

Retraction

Retracted: Transformation of Nonmultiple Cluster Music Cyclic Shift Topology to Music Performance Style

Complexity

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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) Manipulated or compromised peer review

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] J. Li, "Transformation of Nonmultiple Cluster Music Cyclic Shift Topology to Music Performance Style," *Complexity*, vol. 2021, Article ID 5590503, 11 pages, 2021.

Research Article

Transformation of Nonmultiple Cluster Music Cyclic Shift Topology to Music Performance Style

Jing Li 

Institute of Education, Chongqing University of Arts and Sciences, Chongqing 402160, China

Correspondence should be addressed to Jing Li; 20030010@cqwu.edu.cn

Received 2 March 2021; Revised 2 April 2021; Accepted 15 April 2021; Published 26 April 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Jing Li. 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 is an abstract art form that uses sound as its means of expression. It has deeply affected our lives. This paper proposes a method for extracting segment features from nonmultiple cluster music files. We divide each piece of music into multiple segments and extract the features of each segment. The specific process includes nonmultiple cluster music file note extraction, main melody extraction, segment division, and segment feature extraction. The segment feature is extracted from a segment of a piece of music, contains the main melody and accompaniment information of the segment, and can reflect the sequence relationship of the notes. This paper proposes a performance style conversion network based on recurrent neural network and convolutional neural network. The bidirectional recurrent neural network based on Gated Recurrent Unit (GRU) is used to extract different styles of note feature vector sequences, and the extracted note feature vector sequence is used to predict the intensity of a specific style, and the intensity changes of different styles of nonmultiple cluster music are better learned. Through the comparison, the multiclassification strategy of “one-to-the-rest” is selected, and the fuzzy recurrent neural network is applied to the shortcomings of the unrecognizable area. Finally, according to the feature extraction method and the principle of the classifier algorithm studied in this paper, a music style classification system is implemented in the MATLAB environment. Experimental simulation shows that this system can effectively classify music performance styles.

1. Introduction

Music is one of the oldest, most universal and infectious art forms of mankind. It is a special language for humans to express their thoughts and feelings and realize mutual communication through the harmonious and orderly arrangement and combination of various sounds [1]. The creation, expression, understanding, and appreciation of music are one of the most basic spiritual activities of mankind. As the most important carrier of human culture, music has rich cultural and historical connotations, so it has been passed on for thousands of years and still occupies an indispensable position in human life [2]. In the new era, its meaning, form of existence, and means of communication have been given new interpretations with the rapid development of high technology. For the database manager itself, the tags available for retrieval are obtained through manual annotation, and the accuracy of many tags is also

compromised due to unavoidable negligence or subjective judgment of the annotator [3]. With the continuous expansion of music databases and the continuous changes in user needs, such problems will only become increasingly serious. The same problem also exists in music data in other organizational forms, such as a large number of scattered music resources scattered on the Internet. The fundamental reason why music resource providers cannot provide precise search and personalized labels lies in their inability to find effective methods to intelligently and automatically analyze and process the content of music fragments and various related attributes [4].

Effective analysis of music elements is of great significance to music information retrieval technology and music teaching and creation intelligence. For the field of music information retrieval, building as many indexes as possible from different angles is a key prerequisite for effective retrieval of massive music information; at the same time,

measuring the similarity of different music from multiple angles is the core issue of sample retrieval. Among them, global elements such as mode and style can provide an index for music to be retrieved according to different user requirements and can enrich massive database management methods [5]. Global elements such as structure and melody can provide a basis for the measurement of similarity between music segments on the structural characteristics of music and can also provide important features for the realization of humming retrieval [6]. For the field of intelligent music teaching and creation, the recognition of musical elements such as notes, chords, modes, and melody is a necessary part of the automatic translation of full scores. The automatic translation of music scores can directly convert music into music scores. It can be used in many fields such as music teaching visualization, music sight-singing, and ear training, and it has a huge role in improving the autonomy, efficiency, and intelligence of music teaching [7]. At the same time, music score translation is also an important auxiliary technology to improve the efficiency and intelligence of music creation. It can greatly simplify the process of music creation. The composer can complete the process of composing and modifying by playing or humming, changing the current situation. The traditional creation mode of playing and noting and modifying makes it possible to popularize music creation [8].

In the process of extracting the note matrix, this paper studies the format of the nonmultiple cluster music file and the events related to the note information in it and designs a note information extraction scheme based on this. Specifically, the main contributions of this article can be summarized as follows:

- (i) First, in the process of extracting the main melody, this paper studies the Skyline algorithm, and on this basis, it combines the previous research results to propose an MTC main melody extraction algorithm. In the process of segment division, this paper studies and implements a segmentation algorithm based on energy feature vector, which divides the whole piece of music into multiple segments.
- (ii) Second, in the process of extracting the features of the music segment, this paper proposes an algorithm for extracting features using the sample code of the music note set to extract the features of the music segment. The section features include the main melody and accompaniment information of the section and can reflect the sequence of notes. This article introduces the one-dimensional convolutional neural network for processing sequence data, analyzes the recurrent neural network that is more widely used in sequence processing, and describes the structure and training method of the recurrent neural network.
- (iii) Third, this article uses recurrent neural networks and one-dimensional convolutional neural networks to build a performance style conversion network and

a dynamic classifier and discusses the processing of the music database for training and testing. According to the components of the music classification system, the system implemented in the MATLAB environment according to the training module and the test module process is introduced. In the analysis of the experimental process, the classification experiments were carried out on the existing test music library on the effectiveness of introducing the fuzzy membership function into the classifier and the stability of the classification system.

2. Related Work

The melody of music is formed by the music of different pitches arranged horizontally in an orderly manner according to a certain rhythm. It runs through each section and is one of the most basic and most important means of expression in a complete musical form [9]. At the same time, it is also the attribute that gives people the most intuitive impression among the many musical elements. Classical literature points out that the existence of melody allows the listener to distinguish two different musical works [10]. The melody is the part that the listener expresses when he resings the musical work. The melody is the part of the musical work that we can remember and forget for a long time. Most of the tunes of this work may still be recalled. It can be seen that finding the melody part of a work is crucial for analyzing the entire musical work, so melody extraction has always been a focus issue in the field of music signal processing. In the narrow sense of music theory, melody is generally defined as a sequence of musical notes composed of single notes. Therefore, in many studies, a method of finding the most prominent vocal note at each moment to form the melody part is used, which is called melody line estimation [11, 12]. Relevant scholars have proposed a method of using recurrent neural network for classification and recognition to detect melody lines [13]. In this method, each classifier corresponds to a semitone represented by F0, and the classifier group is trained on the melody-labeled polyphonic music audio data, and then the trained classifier group is used to identify whether there is a corresponding F0. The semitone appears, and the final melody line is smoothed by HMM.

