ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num.2 - 2026

The Different Orientations of Western Symphony Orchestras and Chinese National Orchestras in Tonal Balance Control

Linfu Ta[1] *and Mingge Li*[1,*]

[1] National Academy of Music “Prof. Pancho Vladigerov” - Sofia, Sofia, 1142, Bulgaria

SUMMARY: *This study clarifies that timbre features are mainly composed of time-frequency, frequency-domain and cepstrum-domain features, and accordingly constructs a timbre feature dataset, which provides an important data basis for subsequent studies. As the dimensionality of timbre features increases, the correlation between timbre features and instrument labels decreases or becomes redundant. To address this problem, we chose to use principal component analysis to reduce the dimensionality of the timbre feature dataset, thus obtaining the Fisher-PCA and IG-PCA features, based on which we used the HMM classifier to categorize the features, and ultimately designed a timbre recognition model based on the HMM classifier, and used the model to carry out the timbre balance control of Western symphony orchestras and Chinese national orchestras. Difference analysis. The acoustic energy of the bass violin, which is a commonly used instrument in Western symphony orchestras, shows a decreasing trend from low to high frequencies, and the energy is lower when it is above the neighborhood of 21.3 KHz. On the other hand, the energy of the Ma Touqin, one of the instruments commonly used in Chinese folk orchestras, covers a wide range and is more evenly distributed, which indicates that the timbre identification model based on the HMM classifier is able to effectively reveal the differences between the Western symphony orchestra and the Chinese folk orchestra in terms of the control of timbral balance.*

KEYWORDS: *HMM classifier; timbre identification model; western symphony orchestra; Chinese folk orchestra; timbre balance control*

1 Introduction

Timbre is a characteristic of sound, which is an important feature that distinguishes different instruments or human voices [1]. Tonal color control is a way for players to change the timbral performance of their instruments by adjusting playing techniques, strength, speed and other factors to suit the emotional and stylistic demands of the musical work [2, 3]. For Western symphony orchestras and Chinese national orchestras, there is a big difference in timbre balance control orientation.

Western symphony is a traditional form of musical expression, and its charm lies in the fact that the feelings expressed are real and palpable, and the melody is beautiful and melodious [4, 5]. The reason why symphonic music sounds magnificent and inspiring comes from its extremely enriched instrumental composition. A symphony orchestra consists of three main parts: strings, winds, and percussion [6]. These instruments may be played at the same time, may also be sounded separately, which creates a rich change in the auditory effect, so that people can listen to the sound of mountains and rivers, but also the force of the mountains and rivers

*15354855202@163.com
https://doi.org/10.65102/is2026987
Published: 30 April, 2026

[7, 8]. In western symphony orchestras, the timbre balance control method belongs to the category of music performance theory, which fully considers the timbre matching and balance between instruments to achieve harmonious and changeful acoustic effects [9, 10]. While the Chinese national orchestra is a performance orchestra with Chinese national instruments as the main ensemble, the concept of modern national orchestra refers to the national orchestra with strings, plucked, wind and percussion as the main body, with a complete configuration of voices and balanced ratios, which is capable of accomplishing most of the national orchestral works composed or adapted after the 21st century until now [11-14]. The timbre balance control of Chinese national orchestras tends to be more oriented to the pursuit of acoustic balance under the "unity of flavor" than the timbre fusion of Western symphony orchestras, mainly because of the differences between the Chinese and the West in terms of aesthetics, as well as the improvement of the instruments imported into the country, and the control of the performances, etc. [15-18].

With the theoretical support of relevant research literature and data, it is understood that the timbre balance control features of Western symphony orchestras and Chinese national orchestras can be categorized into time-domain features, frequency-domain features, and cepstrum features, and the corresponding timbre feature datasets are constructed. Considering that an increase in the dimensionality of the timbre features will result in a decrease in the correlation or redundancy between the timbre features and the instrument labels, the feature selection and dimensionality reduction are performed using principal component analysis to obtain the Fisher-PCA and IG-PCA features, and then the HMM classifier is used to construct a timbre recognition model, and a timbre recognition model based on the HMM classifier is finally designed. The timbre recognition model is first combined with the timbre feature dataset to carry out the timbre recognition model validation analysis. After validating the validity of the model, the model is used to further explore the differences in timbre balance control between Western symphony orchestras and Chinese folk orchestras, aiming to reveal the different orientations of the Western symphony orchestras and Chinese folk orchestras in the control of timbre balance.

2 Orchestra Tone Balance Control Exploration

2.1 Characteristics Related to Tone Balance Control

2.1.1 Time-frequency characteristics

(1) Team tone balance control feature extraction based on audio signals

The autocorrelation coefficient and the over-zero rate are time-domain features computed directly from the audio signal.

a) Autocorrelation coefficient: used to represent the spectral distribution of the signal $s(t_n)$ in the time domain, which has been shown to provide a good description for classification. From keeping only the first 12 dimensional autocorrelation coefficients $(c \in \{1,\cdots,12\})$, denoted as:

$$xcorr(c) = \frac{1}{xcorr(0)} \sum_{n=0}^{L_n - c - 1} s(n)s(n+c) \tag{1}$$

where L_n is the window length and c is the time lag. In the experiment, we find the mean and variance of the 12-dimensional autocorrelation coefficients for all the frames to obtain a 24-dimensional feature.

b) Cross-zero rate: is the number of times the value of the signal $s(t_n)$ crosses the zero axis. This value tends to be small for periodic sounds and large for noisy sounds.

(2) Energy Envelope Based Feature Extraction

To estimate the start (t_{st}) and end (t_{end}) times of a musical tone, many algorithms rely on applying to the signal energy envelope $e(t_n)$ and inter-values. The logarithmic onset time is defined as follows:

$$LAT = \log_{10}(t_{end} - t_{st}) \tag{2}$$

Estimated onset times, decay times, release times, and logarithmic onset times of musical sounds were extracted. Percussion is distinguished from sustained sounds by this characterization with the following formula:

$$tc = \frac{\sum_{n=n_1}^{n=n_2} e(t_n) \cdot t_n}{\sum_n e(t_n)} \tag{3}$$

where n_1 and n_2 are the first and last values of n.

The approximation is such that the time-energy envelope $e(t_n)$ is higher than a given inter-value. After many empirical tests, the amplitude envelope is a macroscopic examination of a waveform, and a common method of calculating the amplitude envelope is the RMS algorithm. That is:

$$RMS = \sqrt{\frac{1}{L}\sum_{n=0}^{L} x^2(n)} \tag{4}$$

The RMS value is closer to the sensitivity of the human ear auditory system to audio signal intensity transformations. We find the mean and variance of the RMS energy envelope for all frames.

2.1.2 Frequency domain characteristics

For the frequency domain feature extraction in the team tone balance control features firstly, the fast Fourier transform is performed to obtain the STFT energy spectrum and the STFT power spectrum, and then the following features are calculated for the frequency spectrum respectively. $a_k(t_m)$ denotes the amplitude of the spectrum obtained by STFT transform, and $p_k(t_m)$ denotes the normalized value of the spectrum obtained by STFT transform.

