ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Music Visualization Design Method Based on Multi Audio Features of Video Image Processing Technology

Xin Chen[1], *Liqiong Duan*[2,*] *and Xiaowei Liu*[3]

[1] Vocal Music, Sookmyung Women's University, Seoul, 04310, Korea
[2] Admissions Office, Zhangjiakou Open University, Zhangjiakou 075000, Hebei, China
[3] Sangmyung University, Seoul, 04310, Korea

SUMMARY: *With the development of music visualization, people are more and more likely to accept it, and people unconsciously form a different concept of music art. Among them, image processing technology is an important part of music visualization design. In audio multi feature extraction, preprocessing is the first step of speech signal analysis. After a certain preprocessing process, the sound signal can be converted into a data format that can be processed by algorithms and computers. The specific steps are MIDI file note extraction, theme extraction, segment segmentation, and segment feature extraction. In the overall design of music visualization, this paper proposed two design schemes for different music forms. One is for different music characteristics, and the other is to use a single or combined way to enrich the music visual effect. Different main and secondary melodies (pitch, beat, speed, time fractal dimension, etc.) would produce different visual effects. The classification accuracy of metal music, classical music and folk music in this paper was 93%, 92.7% and 93.3% respectively. The visualization method mentioned in this paper is more targeted and has better implementation effect.*

KEYWORDS: *Music Visualization, Multi Audio Features, Video Image Processing Technology, MIDI Files*

1 Introduction

With the advent of the network era, multimedia devices are widely used, and computer technology has also developed. With the continuous development of music visualization technology, many research institutions and companies have begun to pay attention to this aspect. Music visualization has many forms of expression, such as smoke and fire, waves and flames, whose forms of expression are determined by the spectrum. Music understanding is a subjective form composed of mathematical formulas, with strong subjectivity. Sound is the most common, direct and extensive means of communication. It can enhance the expression of emotions on the stage, and also can effectively record the voice of nature. For example, birds are also life forms of many organisms. Therefore, a lot of useful information can be obtained from effective analysis of sound. At a deeper level, video image processing technology can improve the traditional sound analysis methods. Traditional acoustic analysis methods are not only inefficient, but also require a lot of experimental experience. Moreover, through data visualization technology, useful information can be quickly found from the historical data of sound.

*momo_918@163.com
https://doi.org/10.65102/is2026338
Published: 30 April, 2026

The visualization research of music information can improve the effectiveness of music information in the transmission process. Wu Yongmeng believed that most contemporary western performing art practices limit the creative interaction of the audience. The operating system Open Symphony is to explore the interaction between audience and performers in live music performance, supplemented by digital technology. The audience can choose various music "modes" to perform impromptu performances by voting. The technical components include a web-based mobile application, a visual client that displays generated symbolic scores, and a server service for exchanging creative data. The interaction model, application and visualization are designed through an iterative participatory design process [1]. Khulusi Richard believed that digital methods were increasingly used to store, construct and analyze large amounts of music data. In this case, visualization played a crucial role because it helped musicologists and non professional users to analyze data and acquire new knowledge. His investigation focused on the unique relationship between musicology and visualization, classified 129 related works according to the type of visualization data, and analyzed which visualization technologies were used for specific research queries and to complete specific tasks. In addition to scientific references, he also examined commercial music software and public websites, which have contributed to the new concept of visual music data [2]. Lima Hugo B believed that music information research includes all research topics related to modeling and understanding music. Visualization is often used to convey a better understanding of music works, and it has a history and extensive practice to link music with visual elements. He investigated papers related to music visualization. The Infovision community can benefit from identifying trends and possible new research directions in music visualization topics [3]. De Prisco Roberto has proposed a visualization technology that can help people lack strong theoretical knowledge of music. This technique used graphic elements to attract people's attention to possible mistakes in composition. He has developed an interactive system called Visual Melody, which used the proposed visualization technology to promote the understanding of the structure of music works [4]. Jin Yucheng believed that music recommendation systems usually provide a "one size fits all" approach, providing all users with the same user controls and visualization. He studied the impact of personal characteristics on the visual design to enhance the diversity of recommendations and the design to optimize the level of user control while minimizing cognitive load. The results showed that personalized visualization and control elements are beneficial to the complexity of music. When studying the combined effect of control and visualization, the complexity of music would only strengthen the impact of User Interface (UI) on perceived diversity. These results allowed to expand the personalized model of the music recommendation system and provide guidance for the interactive visual design of the music recommendation system, including visualization and user control [5]. The presentation effect of the music visualization method proposed by the above scholars is not very good. This paper introduces video image processing technology to further optimize the research of music visualization.