The mode of music is a very important core concept of the tonal system in music theory. It usually refers to a number of different high and low tones, organized around a stable central tone (tonic) according to a certain mutual relationship. A system is formed; this system is the mode. Each musical work has its own defined mode, which reflects the organization of musical scales and the types of chords that may appear. At the same time, the importance of mode is also reflected in its influence on human perception of music. Each different mode has a different sense of hearing, and this difference in sense of hearing in harmony structure is exactly the most important element attribute of music on a long-term scale [14].

Researchers have proposed a music digest algorithm based on PCP feature representation, and earlier they proposed an algorithm for extracting chorus in music, which has achieved good results on a popular music library with a typical music structure [15]. Relevant scholars have researched and improved this method [16]. By analyzing the relationship between various types of repetitive sections, they can find the repetitive sections in the music and then identify the repetitive sections that have transitions. On the self-built test set, 80% of the choruses have been detected. However, this work is limited to the structural analysis of the chorus. Relevant scholars use key phrase extraction technology to study music abstracts [17]. By extracting segment features frame by frame and clustering them, the results are used to train the various states of the target music to go through the HMM to discover the song structure and then to determine the key segment based on semantics. This method is better than HMM for finding key passages on the 18 Beatles' popular music test sets. Related scholars constructed a frame-by-frame comparison matrix to measure the degree of similarity between music segments and finally used music processing methods to process the subdiagonal lines of the matrix to find similar segments [18].

Music style is a rather subjective concept. It is a global label created by humans to classify and describe music with different listening sensations. Because it is related to many factors such as culture and historical stage, it has no strict definition or classification boundary. But what is certain is that from the current general music style classification method, the same style of music must have some similar musical elements, such as rhythm, mode, structure, and so on. Therefore, music style classification is a macroelement analysis technique related to many musical elements. At present, the mainstream style classification method is to extract timbre features, rhythm features, and pitch content features from music audio and then jointly input the classifiers to obtain the style classification. Related scholars have proposed similar methods on the same database, using rhythm patterns and additional information features derived from them for classification [19, 20].

3. Nonmultiple Cluster Music File Feature Extraction

3.1. Nonmultiplex Music Track Block Composition. The identifier of the track block is "MTrk." In the header block of a nonmultiplex music file, the second parameter defines the number of track blocks in the file. Normally, the first track block after the header block of a nonmultiplex music file records some global information of the file, such as tempo, beat, key signature, etc. If there is only one track block in the entire file, then the global information is followed by the nonmultiplex cluster music time of the track. If there are multiple track blocks, the first track block records the global information [21].

Nonmultiple cluster music events, also called non-multiple cluster music messages, usually consist of a status byte followed by multiple data bytes. In the status byte, the highest bit is always 1, and the lower 4bits are used to

indicate which channel this nonmultiplex music message belongs to. These 4bits are used to indicate 16 possible channels. The other 3 bits are used to indicate the type of this nonmultiple cluster music event. Among them, the non-multiplex cluster music events that this article focuses on include two events: note-on event and note-off event [22]. The note-on event means that the synthesizer begins to play the specified note. The first 4 bits of the status byte are 1001, and the next two bytes represent the note number and keystroke, respectively. The most significant bits are all 0. The note-off event means that the synthesizer stops playing the specified note. The first 4 bits of the status byte are 1000, and the next two bytes represent the note number and key release velocity, respectively. It should be noted that in addition to the note-off event, there is another event that allows the synthesizer to stop playing the note being played, that is, the note-on event with a key velocity of 0. There is no difference in music between these two methods, but when describing a note, the latter saves more bytes. Nonmultiplex cluster music stipulates that when the status bytes of the preceding and following events are the same, the latter status bytes can be omitted.

3.2. The Main Melody Extraction Algorithm. The note matrix of the song will contain the note information in the non-multiplex music file [23–27]. In music, it often contains the main melody and accompaniment melody. The main melody is the main line of the music. It determines the tonality, format, and progress of the music. It is the soul of the music and an important factor in determining the style of music. In order to integrate the main melody information in the features, we need to find the main melody in the nonmultiple cluster music file. The main content of this section is to study the main melody extraction algorithm of the nonmultiple cluster music file.

3.2.1. Skyline Main Melody Extraction Algorithm. We traverse all the notes in the note matrix. For the notes that constitute the polyphony relationship, we only keep the notes with the highest pitch and delete the remaining notes. The polyphonic relationship here means that two notes meet the following conditions:

$$\begin{cases} \text{StartTime}^1 - \text{Duration}^1 > |\text{StartTime}^2 - \text{Duration}^2| \\ \text{StartTime}^1 - \text{StartTime}^2 < 0 \end{cases} \quad (1)$$

In the generated note array, we arrange them in descending order of the starting time. For adjacent notes, the following is satisfied:

$$\begin{cases} \text{Pitch}^1 - \text{Pitch}^2 < 0 \\ \text{StartTime}^1 - \text{StartTime}^2 + \text{Duration}^1 > 0 \end{cases} \quad (2)$$

Then, we make

$$\text{Duration}^1 = |\text{StartTime}^1 - \text{StartTime}^2|. \quad (3)$$

It can be seen from the above steps that the basic idea of the Skyline algorithm is to select the note with the highest pitch when it encounters notes that are played at the same time and discard the remaining notes. In real life, the pitch of the main melody of the music is often higher than the accompaniment melody, so the Skyline algorithm can easily and effectively extract the main melody of the music in most cases. However, the Skyline algorithm still has the following disadvantages:

- (1) If the main melody is temporarily stopped, the Skyline algorithm will use the notes of the accompaniment part as the main melody.
- (2) In modern music, for songs whose main melody is in the bass region, the Skyline algorithm will use the accompaniment as the main melody.

Aiming at the above shortcomings of the Skyline algorithm, this paper proposes a Multitrack Clustering (MTC) theme extraction algorithm. The basic idea of the algorithm is based on the multichannel clustering algorithm. The difference is that the multichannel clustering algorithm implements the main melody extraction through channel clustering, while the MTC implements the main melody extraction through the audio track block clustering.