(1) Spectral energy

The energy of the spectrum is the sum of the amplitudes a_k^2 in time t_m after the Fourier transform, as in the following equation. In the experiment, the STFT energy spectrum and STFT power spectrum of all the frames of the signal are extracted as features and the mean and variance are obtained. That is:

$$E_T(t_m) = \sum_k a_k^2(t_m) \tag{5}$$

(2) Spectral statistical features

For the tone balance control features, the common spectrum-based statistical features include four kinds of spectral center of mass, spectral width, spectral skewness and spectral kurtosis, and the corresponding calculation formulas are shown in Table 1. The STFT energy spectrum and STFT power spectrum of all frames of the signal are extracted from the above four features and their mean and variance are calculated.

Table 1: Spectral statistical characteristics and their calculation formulas

Feature name	Calculation formula	Serial number
Spectral centroid	$\mu_1(t_m)=\sum_{k=1}^{K} f_k p_k(t_m)$	(6)
Spectral width	$\mu_2(t_m)=\left(\sum_{k=1}^{K}\left(f_k-\mu_1(t_m)\right)^2\cdot p_k(t_m)\right)^{1/2}$	(7)
Spectral deviation	$\mu_3(t_m)=\left(\sum_{k=1}^{K}\left(f_k-\mu_1(t_m)\right)^3\cdot p_k(t_m)\right)^{1/2}/\mu_2^3$	(8)
Spectral kurtosis	$\mu_4(t_m)=\left(\sum_{k=1}^{K}\left(f_k-\mu_1(t_m)\right)^4\cdot p_k(t_m)\right)^{1/2}/\mu_2^4$	(9)

(3) Other features of the spectrum

In addition to the spectral energy and statistical features mentioned above, the following six features can be extracted in the frequency domain of the tone balance control feature based on the spectrum of the tone balance control feature. The following features are extracted for the STFT energy spectrum and STFT power spectrum and their means and variances are found.

a) Spectral slope: defined as the mean value of the spectral slope.

b) Spectral descent rate: defined as the average value of a set of slopes when the spectrum is descending.

c) Spectral roll-off: defined as the critical frequency corresponding to a drop in amplitude to 85% of the total spectral energy.

d) Spectral flatness: defined as the ratio of the geometric mean to the arithmetic mean of the spectrum.

e) Spectral crest: defined as the ratio of the maximum value of the spectrum to the arithmetic mean value.

f) Spectral Flux: defined as the frame-to-frame energy fluctuations that transform over time, responding to changes in the spectrum over time. This feature is commonly used in the separation of speech and musical signals.

2.1.3 Characterization of the inverted spectral domain

(1) Inverted Spectral Domain Characteristics

The partial characterization of the resonance peaks varies from person to person and is the key to determining the characteristics of the tonal balance control. The cepstrum coefficients containing the signal quantity $y(n)$ are defined as, and F denotes the discrete Fourier transform. i.e:

$$c(n)=F^{-1}\{\log|F\{y(n)\}|\} \tag{10}$$

There are two Fourier transforms in the above equation, whose computational efficiency is

not very high, and they are not used in the actual instrument recognition.

(2) Linear Prediction Cepstrum Coefficients

The main idea of linear prediction (LP) is to use a linear combination of the sampling values of the past several moments to represent the current moment of sampling.

a) Linear Prediction

The basic principle of linear prediction is to represent the analyzed signal with a model, i.e., the signal is regarded as the output of a certain model, so that the signal can be described by the model parameters, and the linear prediction model is shown in Figure 1. The linear prediction model is shown in Figure 1, where $u(n)$ denotes the model input and $x(n)$ denotes the model output. When $x(n)$ is a deterministic signal, the model input is a sequence of unit shocks; when $x(n)$ is a random signal $u(n)$ can be used as a white noise sequence.

The transfer function $H(z)$ of the model can be written in the form of a rational fraction:

$$H(z) = G\frac{B(z)}{A(z)} \tag{11}$$

Among them:

$$\begin{cases} A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \\ B(z) = 1 + \sum_{k=1}^{q} b_k z^{-k} \\ H(z) = \sum_{k=1}^{p} h(k) z^{-k} \end{cases} \tag{12}$$

where a_k, b_k and the gain factor G are the parameters of the model; and p, q is the order of the selected model.

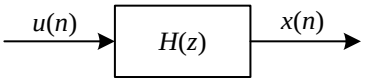


Figure 1: Linear prediction model

b) Linear Predictive Cepstrum Coefficient

Since the frequency response $H\left(e^{jw}\right)$ responds to the frequency response of the channel and the spectral envelope of the analyzed signal, the linear prediction cepstrum coefficient (LPCC) is derived by doing a Fourier inverse transform with $\log\left|H\left(e^{j\omega}\right)\right|$, and it is also considered to contain the spectral envelope information of the signal. envelope information of the signal, so it can be regarded as an approximation of the short-time cepstrum of the original signal.

The system function of the synthesized filter obtained by linear prediction analysis is $H(z) = 1/\left(1 - \sum_{i=1}^{p} a_i z^{-1}\right)$, and its impact response is $h(n)$. A recurrence relation between the inverse spectrum $\hat{h}(n)$ of $h(n)$ and a_i can be obtained by inference, knowing that $\hat{h}(1) = a_1$, one has:

$$\hat{h}(n) = a_n + \sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)a_i\hat{h}(n-i) \qquad 1 < n \le p \tag{13}$$

$$\hat{h}(n) = \sum_{i=1}^{p}\left(1-\frac{i}{n}\right)a_i\hat{h}(n-i) \qquad n > p \tag{14}$$

According to the above equation, the inverse spectrum $\hat{h}(n)$ can be obtained directly from the prediction coefficient a_i. This cepstrum coefficient is obtained according to the linear prediction model and utilizes the minimum phase property of the system function $H(z)$ in the linear prediction, so it avoids the trouble of seeking the complex logarithm in the general homomorphic treatment.

