EURASIP Journal on Audio, Speech, and Music Processing

EMPIRICAL RESEARCH

Open Access

Acoustical feature analysis and optimization for aesthetic recognition of Chinese traditional music

Lingyun Xie[1], Yuehong Wang[1] and Yan Gao[2*]

Abstract

Chinese traditional music, a vital expression of Chinese cultural heritage, possesses both a profound emotional resonance and artistic allure. This study sets forth to refine and analyze the acoustical features essential for the aesthetic recognition of Chinese traditional music, utilizing a dataset spanning five aesthetic genres. Through recursive feature elimination, we distilled an initial set of 447 low-level physical features to a more manageable 44, establishing their feature-importance coefficients. This reduction allowed us to estimate the quantified influence of higher-level musical components on aesthetic recognition, following the establishment of a correlation between these components and their physical counterparts. We conducted a comprehensive examination of the impact of various musical elements on aesthetic genres. Our findings indicate that the selected 44-dimensional feature set could enhance aesthetic recognition. Among the high-level musical factors, timbre emerges as the most influential, followed by rhythm, pitch, and tonality. Timbre proved pivotal in distinguishing between the JiYang and BeiShang genres, while rhythm and tonality were key in differentiating LingDong from JiYang, as well as LingDong from BeiShang.

Keywords Chinese traditional music, Aesthetic, Machine learning, Acoustic features, Feature selection

1 Introduction

Why are listeners so captivated by music? A primary allure is its ability to convey emotions and provide aesthetic enjoyment. Aesthetics evoke a distinct emotional experience, predominantly associated with affective responses [1]. Music theory posits that emotional patterns are crafted by musical elements such as tonality, rhythm, and timbre. The nexus between musical components and emotions was first scrutinized by music psychology, with Hevner pioneering this inquiry in 1936. He devised a musical emotion model and assessed the separate influences of rhythm, harmony, and melody [2]. Since then, Gabrielsson and Lindström have compiled related studies looking at the factors that affect musical expression of emotion, identifying pattern, rhythm, dynamics, articulation, timbre, and phase as pivotal and most frequently examined [3]. For instance, attributes like rhythm, mode, and pitch have been shown to distinguish emotions such as happiness and sadness effectively. Baltes FR delved into music's psychological effects on behavior and motivation [4], while Jonna K. Vuoskoski conducted empirical research on music's influence on emotions and personality [5]. Tuomas Eerola explored how six musical elements-mode, rhythm, dynamic range, articulation, timbre, and pitch range-impact emotional perception [6]. The results revealed that mode was the most important musical factor, followed by rhythm, pitch range, dynamics, articulation, and timbre, and that these factors acted additively rather than interactively.

*Correspondence:
Yan Gao
gyan@ccom.edu.cn
[1] School of Information and Communication Engineering, Communication University of China, Beijing, China
[2] Department of AI Music and Music Information Technology, Central Conservatory of Music, Beijing, China

With recent advancements in artificial intelligence and enhanced understanding of sound quality in human cognition, music emotion research increasingly focuses on the correlation between automatic music emotion recognition and objective acoustical features. George Tzanetakis conducted a comparative study in 2002 on the performance of three feature sets representing timbre texture, melody content, and pitch content for music genre classification [7]. Yang Liu et al.'s deeper exploration into music-emotion connections highlighted BPM (beats per minute), spectral flux's mean and standard deviation, and the first component of MFCC (Mel Frequency Cepstrum Coefficient) as crucial for sentiment classification. The mean and standard deviation of spectral flux, on the other hand, reflect the rate of pitch change and degree of consistency of the song [8]. Xinyu Yang et al. summarized data-driven music emotion recognition methods, mentioning that acoustic features commonly used in music emotion recognition include the first 10 dimensions of the MFCC coefficient, the 14 dimensions of spectrum contrast based on octave, and 4 dimensions of spectral statistical descriptors (Spectral Centroid, Flux, Rolloff, and Flatness), and a 12-dimensions Chromogram [9]. Jun Su et al. extracted 15 basic audio features across timbre, rhythm, pitch for modeling the intrinsic relationship between emotion and music using 8 machine learning methods [10].

Musical aesthetics, a complex emotion, has received less attention than musical emotion detection in previous studies. Aesthetic experience, as a unique form of aesthetic emotion, relies on higher cognitive functions beyond perception. A five-dimensional processing model that Leder et al. created to explain aesthetic enjoyment and evaluation explains the varied aesthetic experiences of modern art [11]. Subsequently, Juslin et al. sought to model music aesthetic judgment using subjective criteria, concluding that listeners have distinct ways of interpreting and experiencing music, and that these distinct ways with aesthetic value can be effectively modeled using concrete methods [12]. Schindler et al., drawing on aesthetic emotion theory, reviewed existing measurement methods across various domains like music, literature, film, painting, advertising, design, and architecture, proposing a new framework for studying aesthetic emotion [13].

The aforementioned studies predominantly focus on Western music. However, the emotional and aesthetic perception of music will differ depending on one's cultural background. Traditional Chinese culture has its own distinct aesthetic characteristics. Similarly, Chinese traditional musical instruments and national music exhibit distinct emotional and aesthetic cognition through their distinctive timbre. Some scholars have conducted related research in this regard. Through experiments, Liu Tao created an emotional loop that corresponded to the emotional cognitive habits of Chinese people towards music [14]. Gao and Xie proposed a classification and quantification method for the aesthetic attributes of Chinese traditional music and painting after conducting experiments to confirm the limitations of the current aesthetic categories on the aesthetic classification of Chinese traditional arts [15], and established a database for aesthetic classification of Chinese traditional music [16]. Jiang Shengyi et al. built a tree-shaped hierarchical structure Chinese sentiment dictionary in the music field based on the improved Hevner model, and they realized lyrics emotion classification using a sentiment vector space model and a sentiment dictionary [17]. Wu Wen et al. investigated the similarities and differences of classical music in emotion classification and discovered that emotion classification feature sets applicable to western classical music could not achieve the same good effect on data from Chinese traditional music [18, 19]. Yi-hsuan Yang et al. conducted a cross-cultural comparison of the emotional expression of Chinese and Western pop songs [20], and established a Chinese pop music emotional database containing 818 songs [21].

