0.0978)	(0.0012,0.0010)	0.03	21	(0.0582,0.1019)	(0.0145,0.0246)	0.02
7	(0.0561,0.0664)	(0.0010,0.0011)	0.06	22	(0.0342,0.0460)	(0.0017,0.0022)	0.03
8	(0.1011,0.1045)	(0.0033,0.0026)	0.04	23	(0.0595,0.0704)	(0.0011,0.0014)	0.01
9	(0.0749,0.0813)	(0.0001,0.0003)	0.01	24	(0.0291,0.0460)	(0.0031,0.0045)	0.05
10	(0.0877,0.1028)	(0.0115,0.0275)	0.01	25	(0.0243,0.0463)	(0.0019,0.0034)	0.05
11	(0.1115,0.1090)	(0.0100,0.0238)	0.01	26	(0.1382,0.1395)	(0.0102,0.0099)	0.01
12	(0.0426,0.0525)	(0.0003,0.0011)	0.03	27	(0.0631,0.0738)	(0.0009,0.0016)	0.05
13	(0.0443,0.0568)	(0.0014,0.0019)	0.02	28	(0.0707,0.0785)	(0.0003,0.0006)	0.07
14	(0.0380,0.0509)	(0.0014,0.0023)	0.07	29	(0.1439,0.3094)	(0.0122,0.0223)	0.01
15	(0.0760,0.0837)	(0.0011,0.0010)	0.03	30	(0.1142,0.1161)	(0.0054,0.0050)	0.02

Fig. 3 shows that the pitch of *XinTianYou* spans a large range and mainly gathers in high tone area. Here we analyze the histogram in details. First, we can see that the overall pitch range is between d-e and c3-d3. Generally speaking, the high pitch of soprano is about c2-c3, and tenor is c1-c2. The clusters fall into pitch range d2-e2 reach nearly 22%, which shows the pitch is gathered in a higher value area. Also we can see that the proportion of altos is the majority part of the histogram. This is because of the limitation of human voice and the need of altos for the preparation of climax. Moreover, the center of cluster ranges from one-lined octave a-b to three-line octave c3-d3, which is across three octaves. This shows that the pitch spans a relatively large range. In addition, the histogram shows some very small values, such as those for slots b2-c3, b-c1 and e2-f2. These values look like 0 because of the visualization of the histogram. The reason for these small values is that the Chinese Five-tone modes do not include semitone relations except that the six and seven tones do have semitone, but the six and seven tones rarely happen in Chinese folk songs. Lastly, we can see that there is even a recognizable value in pitch range c3-d3, which seems too high for human to sing. In fact such high notes are rarely happen in *XinTianYou*. The reason is that LSP may confuse overtones with normal notes and causes this unusual high pitch.

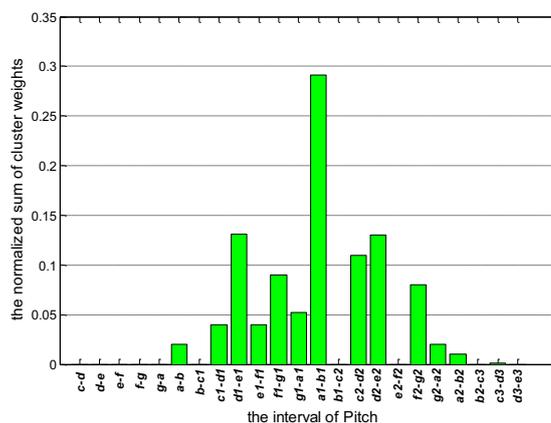
**Table 3.** Pitch and corresponding PF.

Pitch	PF	Pitch	PF
c	8452.703682	a <sup>1</sup>	10540.322318
d	8637.746753	b <sup>1</sup>	10735.355463
e	8841.169540	c <sup>2</sup>	10841.908323
f	8941.068626	d <sup>2</sup>	11043.199109
g	9137.720976	e <sup>2</sup>	11249.289466
a	9338.307723	f <sup>2</sup>	11340.560287
b	9539.031137	g <sup>2</sup>	11544.899454
c <sup>1</sup>	9640.591765	a <sup>2</sup>	11742.506746
d <sup>1</sup>	9839.225916	b <sup>2</sup>	11939.553027
e <sup>1</sup>	10036.915030	c <sup>3</sup>	12042.834835
f <sup>1</sup>	10141.457661	d <sup>3</sup>	12227.168391
g <sup>1</sup>	10343.886322	e <sup>3</sup>	12428.597918

To analyze the pitch, we also compute the feature PF. The correspondence between PF and pitch is shown in Table 3.