(3) Mel frequency cepstrum coefficient

The analysis of Mel Frequency Cepstrum Coefficient (MFCC) is based on human auditory mechanism, i.e., to analyze the spectrum of the tone balance control features based on the results of human auditory experiments.

a) Auditory Mechanism

First, the delineation of the frequency domain of human subjective perception is not linear with the following equation:

$$F_{mel} = 1125\log\left(1+\frac{f}{700}\right) \tag{15}$$

where F_{mel} is the perceived frequency in Mel; f is the actual frequency in Hz. If the spectrum of the tonal balance control is transformed into the perceived frequency domain, the different orientations of the western symphony orchestra and the Chinese national orchestra in the tonal balance control can be better simulated.

b) Mel filter bank

Each filter has a triangular filtering characteristic with a center frequency of $f(m)$, and these filters are of equal bandwidth in the Mel frequency range. The transfer function of each bandpass filter is:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1)-k}{f(m+1)-f(m)} & f(m) < k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \tag{16}$$

where $f(m)$ can be defined in the following way:

$$f(m) = \left(\frac{N}{f_s}\right)F_{mel}^{-1}\left(F_{mel}(f_l) + m\frac{F_{mel}(f_h)-F_{mel}(f_l)}{M+1}\right) \tag{17}$$

where f_l is the lowest frequency in the filter's frequency range; f_h is the highest frequency in the filter's frequency range; N is the length at the Fourier transform; f_s is the sampling

frequency; and the inverse function of F_{mel}, F_{mel}^{-1}, is:

$$F_{mel}^{-1}(\mathrm{b}) = 700\left(e^{\frac{b}{1125}} - 1\right) \tag{18}$$

c) MFCC feature extraction

MFCC feature extraction includes the following steps: preprocessing, Fast Fourier Transform, calculation of spectral line energy, calculation of energy through Mel filter, and calculation of DCT cepstrum. Preprocessing: Preprocessing includes pre-emphasis, frame-splitting and windowing. Speech signal $x_i(m)$, where subscript i denotes the ith frame after framing.

Fast Fourier Transform: FFT transform $X(i,k) = FFT\left[x_i(m)\right]$ for each frame of the signal.

Calculate the spectral line energy: $E(i,k) = [X(i,k)]^2$.

Calculate the energy through the Mel filter: $S(i,m) = \sum_{k=0}^{N-1} E(i,k)H_m(k), 0 \le m < M$;

Calculate the DCT cepstrum: Calculate the DCT discrete cosine transform after taking the logarithm of the energy of the Mel filter, and the MFCC features obtained are shown below:

$$mfcc(i,n) = \sqrt{\frac{2}{M}} \sum_{m=0}^{M-1} \log[S(i,m)] \cos\left(\frac{\pi n(2m-1)}{2M}\right) \tag{19}$$

2.2 Tone Recognition Model

Although the human ear's perception of the timbre of a musical instrument begins at the onset stage, it is not until the sustained stage that a true sense of it can be gained. Therefore, it is more reasonable to judge the timbre of a musical tone in terms of a complete note. In addition, the musical signal is a non-smooth signal, after de-mean and out-of-unity, it needs to be processed in frames, and the same timbre feature extraction is performed on each frame of the signal, and finally the timbre feature sequence is obtained. This section involves three models such as GMM, UBM and HMM, etc. GMM and UBM ignore the temporal information of the timbre feature sequence and only establish the probability distribution to consider the hopping probability of the hidden state, and HMM can make up for the lack of temporal information to a certain extent.

2.2.1 Tone Characterization Dataset

Scholars have been designing timbre balance control features for a long time, which can be roughly categorized into time-domain features, frequency-domain features, and cepstrum features. Among them, the cepstrum features, which are derived from the sound mechanism of musical instruments, are the most abundant and effective, and the structure of the timbre feature set selected in this section is shown in Table 2. Most of the timbre features have already been introduced in the previous section, and only the last three are briefly introduced here.The OBSI and OBSIR features have good performance and focus on describing the energy relationship between musical octaves. The extraction process can be summarized as follows:

(1) Calculate the amplitude spectrum of each signal frame.

(2) Filter the amplitude spectrum using an octave filter bank.

(3) Calculate the energy within each filter and take the logarithm to obtain the OBSI.

(4) The ratio of adjacent OBSI coefficients is extended to another set of feature parameters,

OBSIR.The TimbreTB is extracted using the Timbre Toolbox, which mainly contains the time-domain features and frequency-domain features of each frame of the signal, with a total of 121 dimensions.

Table 2: Timbre feature dataset

Feature name	Feature dimension	Annotation
MFCC	13	The number of MEL filters is 50, the dimension of the DCT is 50, and the first 13 dimensions are taken.
$\triangle$*MFCC*	13	MFCC first-order time difference.
$\triangle\triangle$*MFCC*	13	MFCC second-order time difference.
GTCC	13	The number of Gammatone filters is 50, the dimension of DCT is 50, and the first 13 dimensions are taken.
LPC	13	The full pole model is of order 13.
LPCC	16	The LPCC is directly calculated on the LPC according to equation
OBSI	8	Octave Band Signal Intensities
OBSIR	7	Octave Band Signal Intensities Ratio
Timbre TB	25	Feature set derived from the timbre toolbox

2.2.2 Feature selection and dimensionality reduction

Although the performance of musical instrument recognition usually improves as the dimensionality of the timbre balance control features increases, the opposite effect will occur if the correlation between the timbre features and the instrument labels decreases or there is redundancy. Therefore, timbre balance control feature selection becomes a necessary technical tool to improve the performance of musical instrument recognition.

Tone balance control feature selection can be categorized into four types: (1) Filtered feature selection, which has low algorithmic complexity and flexible operation because it does not require classifier training. (2) Encapsulated feature selection, which requires classifier training and recognition results, with high algorithmic complexity and good robustness. (3) Embedded feature selection, as part of the classifier, efficient but high correlation with the classifier, and needs to be redesigned if the classifier is replaced. (4) Hybrid feature selection, combining filtered and encapsulated, taking into account the advantages of both, but may suffer from overfitting. Considering that multiple classifiers will be examined in this section, two types of encapsulated feature selection are introduced: feature selection based on Fisher's criterion and feature selection based on information gain (IG).

The Fisher criterion-based score will be higher if the intra-class distance of a particular dimensional tone balance control feature is smaller and the inter-class distance is larger. The process of Fisher criterion based feature selection is as follows:

(1) Calculate the timbre balance control features.

(2) Calculate the Fisher score for each dimension of the timbre balance control feature. Namely:

$$F\left(X_j, Y\right) = \frac{\sum_{i=1}^{M} n_i \left(\mu_i^j - \mu^j\right)^2}{\sum_{i=1}^{M} n_i \left(\sigma_i^j\right)^2} \tag{20}$$

where j is the timbre feature dimension label, i is the instrument category, n_i is the number of samples of the i th instrument category, and Y is the label set.

(3) The timbre balance control features are arranged in descending order according to the Fisher score, taking the first K dimensions. The value of K needs to be determined experimentally.

IG is another metric to measure the correlation between timbre balance control features and instrument labels. The IG-based feature selection process is as follows:

(1) Extract the timbre balance control features.

(2) Calculate IG:

$$IG\left(Y \mid X_j\right) = H(Y) - H\left(Y \mid X_j\right) \tag{21}$$

where $H(Y)$ is the information entropy of the label set and $H\left(Y \mid X_j\right)$ is the conditional entropy.

(3) Take the first K dimensions after arranging the tone balance control features in descending order according to the IG score. K can be selected experimentally.

After feature selection, it can only guarantee that the selected features are more relevant to the labels, but cannot guarantee that there is no redundancy among the features. For this reason, Principal Component Analysis (PCA) is used to remove redundancy.