Current research on the aesthetics of Chinese traditional music predominantly resides within the realm of humanities-based artistic aesthetics, with minimal exploration in science and engineering fields. Notably, studies focusing on the automatic recognition of beauty in Chinese traditional music are especially scarce. Ma Xinyu et al. conducted preliminary automatic classification based on five categories from the Chinese traditional music database in 2018 [16, 22]. The aesthetics of national music play a crucial role in Chinese traditional aesthetics, and the intelligent objective analysis is beneficial in providing valuable insights and inspiration for the scientific study of the aesthetics of Chinese traditional music. Different musical aspects, such as timbre, mode, and rhythm, should contribute uniquely to aesthetic perception, especially when seen from a more in-depth perspective. The cognitive patterns of music aesthetics can be enhanced if the association between musical elements and aesthetic classification can be objectively and statistically stated.

In this study, the feature optimization of the automatic classification and the correlation analysis between music aesthetics and musical aspects are conducted using a database of Chinese traditional music with five categories of aesthetic assessments generated in [16]. Figure 1 illustrates the overall workflow. The paper begins with an introduction to the Chinese music aesthetic database and the music feature set, which consists of 447-dimensional low-level physical features. The 44-dimensional features

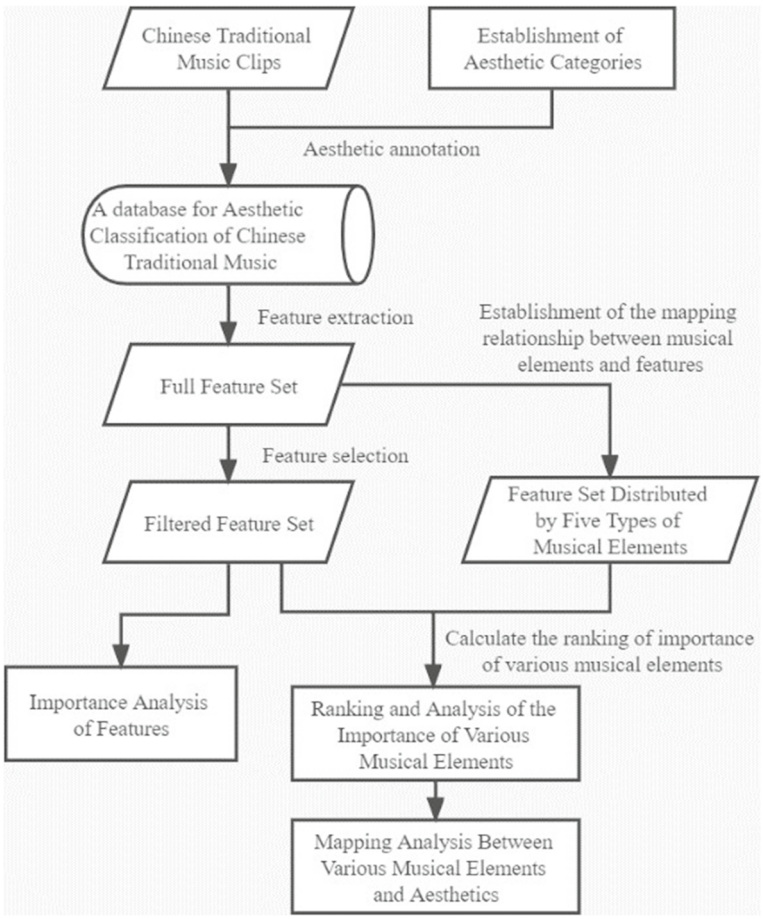

Fig. 1 The overall workflow of this paper

suitable for aesthetic classification of Chinese traditional music are then generated using two feature selection methods: filtering and wrapping. Based on the feature importance coefficient, the association between aesthetic categorization and musical elements is further investigated. This paper's database and research findings can aid in the study and implementation of intelligent information retrieval of Chinese traditional music and the integration of audio-visual scenarios based on aesthetics.

2 Chinese traditional music aesthetic database

The music data we used came from albums performed by well-known artists as well as instrumental music compilation CDs. Stringed, plucked, blowpipe, and percussion were the four types of musical instruments, and each music was primarily performed by one instrument (or solo). A total of 441 classic clips with 17 different instruments were collected. One or 2 segments of about 20–30 s were intercepted from each piece of music to reduce the burden on the labeling staff and to prevent emotional abrupt changes in the entire music [23]. Finally, a Chinese traditional music dataset with 500 clips was gathered, as shown in Table 1, and all clips were in wav format with a bit rate of 1411 kbps.

The division and definition of aesthetic categories should be determined first in the evaluation of aesthetics. We gathered 350 words suitable for evaluating aesthetics by reviewing aesthetics-related literature. Then, using a questionnaire survey and word frequency statistics, 40 words were chosen to evaluate the aesthetics of Chinese traditional music. On this basis, we developed five aesthetic categories through a combination of subjective evaluation and factor analysis. To annotate the dataset, a total of 20 people were recruited. The subjects determined which of the five aesthetic categories the pieces should fall into. Table 2 displays the categories of Chinese traditional music aesthetic attributes and data distribution [16].

Table 1 Musical instrument distribution of Chinese traditional music aesthetic database

	Instrument	Clips	Total
Wind instrument	Flute	47	113
	Vertical flute	22	
	Xun	12	
	Sheng	10	
	Gourd pipe	12	
	Suona	10	
String instrument	Erhu	90	144
	Jinghu	12	
	Banhu	8	
	Matouqin	34	
Plucked instrument	Guqin	26	190
	Zither	67	
	Konghou	12	
	Pipa	47	
Percussion instrument	Drum	47	53
	Chime bell	6	

3 Chinese traditional music feature set

By analyzing and processing the extracted features of music signals, high-level semantic information about music, such as emotion and aesthetic, can be obtained. As a result, the primary task of realizing music intelligent information processing is to obtain relevant features. For the time being, the extractable music signal features consist primarily of low-level physical features and high-level music element features. Low-level physical features are based on sound's time-frequency properties, and include time-domain features, frequency-domain features, and time-frequency domain features; high-level music element features are abstract semantic features that can describe the inherent elements of music [24], such as energy, pitch, timbre, rhythm, and mode. These high-level characteristics are typically represented by a number of low-level features. Marsyas [25], OpenSMILE [26], and MirToolbox [27] are popular tools for extracting features from music or audio signals. These three tools were used in this paper to extract low-level physical features and map them to the corresponding high-level musical elements.

3.1 Energy features

The objective amplitude of the sound and the auditory subjective psychological perception are related to sound energy. So, both objective and subjective viewpoints are used to extract the information linked to energy. Loudness reflects the perceived sound level from a subjective perspective, while RMS (root mean square), sound intensity, and LER (low energy rate) reflect the objective variations of signal amplitude and energy in time domain. MirToolbox was used to extract the LER, while OpenSMILE was used to extract the RMS, intensity, and loudness. In order to quantify the statistical characteristics, the mean, standard deviation, and variance of these four variables were calculated, yielding a total of 12-dimension features.