It can be seen from Table 3 that there are a simple linear relationship between pitch and PF. The difference of PF value is 100 corresponding a semitone, such as e->f, b->c, and 200 corresponding a whole tone such as c->d, d->e, f->g. Based on PF, the GMM clusters results are showed as Table 4.

We also compute a histogram for the PF cluster by using the cluster mean. The weights for clusters are accumulated into 23 different pitch ranges. The result is showed as Fig. 4.



**Figure 4.** The histogram of pitch feature

Based on the distribution of PF, it can be further concluded that the pitch of *XinTianYou* spans in a large range and mainly gathers in high pitch area. The pitch range above d2-e2 takes nearly 23%. The proportion of altos takes a majority part of the data. The pitch range is from a-b to c3-d3, which covers three octaves. This demonstrates again that *XinTianYou* spans a relatively large range. The proportion of the pitch above c<sup>3</sup> is different from the result of LSP. The histogram value for

pitch  $c^3-d^3$  is near to 0. This is because PF feature only consider frequency up to 2400Hz and filters the overtone.

In addition, the histogram values for  $b-c^1$ ,  $b^1-c^2$ ,  $e^2-f^2$ ,  $b^2-c^3$  are the similar to the results by LSP.

**Table 4.** GMM parameters of PF

cluster	$\mu$	$\Sigma$	$\pi$	cluster	$\mu$	$\Sigma$	$\pi$	cluster	$\mu$	$\Sigma$	$\pi$
1	10241.17	5.8891	0.03	11	11044.82	6.2639	0.07	21	9984.156	2.7769	0.001
2	10439.45	2.2246	0.001	12	12066.48	42.7593	0.001	22	9743.09	4.0011	0.01
3	10038.77	1.4168	0.01	13	11421.28	114.4496	0.08	23	10533.91	13.9345	0.001
4	10140.27	1.6062	0.01	14	9839.809	0.9368	0.04	24	10578.82	6.6313	0.001
5	9494.753	80.377	0.02	15	10140.85	0.7101	0.04	25	10730.61	3.2575	0.06
6	10937.21	4.61	0.03	16	10645.72	1.5175	0.05	26	9641.595	1.1343	0.03
7	11218.53	61.4448	0.06	17	10596.56	751.5881	0.10	27	10540.76	0.8913	0.06
8	9936.723	1.6054	0.04	18	9839.517	2.0984	0.01	28	11741.93	0.7763	0.02
9	10332.38	9.3718	0.01	19	10440.42	0.6573	0.03	29	10342.73	1.2871	0.05
10	10036.36	1.1352	0.04	20	11937.94	7.8618	0.01	30	10841.98	0.8463	0.08

### 3.2 General characteristics on rhythm

Rhythm denotes the organization of the length and strength of sound. It includes time-related elements such as tempo and beat. Here we use Tempo and FP to represent the rhythm information.

Tempo is computed by first detecting the peak of the signal curve, then counting the number of beats per time unit. It reflects the rhythm information to some extent. Table 5 shows the tempo values and corresponding speed.

**Table 5.** Speed and corresponding Tempo

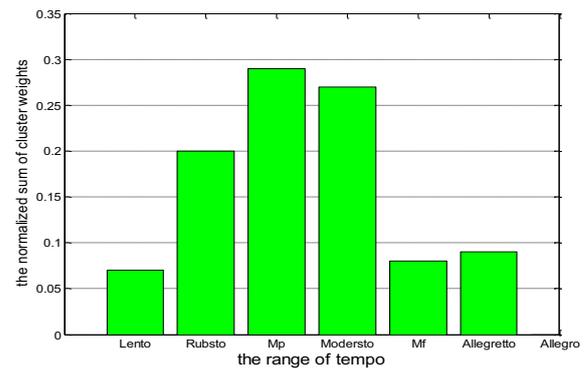
Speed	Tempo	Speed	Tempo
Lento	50.6668	Mf	90.6661
Rubato	60.0928	Allegretto	99.3841
Mp	69.8375	Allegro	109.9577
Moderato	80.7498	Molto mosso	120.1858

We compute the GMM for tempo feature. The parameters of clusters are shown in Table 6 below.