The calculation process of principal component analysis is as follows:

(1) Calculate the selected timbre balance control feature matrix $C = \left\{C_{ij}\right\}, i, j = 1, 2, \cdots, I, I$ is the dimensionality of the timbre features. The computational expression for C_{ij} is:

$$C_{ij} = \frac{\sum_{n=1}^{N_{all}} \left(X_{ni} - \bar{X}_i\right)\left(X_{nj} - \bar{X}_j\right)}{N_{all} - 1} \tag{22}$$

(2) Compute the eigenvalues $\left\{\lambda_1, \lambda_2, \cdots, \lambda_I\right\}$ of the covariance matrix with the eigenvectors $\left\{w_1, w_2, \cdots, w_I\right\}$.

(3) Arrange the eigenvectors according to the descending order of the eigenvalues and retain the first N tone balanced control vectors to form the projection matrix. To avoid discussing the number of dimensions to be retained, the sum of the retained eigenvalues is 99% of the sum of all eigenvalues and at least five dimensions are retained. The projection matrix can be expressed as: $W = \left[w_1, w_2, \cdots, w_N\right]$.

(4) The timbre balance control feature is: $X^* = W^{\mathrm{T}} X = \left[w_1, w_2, \cdots, w_N\right]^{\mathrm{T}} \left[X_1, X_2, \cdots, X_1\right]$.

2.2.3 Shallow Classifier Configuration

In view of the differences in the formation mechanism of musical instrument timbre, the classifier can realize musical instrument recognition by fitting the probability distribution of timbre features. At the same time, the selection of timbre balance control features and the dimensionality reduction strategy also play an important role in improving the recognition performance of the classifier.In this paper, HMM is chosen to complete the shallow classifier configuration in the timbre recognition model.The HMM model contains two types of stochastic processes, the first one is the transfer process between the states in the model, and

the second one is that each state corresponds to an observation.The HMM model consists of five parameters. The specific categorization is as follows:

(1) The number of states N of the HMM model. The state space $S=\{S_1,S_2,\ldots,S_N\}$, and let the state at moment t be q_t. Then $q_t \in S$, i.e., q_t is some value in the set S.

(2) The number of observations M corresponding to each state. Let the set of observations be O, then $O=\{O_1,O_2,\ldots,O_N\}$.

(3) The transfer probability distribution matrix A between the states, with the expression:

$$A=\{a_{ij}\},a_{ij}=P\left(q_{t+1}=S_j \mid q_t=S_j\right),1\le i,j\le N \tag{23}$$

(4) The state-generated observation matrix B with the expression:

$$B=\{b_j(O_k)\},b_j(O_k)=P,1\le j\le N,1\le k\le M \tag{24}$$

(5) The initial state probability distribution π with the expression:

$$\pi=\{\pi_i\},\pi_i=P\left(q_1=S_i\right),1\le i\le N \tag{25}$$

The HMM model is shorthanded by $\lambda=(A,B,\pi)$ and is complete when the above five parameters are determined. Among them, the parameters A and π together determine the Markov chain that determines the transfer probabilities between states, and the parameter B determines the observation probability distribution.

3 Model Validation and Tone Balance Control Exploratory Analysis

3.1 Tone Recognition Model Validation Analysis

Based on the dataset, this paper carries out the validation analysis of the timbre recognition model from the time-frequency features, frequency domain features, cepstrum domain features, feature selection and dimensionality reduction to provide theoretical support for the following exploration and analysis of the Western and Chinese timbre balance control.

3.1.1 Time-frequency characterization

Piano, violin, xylophone, trumpet, flute, mandolin, which are commonly used instruments in western symphony orchestra and Chinese national orchestra, are recognized and analyzed for their time-frequency features using the HMM-based timbre recognition model, and the amplitude envelope of the musical signals reflects the change of their amplitude during the playing time, which can be expressed as the waveforms formed by the sound's “volume-time”. The amplitude envelope of the musical signal reflects its amplitude change during the playing time, which can be expressed as the “volume-time” waveform of the sound, and the monophonic time domain waveforms of different musical instruments are shown in Fig. 2, where (a) to (f) represent the piano, violin, xylophone, trumpet, flute, and mandolin, respectively. Combined with the data performance in the figure, it can be seen that the timbre balance control of piano and violin is more stable and smooth compared with xylophone, trumpet, flute and mandolin, which not only visually demonstrates the changes in the

distribution of time-frequency features of different musical instruments, but also proves the feasibility of the HMM-based timbre recognition model in the analysis of time-frequency feature recognition.

