

Retraction

Retracted: Automatic Classification and Analysis of Music Multimedia Combined with Hidden Markov Model

Advances in Multimedia

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Advances in Multimedia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Y. Chen, "Automatic Classification and Analysis of Music Multimedia Combined with Hidden Markov Model," *Advances in Multimedia*, vol. 2021, Article ID 7824001, 7 pages, 2021.

Research Article

Automatic Classification and Analysis of Music Multimedia Combined with Hidden Markov Model

Yanjiao Chen 

School of Music and Dance, Yulin Normal University, Yulin 537000, China

Correspondence should be addressed to Yanjiao Chen; chenyanjiao@ylu.edu.cn

Received 11 August 2021; Accepted 25 October 2021; Published 8 December 2021

Academic Editor: Kang Song

Copyright © 2021 Yanjiao Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Music multimedia is one of the more popular types of digital music. This article is based on the hidden Markov model (HMM) and proposed this kind of music multimedia automatic classification method. The method not only analyzes the characteristics of traditional music in detail but also fully considers the important characteristics of other music. At the same time, it uses bagging to train two groups of HMMs and automatically classifies them to achieve a better classification effect. This paper optimizes the variable parameters from different aspects such as model structure, data form, and model change to obtain the optimal HMM parameter value. This method not only considers the prior knowledge of feature words, word frequency, and number of documents but also fuses the meaning of the feature words into the hidden Markov classification model. Finally, by testing the hidden Markov model used in this paper on the music multimedia data set, the experimental results show that the method in this paper can effectively perform automatic classification according to the melody characteristics of music multimedia.

1. Introduction

With the continuous development of Internet technology and the improvement of the technological level, different Internet-based music multimedia began to emerge. Music multimedia is one of the more popular types of digital music, so its automatic classification and recognition has become the focus of domestic scholars' research [1–3]. However, how to obtain the specific content source from the initial music multimedia data, lacking the definition of the music content, has become a huge challenge for the current automatic classification of music multimedia because the music multimedia signal belongs to a way of time sequence, which can be concealed according to its concealment. Using the hidden Markov model, at present, other classification methods are relatively simple, and the characteristics of music multimedia are not accurate [4–7].

In this paper, the hidden Markov model is applied to the automatic classification process of music multimedia. This method can be based on the characteristics of the automatic lyrics of the music multimedia, and the lyrics, word frequency, content, and meaning of the lyrics of the music

multimedia are used as the current prior knowledge. It is suitable for the automatic classification of music multimedia, classification process. We use the information gain method to obtain the music content of the music multimedia content, use the hidden Markov model to limit the weight of the lyrics and the semantic information of the music, and integrate the lyrics with high meaning similarity to construct various types of music multimedia content. Automatic classification mode, music multimedia, should be classified automatically [8–10].

2. Hidden Markov Model

The hidden Markov model is usually composed of multiple different states. Because of the continuous change of time, the hidden Markov model can stay in a certain state and can also switch between different states. Each observation vector can get different output frequencies for different states. In this paper, a hidden Markov model with four states $S_1 \sim S_4$ is used for computational processing. The input observation sequence is represented by o_1, o_2, \dots, o_T , and the transition probability between each state or between the states is

represented by a_0 , where each observation sequence is used as each frame MFCC parameters. In this model, the observation sequence o_1, o_2, \dots, o_T is used as the observable input sequence, but the state at each time cannot be directly observed [11, 12].

The time complexity and space complexity of music multimedia automatic processing depend on the dimension of music multimedia automatic feature vector. In order to improve the accuracy of model classification, music multimedia needs to be processed automatically in the descending order, and finally, lyrics are selected to display documents. In the information gain, the measurement standard is to see how much information the average information of the class can bring to the classification system. The more information it brings, the more important the average information of the class is. For the average amount of information of a class, the amount of information will change when the system has it or not, and the difference between the amount of information before and after is the amount of information brought to the system by the average amount of information of this class.