3.2 Pitch features

The fundamental frequency in music is represented by pitch, and the melody is mostly represented by how the fundamental frequency changes over time. The fundamental frequency (F0) and the smoothed fundamental frequency (F0env) were both retrieved by OpenSMILE and used in this study as characteristics reflecting pitch. There are a total of 6 dimensions characteristics because each feature needs to calculate its mean, standard deviation, and variance.

3.3 Timbre features

The frequency spectrum of sound, which is the proportion of fundamental frequency and overtones in musical signals, is the primary determinant of timbre. Timbre features have been frequently employed in music genre and emotion categorization [28], and timbre-based feature parameters have a significant impact on music aesthetic classification. As a reflection of timbre acoustic features, this paper extracted MFCC (Mel-frequency coefficients), SFM (spectral flatness measure), SCF (spectral crest factor), LPC (linear prediction coefficients), LSP (linear spectral pairs), spectral slope, spectral entropy,

Table 2 Categories of Chinese traditional music aesthetic attributes and data distribution

Category	Description	Number of samples
LingDong	Lively, flexible and full of change.	155
JiYang	Exciting, inspiring and strong passion.	117
QingRou	Easy to ease, calm and gentle.	122
ShenChen	Deep, steady and not exposed. Strong and powerful, not slim.	47
BeiShang	Sad, grieved, sorrowful, mournful, and distressed.	59

spectral harmonicity, MCR (mean-crossing rate), ZCR (zero-crossing rate), alphaRatio (the ratio of 1–5 kHz energy to energy less than 1 kHz), spectral centroid, spectral spread, spectral flux, spectral rolloff, and sharpness, a total of 16 features.

Marsyas was able to extract the MFCC, SFM, and SCF from this group. In particular, SFM and SCF had both 24 dimensions, but MFCC had 13 dimensions. Marsyas devised a segmented computation method in which each segment calculates the mean and standard deviation of the related features using 20 frames of data as input. A music data sample contained a mean time series and a standard deviation time series. The mean and standard deviation of these two series were then computed. Finally, a total of $(13+24+24)*4=244$ dimensional features were created by combining the four statistical characteristics of the mean of the mean, the variance of the mean, the mean of the variance, and the variance of the variance.

The remaining 13 features were extracted by OpenSMILE, of which LPC and LSP each had 10 dimensions, and the three statistical features of mean, standard deviation and variance were calculated, a total of $(10+10+11)*3=93$ dimensional features. Thus, the timbre class has 337 dimensional features in total.

3.4 Rhythm features

The speed and movement of a piece of music are described by rhythm, which is an important musical element. Rhythm is comprised of two concepts: tempo and speed. The former relates to the regular alternating movement of strong and weak beats in music, which is the beat combination; the later refers to the rate at which these beats occur. MirToolbox was used to extract the rhythm features in this work, which included fluctuation, beatspectrum, events, event density, tempo, and pulse-clarity. The mean, standard deviation, and variance of each feature had to be calculated, resulting in a total of 18 dimensional features in the rhythm.

3.5 Mode features

A mode is an arrangement of many musical notes with one note serving as the tonic, structured according to a specific interval relationship and varied pitches. It serves as the foundation for the melody and reflects the relationship between the various tones in a piece of music. Chroma (14 dimensions), a histogram of pitch level energy distribution, and KeyStrength, which indicates the significance of tone, were extracted in this study to reflect mode. MIRToolbox was used to extract KeyStrength characteristics, and the mean, standard deviation, and variance were calculated as three statistical features. Marsyas extracted chroma characteristics, and four statistical features were computed. Additionally, in accordance with [29], we used OpenSMILE to extract energy features from 5 sub-bands based on octaves. Table 3 displays the frequency band distribution. This feature, which is categorized as a mode feature, displays the energy distribution of musical signals over various scale ranges. As a result, there were a total of $3+14*4+5*3=74$ features in the mode features.

Overall, 447-dimensional features have been separated into 5 categories, all of which are given in Table 4.

4 Feature selection

To increase the recognition rate in a variety of music information retrieval applications, the feature information was always extracted as much as possible. As a result, the feature dimension expanded and gave rise to numerous unnecessary and redundant features, which negatively impacted the back-end classifier's performance. We preprocessed the data first, used two methods of filtering and wrapping to screen the 447-dimensional features in Table 4, and then compared and decided on the most appropriate feature selection method in order to extract the signal features that can best reflect the essence of music aesthetic classification.

4.1 Classifier selection

An appropriate classifier is a prerequisite for the aesthetic classification of traditional Chinese music. In this study, we selected five commonly used traditional machine learning classification models. We extracted 447 features from a database of 500 pieces of traditional Chinese music, as shown in Table 4, to perform a comparative analysis of the classification results. The five classification models are logistic regression (LR), K-nearest neighbor (KNN), linear support vector machine (SVM), random forest (RF), and extremely randomized trees (ERT). For the sake of simplicity and reproducibility in our initial experiments, we employed the default parameters provided by the scikit-learn library(version 1.2.2) for all the models. For example, the default cost of linear SVM is 1.0. Ultimately, the classification accuracy and F1-score metrics were used to compare the classification performance of different classifiers. The best classifier suitable

Table 3 Sub-band frequency division based on octave

Subband	Freq (Hz)	Octave scale
1	0–200	-F3
2	200–400	G3-F#4
3	400–800	G4-F#5
4	800–1600	G5-F#6
5	1600–3200	G6-F#7

Table 4 Feature set of the Chinese traditional music

Category	Feature name	Dimensions	Extraction tool
Energy	Low energy frame ratio	3	Mir toolbox
	RMS energy, intensity, loudness	6	OpenSMILE
Pitch	Fundamental frequency, fundamental frequency	6	OpenSMILE
	By smoothing		
Timbre	MFCC, spectral flatness measure, spectral crest factor	244	Marsyas
	Linear prediction coefficient, line-spectral pair, spectral slope, spectral entropy, zero-crossing rate, spectral centroid, spectral flux, spectral roll-off, sharpness, spectral harmony, spectral spread, mid-low frequency energy ratio	93	OpenSMILE
Rhythm	Fluctuation, beatspectrum, events, eventdensity, tempo, pulseclarity	18	MirToolbox
Mode	Key strength	3	Mirtoolbox
	Chroma	56	Marsyas
	Octave-based subband energy	15	OpenSMILE
Total		447	

for the aesthetic recognition of traditional Chinese music in this paper was identified based on these evaluations. The results, presented in Table 5, indicate that the extremely randomized trees outperformed other classifiers in both classification accuracy and F1-score. Hence, this classifier will be used for subsequent research.