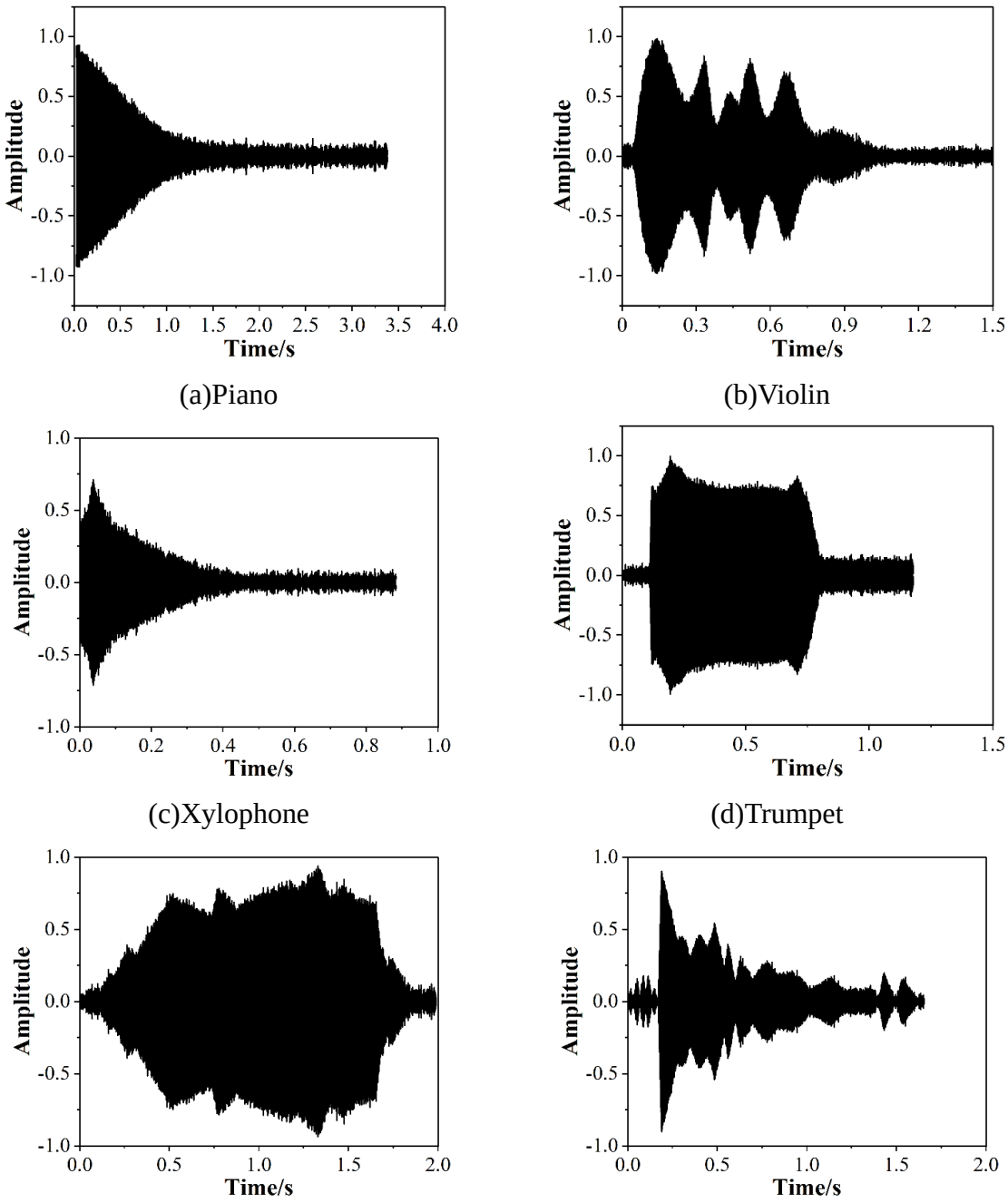


Figure 2: Single-tone time-domain waveforms of different Musical Instruments

3.1.2 Frequency domain characterization

The timbre of a musical instrument is determined by the intensity, distribution and variation of its harmonic components. In musical instruments, the fundamental frequency is usually the main component of the note. However, the vibration of a musical instrument contains not only the fundamental frequency, but also a series of harmonic components, i.e., subharmonics, second harmonics, third harmonics, etc. whose frequencies are integer multiples of the

fundamental frequency. The intensity and frequency distribution of these harmonic components will directly affect the characteristics of the instrument's timbre, and the frequency domain characterization is shown in Figure 3. In the C4-B4 pitch musical notes, the amplitude and number of harmonics of the violin and trumpet are close to each other, and the amplitude decays slower after the 1st harmonic, while the amplitude value of the 1st harmonic is 1. Generally speaking, the amplitude of the lower harmonics of most musical instruments is larger, however, the amplitude of the higher harmonics of a small number of musical instruments is relatively larger. It is the large differences in the harmonic structures of different musical instruments that constitute the differences in timbre between different musical instruments, verifying the effectiveness of the application of the HMM-based timbre recognition model in frequency domain characterization.

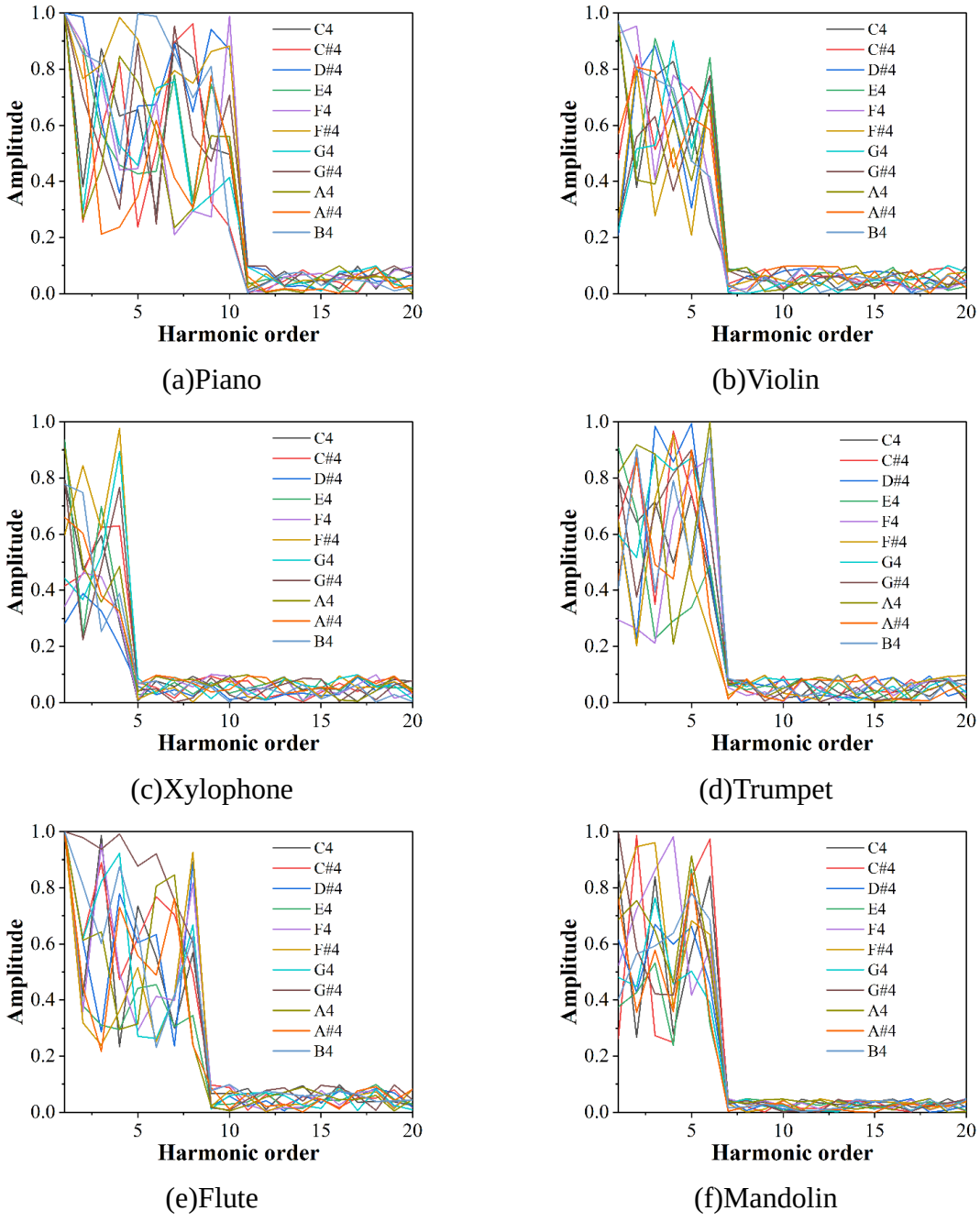


Figure 3: Frequency domain characteristic analysis

3.1.3 Characterization of the inverted spectral domain

After analyzing the time-frequency features and frequency-domain features of the instruments commonly used in Western symphony orchestras and Chinese national orchestras, the HMM-based timbre recognition model is used to analyze the cepstrum features, in which the cepstrum features include LP, LPCC, and MFCC, and the confusion matrix of the LP features is shown in Table 3, the confusion matrix of the LPCC features is shown in Table 4, and the confusion matrix of the MFCC features is shown in Table 5. It can be seen from the data in the table that the recognition rate of LP, LPCC and MFCC features of commonly used musical instruments in Western symphony orchestras and Chinese folk orchestras by the HMM-based timbre recognition model reaches more than 0.9, which indicates that the HMM-based timbre recognition model has more powerful feature learning and feature expression capabilities, and is able to identify the inverted spectral domain features of the timbres of commonly used musical instruments in Western symphony orchestras and Chinese folk orchestras effectively, which has guiding value for the control of timbre balance in Western symphony orchestras and Chinese folk orchestras. It can effectively identify the inverted spectral domain features of musical instrument timbres, which is of guiding value in controlling the timbre balance of Western symphony orchestras and Chinese folk orchestras.