The pitch part adopts the pitch period judgment method. When the difference between the pitch of the next frame and the average of the pitches of all previously stored frames is less than the threshold, it is determined that these frames are frames of one category, and the frame is saved until the difference between the pitch of the next frame is greater than the threshold. All these frames are processed as one frame, and the melting spectrum parameter Mel-cestral (MCEP) is calculated to obtain the 13th MCP. The threshold defined in the experiment is 1. The schematic diagram is shown in Figure 1.

In order to maintain the balance of the data in the subsequent spectrum average (CMS) and the correlation spectrum (PASTA) algorithm, we copy the calculated MCEP data in the corresponding 10 ms frame and perform spectral region filtering on it. The first-order and second-order differences are performed on the filtered parameters to obtain 39-order parameters. Finally, the parameters are Gaussized to improve the recognition rate.

The Markov model uses the Baum-Welch algorithm to solve the Markov parameter estimation problem and solve the hidden Markov training problem. Usually, when using a given sequence of observations $O = o_1, o_2, \dots, o_T$, this method uses the determined $\lambda = (A, B, \pi)$ parameter to ensure that the value of $P(O|\lambda)$ can reach the maximum value.

According to the related definitions of forward probability and backward probability, we know

$$P(O|\lambda) = \sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad 1 \leq t \leq T-1. \quad (1)$$

When using $P(O|\lambda)$ to reach the maximum value, since the training sequence of each experiment is limited, the best method cannot be realized for estimating the parameters. In

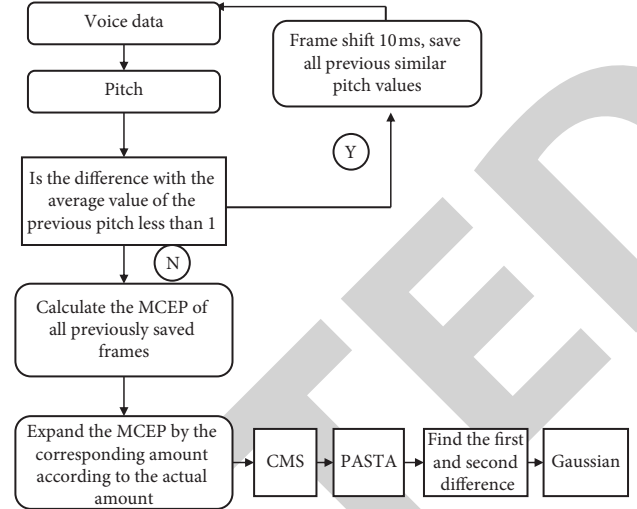


FIGURE 1: Pitch classification and recognition model.

this case, the Baum-Welch algorithm uses $P(O|\lambda)$ with a recursive idea, the part is very large, and finally, the model parameter $\lambda = (A, B, \pi)$ is obtained.

The revaluation formula of the Baum-Welch algorithm is derived by recursion as

$$\bar{\pi}_i = \xi_1(i),$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \xi_t(i)}, \quad (2)$$

$$\bar{b}_{jk} = \frac{\sum_{t=1}^T \sum_{O_t=V_k} \xi_t(j)}{\sum_{t=1}^{T-1} \xi_t(j)}.$$

Among them, $\xi_t(i, j)$ represents the given training sequence O and the model parameter λ , the probability that the Markov chain is in the θ_i state at time t , and the θ_j state at time $t+1$ and $\sum_{t=1}^{T-1} \xi_t(i)$ represents the expected value of the number of transitions from the state θ_i to the state θ_j .

Define the auxiliary function as

$$Q(\lambda, \bar{\lambda}) = \frac{1}{P(O|\lambda)} \sum_{\text{all } S} P(O, S|\lambda) \ln P(O, S|\bar{\lambda}). \quad (3)$$

Among them, λ is the original model parameter $\lambda = (A, B, \pi)$, $\bar{\lambda}$ is the model parameter $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ to be solved, O is the observation sequence used for training and $O = o_1, o_2, \dots, o_T$, and S are the state sequence $S = q_1, q_2, \dots, q_T$.