3.2.2. Multitrack Clustering Main Melody Extraction Algorithm. We first define the note name value of a note:

$$\text{Pitchname} = \frac{\text{Pitch}}{12}. \quad (4)$$

The note name value of a note is divided into 12 moduli, the pitch of the note. It can be seen that there are a total of 12 different note name values. We perform the Skyline algorithm on each audio track block to ensure that there are no polyphonic notes in each audio track block. Then, for each track block, we find its pitch distribution vector. The composition of the pitch distribution vector is as follows:

$$v_i = [v_{i1}, v_{i2}, v_{i3}, \dots, v_{i12}]. \quad (5)$$

Since vectors can also be regarded as coordinate points, then we will perform agglomerative hierarchical clustering operations on these pitch distribution vectors. First, we use Euclidean distance to describe the distance between two vectors. The calculation formula is as follows:

$$D(v_i, v_j) = \left[\prod_{k=1}^{12} (v_{ik} - \bar{v}_{jk})^2 \right]^{1/2}. \quad (6)$$

The flow of the MTC main melody extraction algorithm is shown in Figure 1.

3.3. Segment Feature Extraction. This paper proposes a sampling and coding algorithm for the musical note set of a section. The main idea of the algorithm is to sample the section to generate multiple sampling moments and encode the notes being played at each sampling moment to generate a 128-bit one. The arrays generated at all sampling moments

are combined in chronological order to generate segment features. The specific steps for the sample coding of the musical note set are as follows:

- (1) Find the main melody note with the highest pitch in the section.
- (2) Sample the entire music segment at a certain time interval dt to obtain M sampling moments and initialize the sequence number $i=1$ of the sampling moments.
- (3) For the i th sampling moment, extract the notes being played at that moment, and these notes meet the following conditions:

$$\text{StartTime}^1 + \text{Duration} \geq i \cdot dt > \text{StartTime}^2. \quad (7)$$

- (4) Initialize an encoding array EncodeArr, which is a one-dimensional array with a length of 128, and the elements in it are initialized to zero.
- (5) Traverse all the notes in the note set. If the traversed notes belong to the elements in the main melody note group, then let

$$\text{EncodeArr}[\text{Pitch}] = \frac{\text{Volume}}{\text{Volume}_{\min}}. \quad (8)$$

Otherwise, it means that the note is an accompaniment note; this time, let

$$\text{EncodeArr}[\text{Pitch}] = \theta \left[\frac{\text{Volume}}{\text{Volume}_{\min}} \right]. \quad (9)$$

Among them, θ is the accompaniment coefficient, with a value between 0 and 1, which is used to reduce the influence of accompaniment notes on the classification results.

- (6) In the above steps, we have completed the encoding of the note collections at all sampling moments of the music segment and obtained M encoding arrays. Combining these encoding arrays in chronological order, we obtain the music segment characteristics.

4. Construction of Music Performance Style Conversion Network

4.1. Convolutional Neural Network. Feedforward Neural Network (FNN) is the earliest type of simple artificial neural network invented in the field of artificial intelligence. In this network, parameter values are unidirectionally propagated along the input layer to the output layer. Convolutional neural network is a kind of feedforward neural network, whose neurons can excite a part of the units in the surrounding coverage. All this is due to the convolution operation, which can extract local features and efficiently use data. The working mode of convolutional neural network is shown in Figure 2.

Convolution operations include one-dimensional convolution, two-dimensional convolution, and three-

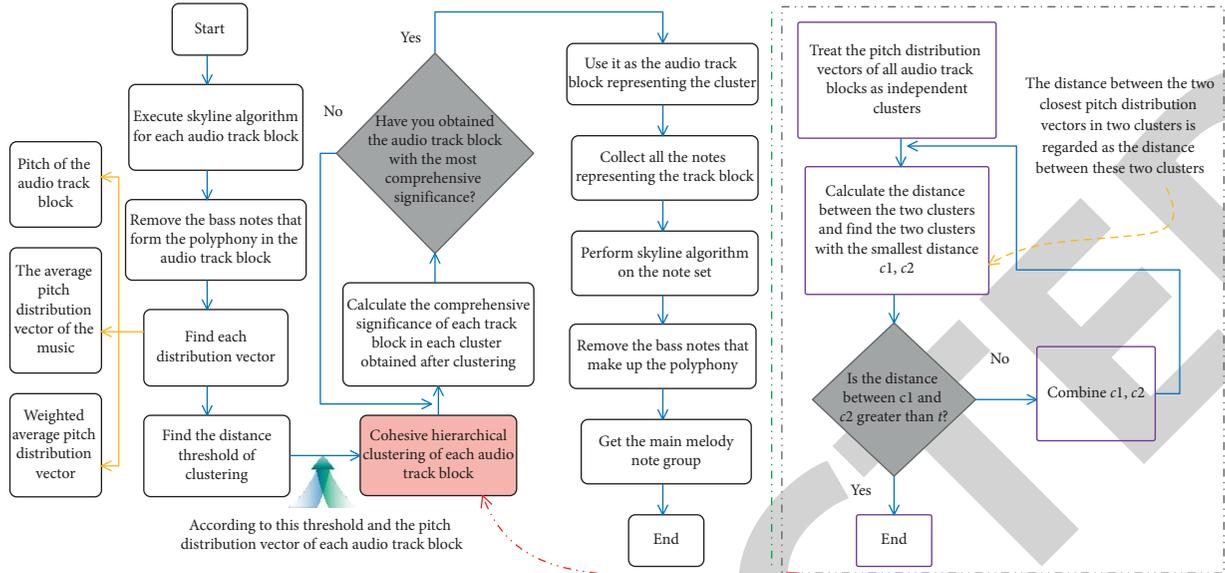


FIGURE 1: MTC algorithm flowchart.