Table 3: LP feature confusion matrix

Musical instrument	Piano	Violin	Xylophone	Trumpet	Flute	Mandolin
Piano	0.912	0.015	0.019	0.016	0.019	0.019
Violin	0.017	0.924	0.02	0.015	0.019	0.005
Xylophone	0.014	0.011	0.915	0.017	0.011	0.032
Trumpet	0.016	0.014	0.018	0.914	0.017	0.021
Flute	0.02	0.015	0.016	0.017	0.905	0.027
Mandolin	0.017	0.017	0.018	0.018	0.01	0.92

Table 4: LPCC feature confusion matrix

Musical instrument	Piano	Violin	Xylophone	Trumpet	Flute	Mandolin
Piano	0.922	0.01	0.018	0.016	0.02	0.014
Violin	0.017	0.921	0.01	0.018	0.011	0.023
Xylophone	0.014	0.015	0.928	0.012	0.014	0.017
Trumpet	0.019	0.013	0.015	0.933	0.014	0.006
Flute	0.012	0.019	0.011	0.019	0.935	0.004
Mandolin	0.012	0.018	0.019	0.013	0.013	0.925

Table 5: MFCC feature confusion matrix

Musical instrument	Piano	Violin	Xylophone	Trumpet	Flute	Mandolin
Piano	0.012	0.012	0.012	0.012	0.012	0.012
Violin	0.009	0.009	0.009	0.009	0.009	0.009
Xylophone	0.001	0.001	0.001	0.001	0.001	0.001
Trumpet	0.01	0.01	0.01	0.01	0.01	0.01
Flute	0.017	0.017	0.017	0.017	0.017	0.017
Mandolin	0.913	0.913	0.913	0.913	0.913	0.913

3.1.4 Feature Selection and Dimensionality Reduction Analysis

Two feature selections were mentioned earlier, which are Fisher and IG, and their dimensionality reduction was carried out using Principal Component Analysis, from which Fisher-PCA and IG-PCA features were obtained. With the support of the timbre feature dataset, the feature selection and dimensionality reduction analysis is carried out using the HMM classifier, and the Fisher-PCA feature confusion matrix is shown in Table 6 and the IG-PCA feature confusion matrix is shown in Table 7. Regardless of the Fisher-PCA and IG-PCA features, the HMM classifier has a recognition accuracy of more than 0.9, which fully verifies the efficacy and effectiveness of the HMM classifier in the timbre recognition model, with a view to promoting the digital research and development of the western symphony orchestra and the Chinese folk orchestra in the control of timbre balance.

Table 6: Fisher-PCA feature confusion matrix

Musical instrument	Piano	Violin	Xylophone	Trumpet	Flute	Mandolin
Piano	0.943	0.006	0.007	0.008	0.005	0.031
Violin	0.011	0.955	0.012	0.011	0.01	0.001
Xylophone	0.011	0.009	0.937	0.007	0.007	0.029
Trumpet	0.006	0.01	0.01	0.948	0.008	0.018
Flute	0.012	0.009	0.012	0.01	0.939	0.018
Mandolin	0.01	0.009	0.01	0.012	0.011	0.948

Table 7: IG-PCA feature confusion matrix

Musical instrument	Piano	Violin	Xylophone	Trumpet	Flute	Mandolin
Piano	0.945	0.014	0.009	0.012	0.005	0.015
Violin	0.014	0.951	0.007	0.01	0.015	0.003
Xylophone	0.011	0.007	0.939	0.005	0.006	0.032
Trumpet	0.008	0.009	0.007	0.944	0.012	0.02
Flute	0.007	0.015	0.011	0.012	0.952	0.003
Mandolin	0.013	0.011	0.006	0.005	0.007	0.958

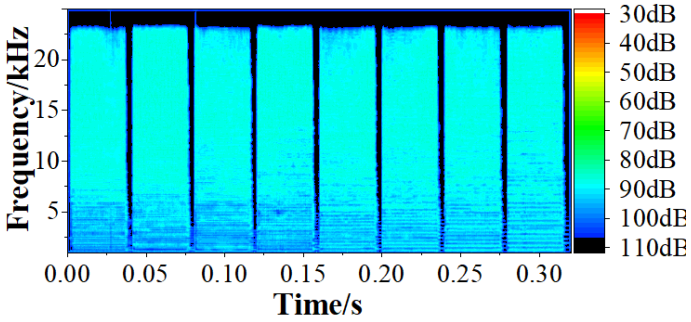
3.2 Analysis of Chinese and Western Tone Balance Control Inquiry

Through the above timbre recognition model validation analysis, it can be seen that there are also many differences between Western instruments and Chinese folk instruments. On the basis of the timbre recognition model based on HMM, a brief comparative analysis of three performance techniques of the horse-head fiddle and the double bass is taken as an example, aiming at revealing the different orientations of the western symphony orchestra and the Chinese national orchestra in terms of timbre balance control.

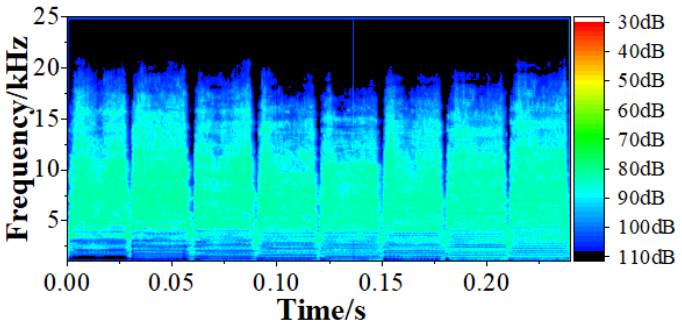
3.2.1 Comparative Analysis of String Pulling Conventional Techniques

On the basis of the timbre recognition model based on HMM, a comparative analysis of the string-pulling conventional techniques was carried out. Figure 4 shows the spectral comparison between the bass violin and the horse-head violin using the string-pulling conventional techniques to play the A3 to A4 scales, of which (a)~(b) are the bass violin and the horse-head violin, respectively. Both belong to the same stringed instruments, and under the premise of the same pitch and strength, it can be found that the bass violin has higher energy in the low and middle frequencies, and the energy within 4.07 KHz is more aggregated, and the energy in the fundamental frequency around 212 Hz-433 Hz is very prominent. On the other hand, the

marimba is relatively dispersed, with relatively high energy within 5452Hz, but much lower than that of the double bass, while the energy near the fundamental frequency is not obvious and hard to detect. On the other hand, the acoustic energy of the double bass shows a decreasing trend from the low frequency to the high frequency, and when it is above the neighborhood of 21.3KHz, the energy is lower and hard to find. On the other hand, the energy of the marimba covers a wide range and is more even, and energy can be found in the spectral range supported by the equipment. Therefore, from the hearing, the bass violin is more prominent bass brings the sense of hearing, the bass gathering force is stronger, the horse-head violin will be relatively weaker, but from the spectrum of the results to analyze, it is more focused on the full frequency range of the more average energy to bring the sense of hearing.