**Table 6.** The parameters of GMM\_Tempo clusters

cluster	$\mu$	$\Sigma$	$\pi$
1	60.3513	1.929	0.2
2	100.2964	3.3397	0.09
3	70.109	1.2834	0.29
4	50.6639	8.1217	0.07
5	80.0148	1.8544	0.27
6	90.002	2.1225	0.08

We build a histogram by using the centroid of each cluster. The vertical axis values of the histogram are the proportion of the clusters falls into each tempo range. The histogram is showed in the Fig. 5.



**Figure 5.** The histogram of rhythm feature

The general characteristics of *XinTianYou* rhythm is that the distribution of tempo spans from Lento to Allegro. The central part of the whole range is dominant while the two endings keep a similar ratio. The values from Mp to Mf are dominated, which is consistent with the frequency of walking. In addition, the figure shows that the value for Allegro is zero. This is because the Allegro is clustered into Allegretto because there are only few instances in *XinTianYou*.

Fluctuation Pattern (FP) is also used to analyze the distribution of speed. The FP values and their corresponding rhythm types are shown in Table 7.

**Table 7.** Speed and corresponding FP

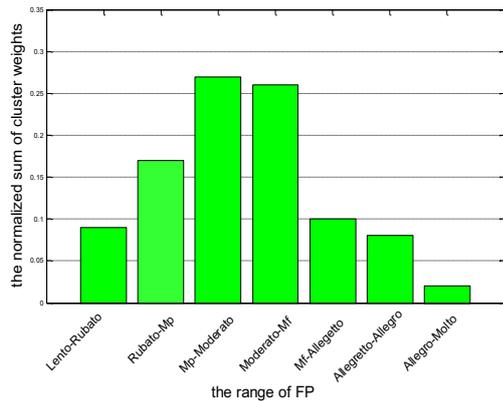
Speed	FP	Speed	FP
Lento	1.704	Mf	3.124
Rubato	2.062	Allegretto	3.395
Mp	2.426	Allegro	3.645
Moderato	2.832	Molto mosso	3.931

The GMM parameters of Tempo clusters are shown in Table 8.

**Table 8.** The parameters of GMM\_FP clusters

$Model_i$	$\mu$	$\Sigma$	$\pi$
1	2.7270	0.1901	0.12
2	3.0051	0.2483	0.17
3	3.4658	0.5831	0.08
4	2.7594	0.4926	0.07
5	2.4353	0.2724	0.08
6	2.9121	0.2015	0.09
7	3.25	0.0902	0.10
8	1.712	0.1655	0.09
9	2.2024	0.1643	0.17
10	3.830	0.8542	0.02

Again we compute the histogram as shown in Fig. 6, from which we can see a similar distribution as Tempo. Compared to Fig. 5, the Allegro to Molto mosso has a larger value. This denotes that there is few Molto mosso in *XinTianYou*.



**Figure 6.** The proportion of *XinTianYou* speed

### 3.3 Analysis on the general characteristics of pitch and rhythm's combination

As introduced in Section 3.1 and 3.2, the general characteristics of pitch of *XinTianYou* is the overall high value and the large span, and the general characteristics of rhythm is that moderato is the dominant part and lento-moderato has a similar ratio as moderato-allegro. In this section, we aim to analyze the combination of these two factors. We propose to use a method named CFRCA (Complete Features Relative Contribution Algorithm) for this purpose. This method is inspired by FRCA (Features Relative Contribution Algorithm) [17] which can be used to compute all features' relative contributions. In the rest of this section we will provide more details of both FRCA and CFRCA.