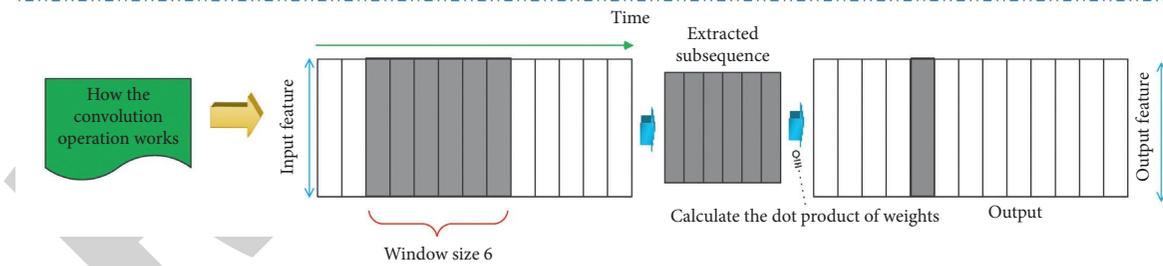
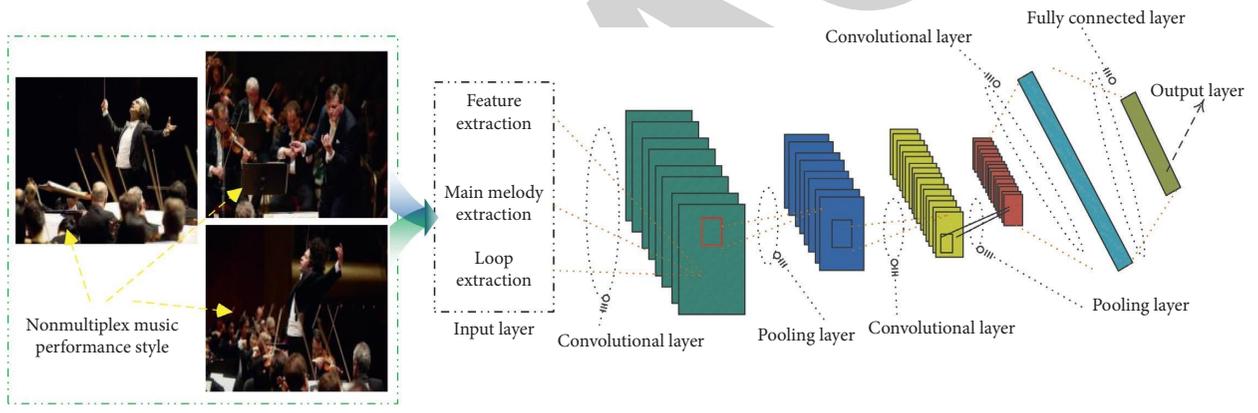


FIGURE 2: Convolutional neural network working mode.

dimensional convolution. The most widely used in the field of music processing is two-dimensional convolution.

One-dimensional convolutional neural networks have also made great progress in the field of machine translation and audio generation. For example, text data can also be processed by one-dimensional convolutional neural networks. The achieved effect can even replace the recurrent neural network, and its calculation cost is smaller and the speed is faster. The calculation formula of one-dimensional convolution is as follows:

$$O(t) = x(t) \otimes h(t) = \sum_i h(i-t) \cdot x(i-1), \quad (10)$$

$x(t)$ is the input sequence, $h(t)$ is the convolution kernel, and $O(t)$ is the output sequence.

In the one-dimensional convolution operation, the convolution kernel slides from the left to the right of the input array. At a certain position, the input subarray covered by the convolution kernel and the convolution kernel are multiplied by the element. The value of the element in the

array is output at the corresponding position. One-dimensional convolution can identify local patterns in the sequence. A one-dimensional convolutional neural network with a convolution window of size 6 can learn shorter fragments. The output of each time step is based on the input sequence.

One-dimensional convolutional neural networks are often used together with hollow convolution kernels, which can expand the receptive field without pooling loss information and obtain multiscale context information at the same time. When the number of holes is d , the new convolution kernel size is

$$n = k - (d + 1) \cdot (k - 2), \quad (11)$$

where k is the size of the original convolution kernel and $d + 1$ is called the expansion ratio.

4.2. Recurrent Neural Network

4.2.1. Ordinary Recurrent Neural Network. In a convolutional neural network, there is a one-to-one correspondence between input and output, and there is no correlation between different inputs. However, for many sequence problems, the overall sequence of the sequence is a very important factor, and different elements before and after are generally related; if only one input is not enough at this time, you need to use a recurrent neural network. Recurrent neural networks are different from feedforward neural networks in that there are loops in the internal connections of the network, and they perform well in dealing with sequence problems.

In the recurrent neural network, the output value of the hidden layer at the current moment not only depends on the input at the moment, but also depends on the output of the hidden layer at the previous moment, and the weight matrix is used to store the output value of the hidden layer at the previous moment versus the hidden layer at the current moment. If the left figure is expanded according to the timeline, at time t , the output of the hidden layer depends not only on the input value of the input layer, but also on the output of the hidden layer at time $t-1$. The cyclic neural network makes it easy to manipulate the vector sequence, and the input and output design of the entire network is flexible.

4.2.2. Bidirectional Recurrent Neural Network. Ordinary recurrent neural networks are one-way; that is, the prediction output at the current moment only considers the input information at the current and past moments. The hidden state of the network propagates from the front to the back, but sometimes, the state at the current moment may also be derived from the future. For example, when it is necessary to predict the missing words in a sentence, what may actually provide useful information is not the phrase before the missing position, but the sentence after the missing word. For such a scenario, it is necessary to add the consideration of future input information on the basis of the ordinary recurrent neural network. This is the Bidirectional

Recurrent Neural Network (BRNN), which considers the previous and next moments for the output at the current moment.

The hidden layer of the bidirectional recurrent neural network can be divided into forward pass and reverse pass. The output of these two parts determines the final result.

4.2.3. BPTT Algorithm. The training of the recurrent neural network uses the Back Propagation Through Time (BPTT) algorithm. The basic principle of the BPTT algorithm is the same as the BP algorithm, which is divided into forward propagation and backward propagation. The steps are as follows:

The first step is to perform forward propagation and calculate the output value of each neuron. The second step is to calculate the error term δ of each neuron through backpropagation, which is represented by the partial derivative of the loss function E with respect to the weighted input of neuron j . The BPTT algorithm will pass in two directions, the upper layer is the 1-1 layer, and the other direction propagates back to the initial moment along the timeline.

The third step is to calculate the gradient of each weight. The calculation formula is as follows:

$$\left\{ \begin{array}{l} \nabla_U E = \frac{\partial E}{\partial U} = \sum_{i=1}^t \nabla U_t \cdot E \\ \nabla_W E = \frac{\partial E}{\partial W} = \sum_{k=1}^t \nabla W_t \cdot E \end{array} \right. \quad (12)$$

4.2.4. Improved Recurrent Neural Network. Based on the original RNN, LSTM saves long-term memory by adding a unit state. Furthermore, it controls the unit state by introducing three gates, namely, the forget gate, input gate, and output gate. The input of the gate is a vector, and the output is a real number between 0 and 1, which is used to describe the amount of information passed. For example, when the output is 0, it means that no information is allowed to pass; when the output is 1, it means that all information can be passed. LSTM uses forget gates and input gates to control the content of the unit state, where the forget gate controls how much the state from the previous moment is retained to the current moment; the input gate controls how much input from the current moment is retained to the cell state. LSTM controls how many unit states are output to the current output of the network through output gates. The calculation formula for the 3 gates is as follows:

$$\left\{ \begin{array}{l} o_t = \sigma[W_o \cdot (h_{t-1}, x_{t-1}) + b_o] \\ f_t = \sigma[W_f \cdot (h_{t-1}, x_{t-1}) + b_f] \\ i_t = \sigma[W_i \cdot (h_{t-1}, x_{t-1}) + b_i] \end{array} \right. \quad (13)$$

Among them, σ represents the sigmoid function, and W and b , respectively, represent the weight matrix and bias term of the corresponding gate.