The Markov model can not only find a good enough state transition path but also quickly calculate the output probability corresponding to the path. At the same time, the amount of calculation required by the method of using the Markov model to calculate the output probability is much less than that in the total probability formula.

Define $\delta_t(i)$ as the maximum probability of o_1, o_2, \dots, o_t along a path q_1, q_2, \dots, q_t at time t and $q_1 = \theta_i$, which is

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_t, q_1 = \theta_i, o_1, o_2, \dots, o_t | \lambda). \quad (4)$$

The recursive form of the hidden Markov model is as follows:

(1) Initialization:

$$\begin{aligned} \delta_t(i) &= \pi_i b_i(o_1), \quad 1 \leq i \leq N, \\ \psi_1(i) &= 0. \end{aligned} \quad (5)$$

(2) Recursion:

$$\begin{aligned} \delta_t(j) &= \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t), \quad 2 \leq t \leq T, 1 \leq j \leq N, \\ \psi_t(j) &= \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T, 1 \leq j \leq N. \end{aligned} \quad (6)$$

(3) End:

$$P^* = \max_{1 \leq i \leq N} \delta_T(i), \quad (7)$$

$$q_T^* = \arg \max_{1 \leq i \leq N} \delta_T(i). \quad (8)$$

(4) Find the state sequence:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad 1 \leq t \leq T-1. \quad (9)$$

Among them, $\delta_t(i)$ represents the probability of accumulating the output value of the i th state at time t , $\psi_t(i)$ represents the continuous state parameter of the i th state at time t , q_t^* is the state at time t in the optimal state sequence, and P^* is the final output probability.

3. Feature Extraction

3.1. Timbre Feature Extraction. For audio with similar pitch and melody, timbre characteristics can be used to distinguish. The feature of timbre is a short-term concept. In order to extract the feature of timbre, the frame is firstly divided [13, 14]. The original audio signal is divided into fixed-length frames, and there is a certain overlap between frames. In order to make the signal smoother, the frame needs to be turned on. The frame length used by this system is 524 (sampling rate 50 Hz about 25 ms), overlapping 26 points, and each frame adds a Hamming window to reduce the influence of edges. After dividing the frames, the fast Fourier can obtain the frequency spectrum of the signal of each frame and can also extract the spectral features [15].

3.1.1. Skewness and Kurtosis. In addition to the mean and variance, common statistics include skewness and kurtosis. Skewness is a measure of the asymmetry of a signal and can represent the relative tendency of tonal and nontonal components in the band. For a sequence x of length N , the mean is $\sum_{n=1}^N (x_n - u)^3$ and the standard deviation is σ . The mathematical expression of skewness is as follows:

$$\text{skewness} = \frac{\sum_{n=1}^N (x_n - u)^3}{(N-1)\sigma^3}. \quad (10)$$

It measures the peak state of the probability distribution of a real random variable. The peak height of the dataset distribution means that there are individual peaks near the average value. The peak can plot the effective dynamic range of the spectrum. The peak formula is as follows:

$$\text{Kurtosis} = \frac{\sum_{n=1}^N (x_n - u)^4}{(N-1)\sigma^4} - 3. \quad (11)$$

3.1.2. Spectral Centroid. The center of the spectrum is the center of gravity of the STFT amplitude spectrum. Determine the most concentrated point of spectral energy, which is related to the most dominant frequency in the signal. The mathematical performance of the spectral center () is as follows:

$$c_t = \frac{\sum_{n=1}^N M_t[n] * n}{\sum_{n=1}^N M_t[n]}, \quad (12)$$

where $M_t[n]$ represents the n th value in the amplitude spectrum of the t th frame signal.