First of all, we normalize the extracted features to map the feature values to the same range. More specifically we use a Standardized deviation based method which is commonly used. The calculation formula is as follows:

$$x'_i = \frac{x_i - \bar{x}}{s} \quad (6)$$

$$\text{where } s = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}.$$

#### FRCA(Features Relative Contribution Algorithm)

For the features sequence  $(x_1, x_2, \dots, x_N)$ , the relative contribution of feature  $x_i$  can be calculated as follows:

$$C_1(i) = \sum_{j>i} J(i, j) - J(i+1, j) + \sum_{j<i} J(i, j) - J(i, j-1) \quad (7)$$

Where  $J(i, j)$  represents the cosine similarity between the  $i$ -th and  $j$ -th features, which can be computed by  $J(i, j) = \prod_{1 \leq g < h \leq M} \cos_{i,j}(g, h)$ ;  $M$  represents the number of

songs.

#### CFRCA (Complete Features Relative Contribution Algorithm)

The result of FRCA, the relative contribution  $C_1(i)$ , just represents the relative importance of feature  $x_i$  in a specific feature arrangement. The relative contribution  $C_2(i)$  of feature  $x_i$  in other sequences can be different from  $C_1(i)$ . To make the contribution measurement independent to the feature arrangement, we put propose to use CFRCA, which considers all permutations of  $N$  features:  $P = \{p_1, p_2, \dots, p_{A_N^N}\}$ , where  $p_i$  represents the  $i$ -th permutation. Then the average of the contribution values computed by all permutations is used for the final contribution:

$$C(i) = \frac{1}{A_N^N} \sum_{k=1}^{A_N^N} C_k(i) \quad (8)$$

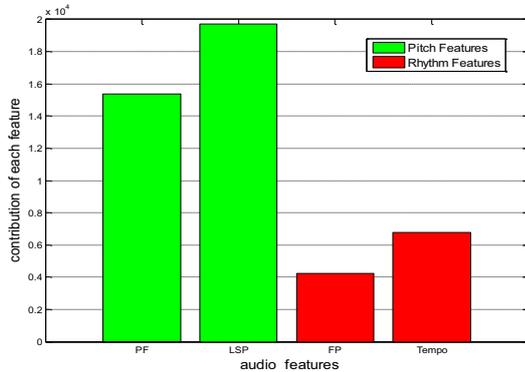
where  $C_k(i)$  represents the relative contribution of  $i$ -th feature in  $k$ -th permutation.

By doing this we can get the contribution of each feature which fully consider the permutations between all features. Fig. 7 shows the pseudo code of CFRCA.

<p><b>Algorithm</b> : Complete Features Relative Contribution Algorithm  <b>Input</b> : Feature set : <math>X = \{x_1, x_2, \dots, x_N\}</math>  <b>Output</b> : The complete relative contribution of feature:  <math>C = \{C(1), C(2), \dots, C(N)\}</math></p>
<pre> 1: for (i = 1; i &lt;= N; i++) 2:   Sum = 0; 3:   for (j = 1; j &lt;= A_N^N; j++) 4:     C_j(i) = relative_contri(i, j);            //compute the ith feature's relative contribution            //in jth permutation 5:     Sum = Sum + C_j(i); 6:   end 7:   C(i) = Sum / A_N^N; 8: end </pre>
<p><b>Procedure relative_contri(i, j)</b></p> <pre> 1: P = {p_1, p_2, ..., p_{A_N^N}}    // Set of full permutation by feature sequence 2: l = find_location(p_j, i);    // find ith feature's location in jth permutation 3: Contri = 0; 4: for (m = 1; m &lt;= N; m++) 5:   if(m &gt; l) 6:     Contri = Contri + [cos(p_j(l, m)) - cos(p_j(l+1, m))]; 7:   else if (m &lt; l) 8:     Contri = Contri + [cos(p_j(m, l)) - cos(p_j(m, l-1))]; 9: end 10: return Contri; </pre>

**Figure 7.** Complete Features Relative Contribution Algorithm

The experimental result of CFRCA is shown in Fig. 8. We can see that the contribution of pitch is relatively higher than rhythm. The contribution of LSP is the highest in all four features, which shows that LSP is the most representative feature of pitch. Although lower than LSP, the contribution of PF is still outstanding. The lowest contribution is from FP. The reason is that FP represents more information of rhythm compare to Tempo and the contribution of rhythm is low [18].



