Retraction Notice

The Editor-in-Chief and the publisher have retracted this article, which was submitted as part of a guest-edited special section. An investigation uncovered evidence of systematic manipulation of the publication process, including compromised peer review. The Editor and publisher no longer have confidence in the results and conclusions of the article.

JC did not agree with the retraction. HW either did not respond directly or could not be reached.

Multimedia music image production combining HMM and song feature tags

Jiayue Cui[a] **and Hongjun Wang**[b,*]
[a]Harbin Normal University, Harbin, China
[b]Harbin Conservatory of Music, Harbin, China

Abstract. In the vast river of art and music, this wonderful flower of art has inherited culture through its expansion in horizontal space and continuation in vertical time since its birth. This research mainly discusses the filming of multimedia music images that integrates hidden Markov model (HMM) and song feature tags. Hidden Markov model is a statistical model based on hidden Markov process, which is mainly used for modeling stochastic processes. The editing structure, editing style, and the time and space performance of the editing should be considered in the editing process. When filming multimedia music images in the study of music history, there are not only the shooting pictures of the main body of the musical things but also the introduction of the environmental information generated by the musical things. Editing needs to consider whether each shot and each picture meets the actual research situation for the musical things to be described. It uses HMM to predict the song category based on the user's listening behavior records and song characteristics and then recommends the corresponding category of music to the user in real-time. Finally, the recommended results of several algorithms are verified on the experimental data set. It compares the difference between the traditional music algorithm and the HMM from different dimensions, such as accuracy, error rate, and recommended time efficiency. The accuracy of HMM for music is predicted to be 95%. The multinote recognition system of the piano uses the characteristics of the hidden Markov double random process to describe the statistical characteristics of the audio stream of notes. The proposed research will help promote the dissemination of music image information. © *2022 SPIE and IS&T* [DOI: 10.1117/1.JEI.31.5.051419]

Keywords: fusion hidden Markov model; song feature tags; multimedia music images; filming plan; music history.

Paper 220091SS received Feb. 23, 2022; accepted for publication May 18, 2022; published online Jun. 17, 2022; retracted Jun. 24, 2023.

1 Introduction

The multimedia system is a rich resource library, which can realize the fair distribution of educational resources. Music majors can view the shared music resources in the multimedia system, and there is no difference between good and bad conditions. In this way, every student has equal educational opportunities, which helps to improve students' music literacy and teaching quality in an all-around way. This paper studies the automatic arrangement of chords. The feature part adopts a high-level contour feature based on tone and considers the high frequency overtone problem in music and the negative impact of high feature redundancy brought by multiple timbres improved.

Music is a hands-on course that requires an in-depth discussion of experiences between students and between students and teachers. It can be improved quickly by design and realization of singing or performance level and personal performance level of multimedia for music teaching. The combination of traditional teaching methods and informatization methods can learn from each other's strengths and can improve students' interest in learning through some video and audio.[1] It is of great significance to use multimedia music image technology to record the musical, cultural relics left in the past dynasties in the study of music history. This is because these cultural relics themselves with the passage of time, and the existing environment will change and

*Address all correspondence to Hongjun Wang, xieruifeng@hrbnu.edu.cn

their authenticity will be attenuated. It also has invaluable academic value to use multimedia music image technology to record the live performance of various music works today. Traditional music teaching focuses on emotional expression and timbre. China's traditional music culture is different from other types of music, and it pays more attention to the timbre of music.

Made a close connection with the theme of the article: study the principle of speech recognition based on hidden Markov model (HMM), find out the commonalities and differences between speech recognition and multinote recognition, and establish multinote HMM acoustic model and multinote intermodel. The present research study proposes the content that multimedia music images in music history research should carry: including the environmental information of musical things, the ontology information of musical things, the performance and performance information of musical things, and the research information of musical things. This is considered the only method that brings no regrets for future generations. Aiming at the cold start problem in the music recommendation process, this paper proposes a new music recommendation algorithm. The algorithm uses song feature tags to recommend popular songs to new users. At the same time, in music recommendation, recommendation systems often need to continue song recommendation at the end of the playlist. Faced with the user's favorite playlist, the system needs to dig out the user's interest deviation from it. In the field of communication, HMM is generally used to predict the behavioral probability of time series. Based on this principle, this paper combines the user's listening behavior record and song characteristics to predict the song category (HTR algorithm) and then recommend the corresponding song to the user in real-time. Although the traditional collaborative filtering algorithm can easily mine the user's hobbies, it has the problems of data cold start and sparsity.

2 Related Work

When users use the search system to obtain information, they will find that the traditional search system cannot dig out the user's potential needs, and the user needs to participate in the information query process. Amini believed that watermark detection is a method to verify the existence of a watermark in a watermarking scheme used for copyright protection of digital data.[2] Rashidi considered that using Android devices, users can install third-party applications from various open markets. This can cause security and privacy issues because third-party applications can be malicious.[3] To quickly detect early bearing failures, Li developed bearing condition monitoring indicators based on self-organizing maps.[4] Carrer believed that radar detectors are unique underground surveys for ground and space applications. The authors proposed a new technology based on local-scale HMM to detect layer boundaries[5] automatically. Based on the output of a mismatched HMM test filter, an HMM filter was applied to observations not generated by its model. In another study, Jameset al. proposed a change detection method for dependent processes.[6] According to the MIDI file and the corresponding audio file, construct training, test waveform data, and corresponding annotation files, the recommendation algorithm based on the song's content will first analyze the song description with a higher score in each user and use it as the user's interest data. The system calculates the similarity between the higher-rated songs and the remaining songs in the resource library. Finally, it recommends the corresponding song collection to each user according to the degree of similarity. With the development of information and multimedia technology, music digitization is widely used in various media, such as radio broadcasting, digital storage, network, and so on. In this way, efficiently retrieving and managing music in a large amount of music has become the focus of research and development in recent years.