3.1.3. Spectral Flux. The spectral flux F_t is defined as the second moment of the difference between the spectra of two adjacent frames, which can be used to measure the speed of the energy spectrum change and determine the timbre of the audio signal. The mathematical performance of spectral flux () is as follows:

$$F_t = \sum_{n=1}^N (N_t[n] - N_{t-1}[n])^2, \quad (13)$$

where $N_t[n]$ and $N_{t-1}[n]$, respectively, represent the n th value in the normalized amplitude spectrum of the t th frame and the $t-1$ th frame signal.

3.1.4. Spectral Rolloff. Spectral rolloff R_t is defined as the spectrum that measures the bandwidth of the audio signal and concentrates 79% of the energy on the first point less than the frequency:

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^N M_t[n], \quad (14)$$

where $M_t[n]$ represents the n th value in the amplitude spectrum of the t th frame signal.

3.1.5. Spectral Spread. This concept is mentioned in Lench's "Audio Content Analysis" [14, 15]. Spectrum spread can be referred to as a temporary bandwidth and describes the concentration of the energy spectrum near the center of the spectrum. Spectral propagation can be regarded as the standard deviation of the energy spectrum near the center of the spectrum. It is defined as follows:

$$\text{spread} = \sqrt{\frac{\sum_{n=1}^N (n - c_t)^2 \cdot |M_t[n]|^2}{\sum_{n=1}^N |M_t[n]|^2}}, \quad (15)$$

where C_t is the center of the spectrum and the amplitude spectrum M_t of the signal.

3.1.6. Spectral Flatness. The flatness of the spectrum is also used to describe the shape of the spectrum and is provided to quantify the melody of the sound. In the formula of spectral flatness (F),

$$F = \frac{\exp\left(\frac{1}{N} \sum_{n=1}^N \ln M_t[n]\right)}{\left(\frac{1}{N} \sum_{n=1}^N M_t[n]\right)}, \quad (16)$$

where M_t is the amplitude spectrum of the signal.

3.1.7. Zero Crossing. Measure the number of times the audio signal passes the zero value in a given time interval Z_t . For natural noise, this value is random. Therefore, for silent segments, since the zero-crossing rate is higher than that of voiced segments, it is often used to distinguish whether the audio has sound. The zero-crossing rate (Z_t) formula is as follows:

$$Z_t = \frac{1}{2} \sum_{n=1}^N |\text{sgn}(x[n]) - \text{sgn}(x[n-1])|, \quad (17)$$

where $\text{sgn}(x)$ is a symbolic function.

3.1.8. Mel Frequency Cepstrum Coefficient (MFCC). MFCC is triggered by human auditory characteristics and is based on the audio characteristics of STFT. According to the characteristics of human hearing, the perception of human ears becomes a logarithmic relationship with the actual changes in the size and amplitude of the sound. After taking the logarithm of the amplitude spectrum, the FFT coefficients are divided into multiple frequency bands and smoothed according to the Mel frequency. Since the Mel spectrum vectors obtained in this way are highly correlated, in order to eliminate their correlation, transformation is required, and DCT is used for processing here. The overall MFCC extraction process is shown in Figure 2.

Since the first coefficient of MFCC represents the DC component in the signal, it is often truncated in practical applications. MFCC has the differential component AMFCC again and twice and has a certain meaning. As a result of the experimental test, the system finally extracted 12 times (with the 0th coefficient removed) MFCC and its 1st difference component as the feature vector.

3.2. Melody Feature Extraction. The melody of music multimedia can see that the melody of music multimedia changes with time. The characteristics of melody are mainly composed of the characteristics, rhythm, and speed of music multimedia. Using the distribution histogram, the characteristics of the music multimedia melody can be extracted,

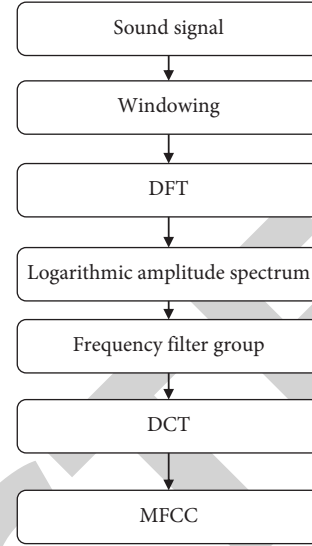


FIGURE 2: Extracting MFCC from audio.