**Fig. 8.** The contribution of each feature by using CFRCA

In conclusion, from this section we find that *XinTianYou* has the following general characters: first, the pitch of spans a large range and mainly gathers in the high tone area. Second, the distribution of tempo spans from Lento to Allegro. The central part of the whole range is dominant while the two endings keep a similar ratio. Third, the pitch feature is more representative than the rhythm feature.

#### 4 Recognition of *XinTianYou* based on GMM

In this section we use the GMM computed in Section 3.1 and Section 3.2 to recognize *XinTianYou*.

We randomly select 80 from the overall 109 songs for the training set and the remaining 29 songs and another 30 non-*XinTianYou* folk songs as test set. To measure the similarity between the test data and the GMM models, we use the following likelihood equation:

$$P(n) = \frac{1}{N_{frames}(n)} \sum_{i=1}^{N_{frames}(n)} \sum_{k=1}^K \tau_k \cos(x_i, \mu_k) \quad (n=1, 2, 3, \dots, 80) \quad (9)$$

Given a test song, we compare each frame of it to each GMM center and compute the cosine similarity. Then we compute the mean of all these similarities and use it for the final measurement.

To make our results take the form of probability, we normalize the value of the likelihood value for all 80 training data. More specifically each similarity value is divided by the maximum value. As a result the probability interval of *XinTianYou* is  $[\min(P(x))/\max(P(y)), 1]$ ,  $x=1,2,3,\dots,80$ ;  $y=1,2,3,\dots,80$ . The similarity distribution can be got by sorting the  $P(n)$  values of the training set.

This make the  $P(n)$  value for test data is higher. We then compute  $P(n)$  for test data. We consider the test song is more likely to be *XinTianYou* if  $P(n)$  is higher.

In the experiment, we use the cross-validation that randomly selected five combinations of training set (80 pieces) and test set (29 pieces of *XinTianYou* and 30 pieces of non-*XinTianYou*). Table 9 shows the similarity range and probability interval of training sets. Table 10 shows the true positives (the number of 29 pieces of *XinTianYou* correctly classified as *XinTianYou*), and Table 11 shows the false positives (the number of 30 pieces of non-*XinTianYou* incorrectly classified as *XinTianYou*).

**Table 9.** The similarity range and probability interval of training sets

Experiment No.	Similarity range	Probability interval
i	(0.999381,0.999774)	(99.9602%,1)
ii	(0.999377,0.999791)	(99.9585%,1)
iii	(0.999574,0.999777)	(99.9796%,1)
iv	(0.999483,0.999897)	(99.9586%,1)
v	(0.999466,0.999768)	(99.9698%,1)

Probability Interval is small here because there exists general characteristics in pitch and rhythm of *XinTianYou*, which make different folk songs of test sets substituted into the model can also get high probability.

**Table 10.** The recognition of 29 pieces of *XinTianYou* in test sets

Experiment No.	Test sets Probability interval	True positives	Accuracy
i	(99.9725%, 100.0004%)	27	93.1034%
ii	(99.9766%, 100.0005%)	28	96.5517%
iii	(99.9601%, 99.99994%)	26	89.6551%
iv	(99.9703%, 100.0002%)	27	93.1034%
v	(99.9583%, 100.0000%)	26	89.6551%
<b>Average</b>		26.8	92.4138%

**Table 11.** The recognition of 30 pieces of non-*XinTianYou* in test sets

Experiment No.	Test sets probability interval	False positives	False identification rate
i	(79.9855%, 99.9607%)	2	6.67%
ii	(81.0654%, 88.3482%)	0	0%
iii	(81.9811%, 99.9810%)	3	10.00%
iv	(83.9703%, 99.9601%)	4	13.33%
v	(78.8953%, 99.9712%)	2	6.67%
<b>Average</b>		11	7.33%

Probability Interval is small in Table 9. This again proves that general characteristics in pitch and rhythm of *XinTianYou*. From Table 10 we can see that there are 26.8 true positives on average and the average accuracy is 92.4138%, and Table 11 shows the False identification rate is only 7.33% on average. The high accuracy and the low False identification rate shows that our method can recognize *XinTianYou* quite effectively. Besides, we can

see some the testing interval boundaries are greater than 100% in the Table 10. This is because we use the training sets' probability interval as references, and the resulting probability interval of test sets can be out of this range.