4.3. Playing Style Conversion Network. In this paper, the note matrix and velocity matrix are obtained from nonmultiplex cluster music pieces, and the encoder in the pretrained

autoencoder is further used to extract the musical implicit style from the note matrix. Music is a sequence of notes that changes over time. Based on the analysis of network models commonly used to deal with sequence problems in the previous sections of this article, this article will use a combination of recurrent neural networks and convolutional neural networks to build a performance style conversion network.

Since the output of the network contains many different styles, this is a multioutput model, and the use of a shared layer can reduce the learning parameters of the network. Recurrent neural network is widely used to deal with sequence problems. It uses the past memory state and current input to predict future information because it only considers the past information, which is not enough to fully understand the music context. At the same time, due to the comparison of GRU, the LSTM structure is simpler, so the shared layer of the performance style conversion network designed in this article uses a bidirectional GRU layer, which takes into account the past and future information. For each style of specific learning, this article uses a one-dimensional convolutional neural network from the shared layer. Regression prediction velocity matrix is performed in the output musical note feature vector sequence that already contains context information.

Figure 3 shows the structure of the performance style conversion network model proposed in this paper, which can be divided into three parts: input layer, hidden layer, and output layer.

The input of the input layer is the music implicit style extracted by the encoder part of the pretrained autoencoder. We first pretrain the autoencoder model to obtain the weight of the encoder part. After that, we build the encoder architecture, load the trained weights, freeze each layer of the encoder, and use the output of the encoder as the input of the performance style conversion network.

The hidden layer is used to learn the relationship between the implicit style of music and the real strength matrix. The shared two-way GRU layer is used to reduce the training parameters of the network, obtain the sequence of note feature vectors, and use the same sublayer for each style. In the sublayer, 3 stacked one-dimensional convolutional layers are used to learn the velocity distribution according to the note feature vector sequence. After each one-dimensional convolutional layer, batch normalization is also used to make the prediction effect closer. The true force distribution ensures the nonlinear expression ability of the model.

The output layer is used to predict the force matrix, using a fully connected layer and using a wrapper to apply the shared weight to each time step.

In order to judge whether the velocity matrix predicted by the performance style conversion network conforms to a specific style, a velocity classifier is used to classify the generated velocity matrix. LSTM Fully Convolutional Networks (LSTM-FCN) contains two branches. The first branch is implemented by a layer of LSTM-based recurrent neural network, and Dropout is used to randomly discard some neurons; the second branch is implemented by three layers of convolutional layers, each of which is composed of

a one-dimensional convolutional layer and batch normalization layer. The output of the second branch is connected to the output of the first branch after the average pooling layer, as the input of the final fully connected layer.

5. Simulation Experiment of Music Performance Style Conversion

5.1. Music Style Database Construction. The music database in the experiment contains 5 music style categories: dance, folk music, jazz, rock, and lyric. These songs are all in MP3 format downloaded from music websites. The music collections used to train and test the cyclic neural network are subjectively selected by humans. The cyclic neural network uses the known music style to obtain its feature vector to identify the unknown. The selection of music used for training samples is very important. There are more than 600 songs of each style, a total of more than 3,000 songs, and 50 students are invited to be divided into 10 groups, with 5 people in each group, and these 5 people were allowed to classify and label about 300 songs in this group. When the classification label is uncertain, the labeler is allowed to listen repeatedly until the label is correctly given. After the labeling is over, each song has 5 classification tags. Only when the classification tags of a song are gathered in the same music category, we will put the song into the music database. The composition of the finally obtained music training library and test library is shown in Figure 4.

Before categorizing the music files, the music in the music sample library should be formatted. We use the Format Factory format conversion tool to convert the MP3 format in the database to WAV format. Since the time of a whole song is relatively long, the data for extracting the feature vector will be relatively large, so 45 seconds of each song is intercepted as a music sample, and one of the channels is taken, and all the sampling rates are converted to 18 KHZ.

5.2. Experimental Results and Analysis. When the sample space is linearly inseparable, the slack variable ξ_i is introduced into the cyclic neural network, which allows a small amount of errors in the classification surface. When the fuzzy membership function is introduced in the experiment, the RBF kernel function is used, and the selected slack variable is $\xi_i = 1$. When the value is too small, the classification accuracy will decrease; that is, the training model is in an underlearning state. When the size is too large, the training model is in a situation of overfitting. By comparing and analyzing the influence of the change of the slack variable on the system, the slack variable chosen in this experiment is $\xi_i = 1$.

In order to verify the effectiveness of using the fuzzy membership function MF in the recurrent neural network, the MFCC and RASTAP-PLP feature parameters are also used, and the RBF kernel function of the recurrent neural network classifier is selected for classification, and the MF function is used respectively. The neural network classifier and the recurrent neural network classifier without MF