Video image processing technology promoted the development of video image processing system. Birajdar Gajanan K proposed a music classification feature extraction method based on generalized Gaussian distribution. Compared with traditional short-time Fourier transform analysis that provided unified frequency resolution, spectrogram visualization provided superior temporal resolution at high frequencies and better spectral resolution at low frequencies [6]. Wang Bryan believed that although recent work has made great progress in automatic music generation in the field of symbols, few people try to build an AI (Artistic Intelligence) model that can reproduce real music audio from music scores. He proposed a depth convolution model, which learned the mapping of music score to audio between the symbolic representation of music (called piano sound) and the audio representation of music

(called spectrogram) in an end-to-end manner. He refined the results by adding overtones and spectral textures of timbre [7]. de la Fuente Carlos believed that optical music recognition and automatic music transcription represent the research field of obtaining structured digital representation from music score images and recordings, respectively. Although these fields have traditionally developed independently, the fact that these two tasks may share the same output representation raises the question of whether they can be combined in a synergistic manner to take advantage of the single transcriptional advantage described by each mode. To evaluate this assumption, he proposed a multimodal framework, which considered the local alignment method [8]. Herremans Dorien believed that people are increasingly interested in using deep neural networks to handle tasks in the audio and music fields. He believed that computer capture and quantification of music structure is a difficult task. Recently, other researchers have tried to use deep learning models to learn features and relationships that allow him to complete tasks such as music transcription, audio feature extraction, emotion recognition, music recommendation and automatic music generation. He aimed to introduce a series of studies to improve the latest level of music and audio in the field of intelligent machines [9]. Pandeya Yagya Raj believed that the widely spread online and offline music videos are one of the rich sources of human emotion analysis, because they integrate the inner feelings of the musicians through song lyrics, instrument performance and visual expression. He first built a balanced music video emotion dataset, including the diversity of regions, languages, cultures and musical instruments. He also tested this dataset on four unimodal and four multimodal convolutional neural networks for music and video [10]. However, the precision of the video image processing technology proposed by the above scholars is not high.

This paper discussed the relationship between visual elements and data features in audio visual processing from the perspective of music visualization. When multiple complex sound structures appear at the same time, people should not only analyze their characteristics, but also associate them with other music elements. On this basis, this paper designed the visual products of emotion and theme with specific design examples. The validity, feasibility and superiority of the music visualization design scheme were verified by a series of objective indicators such as the experimental results and the accuracy of product visualization. This paper systematically collated the relationship between theme and emotion, which was verified by experiments. The basic features such as music score and tone have some basic benchmarking with the emotional features of music works, and there is a corresponding correlation between the forms of expression. Melody plays a very important role in music works. The melodic image of songs vividly reflects people’s subjective feelings, because the direction and dimension of music not only reflect the height of the voice, but also the length of the tone. Using the music classification method in this paper, the accuracy rate was 93.2%.

2 Music Visualization Design Method

2.1 Audio Multi Feature Extraction

Audio visualization refers to presenting sound in a non subjective way. Its main content is to analyze and evaluate the audio data, which provides a basis for future visual analysis. There are also many categories of music types, and their specific visual processing involves the comparison between sound quality, timbre, volume, rhythm, and related visual elements [11]. There are many elements of visualization, such as basic geometry, particle effects (water flow, fire), and three-dimensional characters and scenes generated by computer calculation. With proper visualization technology, audible speech can be turned into visible speech. At the same

time, with the progress of technology, audio data visualization becomes more practical in the following aspects, including:

The analysis of audio data is simplified, so that users can easily identify the connection and difference between sound and its inherent characteristics;

Hearing impaired people can directly feel the content of sound when playing music, and adults may have biased views on music. For example, when playing music, they would have a visual feeling, thus "seeing" music [12];

In the musical fountain, the combination of color light source and painting is used to express the high sound quality;

It avoids the noise confusion caused by multiple sound sources. Music mashups can bring different feelings to people. For example, in voice monitoring, listeners can convert auditory content into visual content and see these information. In this way, managers can deal with different audio situations in different places at the same time and quickly find them [13]. The monitoring software can track the noise in hospitals, libraries and other places.

Good experience is an indispensable part of excellent scene design. For example, advertising music can give users the most intuitive feeling even if it does not represent good design [14]. It can be said that only neurons can feel the world, because it is the response of the neural network to external stimuli. In the analysis of digital music signal, feature extraction must be adopted, but due to different final purposes, the required features are also different. Therefore, this paper first introduces the speech signal preprocessing in detail, and then analyzes various features and classifies them. The details are as follows:

Preprocessing: the embodiment of the musical thinking concept makes the music content richer and more diversified [15]. The preprocessing of sound signal is the first step of sound analysis. After a certain preprocessing process, the sound signal can be converted into a data format that can be processed by algorithms and computers. The whole process can be roughly divided into several steps, such as sound signal filtering, pre-emphasis processing, framing and window processing, and mute frame recognition processing.

Time domain feature extraction: in the process of extracting the features of audio signals in the music retrieval system, people often think of using time as the clue of feature extraction of information, that is, feature extraction of audio signals in the time domain [16].

Frequency domain feature extraction: in the frequency domain, to realize the research on the frequency domain characteristics of audio signals, which is related to music images, it is necessary to convert waveform signals in the time domain into signals in the frequency domain. At present, Fourier analysis is the most widely used and effective method in the world [17]. This method can realize the signal transformation very well, and the speed is very fast.