3 Song Recommendation Based on Fusion HMM

3.1 *Multimedia Music Image Production*

SIFT is a new image compression scheme based on local feature descriptors. SIFT descriptor has scale and rotation invariance to a picture and is widely used in image information retrieval. In this paper, the SIFT method is used to extract the features of each direction and each gradient of the local image block and then combined with the SPM method. Different scale scale-spaces integrate the feature vector of each image block into a feature vector as the feature of the whole image. Thus, the auditory image characteristics of chords.[7] In the music-influenced filming process, HMM uses historical preference data to calculate the distance between users and then uses the weighted score value of the target's nearest-neighbor users to evaluate the song to predict the target user's preference for a specific song.

Mass spectrometer:

$$C_t = \sum_{n=1}^{N} M_t[n] \cdot n / \sum_{n=1}^{N} M_t[n]. \tag{1}$$

Among them, $M_t[n]$ represents the magnitude of the Fourier transform at binary n and frame t. A high mass spectrum indicates that the sound is brighter in the high-frequency range.

The spectrum roll-off represents the spectrum distribution below the frequency R_t85%, and it is also a parameter that reflects the shape of the spectrum. Its expression is as follows:[8]

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \cdot \sum_{n=1}^{N} M_t[n]. \tag{2}$$

The spectral fluctuation represents the mode of the difference in the regularized spectral sequence distribution. Its expression is as follows:

$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2. \tag{3}$$

Among them, $N_t[n]$, $N_t - 1[n]$ represent the regularized spectrum of the t'th frame and the t'1th frame, respectively.

Zero crossing rate:[9]

$$Z_t = \frac{1}{2} \sum_{n=1}^{N} |\sin(x[n]) - \sin(x[n-1])|. \tag{4}$$

The square-integrable function $\phi(t)$ satisfies the condition:

$$\int_{-\infty}^{+\infty} |\hat{\varphi}(\omega)|^2 |\omega|^{-1} \mathrm{d}\omega < +\infty. \tag{5}$$

For the discrete case, the wavelet sequence is[10]

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j} t - k), j, k \in Z. \tag{6}$$

For any function $f(t) \in L^2(R)$, the continuous wavelet transform is[11]

$$W_f(a,b) = \langle f, \psi_{a,b} \rangle = |a|^{-1/2} \int_R f(t) \overline{\psi\left(\frac{t-b}{a}\right)} \mathrm{d}t, \tag{7}$$

$$\psi_{a,b} = r \int_R f(t) \overline{\psi\left(\frac{t-b}{a}\right)} \mathrm{d}t. \tag{8}$$

The process of making multimedia music images is shown in Fig. 1. For the monophonic signal input by the microphone recording, quantization noise and aliasing interference will be generated when converted from quantization to digitization.[12,13] The pre-emphasis process is realized using a pre-emphasis digital filter with 6 dB/octave to improve high-frequency characteristics. The input layer to the hidden layer is fully connected, but the weight is fixed at 1. There is no need to prespecify the weight between the hidden layer and the output layer, and

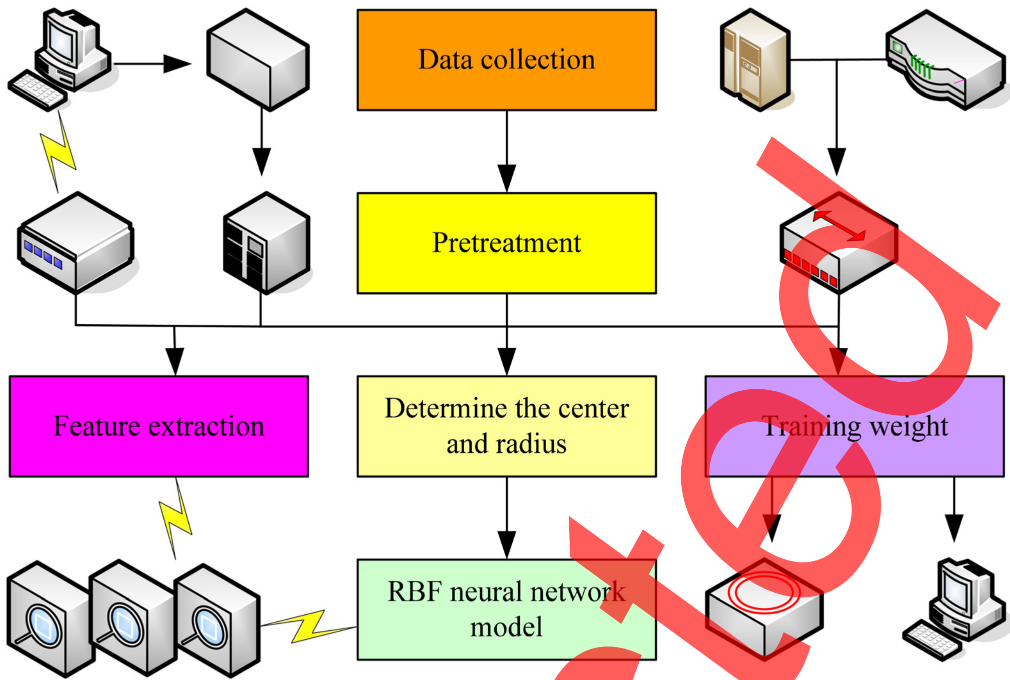

Fig. 1 Multimedia music image production process.

it can be directly calculated in the next step. In addition, each node of each output layer sets a Sigmoid function, which is used to normalize the result of the linear operation to between 0 and 1, which is convenient for classification processing.