4.2 Data preprocessing

Because the 447-dimensional initial feature set acquired above was extracted using several tools and each feature's physical meaning varied, there were significant dynamic range or dimension discrepancies between the feature values, which might significantly impair machine learning performance. The category discrimination of the feature dimension can be expanded and the classification accuracy increased by appropriately preprocessing these data. Upper and lower limit range compression, normalization, maximum normalization, etc. are frequently used data preparation techniques. The initial feature set obtained in the Section 3 was preprocessed using techniques from the Python-based Scikit-Learn toolkit [30], and extremely randomized trees were used as the classifier. The corresponding aesthetic recognition rate was calculated using a 5-fold cross-validation, and the results are displayed in Table 6.

Table 5 Evaluation results of different classifiers

Classifier	Accuracy	F1-score
LR	65.6%	0.624
KNN	62.5%	0.572
SVM	65.0%	0.613
RF	63.8%	0.594
ERT	67.6%	0.655

It can be seen that after data preprocessing, the aesthetic recognition rate improved, with quantile transformation being the most improved preprocessing method. The quantile transformation is a nonlinear transformation that maps data to a uniform or normal distribution within the range of 0–1 based on the quantile range obtained from statistics, in this case the normal distribution map. In combination with the recognition rate results in Table 6, quantile transformation (normal distribution mapping) was used as the data preprocessing method in this paper's subsequent works.

4.3 Feature selection

Wrapping, filtering, and embedding are three common feature selection methods [31]. The feature selection evaluation criterion for the wrapping method is the classifier's performance, the filtering method has nothing to do with the subsequent classifiers, and the embedding method combines feature selection and classifier training. Because it is difficult to obtain quantifiable feature importance using the embedding method, this paper used the wrapping and filtering method for feature selection.

Table 6 Aesthetic recognition rate with different preprocessing methods

Data preprocessing	Aesthetic recognition rate
Unprocessed raw data	67.6%
Normalization	69.2%
Upper and lower limit range transformation (0,1)	69.0%
Normalize by maximum value	69.1%
Quantile transformation (normal distribution mapping)	70.4%
Standardization	69.1%

4.3.1 Wrapping feature selection

Feature search is used in the wrapping approach. The fundamental concept is to continuously choose subsets from the initial feature set and assess each subset based on how well the same classifier and validation technique perform until the best subset is chosen. This study adopted the RFE (recursive feature elimination) method, which uses a classification model with a feature importance evaluation function to train the classifier's initial feature set and determine the importance coefficient of features using correlation or contribution values as attributes [32]. Every time, one or more of the least significant features were deleted. This process was then repeated recursively on the feature set until the most advantageous features could provide a better recognition rate.

The fundamental classifier was the extremely random tree. Figure 2 depicts the screening procedure, and the 447-dimensional original feature set yielded 44 features that may be used to categorize the aesthetics of Chinese traditional music. The number of characteristics is plotted along the horizontal axis, while the cross-validation classification accuracy is plotted along the vertical axis. The top position was attained at 44 features, and the greatest recognition rate was 72%. The remaining 403 elements may be viewed as redundant or unimportant features for aesthetic classification after that point, as the recognition rate then marginally declined and remained consistent. Table 7 displays the 44 features.

4.3.2 Filtering feature selection

The filtering method computes the attributes of the feature and the relationship between each feature, measures the feature's redundancy, and filters out the useless features. The method includes the evaluation of correlation attributes, information entropy, and distance. This paper compared three different measurement methods and chose the one with the highest recognition rate for feature induction and analysis. The three measurement criteria are described below.

Correlation evaluation: This method is mainly selected by comparing the correlation between a single feature and a category, using the Pearson correlation coefficient (Pearson r), and the formula is shown in Eq. (1). The higher the value, and the better the feature is for classification.

$$r_{XY} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \tag{1}$$

Gain ratio evaluation: To determine whether a feature has a good classification effect, this method computes the change in information entropy of the category before and after the feature is selected, and then computes the ratio of the change value to the information entropy of the feature itself. The greater the gain ratio, the more obvious the feature to improve classification effect. The information entropy H is given by formula (2), where m is the number of categories C in the sample set, and pi is the probability that the data belongs to the *i*th category.

$$H = -\sum_{i=1}^{m} p_i \log_2 p_i \tag{2}$$

The calculation of the information gain ratio is shown in formula (3).

$$GainR(Class, Attribute) = (H(Class) - H(Class|Attribute))/H(Attribute) \tag{3}$$

H (Class) denotes the class's information entropy, H (Attribute) denotes the feature attribute's internal information entropy, and H (Class | Attribute) denotes the class's conditional information entropy under the known features.

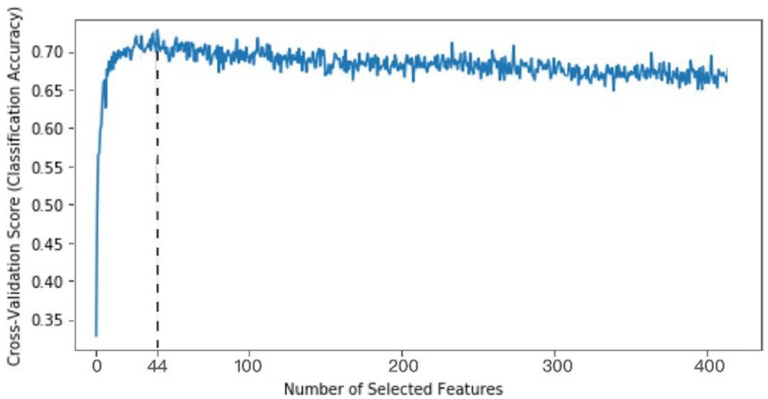

Fig. 2 Feature screening using RFE

Table 7 Filtered list of features

Feature category	Dimension	Specific dimensions
SFM	16	SFM4_std_mean, SFM5_std_mean, SFM6_std_mean, SFM6_mean_mean, SFM7_std_mean, SFM8_std_mean, SFM8_mean_mean, SFM9_std_mean, SFM9_mean_mean, SFM10_std_mean, SFM10_std_mean, SFM11_std_mean, SFM11_mean_mean, SFM12_std_mean, SFM13_std_mean, SFM14_std_mean
MFCC	8	MFCC0_std_mean, MFCC1_std_mean, MFCC1_std_std, MFCC1_mean_mean, MFCC9_std_mean, MFCC10_std_mean, MFCC11_std_mean, MFCC12_std_mean
SCF	10	SCF4_std_std, SCF4_std_mean, SCF5_std_mean, SCF5_std_std, SCF6_std_mean, SCF7_std_mean, SCF9_std_mean, SCF10_std_mean, SCF11_std_mean, SCF12_std_mean
Events	3	Events_mean, Events_std, Eventdensity_mean
F0	1	F0_mean
Fluctuation	2	Fluctuation , Fluctuation_std
Chrome	1	PeakRatio_Average_Chroma_A
LSP	2	lspFreq [3] _stddev , lspFreq [4] _stddev
PuleseClarity	1	Pulseclarity_mean