(1) Feature extraction of multi track MIDI files

The feature extraction of MIDI files is a key link in MIDI music classification. Previous research mainly extracted from the statistical characteristics of main melody notes. However, in the main melody, in the process of music generation, the sequence of notes plays a great role in the characteristics of the melody, and the statistical characteristics cannot fully reflect the sequence relationship of the main melody [18]. Secondly, music itself does not have accompaniment information. In addition, a large number of works often appear repeatedly. If the whole song is taken as an example, it would cause redundancy of samples. This paper extracts fragment features from MIDI files. The pictures and music are converted into several fragments and the features of each fragment are extracted. The steps are note extraction, theme extraction, and fragmentation and fragment feature extraction of MIDI files [19]. The MIDI file notes are mainly used to extract music notes for later processing. Automatic music labeling is helpful to the standardization of music labels. The distinctive feature of a music

segment is that it is composed of a certain segment of a music work. It includes the primary melodic and accompaniment information of the piece and reflects the sequential relationships between the notes [20].

1) MIDI track block composition

Audio track block is an important part of MIDI file. It can be recorded through multiple parallel data, so that it can generate multiple audio tracks. The identification code of the audio track block is “MTrk”. In the header block of a MIDI file, its penultimate arguments serve to define the number of track blocks in the file. Generally speaking, in the first track behind the MIDI file header block, there would be overall information about the track block, such as speed, beat, and modulation.

2) Multi track MIDI file note extraction

A note must include four characteristics: pitch, start time, duration and intensity. Therefore, a note is represented by a vector containing four elements. The first element of this vector is the pitch of the note; the second element is the number of the note; the fourth element is the intensity of the note expressed by the force of the keys.

The steps to extract the note matrix from the MIDI file are as follows:

Step1: The number of ticks per second in the MIDI file is calculated.

Step 2: The audio track data in the MIDI file is extracted, and the audio track blocks that do not contain the audio track opening events are removed. The audio track is initialized according to the number of audio track blocks in the collection.

Step 3: For each track block in the track data group, in each non channel, a matching note open event is found, and a note vector is generated. This vector is added to the corresponding note matrix.

(2) Segment feature extraction

This article introduces the algorithm for sampling and encoding music fragment note sets. The basic idea is to produce several sampling instants by sampling music fragments. Then the tunes in each sampling time are coded to produce 128 one-dimensional arrays. Finally, the sequences generated during the sampling are combined according to the time series to reflect the characteristics of the music segment.

The detailed procedures for sampling and incoding the segment annotation set are as follows:

1) It is determined that the main melody in each part has the largest tone intensity.

2) The whole piece of music is sampled at certain intervals.

3) The i-th sampling time is set to extract the music played at this time.

4) The encoding array is initialized. It should be a one-dimensional row with a 128 length. All of its elements are zero initialized.

5) All notes are traversed in a group of notes, and a group of notes is encoded.

2.2 Multi Track Clustering Theme Extraction and Visualization

(1) Main melody extraction

Based on Skyline algorithm, this paper proposes a theme extraction method of Multi Track Clustering (MTC) based on multi-channel clustering. The main idea is to refer to multi-channel clustering algorithm. MTC mainly uses channel clustering method to extract theme, and MTC uses track block clustering method to extract theme, as shown in Figure 1.

To sum up, the steps of MTC theme extraction algorithm are as follows:

Step1: Skyline operation is performed on each track block to remove the bass tracks that make up polyphony.

Step 2: The pitch distribution vector for each track, the mean pixel distribution vector for the song, and the vector of weighted mean pixel distributions are calculated.

Step 3: The distance threshold of clustering is calculated, and each track block is clustered according to this threshold and the pitch distribution vector of each track.

Step 4: Each cluster of information is clustered, represented by the track block with the highest comprehensive significance, and the overall significant characteristics of each track are calculated respectively.

Step 5: All the notes that stand for track blocks are collected, and a Skyline operation is run on this set of notes to take away the bass notes that form the polyphony, resulting in a master set of notes.