3.2 *Fusion HMM*

The multinote acoustic modeling process is similar to the phoneme modeling, but the parameters in the HMM are different; the multinote modeling is done by counting the number of consecutive times between multinote and multinote or knowing the current multinote calculates the probability of the next polynote occurring. The actual problem is often more complicated than described by the Markov chain model. What people can observe is often not one-to-one corresponding to the state, but connected by a set of probabilities, and such a model is called HMM. If for any $n = 1,2,3,\ldots$ and any $t_0, t_1, \ldots, t_n \in T$ and any real numbers x, y, the formula is as follows:[14]

$$p\{\xi(t_n) \le y|\xi(t_{n-1}) = x, \ldots, \xi(t_0) = x_0\} = p\{\xi(t_n) \le y|\xi(t_{n-1}) = x\}, \tag{9}$$

$$\xi_1(t_n) = \chi/T. \tag{10}$$

First, we must define a forward variable $\alpha_t(i)$:

$$\alpha_t(i) = P(O_1, O_2, \ldots, O_t; q_t = S_i|\lambda). \tag{11}$$

That is, given the condition of the model λ, the sequence $\{O_1, O_2, \ldots, O_t\}$ will be in the Si state at time t. The following are the steps of iterative calculation of forward variables:[15]

(1) Initialization:

$$\alpha_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N, \tag{12}$$

$$\alpha_2(i) = \pi_i b_i(O_2), \qquad 1 \le i \le N, \tag{13}$$

$$\alpha_3(i) = \pi_i b_i(O_3), \qquad 1 \le i \le N. \tag{14}$$

(2) Iterative calculation:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i)a_{ij}\right] b_j(O_{t+1}), 1 \le t \le T-1, \qquad 1 \le j \le N. \tag{15}$$

(3) Final calculation:[16]

$$p(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i). \tag{16}$$

Among them, a_{ij} is an element in the state transition matrix, and $b_j(O_t)$ is an element in the observation symbol matrix.

Through the frequency spectrum frame mapping program of audio data, each frame of the signal is composed of only 12-dimensional feature vectors, which greatly simplifies the data of each frame. The playback time of an audio frame is equal to (length of the frame in bytes)/ sampling frequency. One of the most useful features of this feature vector is that it can encode the chords contained in a given song. Therefore, two audio frames with similar harmony content will have the same feature vector. With the help of this feature vector, the correlation between the feature vectors of two audio frames can be calculated and measure the similarity of the two audio frames accordingly. Spectrum is the abbreviation of frequency spectral density, which is the distribution curve of frequency. Complex oscillations are decomposed into harmonic oscillations with different amplitudes and frequencies, and the graph in which the amplitudes of these harmonic oscillations is arranged according to frequency is called the frequency spectrum.

Assuming that the short-term energy of the nth frame of speech signal $x_n(m)$ is represented by E_n, the equation is[17]

$$En = \sum_{m=0}^{N-1} x_n^2(m). \tag{17}$$

For periodic or random digital speech signals, $R(k)$ is defined as

$$R(k) = \lim_{N\to\infty} \frac{1}{2\ N+1} \sum_{m=-N}^{N} x(m)x(m+k). \tag{18}$$

The width determines the size of the receptive field and affects the accuracy of the network. The width selection needs to make the sum of the receptive fields of all RBF hidden layers cover the entire training sample space. To classify the sample space to form n classes, then[18]

$$\sigma_j = \sqrt{\frac{1}{n}\sum_{j=1}^{n} [(\vec{X} - \vec{W}_{1j})^T (\vec{X} - \vec{W}_{1j})]^2}, \tag{19}$$

To make the image key points have strong robustness and rotation invariance characteristics, it is necessary to determine the direction parameters of the key points according to the gradient distribution characteristics of the pixels:[19,20]

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}, \tag{20}$$

$$\theta(x,y) = \tan^{-1}(L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)). \tag{21}$$

For the continuous HMM, the observation density function is a mixed Gaussian density function. To determine the number of functions M, if the training data are different, the number of functions is different. When the amount of training data increases, M also increases, and at the same time, the amount of calculation increases. The covariance matrix is a full symmetric matrix.