Relief evaluation: It primarily computes the weight of each feature using the ReliefF algorithm [33]. It randomly selects a sample R from the training sample set each time, then uses the Euclidean distance to calculate the k-nearest neighbor samples of R from the same sample set, then searches for k nearest neighbor samples from the different R sample sets, and finally calculates the weight value of each feature, as shown in formula (4):

$$\begin{aligned} W(A) &= W(A) - \sum_{j=1}^{k} diff(A,R,H_j)/(mk) \\ &+ \sum_{C \notin class(R)} \left[\frac{p(C)}{1-p(class(R))} \sum_{j=1}^{k} diff(A,R,M_j(C)) \right] /(mk) \end{aligned} \tag{4}$$

In the formula, m refers to the sampling times, k is the number of nearest neighbor samples, Mj(C) denotes the j-th nearest neighbor sample in the class C⊄class(R), and diff (A, R1, R2) denotes the difference between samples R1 and R2 on feature A. This value is significant because it subtracts the difference between the corresponding features of the same category and plus the difference between the corresponding features of different categories. The greater the value of W(A), the better the feature's classification ability. Finally, the appropriate features are sorted and filtered based on the weight value, yielding an evidence-based feature importance ranking.

The number of features selected by filtered feature selection was 44, the same as wrapped feature selection, to facilitate comparison and analysis. The basic classifier was random forest, and the results were evaluated using a five-fold cross-validation. Finally, Table 8 displays the aesthetic recognition results of Chinese traditional music using the three feature selection methods described above.

Table 8 shows that after removing redundant features, the aesthetic recognition rate of the three feature selection methods improved to some extent when compared to the full features, with ReliefF having the best effect.

4.3.3 Comparison of feature selection methods

The wrapping method based on RFE achieved a 72% aesthetic recognition rate, while the filtering method based on the ReliefF algorithm achieved a 70.2% recognition rate. As a result, in the following aesthetic classification and feature analysis, this paper used the wrapping method based on RFE for feature selection. ...

5 Aesthetic classification and characteristic analysis of Chinese traditional music

Wrapped feature selection was used to reduce the 447-dimensional initial features to 44-dimensional features. In this section, we quantified the importance of each physical feature tested by the classifier, and based on this, as well as the correspondence between musical elements and physical features in Table 4, we further analyzed the more abstract musical elements (such as timbre, mode, rhythm, and so on) to assess their importance to aesthetic classification.

5.1 Importance coefficient of physical features

Despite the fact that the wrapping feature selection method does not provide the value of the corresponding feature importance, some classifiers can calculate the

Table 8 Aesthetic recognition rates of different filtering feature selection methods

Feature selection	Aesthetic recognition rate
Correlation attribute evaluation	67.4%
Gain ratio attribute evaluation	68.2%
Relief attribute evaluation	70.2%
Full features (unscreened)	66.6%

importance coefficient of the feature used in the classification. We used the extremely random tree classifier with the filtered 44-dimensional features to classify Chinese traditional music aesthetics, and we used 5-fold cross-validation to determine the importance coefficient of each feature. The importance coefficients of each feature category were then calculated using the category relationship in Table 7 and the coefficient values of the same category of features, as shown in Table 9.

5.2 Analysis of the importance of musical elements

Although the significance of specific time-frequency physical features is clear, its true implications are not always clear. Musical components like timbre and rhythm are especially better for categorizing musical aesthetics since they can be interpreted more generally. The relationship between musical components and the aesthetic categorization of Chinese traditional music will be further explored in this section.

The quantification findings of the significance of musical elements to the aesthetic categorization can be derived by adding the important coefficients of the characteristics of comparable musical elements in Table 9 as shown in Fig. 3. The findings demonstrated that timbre, which accounts for 78.2% of the classification's aesthetics of Chinese traditional music, was followed by rhythmic elements, which account for 16.3%; pitch and mode aspects, which account for 3.2% and 2.3%, had relatively less influence.

The aforementioned analysis was conducted from the standpoint of the significance of the overall contrast between the five types of aesthetics; nevertheless, it does not adequately convey the significance of musical aspects to a particular aesthetic category. In order to determine the significance of each musical element to each aesthetic category, we used the following methodology: we first automatically classified aesthetics by removing the features of a particular class of musical elements, then we compared and observed the changes in the confusion matrix, analyzing the classification results to how which aesthetics were influenced by each musical element.

Figure 4 depicts the initial confusion matrix that was created following classification using the 44-dimensional filtered features, where letters "a" stand for the LingDong category, "b" for JiYang, "c" for QingRou, "d" for ShenChen, and "e" for BeiShang. The original category is represented by the first column on the left of the matrix, and the classified category is represented by the first row from the top. The figures in the following confusion matrices employ the same representation. In the instance of complete 44-dimensional features, QingRou was the type of aesthetics that was most likely to be mistaken for one of the other four types, whether it was mistaken for another category or another category was mistaken for it. Additionally, there was some misunderstanding regarding JiYang and LingDong. In contrast to QingRou, ShenChen and BeiShang were less perplexed by other categories.

When the timbre features were removed, it was clear from the confusion matrix in Fig. 5 and the classification accuracy rate in Fig. 6 that the aesthetic recognition rate had dropped from 72% to 44.2%, a drop of nearly half, indicating that the timbre features played an important role in the overall classification. More specifically, from the standpoint of various aesthetic categories, JiYang and BeiShang's recognition rates were the most obvious, showing a sharp decline, almost to 17.10% and 6.80%, respectively, and the degree of confusion between the two categories increased significantly, with the number of confusions increasing from 3 to 98. LingDong's recognition rate decreased but not significantly, while ShenChen's recognition rate decreased to some extent. The recognition rate of QingRou has decreased slightly, while the number of misidentifications as JiYang and ShenChen has increased significantly. It was clear that the absence of timbre features had less of an impact on the

Table 9 Importance coefficients of features in aesthetic classification

Feature	Importance coefficient	Musical elements	Feature	Importance coefficient	Musical elements
SFM	0.365	Timbre	Chroma	0.023	Mode
MFCC	0.173	Timbre	LSP	0.042	Timbre
SCF	0.202	Timbre	Fluctuation	0.047	Rhythm
Events	0.098	Rhythm	Pulseclarity	0.018	Rhythm
F0	0.032	Pitch			