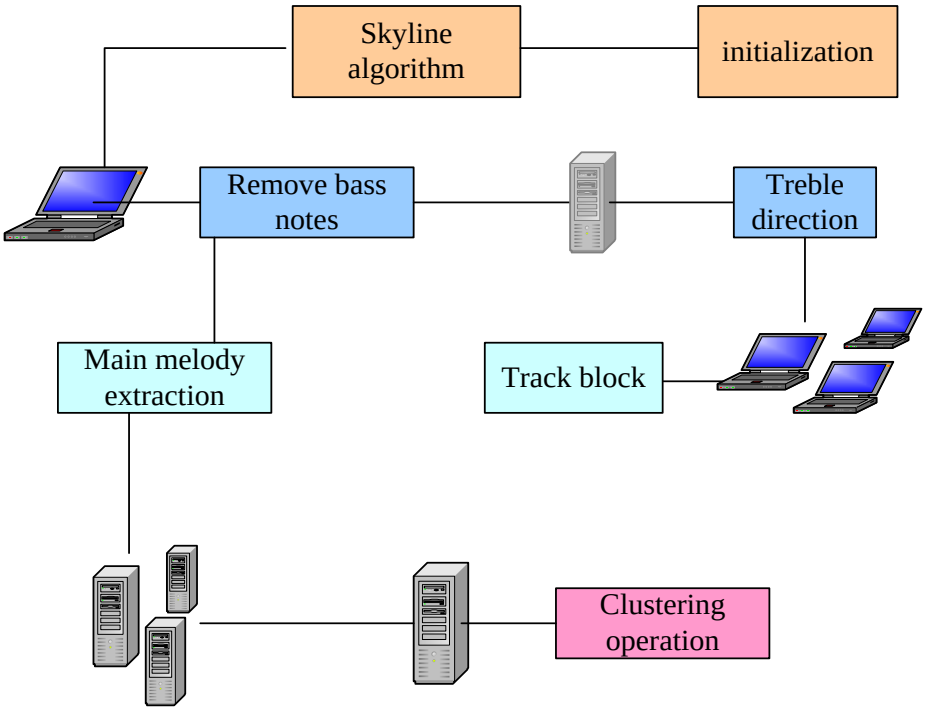

Figure 1: MTC uses track block clustering method to extract main melody

Track position update frequency $2\mathrm{pRp}$ is:

$$2\mathrm{pRp} = (\mathrm{PTD} - \mathrm{M}) = 7\mathrm{D}_2 \tag{1}$$

$$\mathrm{N(d)} = 1 + 3\mathrm{d}(\mathrm{d} + 1) \tag{2}$$

$\mathrm{N(d)}$ is the number of paging cells.

$$2\mathrm{pT1} = 2\mathrm{p}\pi_1 + \mathrm{P}\pi_2 \tag{3}$$

$$2\mathrm{p}\pi = \mathrm{P}\,\pi\mathrm{a} + \mathrm{p}\pi\mathrm{B} \tag{4}$$

(2) Music visualization

Music visualization is a new technology, which needs more and more complex methods to achieve. In order to achieve the maximum immersion effect, the visualization of music must use virtual reality technology. The virtual characters and scenes are used to change the behavior of the characters through the extracted music, so as to realize the change of the scene.

Some systems use advanced visualization technology to create virtual characters of solid

size, and users are forced to immerse themselves in them. Nowadays, virtual reality is closely connected with 3D (three-dimensional) technology, which has a pleasing effect.

1) Because most visualization effects are based on a single feature, this paper adopts a visualization technology based on multiple features. The music in the music library is filtered by several features, and a feature library is constructed. Then, according to the specific requirements, the visual processing is carried out. The visual design of various music forms is realized.

2) The current visual effect is to show the main melody and sub melody at the same time, which sometimes cannot convey the theme of music well. This paper introduces a visual representation method based on main melody and sub melody. Firstly, the features of structured audio MIDI are extracted and specific features are constructed. Secondly, the main melody is visualized, and the auxiliary melody is also visualized. The background is also weakened, so as to achieve the visual effect of various music forms that highlight the theme.

The visual design of music is shown in Figure 2.