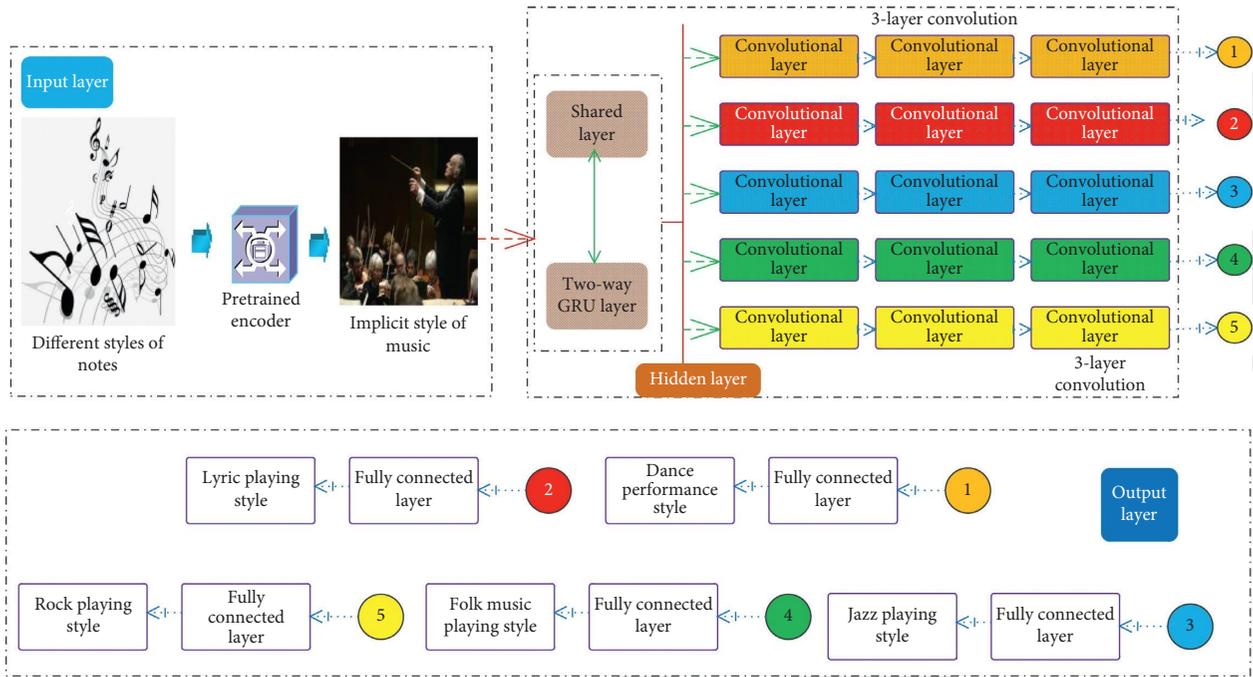


FIGURE 3: Performance style conversion network.

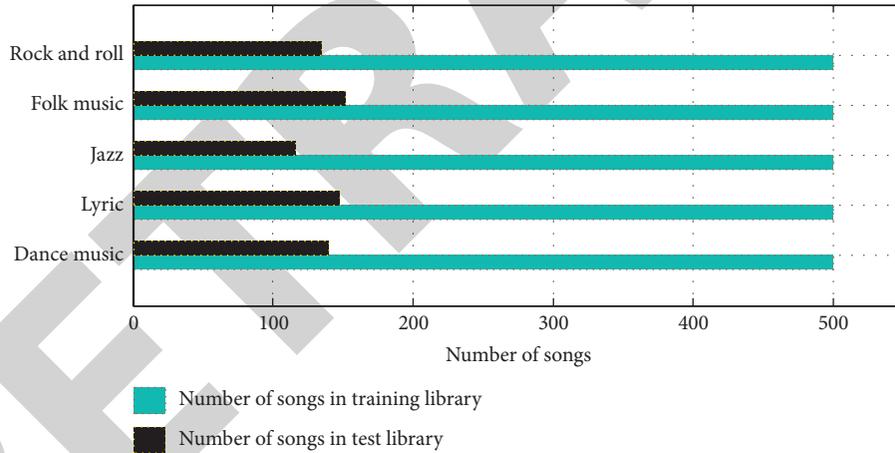


FIGURE 4: Music training and testing database.

function perform classification experiments on a database containing 5 music styles. It can be seen from Figure 5 that the classification results of the five kinds of music after using the fuzzy membership function are relatively ideal, while the classification without the fuzzy membership function is worse. The effectiveness of introducing the fuzzy membership function reduces the range of the unrecognizable area to a certain extent and improves the accuracy of classification. After introducing the fuzzy membership function, the accuracy of music style classification is relatively improved; that is, the cyclic neural network enhances the generalization ability through fuzzy classification.

The subsequent experimental structure of this paper is based on the improved cyclic neural network classifier. In order to further discuss the effectiveness and stability of the

classifier, three music test libraries A, B, and C are formed after the integration of the music test library. We choose to use the RBF kernel function to perform classification tests on these three music libraries, and the average classification error is shown in Figure 6.

It can be seen from the classification results that the classification accuracy of test library C finally reached more than 98%, while the final classification accuracy of test library B was less than 97%, indicating that the music of test library C is more representative. The gap between each style of music is relatively obvious, which shows that the music style classification system can effectively classify unknown music, and the classification accuracy will vary with the representativeness of the tested music library. When selecting the dimensionality that can represent the

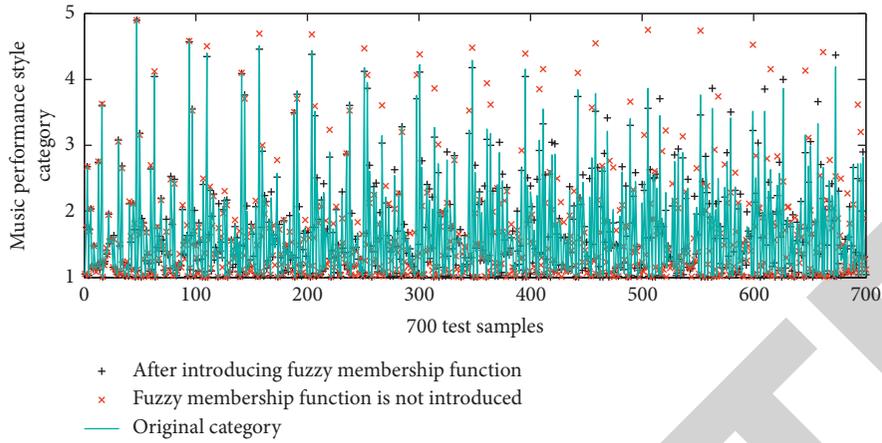


FIGURE 5: Classification results of 700 test samples before and after citing the fuzzy membership function.