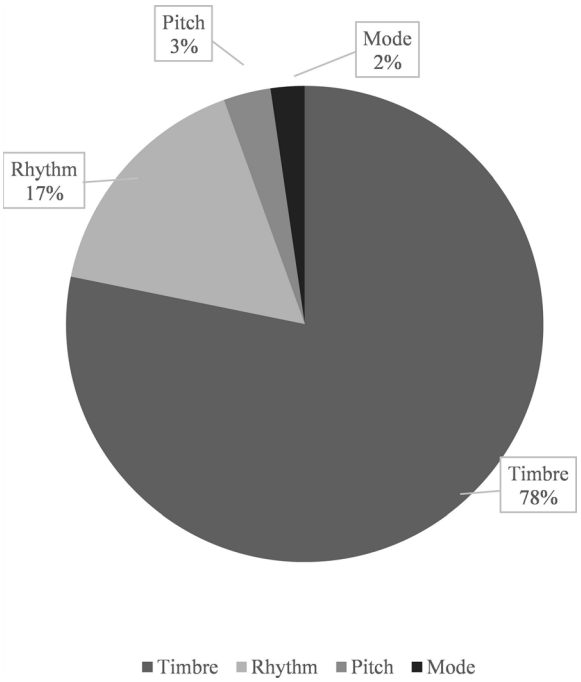

Fig. 3 The importance of musical elements to the classification of Chinese traditional music aesthetics

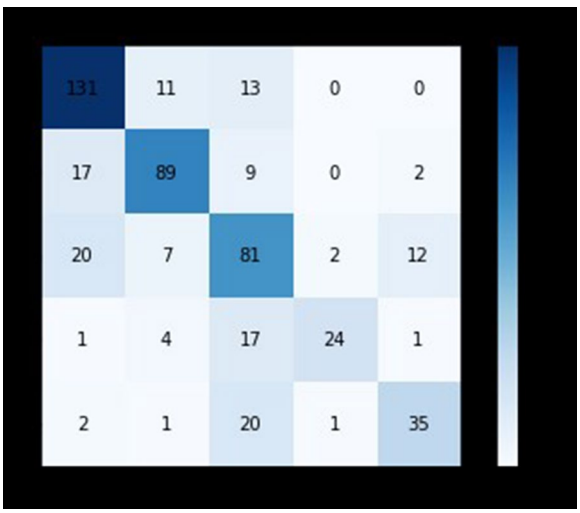

Fig. 4 Initial confusion matrix with 44 filtered features

LingDong aestetic category; it had a significant impact on the identification and mutual distinction between JiYang and BeiShang, and a greater impact on QingRou's misidentification as ShenChen and JiYang.

Figures 7 and 8 show the results of confusion matrices after removing the features of rhythm and mode. The aesthetic recognition rate after removing rhythm features dropped dramatically to 61.6%, while the

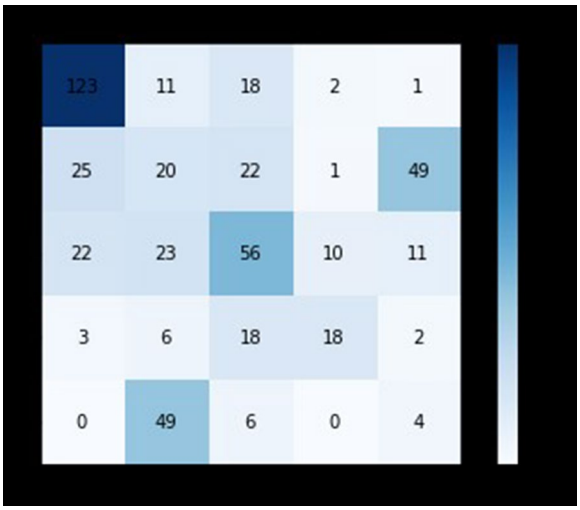

Fig. 5 Confusion matrix without timbre features

recognition rate after removing mode features remained at 67.8%. In comparison to the initial matrix in Fig. 4, rhythm and mode had little effect on ShenChen and BeiShang recognition, but the recognition rates of the other three categories all decreased to some extent, and the decline was even greater after the rhythm feature was removed. After removing the rhythmic feature, the number of confusions between the LingDong and the JiYang increased significantly, from 28 to 54, while the influence of the mode features only increased to 38. The number of misunderstandings between BeiShang and LingDong increased significantly, from 2 to 11 (rhythm) and 9 (mode), respectively. The analysis above shows that the impact of rhythm and mode had similar effects on the recognition of aesthetic categories: they had a certain impact on the identification and differentiation of LingDong and JiYang aesthetics, and they also contributed to distinguishing LingDong and BeiShang aesthetics. However, the influence of rhythmic features was greater than that of mode.

Figure 9 shows the confusion matrix obtained by removing the pitch features; the aesthetic recognition rate was reduced to 69.2%, which is not significantly lower. Individual recognition rates of various aesthetics decreased slightly, but there was no significant change in the degree of confusion with each other. It showed that the contribution of pitch features to aesthetic classification was relatively balanced, with a minor impact.

The influence of each musical element on each aesthetic category could be obtained based on the results of the confusion matrix analysis, which could be summarized as follows: (1) Timbre had a great influence on the aesthetic recognition rate, and it was also an important factor in the recognition and differentiation