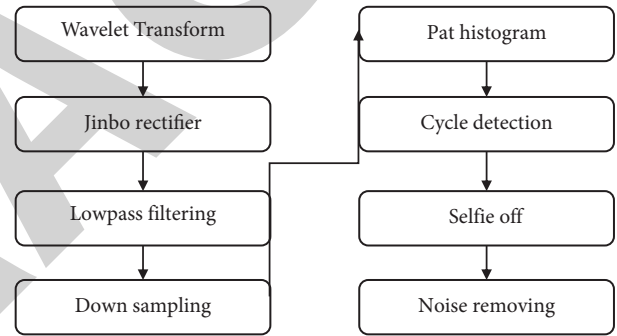


FIGURE 3: The calculation of the histogram.

that is, the distribution histogram can be obtained according to the wavelet decomposition of the signal, and at the same time, the relevant high-pass and low-pass filtering can be performed on the signal in the domain.

According to the histogram, 6 different types of eigenvalues can be obtained. From the first peak in the histogram, it can be expressed according to the period of the peak (using bpm per minute). The sum of the ratio corresponds to the width of the first two peaks and the overall amplitude of the histogram. The operation of the distribution histogram is shown in Figure 3.

4. Automatic Classification of Music Multimedia

4.1. HMM Training. Since only the observation sequence O of the HMM can be obtained in advance, it is considered how to use the sequence to estimate the parameter input of the HMM. Generally, based on the maximum likelihood criterion, the Baum-Welch algorithm can be used to find the local optimization [13]. From the obtained observation sequence, the Kmeans clustering algorithm is used to cluster them into M classes%. After the initial parameters are input,

the Baum–Welch algorithm outputs a set of new parameters A , and the observation sequence, $O = o_1 o_2 \cdots o_T$:

$$P(O|\lambda) \geq P(O|\lambda). \quad (18)$$

The calculation method is the forward-backward algorithm. $P(O|\lambda)$, for the HMM parameter λ and state i , defines the forward probability $\alpha_t(i)$:

$$\alpha_t(i) = P(o_1 o_2 \cdots o_T, q_t = i | \lambda). \quad (19)$$

$\alpha_t(i)$ is the probability ($o_1 o_2 \cdots o_t$) that the parameter A generates a sequence and the o_t state is at time t .

$P(O|\lambda)$ can be calculated by the following forward algorithm.

(1) Initialization:

$$\alpha_1(j) = b_j(o_1), \quad 1 \leq j \leq N. \quad (20)$$

(2) Recursion:

$$\alpha_t(i) = b_i(o_t) \left[\sum_{j=1}^N \alpha_{t-1}(j) \alpha_{ji} \right], \quad 1 \leq t \leq T, 1 \leq i \leq N. \quad (21)$$

(3) Termination:

$$P(O|\lambda) = \sum_{j=1}^N \alpha_T(j). \quad (22)$$

Define $\beta_t(i)$ and $\xi_t(i, j)$ call at the same time:

$$\beta_t(i) = P(o_{t+1} o_{t+2} \cdots o_T, q_t = i | \lambda), \quad (23)$$

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda). \quad (24)$$

The calculation method $\xi_t(i, j)$ from forward to backward can be expressed as [2]

$$\begin{aligned} \xi_t(i, j) &= \frac{P(q_t = i, q_{t+1} = j | O, \lambda)}{P(O|\lambda)} = \frac{\alpha_t(i) \alpha_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i) \alpha_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) \alpha_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}. \end{aligned} \quad (25)$$

The probability that the system is in the m th mixed component $\gamma_t(i, m)$ of state i at time t is

$$\gamma_t(i, m) = \left[\frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)} \right] \left[\frac{\mu_{j,m} N(o_t, \mu_{j,m}, \Sigma_{j,m})}{\sum_{m=1}^M \omega_{j,m} N(o_t, \mu_{j,m}, \Sigma_{j,m})} \right]. \quad (26)$$

4.2. Automatic Classification Using HMM. First, train various types of HMMs using labeled training kits. Set the type to be automatically classified, $k = \{1, 2, \dots, K\}$. The various types are for model parameters, λ_k . Use the maximum likelihood criterion to maximize the posterior probability.