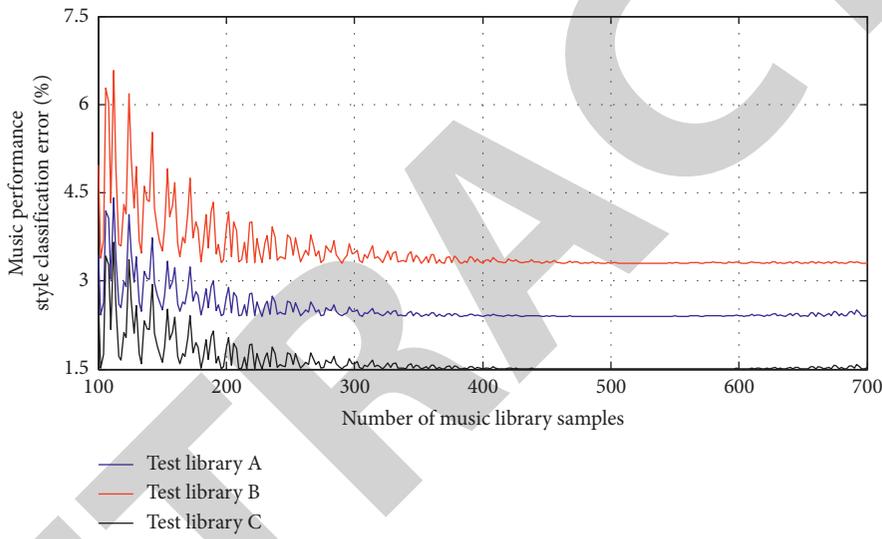


FIGURE 6: Classification error of three music libraries.

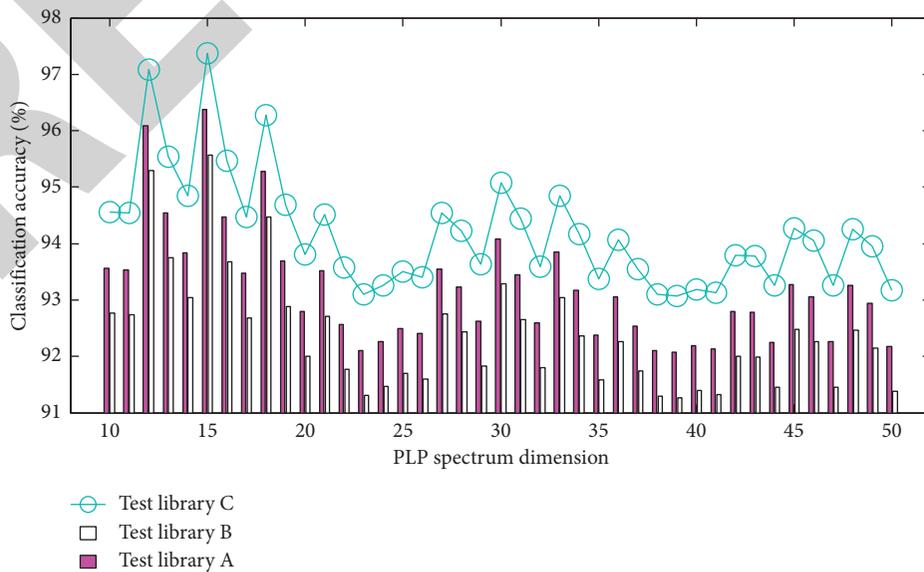


FIGURE 7: Classification accuracy of different PLP spectrum eigenvalues.

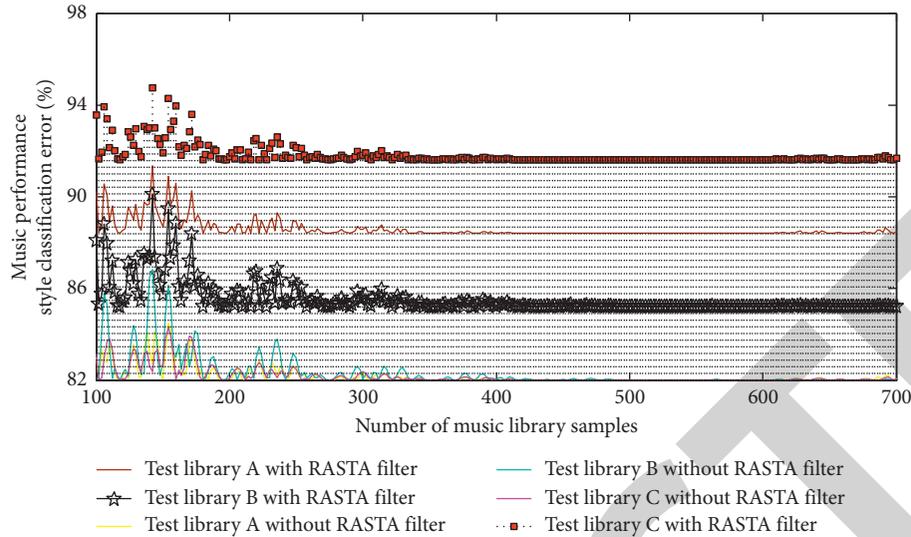


FIGURE 8: Comparison of PLP parameters processed by RASTA filter.

characteristic value of the music signal, a filter can be made to represent the music signal as much as possible while reducing the computational complexity. The classification accuracy of different PLP spectrum eigenvalues is shown in Figure 7. It can be seen that the classification accuracies of music test library A, B, and C are all above 91%.

Under the same other conditions, this paper conducts a comparative experiment on whether the PLP cepstrum and PLP spectrum of the music signal are processed by the RASTA filter on the classification results. The comparative experiment results are shown in Figure 8.

The classification results of PLP parameters that are not processed by the RASTA filter are very bad. It can be seen that in the process of system implementation, simply using PLP parameters cannot produce good results. The RASTA filter suppresses the slowly changing elements of each spectral component in the short-term auditory spectrum before linear prediction analysis. In addition, the content of the high-frequency modulated frequency spectrum produced by the speaker conveys very little voice information, so the difference shown in the automatic recognition process is not critical. Therefore, the band-pass filter will not only attenuate high-frequency changes, but also suppress the rapidly changing parts of the frequency spectrum in the speech. The high-pass part of the RASTA filter will remove the slow-changing elements, and the low-pass part will remove the fast-changing components.

6. Conclusion

This paper introduces in detail the main process of extracting section features from nonmultiplex cluster music files, including nonmultiplex cluster music file note extraction, MTC algorithm theme extraction, and segment division algorithm based on energy feature vectors. We divide the music into multiple segments and use the segment note set sampling coding algorithm to extract the segment features.