(a)Double bass



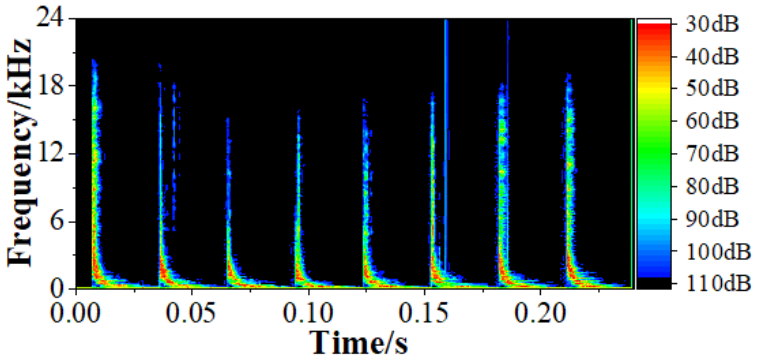
(b)Horsehead fiddle

Figure 4: Comparison of the spectrum of conventional string drawing techniques

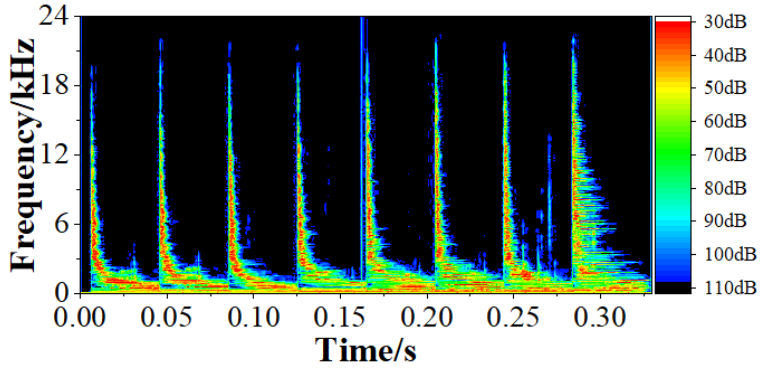
3.2.2 Comparative Analysis of Plucked String Techniques

On the basis of the HMM-based timbre recognition model, a comparative analysis of pizzicato techniques was carried out, and the spectral comparison of pizzicato techniques is shown in Figure 5. There are also some differences between the two instruments. Under the premise of the same pitch and strength, the spectral energy distribution of the two is basically the same as that of the conventional string-plucking technique, and the most intuitive difference is the acoustic spectral envelope. Comparing the pizzicato technique of the double bass with that of the pizzicato under the same premise of pitch and strength, from the observation of the spectral envelope, the starting speed of the vibration phase of the two is very rapid, and in terms of triggering energy, the range of instantaneous triggering of the pizzicato is still more than that of the double bass. In terms of trigger energy, the horsehead violin still has more instantaneous trigger range than the double bass. In terms of energy near the fundamental frequency, the double bass has more energy and a more aggregated range. The energy near the fundamental

frequency is lower than that of the double bass, and the range of the higher energy is more scattered. In the release stage, the two have the same pattern of decay from high to low frequencies, but obviously, the release time of the horse-head fiddle is much longer. Therefore, the playing of the horse-head fiddle has its own merits when compared with the plucking of the double bass.



(a)Double bass

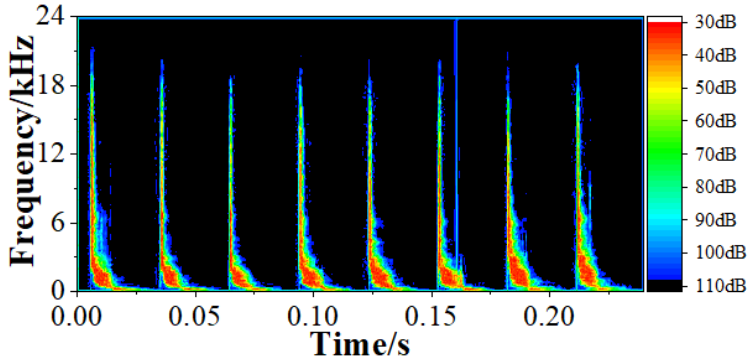


(b)Horsehead fiddle

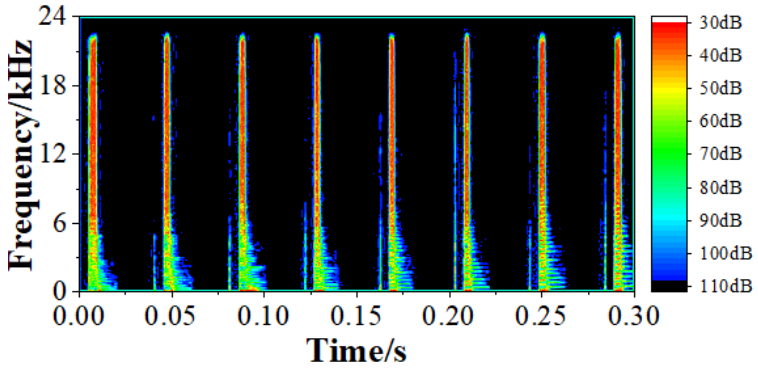
Figure 5: Comparison of the spectrum of plucking techniques

3.2.3 Comparative analysis of break techniques

Using the HMM-based timbre recognition model, a comparative analysis of the broken technique was carried out. The spectral comparison of the broken technique of the double bass and the horsehead violin is shown in Fig. 6, including the samples of the other instruments. Under the premise of the same pitch and strength, the results of the spectral and FFT analyses are basically the same as those of the conventional technique, because only the time value of the conventional technique of the instrument is changed, and there is no qualitative change of the timbre, etc. The results of the instantaneous trigger frequency are basically the same as those of the conventional technique. The triggering phases of the marimba and double bass were shorter, and the instantaneous triggering frequency results and other aspects were basically the same as those of the conventional technique. In the release phase, the decay rate of each harmonic of the horsehead violin is still very fast, which objectively proves that the string vibration time of the horsehead violin is lower than that of the double bass.