At the same time, the Beggs formula must be considered. We have

$$\bar{k} = \arg \max_{1 \leq k \leq K} P(O|\lambda_k) = \arg \max_{1 \leq k \leq K} \frac{P(O|\lambda_k) P(\lambda_k)}{P(O)}. \quad (27)$$

Assuming that the prior probabilities of all types $P(\lambda_k)$ are the same and $P(O)$ is independent of k , the criterion is simply

$$\bar{k} = \arg \max_{1 \leq k \leq K} P(O|\lambda_k). \quad (28)$$

Due to its small size, overflow will occur below floating point during computer operations, $P(O|\lambda_k)$. Therefore, the logarithm is usually taken. For multiple observation sequences, we only need to weigh each formula [13].

5. Construction of Automatic Classification System of Music Multimedia

5.1. Audio Feature Extraction. The processing of timbre features divides the input signal into 512 points in length through preprocessing and superimposes them on a frame of 256 points and then performs spectral energy, spectral center, spectral beam, spectral rolloff, spectral spread, and spectral flatness on each analysis window. Due to the zero-crossing rate and the total 31-dimensional vectors of the first 12MFCC and AMFCC, the average value and dispersion of the feature vector of each analysis window are obtained in the structure window. According to the low energy value obtained by the calculation, it is necessary to perform the window value of different structures. We extract the dimensional feature vector.

5.2. Automatic Classifier Training. For the adopted music multimedia feature system, it is necessary to extract music feature values of different types of music multimedia and optimize the hidden Markov models obtained under different types of training according to the feature vectors used. Determine the structural characteristics of the initial state and hidden Markov model in different states, and connect all the connections with the left-right structure. In the process of automatic classification of music multimedia, it is necessary to perform TC sequence operations in the audio feature extraction process to calculate the probability values of observation sequences composed of different types of HMM models, and at the same time, we perform automatic classification into the type with higher probability.

In a system that effectively combines TC and RC at the same time, according to different types of music multimedia types, the music multimedia features are extracted from timbre and rhythm features, and different types of feature values are used in turn for corresponding classification training. In the automatic classification process, it is necessary to input the TC and RC sequences corresponding to the audio at the same time, calculate the sum of the probabilities of the observation sequences formed by the HMM under different types, and sort them according to the level. The structure of the system is shown in Figure 4.

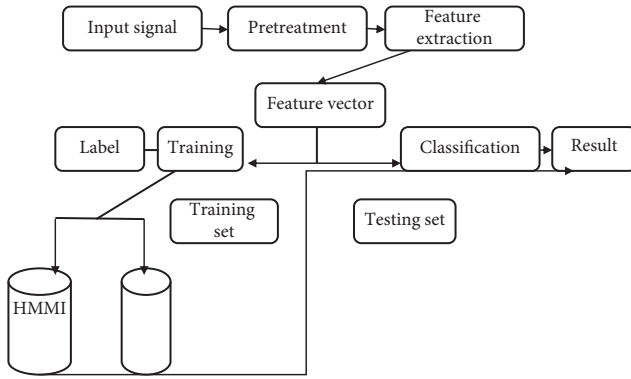


FIGURE 4: Structure diagram of audio automatic classification system.