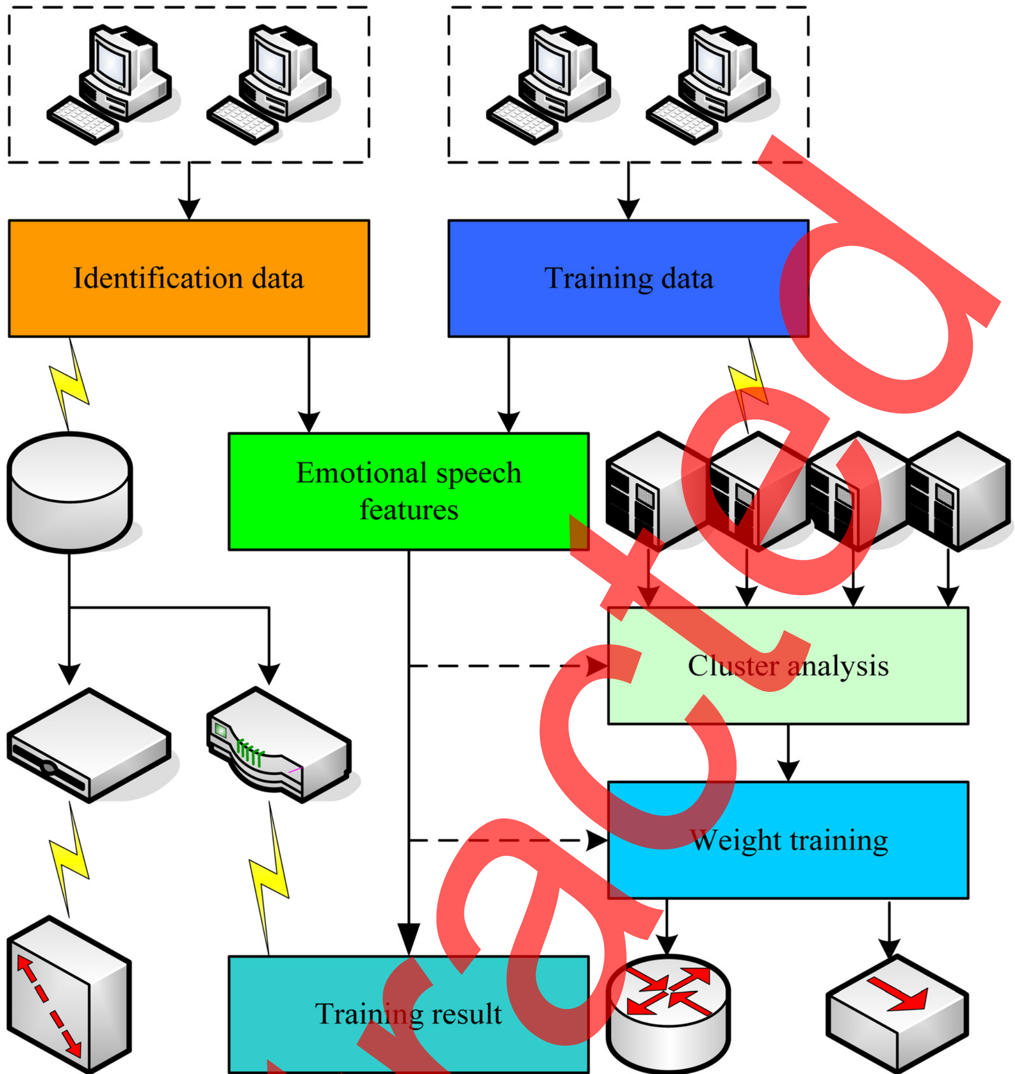

Fig. 2 RBF training and recognition process.

This paper uses a diagonal matrix, assuming that the dimensions of the parameter vector are independent of each other. This can reduce the complexity of the calculation and the storage requirements; the requirements for training data are not high. Using the segmented K-means algorithm, the obtained state sequence is statistically analyzed.[21] It clusters the speech signal frames into M categories, so the observed value probability B is re-evaluated, and a new model is obtained. In this class, the covariance matrix and mean vector of the speech signal frame vector are calculated as the covariance matrix and mean vector of the Gaussian probability density function of this class, and M Gaussian probability density functions are obtained. The RBF training recognition process is shown in Fig. 2.

3.3 *Song Feature Extraction*

To prevent the likelihood function from overflowing, this paper uses logarithmic values to calculate. The equation is as follows:[22]

$$\delta_{t+1}(j) = \{\max_j [\delta_t(i) + \log(a_{ij})]\} + \log[b_j(O_{t+1})], \tag{22}$$

$$\delta_n(j) = \delta_n(i)/\phi. \tag{23}$$

Assuming that the input signal is $x(n)$, the short-time Fourier transform of the signal is defined as[23]

$$X_n(e^{jw}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-jwm}. \tag{24}$$

The DoG equation is as follows:[24]

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma). \tag{25}$$

3.4 *Experimental Data Set*

The recommended algorithm program of this experiment is written in Java language. The recommendation algorithm proposed in this paper mainly implements the model parameter construction of M-HTR (HMM in MTR algorithm) and the Viterbi algorithm for decoding problems. The final recommendation algorithm code will be added to the Mahout recommendation framework. The experimental platform uses a Windows10 operating system, 8G memory, inteli7 processor, and the recommended framework version of Mahout is 0.6.

3.5 *Database Selection*

The selection of the database is one of the important links of the design experiment, and it plays an important role in evaluating the performance of the chord recognition system. Different research teams choose different feature extraction and recognition models, and the resulting chord recognition rates are also different. At the same time, the difference of the recognition object will also cause the difference of the recognition result, and there is no comparability between methods in this way. According to the knowledge of music theory, there are many types of chords. But in reality, many types of chords are rarely used in Western music. Therefore, this article only considers the 24 types of chords that frequently appear in music, namely major and minor. To establish the training sample database, 24 chords were intercepted according to the chord label file. Each type contains a different number of chord samples, but almost all types of genres are included. This article first uses dynamic programming to detect the beats of these four songs and then uses cool editor software to segment the songs according to the detected beat time position. Finally, a complete song is divided into audio data composed of many chord segments.

3.6 *Building M-HTR*

Since the user's rating data are known, the maximum likelihood estimation method can be used to estimate the parameters of M-HTR. First, the recommendation algorithm will collect the user's song playlist, and the order of the songs in the list will be sorted according to the time the user added the song list. The user's actions on each song will also be recorded in the model. When constructing the parameters of the state transition matrix, the algorithm will calculate the transition times of any two adjacent song feature tags in the user's collection. Thus, the transition probability value of each tag state is calculated. It calculates the value of the behavior observation probability matrix for the behavior actions that contain this label when each kind of label appears.

3.7 *Realizing Song Recommendation Based on M-HTR*

When recommending songs to the user's existing playlist, the user's song history score can be divided into several favorite categories (single loop, sharing, like, active playback, listening, skip) according to the score (1 to 100). According to the sequence of time, the user's rating of each song is converted into a favorite category and used as an observation sequence, and the user's implicit interest tag is used as the implicit state sequence of M-HTR.

Table 1 Set playlist attenuation factor display results.

Number of album songs	Attenuation factor
0 to 10	0.90
10 to 20	0.91
20 to 30	0.92
30 to 40	0.93
40 to 50	0.94
50 to 60	0.95
60 to 70	0.96
70 to 80	0.97

4 Multimedia Music Image Production Results

In each playlist recorded by the user, the most recently recorded song by the user will be in the item section of the playlist. Songs closer to the bottom can be presumed to deviate from the user's interest. Therefore, the playlist attenuation factor can be set according to the number of playlist music and the ranking of each song. The display results are shown in Table 1.

Through fields such as playlist attenuation factor, relative song ranking, and song collection time, users can quantify the score of each piece of music. The resulting user rating information is shown in Table 2.

"The Bells of Geneva" enters the first phrase of the main melody (bars 5-11 of the whole song), for example, as shown in Fig. 3. The music score marked "doliss" and the soft tone "una corda," the quiet and soft melody opened the prelude to the whole song.[25]

In this part, it will select the following four authoritative performance versions as the objects of visualization research. The selected four authoritative performance versions are shown in Table 3.

Four performers (Barenboim performance version, Pires performance version, Polini performance version, and Rubinstein performance version) have performed local rhythm expansion and contraction here. The relatively free rhythm processing can be felt from the interpretation of the four performance versions.[26] The characteristics of the four performance versions are shown in Fig. 4.

The HMM algorithm is used to process speech signals. Because of the particularity of the music signal and the process of averaging in codebook training, canceling the smoothing can

Table 2 User rating information.