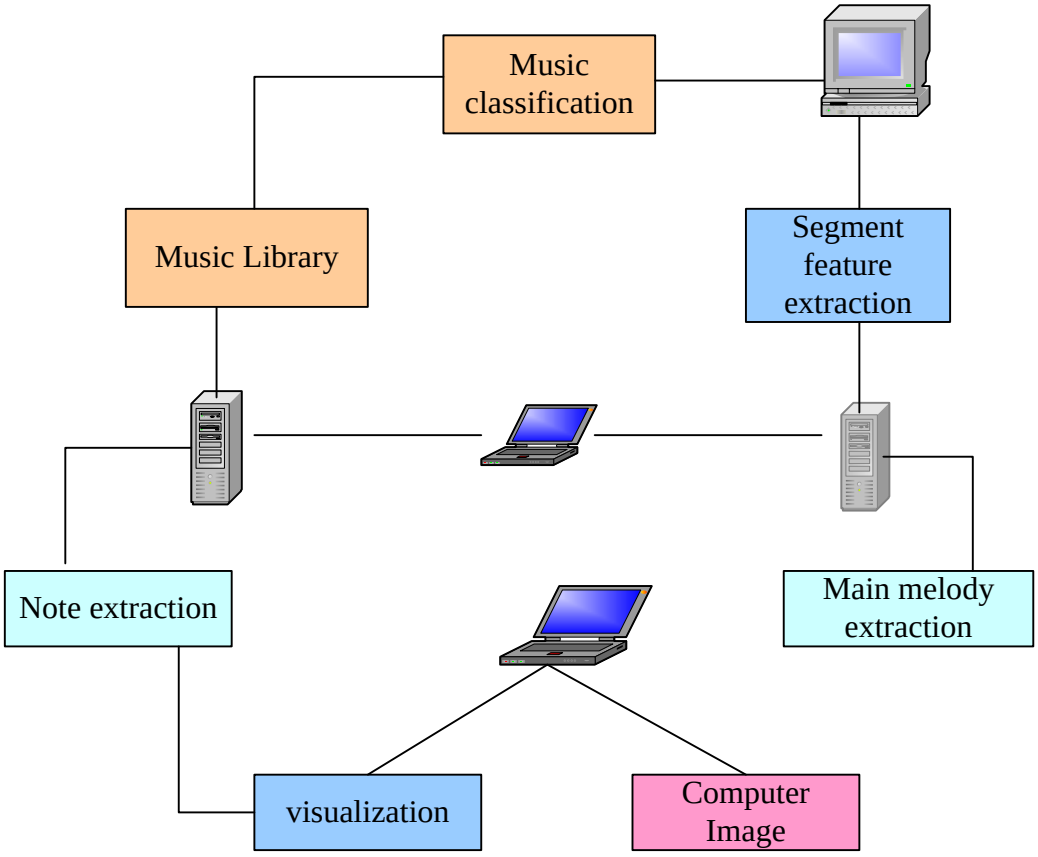

Figure 2: Music visualization design

The transmission cost of music visualization signal Nt is:

$$\mathrm{T} = 1/\mathrm{MT_P} \tag{5}$$

$$Nt = Nsp \times \lambda \times NBpage + LUR \times \mathrm{D_L} \tag{6}$$

T is the transmission cycle. By weighting the coverage time $\mathrm{f_t}$ and the elevation angle σ of the music visualization signal, the cost function C is obtained as follows:

$$C = w\frac{\beta_m}{\beta} - w_0\frac{T_m}{T} \tag{7}$$

$$f_t = \arccos\left(\frac{r+h}{s} * \sin(\epsilon)\right) \tag{8}$$

$$n_\theta = e + \text{arctg}(\cos(i) * \text{tg}(\sigma)) \tag{9}$$

3 Evaluation Results of Music Visualization

This paper uses Python programming and Keras to make a classification experiment on Tensorflow's MIDI music style. In the experiment, the non repeated random sampling method was used. 80% of MIDI files of various music genres were used as samples, and the remaining 20% were used as check groups. The training set and check group were independent and non overlapping. Table 1 shows the number distribution of different types of MIDI files in the training set and verification set.

Firstly, the MIDI music was classified by using BP (Back Propagation) neural network. By extracting two different feature sets, the BP neural network was established, and the role of this method in the classification experiment was discussed. On this basis, this paper designed a new MIDI music classification algorithm based on deep learning, and discussed the role of attention mechanism in the network model and the impact of different segmentation methods on the classification effect through experiments.

Table 1. Number distribution of different types of MIDI files in training set and verification set

Style of music	Training set	Validation set	Difference value
Classical	213	85	128
Rural	210	83	127
Dance Music	248	88	160
Nongovernmental	205	76	129
Metal	233	90	143
Rock Music	220	72	148

In the experiment, this paper selected the test results with the highest accuracy of the test set obtained during the training, and compared them with the accuracy as the main indicator. Through the comparison between Experiment 1 and Experiment 2 (based on music feature task classification), it was concluded that the music reconstruction algorithm based on emotional feedback was used in BP neural network classification experiment, and the classification effect of this method was much worse than that of Experiment 2.

In Experiment 3, MIDI documents were divided into different segments; different segments were used as analysis units; different segments were processed differently; different features were combined to form new different segments. Through Bi-GRU (Gate Recurrent Unit), people can get deeper music expression from the input segment feature sequence. The classification effect was better than the traditional MIDI music classification method based on BP neural network. The comparison between Experiment 1 and Experiment 4 is shown in Figure 3 (the comparison between Experiment 1 and Experiment 2 is shown in Figure 3 (a), and the comparison between Experiment 3 and Experiment 4 is shown in Figure 3 (b)).