6. Test Results and Analysis

6.1. Automatic Classification System Based on Timbre Characteristics. Automatic classification considering only music feature values is a traditional method. Considering that the state of the hidden Markov model is set to 4, 5, and 6 states, the calculated Gaussian mixture distribution values are 1 and 2 in turn, taking into account at the same time to all the connections and the actual application of the left-right model. Table 1 shows the relationship between the automatic classification accuracy and the number of HMM states calculated by the HMM model under different types and the relationship between the two and the number of mixed Gaussian distributions. For the different state values in the table, the automatic classification accuracy is only 1%. It can be seen that there is little correlation between the automatic classification accuracy and the number of states. For the mixed Gaussian distribution value, the intermediate value of 2 can achieve better results, and the accuracy can also be increased to 62.2%, mainly due to the inability to accurately calculate the HMM characteristics under the Gaussian distribution, but the components greater than three or more lack correlation training samples. The result obtained under the HMM model is left-right better, but the accuracy of the related connection model in this system is relatively high.

In the actual category of the first element table of the column, the column element represents the frequency of being classified into the corresponding category of the row. Table 2 is the result of specific automatic classification confusion in the best case, that is, HMM is connected to the automatic classification result. As a result, for music multimedia with relatively clear features such as Classical, Hip-hop, Jazz, Metal, and Pop, HMM can achieve higher accuracy ($\geq 68\%$), and the accuracy of classical reaches 88%. The features of Blues, Reggae, Rock, etc., are not clear, so the accuracy of automatic classification is very low. In addition, it should be noted that a lot of music multimedia that is not a rock genre is classified as a rock genre. The reason for this result may be due to the extraction of sound features that cannot distinguish these types of music multimedia well. The experiment shows that only 70% of people have not been trained to automatically classify music multimedia.

TABLE 1: Automatic classification results of different HMM structures (features of timbre only).

NC	NS		
	4	5	6
(a) Fully connected (%)			
1	60.2	61.3	63.2
2	63.6	64.4	63.3
3	60.3	59.1	58.3
(b) Left-right (%)			
1	59.1	59.8	57.2
2	61.3	63.4	58.8
3	58.4	57.4	51.8

Note. NS represents the number of states of the HMM and NC represents the number of output components of the mixed Gaussian.

TABLE 2: Confusion matrix of GTZAN dataset (using timbre characteristics).

	Bl	Cl	Co	Di	Hi	Ja	Me	Po	Re	Ro
Bl	50	1	9	4	4	5	5	1	6	7
Cl	5	90	2	1	1	7	1	1	1	1
Co	19	2	59	3	2	3	1	3	7	20
Di	4	2	5	56	9	5	6	7	15	15
Hi	1	1	1	5	71	1	2	3	18	4
Ja	4	9	2	1	3	72	1	1	3	4
Me	6	1	7	4	2	5	77	2	2	9
Po	1	1	3	13	8	4	2	78	4	5
Re	1	0	0	2	4	1	0	0	46	2
Ro	18	2	21	20	5	6	14	13	7	42

TABLE 3: Automatic classification results under different HMM structures.

NC	NS		
	4	5	6
(a) Fullyconnected (%)			
1	63.4	58.6	61.8
2	58.6	62.6	63.8
3	62	60.4	59.2
(b) Left-right (%)			
1	63.2	58.4	61.6
2	58.5	62.5	63.7
3	61.5	59.9	58.7

Note. NS represents the number of states of the HMM and NC represents the number of output components of the mixed Gaussian.

6.2. Automatic Classification System Based on Timbre and Melody Characteristics. The sound quality and melody characteristics are automatically classified in the HMM model as a classification method. Set the number of states under the corresponding timbre characteristic HMM model to 5, and output all connected models with a mixed Gaussian distribution value of 2. The number of states of melody characteristics and the number of Gaussian mixture distributions need to be reconfirmed. The test results obtained are shown in Table 3. It can still be found that the effect of all the coupled models is better. For a mixture of Gaussian

TABLE 4: Comparison of accuracy of different automatic classification algorithms.

Classifier	Classification accuracy (%)
LPNTF	93.6
Visual	91
ELM	86.35
OSC + SVM	85.29
GSV + Gabor filters	83.3
DWCH + SVM	79.7
HMM	68
HMM	64.4
GMM and KNN	62.2

distributions with a state number of 6, the accuracy of the 4 model reaches 66.8%. The accuracy here is higher than just considering the tone.