User serial number	Song number	Song recording timestamp
1	595	1333286336
2	586	1336033313
3	627	1306363363
4	23	1303333633
5	61	1336331363
6	63	1336863106
7	595	1636186336
8	586	1333286336

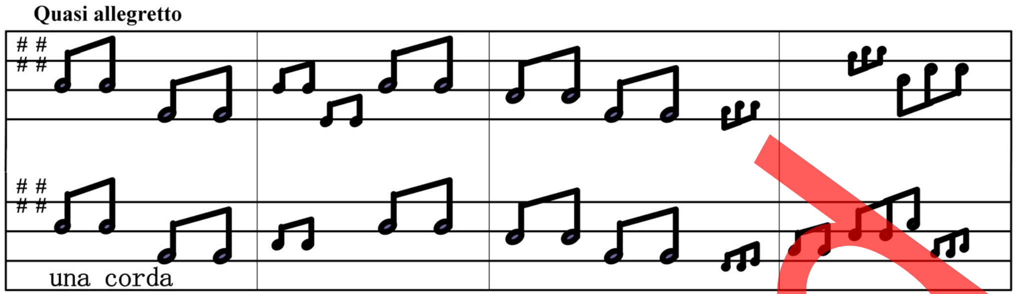

Fig. 3 An example of the score of the first phrase (bars 5-11 of the whole song) that enters the main melody of “The Bells of Geneva.”

Table 3 Four authoritative performance versions selected.

Performer	Publish time	Album name	Record source
Artur Rubinstein	1990	The Chopin Collection: The Noc turnes	RCA 89563 (2)
Daniel Barenboim	1999	Chopin: Nocturnes	DG 463057-2
Maria-Joao Pires	1996	Chopin: Nocturnes	DG 447096-2
Maurizio Pollini	2005	Polilini Chopin :Nocturnes	DG 477571-8

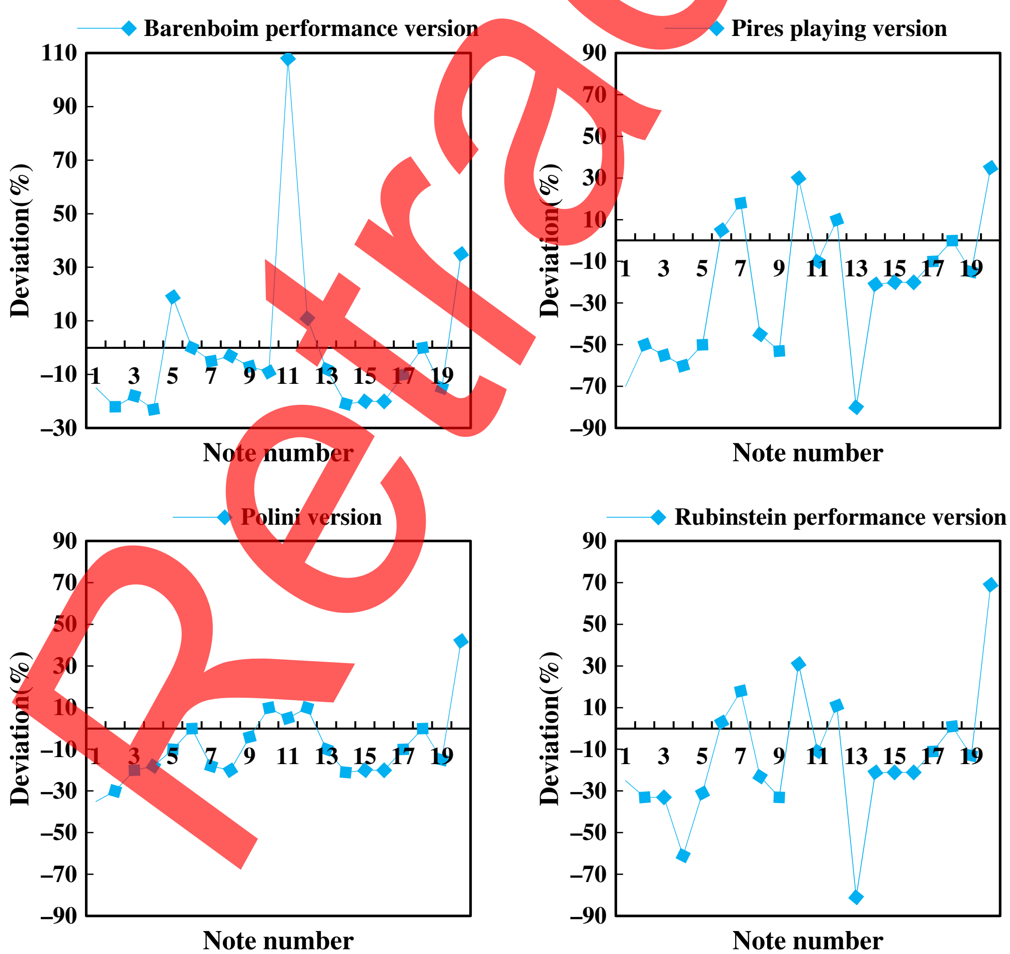

Fig. 4 The characteristics of the four performance versions.

Table 4 Speech signal processing results of HMM algorithm.

Timbre	Number of notes	Correct number	Recognition rate
Piano	61	55	90.2%
Violin	22	22	100%
Oboe	58	58	100%
Total	141	135	95.7%

better meet the requirements of the music signal. The experimental results prove that HMM can well solve the basic problem of music recognition. The result of HMM algorithm speech signal processing is shown in Table 4.

The comparison of the HMM algorithm system itself is shown in Table 5.

Among the audio data synthesized from these MIDI music files, the time length of the classical music data set is 15.3435 h, with 252,588 frame features. The distribution of chords and keynotes in the classical music data set is shown in Fig. 5.

The time length of the Beatles music data set is 9.4756 h, with the characteristic of 155,889 frames. The distribution of chords and keynotes in the Beatles music data set is shown in Fig. 6.