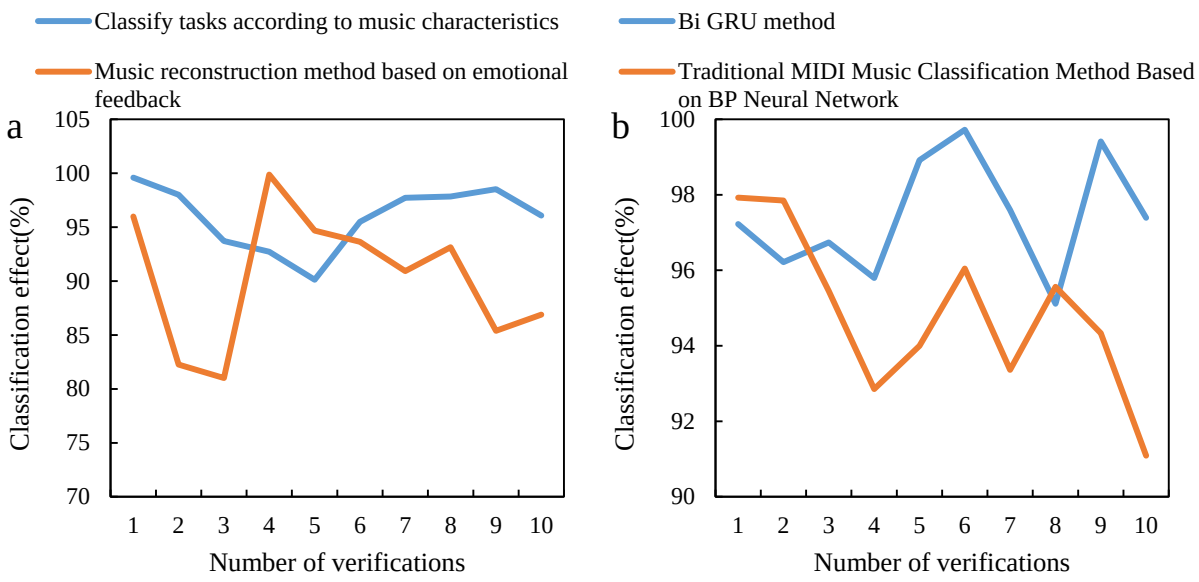

Figure 3: Comparison of Experiments 1 to 4

It can be seen from the comparison between Experiment 6 and Experiment 5 that in Bi GRU, the attention mechanism of Bi GRU was introduced, and the weight of attention was distributed on each segment to increase the attention to certain segments, which can better reflect the characteristics of music, better reflect the style of music, and thus improve the accuracy of classification. In experiment 5, the music classification method in this paper was used. The accuracy rate of 93.2% was achieved, and the classification effect was the best. The comparison between this method and the attention mechanism method is shown in Figure 4.

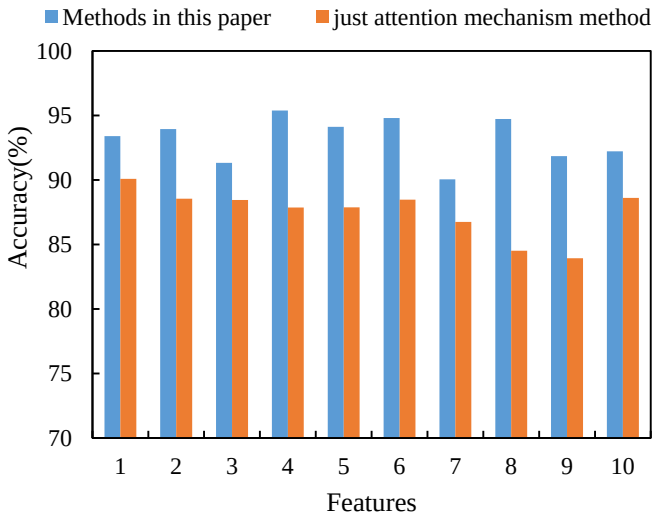

Figure 4: Comparison between this method and attention mechanism method

The change process of accuracy and loss function during network model training in this paper is shown in Figure 5 (accuracy change is shown in Figure 5 (a), and loss function change is shown in Figure 5 (b)). In the early and middle stages of training, with the increase of training times, the accuracy of training set and check set was gradually improved, and the corresponding loss function was also gradually reduced, indicating that the network model was being optimized.

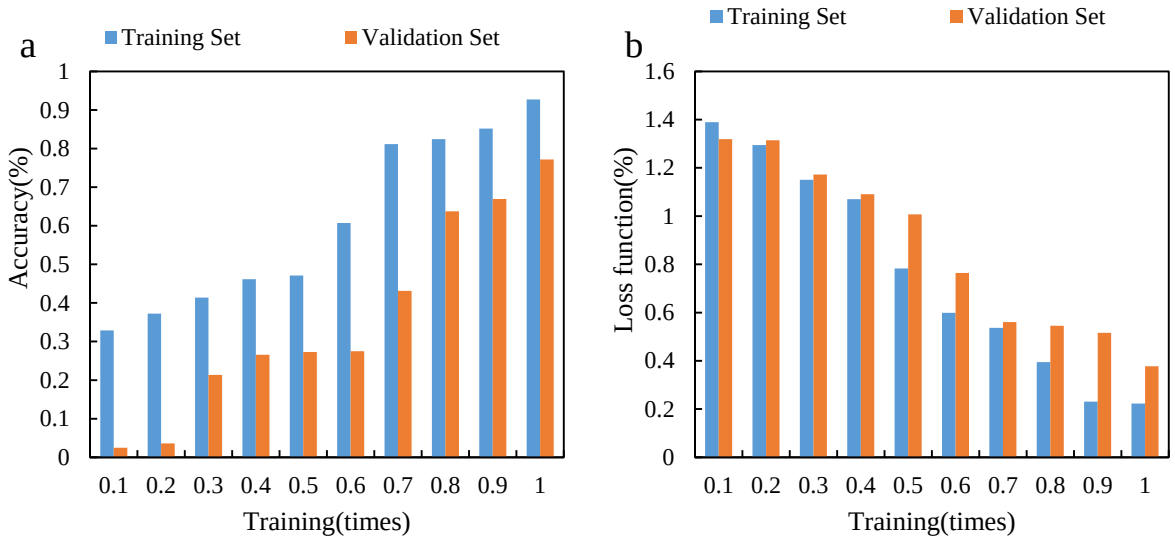

Figure 5: The change process of accuracy and loss function during the training of this network model