## 5 Conclusions

In this paper we use pitch and rhythm features to analyze *XinTianYou*, a traditional Chinese folk song style. The Gaussian Mixture Model is used for the clustering analysis and finding the general characters of *XinTianYou* data. We analyze the results with consideration of music theory, and also propose a method to recognize *XinTianYou*.

The experimental results show that the pitch of *XinTianYou* spans a large range, which is over 3 octaves. High pitch values take a proportion about 22%. The experiment of rhythm of *XinTianYou* shows that moderato is the dominant part while lento-moderato keeps a similar ratio as moderato-allegro. The pitch features of *XinTianYou* are more representative than rhythm features. Moreover, by using our method the average accuracy of recognition *XinTianYou* reaches 92.4%.

"General characteristics" provides a new way of looking into the commonality or gene of Chinese folk songs. The method we use in this paper can also be applied to other styles of music other than *XinTianYou*. In the future we plan to explore other types of Chinese folk song styles, and look for the general characteristics between them.

## References

1. Integration of Chinese folk songs • Shaanxi Juan. The Editorial Committee, Integration of Chinese folk songs: Shaanxi Juan [M]. Chinese ISBN center, (1994).
2. Chen Haoyue. Characteristics and style of Shaanxi folk songs [J]. Stage (journal of Peking university), (2014), 3:238-239.
3. Zhang Zhentao. Into the modern northern shaanxi folk songs [J]. Chinese musicology, 2012, (4):75-86.
4. Liu Dajian. Three elements of constituting the style of XinTianYou [J]. Art research, (2008), (1):42-44.
5. Fu Z, Lu G, Ting K M, et al. A Survey of Audio-Based Music Classification and Annotation [J]. Multimedia IEEE Transactions on, (2011), 13(2):303 - 319.
6. Li T, Ogihara M, Li Q. A comparative study on content-based music genre classification[C]//Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2003: 282-289.
7. Sturm B L. Classification accuracy is not enough [J]. Journal of Intelligent Information Systems, (2013), 41(3): 371-406.
8. Jouviet, D., Vincent, E., Dung Tran. Investigating stranded GMM for improving automatic speech recognition [J]. //Hands-free Speech Communication and Microphone Arrays (HSCMA), 2014 4th Joint Workshop on IEEE, (2014): 192-196.
9. Dileep A, D D A, Dileep A D, et al. Class-specific GMM based intermediate matching kernel for classification of varying length patterns of long duration speech using support vector machines[J]. Speech Communication, (2014):126-143.
10. Qiao Jianzhong. It should be emphasized that the literature value of "integration" "Integration of Chinese folk songs" [J]. People's music, (1986), (2):38-41.
11. MIN Gang; ZHANG Xiongwei; YANG Jibin; CHEN Yanpu. Sparse representation and performance analysis for LSP parameters via dictionary learning [J]. Journal of PLA University of Science and Technology (Natural Science Edition), (2014):121-126.
12. Tolonen, Karjalainen. A Computationally Efficient Multipitch Analysis Model, IEEE Transactions on Speech and Audio Processing, 2000, 8(6):708-716.
13. Lartillot O. Mirtempo: tempo estimation through advanced frame-by-frame peaks tracking [J]. Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX), Utrecht, Netherlands, (2010):1-2.
14. Bosteels K, Kerre E E. Fuzzy audio similarity measures based on spectrum histograms and fluctuation patterns [M]//Computational Intelligence in Multimedia Processing: Recent Advances. Springer Berlin Heidelberg, (2008): 213-231.
15. Tzanetakis G, Cook P. Marsyas: A framework for audio analysis [J]. Organised sound, (2000), 4(03): 169-175.
16. Olivier Lartillot, Petri Toivainen. A MATLAB TOOLBOX FOR MUSICAL FEATURE EXTRACTION FROM AUDIO [M]. Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10-15, (2007):237-244
17. Zhen B, Wu X, Liu Z, et al. On the Importance of Components of the MFCC in Speech and Speaker Recognition [J]. ACTA SCIENTIARUM NATURALIUM-UNIVERSITATIS PEKINENSIS, (2001), 37(3): 371-378.
18. GOTO M. A chorus-section detecting method for musical audio signals [A]. Hongkong: IEEE Press, (2003).437-440.