When testing the chord recognition system, two kinds of music audio are used to test, namely classical music and Beatles music. The test collection of classical music is Bach's Prelude in C major and Prelude in F major. The test collection of Beatles music is the two Beatles songs, LoveMeDo (C major) and Anna (G major). Classical music and the test music in the Beatles music test set are not included in the training set of the model. The specific conditions of the test set are shown in Table 6.

Table 5 Comparison of HMM algorithm system itself.

Timbre	Polyphonic music	Polyphonic music
Piano	1	2
Complete system	18%	9.9%
Fenobo	6.2%	17%
No filtering	22%	20%

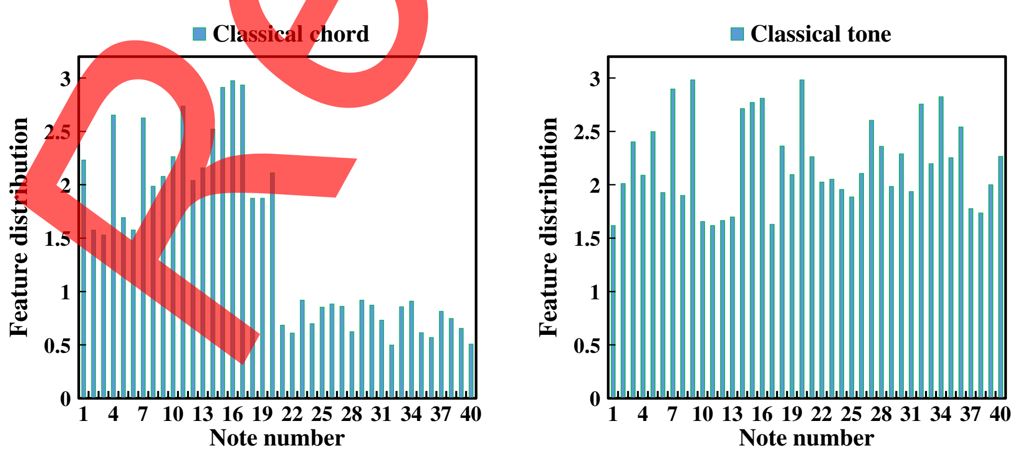

Fig. 5 The distribution of chords and keynotes in the classical music data set.

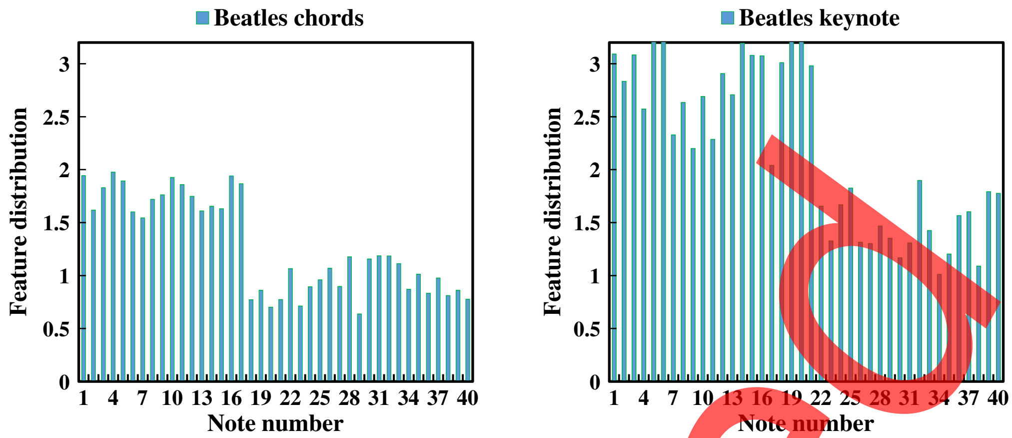

Fig. 6 The distribution of chords and keynotes in the Beatles music data set.

Table 6 Details of the test set.

Timbre	Test music	Keynote	Number of frames
Classical music	Bach Prelude in C Major	C major	478 frames
	Bach Prelude in F Major	F major	353 frames
Beatles music	Anna	C major	1072 frames
	Love Me Do	G major	810 frames

It uses 20 pieces of music (15 classical music, 5 Beatles music) to test the accuracy of keynote estimation. This article accurately estimates the accuracy of 20 of the music, and the accuracy rate is 95%. The accuracy of the pitch estimation of 20 music pieces is shown in Fig. 7.

The postprocessing of multimedia images will project virtual images produced by 3D modeling into the air, creating a virtual scene that looks like an illusion and a real one. There is also a combination of virtual images and touch screens to experience the fun of interaction in the dream space composed of virtual and reality. With the help of multimedia image postprocessing, music visualization can be successfully realized. The principle of multimedia image post-processing is shown in Fig. 8.

The scenery is a place where one must think about during the shooting of video work. It is usually divided into five categories: long-range, panoramic, medium-range, close-up, and

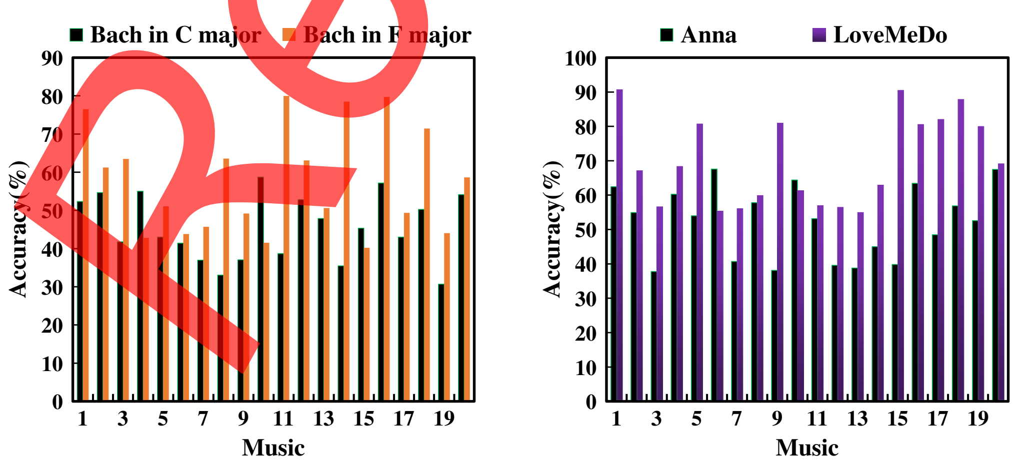

Fig. 7 20 accuracy of pitch estimation for a piece of music.