(a)Double bass



(b)Horsehead fiddle

Figure 6: Spectrum comparison of Staccato technique

3.2.4 Generalization of results

After several sets of spectral comparisons with the horse-head fiddle and double bass for example, it can be argued from this dimension that: as the double bass emphasizes more on the fundamental frequency, the energy is gathered, and the spectral distribution is regular and orderly, and it is observed in the acoustic spectral envelope that the release time is longer, which enables the listener to perceive the corresponding pitches more easily, so that, in the ensemble of a more massive orchestra compilation, it is able to better play the double bass instrument in the original job. The Ma Touqin emphasizes the more average energy in the frequency domain, and the energy is more dispersed, and in the plucking and breaking techniques, although the sound spectrum envelope transient triggers higher energy, the release time is too short, which makes the listener feel that the sound is more dry and lacks the sense of reverberation, and therefore, in the ensemble of the larger compilation, its play will be much inferior. Based on the results of the investigation and analysis on the control of timbre balance between China and the West, it reveals the different orientations of the Western symphony orchestra and the Chinese national orchestra in the control of timbre balance.

4 Conclusion

With the continuous development of international communication, there are relatively few studies using artificial intelligence technology to explore the different orientations of western symphony orchestras and Chinese national orchestras in terms of timbre balance control. In this

paper, the HMM classifier is used to construct a timbre recognition model, and the model is used to explore the different orientations of the timbre balance control between Western symphony orchestras and Chinese national orchestras.

(1) The recognition rate of LP, LPCC and MFCC features of instruments commonly used in Western symphony orchestras and Chinese folk orchestras by HMM classifier reaches more than 0.9, which verifies that the HMM-based timbre recognition model can accurately recognize the inverted spectral domain features of the timbres of instruments commonly used in Western symphony orchestras and Chinese folk orchestras, and facilitates the investigation of the timbre balance control of Western symphony orchestras and Chinese folk orchestras. The exploration work is carried out.

(2) Under the guidance of HMM-based music recognition modeling theory, the bass violin commonly used in Western symphony orchestras has higher energy in the low and middle frequencies, and the energy within 4.07 KHz is more aggregated, and the energy in the fundamental frequency around 212 Hz-433 Hz is very prominent. On the other hand, the horse-head fiddle commonly used in Chinese folk orchestras is relatively scattered, with relatively high energy within 5452 Hz, but much lower than that of the double bass, and the energy near the fundamental frequency is not obvious, which fully demonstrates the difference between the Western symphony orchestra and the Chinese folk orchestra in the control of tonal balance.

About the Author

Linfu Ta (First author) was born in Huhhot, Inner Mongolia, P.R. China, in 1990. He obtained a master's degree from Northwest Normal University in China. He is currently studying at National Academy of Music “Prof. Pancho Vladigerov”, Sofia. His main research direction is the chinese national orchestra and western symphony orchestra.

Mingge Li (Corresponding author) was born in Huhhot, Inner Mongolia, P.R. China, in 1992. She obtained a master's degree from the National Academy of Music “Prof. Pancho Vladigerov” in Sofia, Bulgaria. She is currently pursuing a Ph.D. at the National Academy of Music “Prof. Pancho Vladigerov”. Her research direction is Chinese traditional choir and traditional musical instruments.

References

[1] Van Elferen, I. (2021). The Vibrant Aesthetics of Tone Color. The Oxford Handbook of Timbre, 69.

[2] Kim, J. W., Bittner, R., Kumar, A., & Bello, J. P. (2019, May). Neural music synthesis for flexible timbre control. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 176-180). IEEE.

[3] Kobayashi, K., Toda, T., Doi, H., Nakano, T., Goto, M., Neubig, G., ... & Nakamura, S. (2014). Voice timbre control based on perceived age in singing voice conversion. IEICE TRANSACTIONS on Information and Systems, 97(6), 1419-1428.

[4] Kawabata, M. (2023). The new “yellow peril” in “western” European symphony orchestras. Voices for Change in the Classical Music Profession: New Ideas for Tackling Inequalities and Exclusions, 159-171.

[5] Hess, J. (2018). Interrupting the symphony: unpacking the importance placed on classical

concert experiences. Music Education Research, 20(1), 11-21.

[6] Tamburri, L., Munn, J., & Pompe, J. (2015). Repertoire conventionality in major US symphony orchestras: Factors influencing management's programming choices. Managerial and Decision Economics, 36(2), 97-108.

[7] Bucur, V. (2022). Organology of Percussion Instruments for the Classic Symphony Orchestra. In Handbook of Materials for Percussion Musical Instruments (pp. 41-101). Cham: Springer International Publishing.

[8] Prado-Guerra, A., Paniagua Bermejo, S., Calvo Prieto, L. F., & Llorente, M. S. (2020). Environmental impact study of symphony orchestras and preparation of a classification guide. International Journal of Environmental Studies, 77(6), 1044-1059.

[9] Wenmaekers, R., & Hak, C. (2015). A sound level distribution model for symphony orchestras: Possibilities and limitations. Psychomusicology: Music, Mind, and Brain, 25(3), 219.

[10] D'Orazio, D., Fratoni, G., & Garai, M. (2020). Enhancing the strength of symphonic orchestra in an opera house. Applied Acoustics, 170, 107532.

[11] Wu, T., & Woramitmaitree, N. (2023). TEACHING KNOWLEDGE OF THE CHINESE NATIONAL ORCHESTRA AT THE SICHUAN CONSERVATORY OF MUSIC IN CHINA. International Online Journal of Education & Teaching, 10(3).

[12] Bibu, N., Brancu, L., & Teohari, G. A. (2018). Managing a symphony orchestra in times of change: Behind the curtains. Procedia-Social and Behavioral Sciences, 238, 507-516.

[13] Wang, Y. (2024). The development of the Chinese symphony orchestra after 1949. Frontiers in Art Research, 6(2), 67-74.

[14] Yu, F., & Mat, R. C. (2024). The Evolution of Chinese Orchestra: The Role of Li Delun's Conducting Art in the History of Chinese Orchestra Conduction. Cultura: International Journal of Philosophy of Culture and Axiology, 21(4), 86-105.

[15] Yusof, R. Z. R. (2025). The Evolution and Role of the Clarinet in Symphony Orchestras: A Comparative Analysis of Historical and Contemporary Perspectives. Art and Theory, 1(1), 1-6.

[16] Liu, J., Wang, S., Xiang, Y., Jiang, J., Jiang, Y., & Lan, J. (2022). Comparison and analysis of timbre fusion for chinese and western musical instruments. Frontiers in Psychology, 13, 878581.

[17] Kurbanova, M. M. (2025). COMPARATIVE ANALYSIS OF EASTERN AND WESTERN ORCHESTRAL TRADITIONS. European Review of Contemporary Arts and Humanities, 1(5), 55-59.

[18] Huang, J., & Huang, J. (2025). The integration of Chinese traditional instruments in contemporary symphony ensembles. Música Hodie, 25.