After training the network model, the fuzzy matrix was used to evaluate the classification effect of the network model. Among them, the recognition accuracy of metal music, classical music and folk music was 93%, 92.7% and 93.3% respectively. There were some incorrect distinctions between dance music and country music. Because some country music can be used as the accompaniment of country music, its form was similar to dance music, but some music was wrongly classified as country music. There was a certain deviation between dance music and metal music, because they both payed attention to rhythm and have similarities. In a word, the above five types of MIDI music can be better classified by using the method proposed in this paper, and the accuracy basically meet the requirements. The classification effects of different genres of music are shown in Figure 6 (the classification effects of metal music, classical music and folk music are shown in Figure 6 (a), and the classification effects of dance music and country music are shown in Figure 6 (b)).

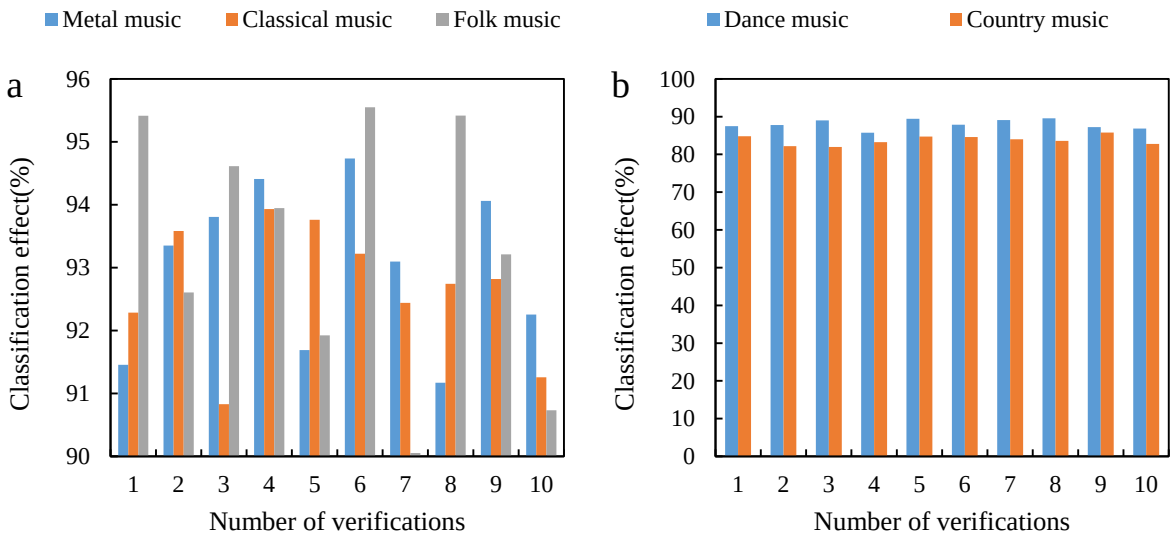

Figure 6: Music classification effect of different genres.

The similarity between Skyline algorithm and MTC algorithm varied with the number of track blocks, as shown in Figure 7. The similarity between the topic extracted by Skyline algorithm and the standard topic would decrease with the increase of the number of audio tracks, and the MTC algorithm had no significant reduction. Therefore, this paper used MTC algorithm to extract music theme.

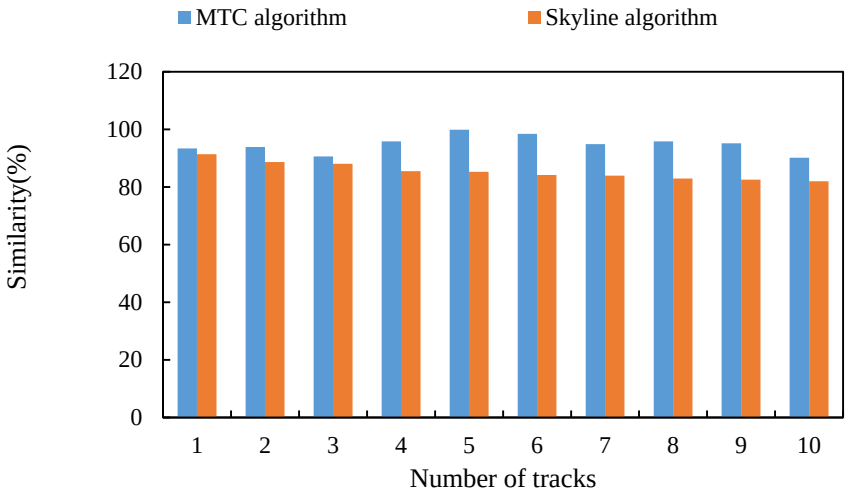

Figure 7: The similarity between Skyline algorithm and MTC algorithm changes with the number of track blocks

Using the classification model proposed in this paper, its classification effect was significantly higher than that of the emotional model based on brainwave music, while the classification effect of the emotional model based on brainwave music was not satisfactory. The main reason was that the emotional mode based on brainwave music was extracted from the main melody of the work, and did not include the accompaniment information of the music. The characteristic of this paper was to include the accompaniment information. In addition, the emotional model based on brainwave music cannot fully reflect the information of the theme, and cannot distinguish the popular style from the classic style. In this method (the method in this paper), the input was all the features of a piece of work, and the features of each segment were sampling codes of a segment, basically including all the information of this part. In other words, the input of this method almost contained all the information of the whole work. In addition, this paper used GRU (Gate Recurrent Unit) as the classification network, which was a RNN (Recurrent Neural Network) method specially used for sequence data processing. For segment features with sequence attributes, its classification performance was good. The classification accuracy of each style of the two methods in the test set is shown in Figure 8.