Table 4 is a comparison of the accuracy of various automatic classification algorithms. Compared with the latest research results, the accuracy of this thesis system is still far away.

7. Conclusions

This paper uses the hidden Markov model to automatically classify music multimedia and improves the accuracy of classification by obtaining the optimal parameter values of the HMM model. Finally, an example analysis result shows that the introduction of the hidden Markov model into the independent classification of music multimedia can achieve accurate and rapid classification. At the same time, combined with the melody characteristics of the music multimedia, the accuracy of the music multimedia classification of the HMM model can be increased to 67.9%. Therefore, the independent classification based on the characteristics of music multimedia is mainly because it is easier to distinguish different types of samples according to the characteristics of music multimedia.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest

References

- [1] D. Deng, X. Li, C. Liu et al., "Systematic investigation on the turning point of over-inflammation to immunosuppression in clp mice model and their characteristics," *International Immunopharmacology*, vol. 42, pp. 49–58, 2017.
- [2] W. Jiao and Z. Xu, "A distributed runoff model for the mountainous region of the heihe river basin (China) based on the spatial modeling environment (sme) i: model structure and equations," *Environmental Earth Sciences*, vol. 68, no. 3, pp. 881–888, 2013.
- [3] G. K. Birajdar and M. D. Patil, "Speech/music classification using visual and spectral chromagram features," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, 2020.
- [4] M. Teräs, "Challenges of intercultural vocational education and training: developing a strand model in the change laboratory," *Astronomische Nachrichten*, vol. 328, no. 10, pp. 1066–1069, 2014.
- [5] L. Xia, M. U. Farooq, and I. M. Bell, "High level fault modeling of analog circuits through automated model generation using Chebyshev and Newton interpolating polynomials," *Analog Integrated Circuits and Signal Processing*, vol. 82, no. 1, pp. 265–283, 2015.
- [6] L. A. Liikkanen and P. Åman, "Shuffling services: current trends in interacting with digital music," *Interacting with Computers*, vol. 28, no. 3, pp. 352–371, 2016.
- [7] M. Simoni, "Tara rodgers: pink noises: women on electronic music and sound," *Computer Music Journal*, vol. 35, no. 3, pp. 102–103, 2011.
- [8] A. Srivastava, O. Mangla, and V. Gupta, "Study of la-incorporated hfo2 mim structure fabricated using pld system for analog/mixed signal applications," *IEEE Transactions on Nanotechnology*, vol. 14, no. 4, pp. 612–618, 2015.
- [9] A. S. Karar and J. C. Cartledge, "Generation and detection of a 56 gb/s signal using a dml and half-cycle 16-qam nyquist-scm," *IEEE Photonics Technology Letters*, vol. 25, no. 8, pp. 757–760, 2013.
- [10] P. Doornbusch, "Gerhard nierhaus: algorithmic composition: paradigms of automated music generation," *Computer Music Journal*, vol. 34, no. 3, pp. 70–74, 2010.
- [11] Y. Ma, "Research on the arrangement and visual design of aerobics under the new situation," *International Core Journal of Engineering*, vol. 5, no. 9, pp. 170–173, 2019.
- [12] J. L. Martin, "Writing about music: the selection and arrangement of notation in jazz students' written texts," *British Journal of Music Education*, vol. 35, no. 1, pp. 73–89, 2017.
- [13] Q. Q. Zhao and Q. Qi, "The application of computer multimedia technology in aerobics music orchestration," *Applied Mechanics and Materials*, vol. 651–653, pp. 1958–1961, 2014.
- [14] J. H. Foleiss and T. F. Tavares, "Texture selection for automatic music genre classification," *Applied Soft Computing*, vol. 89, no. C, 2020.
- [15] H. C. Ceylan, N. Hardalaç, A. C. Kara, and F. Hardalaç, "Automatic music genre classification and its relation with music education," *World Journal of Education*, vol. 11, no. 2, p. 36, 2021.