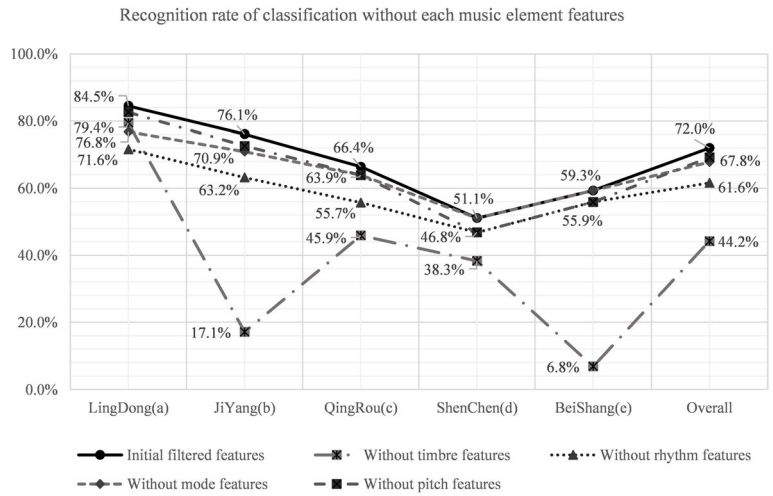

Fig. 6 Recognition rate of classification without each music element features

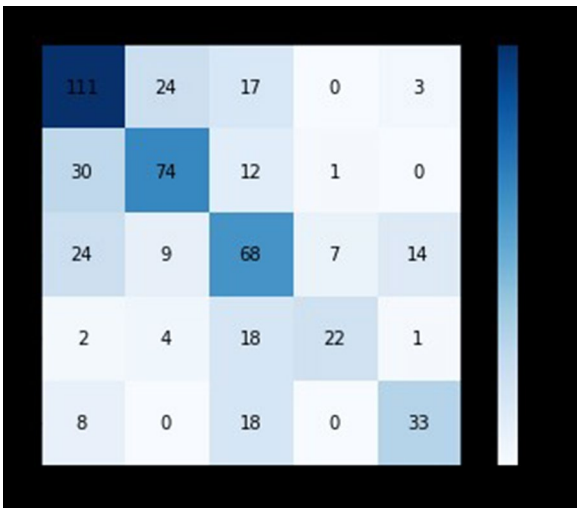

Fig. 7 Confusion matrix without rhythm features

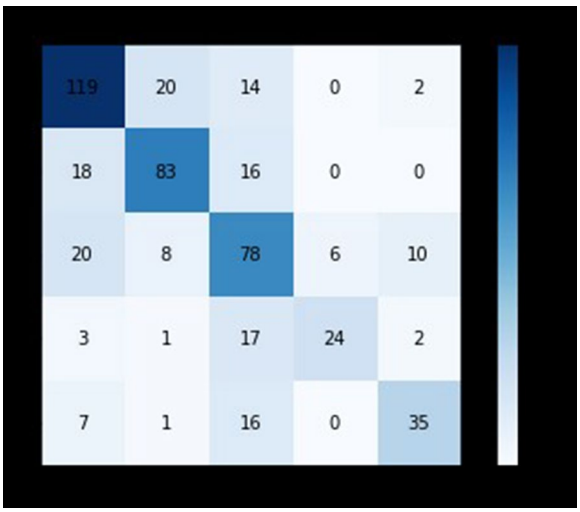

Fig. 8 Confusion matrix without mode features

of LingDong and JiYang, and it also affected the distinction between LingDong and BeiShang; (2) Rhythm had a great influence on the aesthetic recognition rate, and it was an important factor in the recognition and differentiation of LingDong and JiYang, and it also affected the distinction between LingDong and BeiShang; (3) Mode had a similar influence on aesthetic recognition rate and aesthetic categories as rhythm, but its influence degree was relatively light; (4) Pitch had only a minor influence on each aesthetic category's recognition rate and did not contribute to their mutual differentiation.

6 Conclusions

Based on an annotated database of Chinese traditional music with five aesthetic categories, this paper extracted 447 dimensions of music features from five musical elements, including energy, pitch, timbre, rhythm, and mode, and conducted the following research:

(1) The 447-dimensional features were screened using two feature selection methods: wrapping and filtering, yielding 44-dimensional features with a significant impact on aesthetic classification, increasing the accuracy of aesthetic recognition from 66.6% to 72.0%.

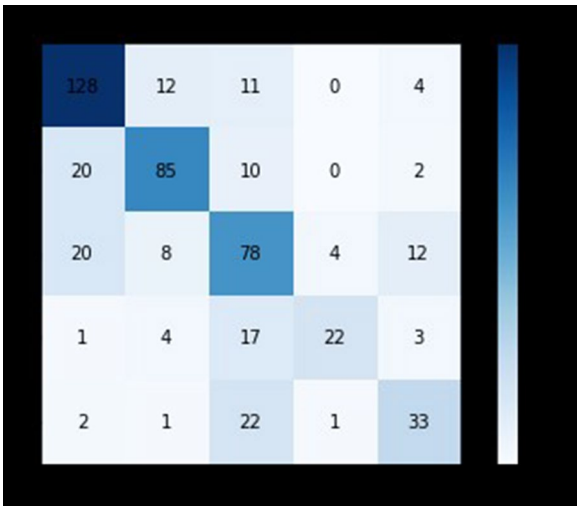

Fig. 9 Confusion matrix without pitch features

(2) The extreme random tree classifier was used to classify the filtered 44-dimensional physical features for aesthetic classification of Chinese traditional music, and the importance coefficients of each feature were calculated. The results revealed that SFM, SCF, MFCC, and Events were the most important first four types of features, with importance proportions of 36.5%, 20.2%, 17.3%, and 9.8%, respectively.

(3) The importance of musical elements for aesthetic classification was quantified when combined with the mapping relationship between physical features and musical elements. The findings revealed that timbre, rhythm, pitch, and mode are the four key musical elements that influence the recognition rate of Chinese traditional music's aesthetic classification, accounting for 78.2%, 16.3%, 3.2%, and 2.3%, respectively.

(4) The influence of each musical element on aesthetic classification was thoroughly investigated by removing relevant features and combining them with a confusion matrix. The findings revealed that the recognition and mutual distinction between the aesthetics of JiYang and BeiShang were primarily determined by timbre; in the absence of timbre features, QingRou would significantly increase the number of misidentified as JiYang and ShenChen; the recognition and distinction between LingDong and JiYang were primarily determined by rhythm, and secondarily by mode; the distinction between LingDong and BeiShang was primarily affected by rhythm and mode; the pitch had an effect on the overall recognition rate of aesthetics, but had less effect on the mutual distinction of different aesthetics.

This paper provided a useful reference for further information retrieval of Chinese traditional music and related music intelligent processing by conducting a comprehensive and in-depth analysis and research on the importance of low-level physical features and high-level musical elements to the classification of Chinese traditional music aesthetics. Of course, research into Chinese traditional music aesthetics is ongoing. Because of the ambiguity and polysemy of aesthetics, multi-label recognition of music aesthetics has a higher practical value, and this will be the next research direction. Furthermore, the database must be expanded in order to address the issue of category imbalance.

Acknowledgements
The producing of experimental stimuli was helped with Xinyu Ma from Communication University of China. Ying Zhan and Yuehong Wang from Communication University of China participated in the path design and feature extracting. The authors would like to express their sincere gratitude to the above students. Finally, we would like to thank the anonymous reviewers for their helpful feedbacks and remarks that improved this paper.

Authors' contributions
YG performed the conceptualization, methodology, formal analysis, writing — original draft. LYX performed the methodology, project administration, supervision, validation, writing — review and editing.