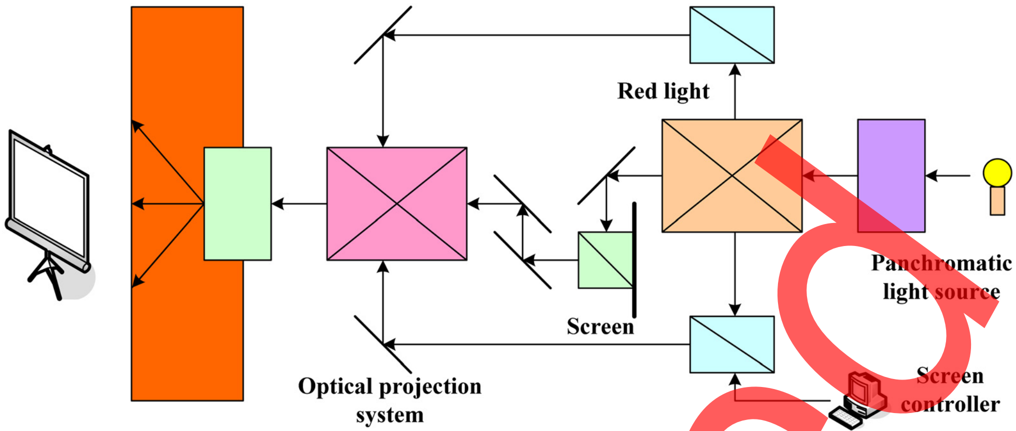

Fig. 8 The principle of multimedia image postprocessing.

Fig. 9 Display of different scenes during the shooting of music images (part of the picture comes from Baidu).

close-up. The classification of these five scenes has its different screen structure. For example, in the image design of Taosi Earth Drum, the introduction of the Loess Plateau, the Yellow River Basin and the environment. It needs to use a long-range or panoramic lens to show the full picture of the environment and also highlights the majestic and magnificence of the Yellow River Basin. The display of different scenes during the shooting of music images is shown in Fig. 9.

5 Discussion

The specific research work of music iconography mainly starts from the history of musical instruments, the history of performance, the life of musicians, and related video materials. It can also be directly applied to the study of stringed music, with the history of stringed instruments being the focus of the study. It searches for information from texts, paintings, bipaintings, and other aspects, combined with the textual verification of music-related characters and events in the history books, gradually establishes the vertical development context of stringed instruments and music history, and gradually enriches it on the cross-section. This has a very important role in promoting the development and maturity of music history. From the aspect of historical research, music culture is regarded as an important aspect, which supplements an important part for enriching the development of string music. Judging from these historical data, the rapid development history of music iconography is traceable. Our research focus in the field of music research can be adjusted and changed promptly as the theoretical data changes in different periods.[27]

According to the available data, the concept of the subject of iconography is not determined all at once. In the long-term development, it gradually emerged from the history of art. The development of Chinese music iconography as an independent subject began in the 20th century. Under the influence of foreign music iconography, it has absorbed the beneficial factors, combined with the actual situation of Chinese music culture investigation and gradually formed its characteristics. Although the subject theory of Chinese music iconology has matured relatively late, the work of using images to explain music culture in the history of Chinese music has already begun.[28]

The history of music visual communication design involves many fields, such as the evolutionary history of music storage media, the history of packaging, the history of poster design, and the history of the development of music media. The main research goal of this paper is the visual communication design language of music. Therefore, the development and evolution of music visual communication design and the form of music communication are not discussed too much.[29]

This research will focus on the aesthetics and dissemination of Chinese music in society under visual culture. It analyzes the current situation, information feedback, and specific aesthetic value of the visual communication of Chinese music from the perspectives of interpretation of the communication form, meaning generation, and power operation. Therefore, the research object set has the characteristics of visual culture, musicology, and communication in connotation. In addition, this characteristic will permeate various musical activities in the medium. It is connected with the audience's music acceptance and social feedback and then outlines the current visual landscape of Chinese music communication.[30]

6 Conclusion

From a disciplinary point of view, Chinese music history is often described as "dumb music history." This is due to the problem of acoustic fracture caused by the loss of a large number of ancient music scores. In this case, sorting out many images of ancient musical instruments in the world so the context of the musical instruments is displayed in front of readers is a way to make up for it. The accuracy and authenticity of music images are often affected by the style and randomness of the painters. How to conduct empirical comprehensive combining of such images is also an aspect that this article attempts to discuss. The classification of music images varies according to different provinces, and there is no clear classification standard. The continuous update and development of science and technology make music iconography adapt to it in terms of the research scope, research methods, and research methods and apply the advantages of multimedia music image technology to the collection, storage, dissemination, and research not only conform to the development principle of things but also will be the trend of future development and make due the contributions to the protection of human cultural heritage.

Acknowledgments

This work was supported by the National Social Science Fund Art Project; Project Number: 2020BD03601. Project category: Normal project; Authorization Number: 20BD053. Project name: The production and development of Russian vocal art in China in 20th. The authors declare no conflict of interest.

References

1. X. Gong et al., "Chmusic: a traditional Chinese music dataset for evaluation of instrument recognition," *ICBDT*, 184–189 (2021).
2. M. Amini, M. O. Ahmad, and M. Swamy, "A new locally optimum watermark detection using vector-based hidden Markov model in wavelet domain," *Signal Process.* **137**(August), 213–222 (2017).
3. B. Rashidi, C. Fung, and E. Bertino, "Android resource usage risk assessment using hidden Markov model and online learning," *Comput. Secur.* **65**(March), 90–107 (2017).