In the performance style conversion modeling, a performance style conversion network based on recurrent neural network and convolutional neural network is proposed. The two-way recurrent neural network based on GRU is used to extract the note feature vector sequence, and the one-dimensional convolutional neural network is used. The extracted note feature vector sequence is used to predict the intensity, and the intensity changes of different styles of nonmultiple cluster music can be better learned. For the realization of the music style system in the MATLAB environment, this paper implements the music style classification system in MATLAB language and mainly analyzes the integration of music feature preprocessing, feature extraction, and learning of cyclic neural network classifiers in the MATLAB environment. We studied and tested a music database containing five music styles and conducted comparative experiments on the effectiveness of introducing fuzzy membership functions, the stability of the classification system, and the functionality of the RASTA filter. Experimental data shows that the music style classification system designed in this paper can realize the style classification of unknown music. However, from the perspective of the technical level of this technical system, it is still not possible to use signal processing alone to perfectly solve the fundamental frequency extraction problem of polyphonic music and the accurate positioning of the starting point of the notes of complex music. This will reduce the reliability and efficiency of the whole score translation of the passage. Therefore, we still need to study the acoustic characteristics and musical structure of music in depth based on the existing technology, so as to propose a more complete technology to automatically analyze and identify its key attributes.

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 that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Chongqing University of Arts and Science Special Project of Ideological and Political Curriculum 2020: Exploration and Practice of Ideological and Political Theory of Music Course in Normal Major (no. 20200204S).

References

- [1] O. Cifka, U. Simsekli, and G. Richard, "Groove2Groove: one-shot music style transfer with supervision from synthetic data," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2638–2650, 2020.
- [2] C. Jin, Y. Tie, Y. Bai, X. Lv, and S. Liu, "A style-specific music composition neural network," *Neural Processing Letters*, vol. 52, no. 3, pp. 1893–1912, 2020.
- [3] S. B. Lie, "Music that tears you apart: jazz manouche and the qualia of ethnorace," *Ethnomusicology*, vol. 64, no. 3, pp. 369–393, 2020.
- [4] A. M. Syarif, A. Azhari, S. Suprpto, and K. Hastuti, "Human and computation-based music representation for gamelan music," *Malaysian Journal of Music*, vol. 9, pp. 82–100, 2020.
- [5] D. Margounakis, G. Tsotakos, and A. Floros, "On digitising the Greek music tradition: simulation of the Cretan lyre for mobile devices," *International Journal of Arts and Technology*, vol. 12, no. 2, pp. 103–117, 2020.
- [6] A. K. Yanchenko and P. D. Hoff, "Hierarchical multidimensional scaling for the comparison of musical performance styles," *Annals of Applied Statistics*, vol. 14, no. 4, pp. 1581–1603, 2020.
- [7] T. V. Vu and D. N. Tran, "Demystifying perceptual learning style preferences of Vietnamese university freshmen in English academic achievement," *International Journal of Learning, Teaching and Educational Research*, vol. 19, no. 9, pp. 15–38, 2020.
- [8] D. Armstrong, "Hot collecting off the record: ralph J. Gleason's start in music journalism," *Jazz Perspectives*, vol. 12, no. 2, pp. 247–282, 2020.
- [9] L. De Mora, A. A. Sellar, A. Yool et al., "Earth system music: music generated from the United Kingdom Earth System Model (UKESM1)," *Geoscience Communication*, vol. 3, no. 2, pp. 263–278, 2020.
- [10] M. M. Sharp, "If you're ready, I am ready (but the wait is harming us both) individual risks in institutional conversions," *Pastoral Psychology*, vol. 69, no. 4, pp. 445–463, 2020.
- [11] L. Turchet, J. Pauwels, C. Fischione, and G. Fazekas, "Cloud-smart musical instrument interactions," *ACM Transactions on Internet of Things*, vol. 1, no. 3, pp. 1–29, 2020.
- [12] M. Salis, "When did amphitrite cross the tyrrhenian sea? some reflections on the appearance of grisaille, myth and the classical renaissance style in the periphery of the crown of aragon," *IKON*, vol. 13, pp. 109–120, 2020.
- [13] T. Roy, "Music and society in late colonial India: a study of Esraj in Gaya," *The Journal of Asian Studies*, vol. 79, no. 1, pp. 25–49, 2019.
- [14] S. Brautbar, P. La Chapelle, and J. Hutchings, "The valley of the dry bones," *Journal of Popular Music Studies*, vol. 32, no. 2, pp. 191–213, 2020.
- [15] O. Smith, "Waringin: recording a composition with gamelan salukat, a crossroads of music and culture," *Nota Bene: Canadian Undergraduate Journal of Musicology*, vol. 13, no. 1, pp. 61–83, 2020.
- [16] A. Starkey, K. Steenhauer, and J. Caven, "Painting music: using artificial intelligence to create music from live painted drawings," *Drawing: Research, Theory, Practice*, vol. 5, no. 2, pp. 209–224, 2020.
- [17] K. Mack, "She's a country girl all right," *Journal of Popular Music Studies*, vol. 32, no. 2, pp. 144–161, 2020.
- [18] S. O'Shea, "Activate, collaborate, participate: the network revolutions of riot grrrl-affiliated music worlds," *Punk & Post-Punk*, vol. 9, no. 2, pp. 309–325, 2020.
- [19] S. Ferrand, F. Alouges, and M. Aussal, "An electronic travel aid device to help blind people playing sport," *IEEE Instrumentation & Measurement Magazine*, vol. 23, no. 4, pp. 14–21, 2020.
- [20] I. Swart, "Benefits of music education to previously disadvantaged South African learners: perspectives of music teachers in the greater Tshwane Metropolis," *International Journal of Music Education*, vol. 38, no. 1, pp. 52–65, 2020.
- [21] Q. Meng, S. Zhang, and J. Kang, "Effects of typical dining styles on conversation behaviours and acoustic perception in restaurants in China," *Building and Environment*, vol. 121, pp. 148–157, 2017.
- [22] B. Green, "'I always remember that moment': peak music experiences as epiphanies," *Sociology*, vol. 50, no. 2, pp. 333–348, 2016.
- [23] K. T. Smith, "Mobile advertising to digital natives: preferences on content, style, personalization, and functionality," *Journal of Strategic Marketing*, vol. 27, no. 1, pp. 67–80, 2019.
- [24] T. Tsai and K. Ji, "Composer style classification of piano sheet music images using language model pretraining," pp. 1–8, 2020, arXiv:2007.14587.
- [25] J. Wen, J. Yang, B. Jiang, H. Song, and H. Wang, "Big data driven marine environment information forecasting: a time series prediction network," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 4–18, 2021.
- [26] J. Yang, J. Zhang, and H. Wang, "Urban traffic control in software defined internet of things via a multi-agent deep reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.
- [27] J. Yang, M. Xi, B. Jiang, J. Man, Q. Meng, and B. Li, "FADN: fully connected attitude detection network based on industrial video," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2011–2020, 2021.