Funding
This work was supported by the National Social Science Foundation of art major project, 21ZD19; and the National Social Science Fund for Special Research Program, 22VJXG012.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 18 May 2023 Accepted: 28 December 2023
Published online: 02 February 2024

References

1. W. Menninghaus, V. Wagner, E. Wassiliwizky, I. Schindler, J. Hanich, T. Jacobsen, S. Koelsch, What are aesthetic emotions. Psychol. Rev. **126**(2), 171–195 (2019)
2. K. Hevner, Experimental Studies of the Elements of Expression in Music. Am. J. Psychol. **48**(2), 246–268 (1936)
3. A. Gabrielsson, E. Lindström, in *Series in affective science. Handbook of music and emotion: Theory, research, applications*, eds. by P.N. Juslin, J.A. Sloboda. The role of structure in the musical expression of emotions (Oxford University Press, New York, 2010), pp. 367-400
4. F.R. Baltes, J. Avram, M. Miclea, A.C. Miu, Emotions induced by operatic music: psychophysiological effects of music, plot, and acting: a scientist's tribute to Maria Callas. Brain Cogn. **76**(1), 146–157 (2011)
5. J.K. Vuoskoski, T. Eerola, The role of mood and personality in the perception of emotions represented by music. Cortex. **47**(9), 1099–1106 (2011)
6. T. Eerola, A. Friberg, R. Bresin, Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. Front. Psychol. **4**, 487 (2013)

7. G. Tzanetakis, P. Cook, Musical genre classification of audio signals. IEEE Trans. Speech Audio Process. **10**(5), 293–302 (2002)
8. Y. Liu, Y. Liu, Y. Zhao, K.A. Hua, What Strikes the Strings of Your Heart? - Feature Mining for Music Emotion Analysis. IEEE Trans. Affect. Comput. **6**(3), 247–260 (2015)
9. X. Yang, Y. Dong, J. Li et al., Review of data features-based music emotion recognition methods. Multimedia Systems. **24**(4), 365–389 (2017)
10. J. Su, P. Zhou, Machine Learning-based Modeling and Prediction of the Intrinsic Relationship between Human Emotion and Music. ACM Trans. Appl. Percept. **19**(3), 1–12 (2022)
11. H. Leder, B. Belke et al., A Model of Aesthetic Appreciation and Aesthetic Judgment. Br. J. Psychol. **95**, 489–508 (2004)
12. P.N. Juslin, L.S. Sakka et al., No accounting for taste? Idiographic models of aesthetic judgment in music. Psychol. Aesthet. Creat. Arts. **10**(2), 157–170 (2016)
13. I. Schindler, G. Hosoya, W. Menninghaus, U. Beermann, V. Wagner, M. Eid et al., Measuring aesthetic emotions: a review of the literature and a new assessment tool. PLoS ONE. **12**(6), e0178899 (2017)
14. T. Liu, Research on Music Emotion Cognitive Model and Interactive Technology (Zhejiang University, 2006)
15. G. Yan, L. Xie, in *proceedings of Asian conference on affective computing and intelligent interaction (ACII Asia)*. Aesthetics - Emotion Mapping Analysis of Music and Painting. (The Institute of Electrical and Electronics Engineers(IEEE), Beijing, 2018)
16. X. Lingyun, G. Yan, A database for aesthetic classification of Chinese traditional music. Cogn. Comput. Syst. **4**(2), 197-204(2022)
17. J. Shengyi, Y. Yao, L. Jingxin, Chinese Music Sentiment Dictionary Construction and Sentiment Classification Method Research. Comput. Eng. Appl. **50**(24), 118–121 (2014)
18. W. Wen, L. Xie, in *proceedings of Congress on Image and Signal Processing*. Discriminating Mood Taxonomy of Chinese Traditional Music and Western Classical Music with Content Feature Sets. (The Institute of Electrical and Electronics Engineers (IEEE), Sanya, 2008)
19. Z. Zhao, L. Xie, J. Liu, W. Wu, in *proceedings of International Conference on Signal Processing Systems*. The analysis of mood taxonomy comparision between Chinese and Western music. (The Institute of Electrical and Electronics Engineers(IEEE), Dalian, 2010)
20. X. Hu, Y.-H. Yang, Cross-dataset and cross-cultural music mood prediction: a case on Western and Chinese pop songs. IEEE Trans. Affect. Comput. **8**(2), 228–240 (2017)
21. X. Hu, Y.-H. Yang, The mood of Chinese pop music: Representation and recognition. J. Assoc. Inf. Sci. Technol. **68**(8), 1899–1910 (2017)
22. M. Xinyu, G. Yan, M. Zihou, in *Proceedings of National Acoustics Congress*. Automatic recognition of aesthetics of Chinese traditional music. The Acoustical Society of China. Proceedings of the 2018 National Acoustical Congress (Beijing, 2018)
23. W. Yuehong, X. Lingyun, in *Proceedings of National Acoustical Congress*. Correlation analysis between aesthetic judgment and time duration of Chinese traditional music (Shenzhen, 2019)
24. Z. Yan, Research on Music genre classification based on acoustic features and musical features (Jiangnan University, 2014)
25. G. Tzanetakis, P. Cook, MARSYAS: a framework for audio analysis. Organised Sound. **4**(3), 169–175 (2000)
26. F. Eyben, M. Wöllmer, B. Schuller, in *Proceedings of the 18th ACM international conference on Multimedia (MM '10)*. openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor (NY, USA, 2010)
27. O. Lartillot, P. Toiviainen, in *proceedings of the International conference on digital audio effects (DAFx)*. A Matlab Toolbox for Musical Feature Extraction from Audio (Bordeaux, 2007)
28. M. Barthet, G. Fazekas, M. Sandler, in *International Symposium on Computer Music Modeling and Retrieval*. Music emotion recognition: from content to context-based models. In: Aramaki, M., Barthet, M., Kronland-Martinet, R., Ystad, S. (eds.) From Sounds to Music and Emotions. CMMR 2012. Lecture Notes in Computer Science, vol 7900. (Springer, Berlin, Heidelberg, 2013)
29. D. Jiang, L. Lu, H. Zhang, et al., in *proceedings of International conference on multimedia and expo*. Music type classification by spectral contrast feature. (The Institute of Electrical and Electronics Engineers (IEEE), Lausanne, 2002)
30. Pedregosa et al., Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
31. Z. Zhihua, Machine Learning (Tsinghua University Press, 2016)
32. X. Huang, L. Zhang, B. Wang et al., Feature clustering based support vector machine recursive feature elimination for gene selection. Appl. Intell. **48**(3), 594–607 (2018)
33. I. Kononenko, Estimating attributes: analysis and extensions of RELIEF. european conference on machine learning, 1994: 171-182. Hevner K . Experimental Studies of the Elements of Expression in Music. Am. J. Psychol. **48**(2), 246-268 (1936)

Publisher's Note