4. Z. Li et al., "Data-driven bearing fault identification using improved hidden Markov model and self-organizing map," *Comput. Ind. Eng.* **116**(February), 37–46 (2017).
5. L. Carrer and L. Bruzzone, "Automatic enhancement and detection of layering in radar sounder data based on a local scale hidden Markov model and the Viterbi algorithm," *IEEE Trans. Geosci. Remote Sens.* **55**(2), 962–977 (2017).
6. J. James, J. J. Ford, and T. L. Molloy, "Change detection for undermodelled processes using mismatched hidden Markov model test filters," *IEEE Control Syst. Lett.* **1**(2), 238–243 (2017).
7. L. Lu et al, "Recognizing human actions by two-level Beta process hidden Markov model," *Multimedia Syst.* **23**(2), 183–194 (2017).
8. F. A. Raposo, D. Martins de Matos, and R. Ribeiro, "Assessing kinetic meaning of music and dance via deep cross-modal retrieval," *Neural Comput. Appl.* **33**, 14481–14493 (2021).
9. K. Li et al., "Method to detect sleep apnea based on deep neural network and hidden Markov model using single-lead ECG signal," *Neurocomputing* **294**(June 14), 94–101 (2018).
10. W. N. Robinson and A. Aria, "Sequential fraud detection for prepaid cards using hidden Markov model divergence," *Expert Syst. Appl.* **91**(January), 235–251 (2018).
11. Y. An, M. Han, and Y. Park, "Identifying dynamic knowledge flow patterns of business method patents with a hidden Markov model," *Scientometrics* **113**(4), 783–802 (2017).
12. X. Li et al., "Exploring the innovation modes and evolution of the cloud-based service using the activity theory on the basis of big data," *Clust. Comput.* **21**(1), 907–922 (2018).
13. M. Zareapoor, D. K. Jain, and J. Yang, "Local spatial information for image super-resolution," *Cognit. Syst. Res.* **52**, 49–57 (2018).
14. T. Chadza, K. G. Kyriakopoulos, and S. Lambotharan, "Analysis of hidden Markov model learning algorithms for the detection and prediction of multi-stage network attacks," *Future Gener. Comput. Syst.* **108**(July), 636–649 (2020).
15. H. Wang, T. Asefa, and A. Sarkar, "A novel non-homogeneous hidden Markov model for simulating and predicting monthly rainfall," *Theor. Appl. Climatol.* **143**(1), 627–638 (2021).
16. D. Acharjee, S. P. Maity, and A. Mukherjee, "Hidden Markov model a tool for recognition of human contexts using sensors of smart mobile phone," *Microsyst. Technol.* **23**(3), 571–582 (2017).
17. T. Petkovic et al., "Human intention estimation based on hidden Markov model motion validation for safe flexible robotized warehouses," *Rob. Comput.-Integr. Manuf.* **57**(June), 182–196 (2019).
18. Z. Tao et al., "A novel method for tool condition monitoring based on long short-term memory and hidden Markov model hybrid framework in high-speed milling Ti-6Al-4V," *Int. J. Adv. Manuf. Technol.* **105**(7-8), 3165–3182 (2019).
19. J. Li, X. Chen, and Z. Li, "Spike detection and spike sorting with a hidden Markov model improves offline decoding of motor cortical recordings," *J. Neural Eng.* **16**(1), 016014 (2019).
20. R. Sinhal et al., "Machine learning based blind color image watermarking scheme for copyright protection," *Pattern Recognit. Lett.* **145**, 171–177 (2021).
21. P. F. Shan, "Image segmentation method based on K-mean algorithm," *EURASIP J. Image Video Process.* **2018**, 81 (2018).
22. Z. Chen et al, "Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy," *Reliab. Eng. Syst. Saf.* **184**(April), 123–136 (2019).
23. R. Amoros et al., "A continuous-time hidden Markov model for cancer surveillance using serum biomarkers with application to hepatocellular carcinoma," *Metron* **77**(2), 67–86 (2019).
24. C. Zheng and W. Cao, "Hybrid high-order semantic graph representation learning for recommendations," *Discov. Internet Things* **1**, 17 (2021).
25. A. K. Ghosh, A. Ullah, and A. Kubo, "Hidden Markov model-based digital twin construction for futuristic manufacturing systems," *Artif. Intell. Eng. Des. Anal. Manuf.* **33**(3), 317–331 (2019).
26. B. Mathieu, A. Crouzil, and J.-B. Puel, "Interactive segmentation: a scalable superpixel-based method," *J. Electron. Imaging* **26**(6), 061606 (2017).

27. E. Yavuz and V. Topuz, "A phoneme-based approach for eliminating out-of-vocabulary problem of Turkish speech recognition using hidden Markov model," *Int. J. Comput. Syst. Sci. Eng.* **33**(6), 429–445 (2018).
28. S. Ghosh, D. Ghosh, and D. K. Mohanta, "Functional uncertainty analysis of phasor measurement unit using fuzzy hidden Markov model," *IETE J. Res.* **64**(1), 100–107 (2018).
29. Y. Zheng et al, "Student's t-hidden Markov model for unsupervised learning using localized feature selection," *IEEE Trans. Circuits Syst. Video Technol.* **28**(10), 2586–2598 (2018).
30. X. Liu et al, "Hidden Markov model based signal characterization for weak light communication," *J. Lightwave Technol.* **36**(9), 1730–1738 (2018).

Jiayue Cui received her doctor's degree in musicology and dancology from Harbin Conservatory of music, China. Now, she works in the College of Master as an associate professor, Harbin Normal University. She has written three academic books and published more than 10 papers in national and provincial core journals, 2020 National Social Science, Art and Science Planning Project first participant. She has completed 11 provincial-level scientific research projects and won the second prize of Heilongjiang Provincial Excellent Achievements in Art and Science Planning.

Hongjun Wang received his mastery of professor, the tutor of doctor, bass singer, the deputy director of first academic committee of Harbin Conservatory of music, and the deputy director of vocal opera department. The 4th member of the CPPCC, the national social artificial science planning project essayist, director of Chinese Music Oral History Research Association, member of Heilongjiang Musician Association, and professor of Lingnan Normal University.