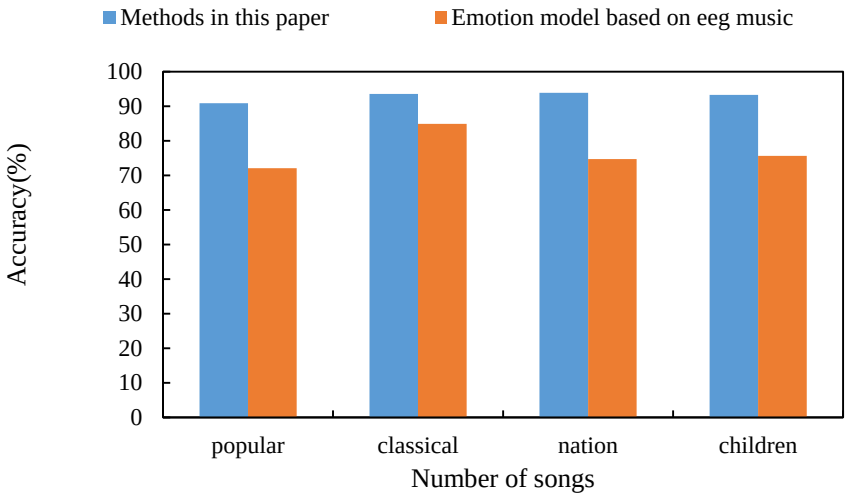

Figure 8: Similarity of the two algorithms changes with the number of track blocks

In the space with unit beat time as the vertical axis, there would be a climax when the long beat, drag beat and speed slowed down temporarily. Taking Rain in June as an example, the peak of red dots was a tone with a higher pitch in the phrase and a longer duration. In the visual performance of music, the long tone was the most important. Therefore, using this way of expression can emphasize this effect. The singing results of different singers are shown in Figure 9.

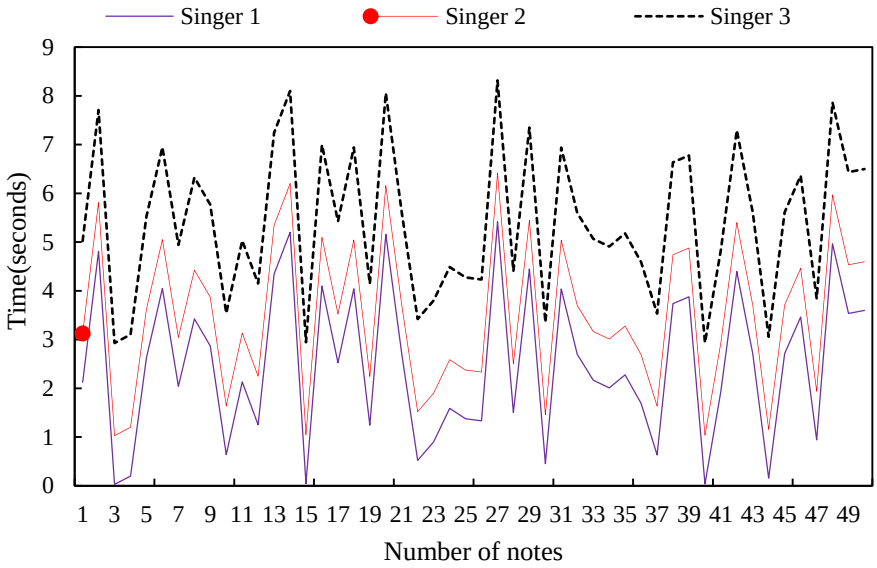

Figure 9: Singing results of different singers

Music information is the specific manifestation of music, and music acoustics is the scientific basis of music. The sound of music cannot be seen directly, so it would have some vague feelings. Although it has no impact on the feeling of art, there are many problems in the music discipline and music teaching. The visualization of music information refers to the visual representation of music information according to the basic principles of music acoustics, so as to achieve an intuitive effect, overcome the fuzzy phenomenon in music teaching and scientific research, and improve the accuracy and purpose of music information

transmission. In fact, the collection and analysis of sound and non sound information of music makes various details of music more diverse and detailed, and can convert music information into data, and use terminal oriented technology to transfer music information to two-dimensional or three-dimensional space, thus enriching the visual form and content of music. Before the 20th century, most of the visualization of music information was not through “data” as a medium, but through music, electricity, light and other forms.

With the rapid development of computer image and video processing technology, people can obtain massive information more conveniently and quickly. As a traditional way of entertainment, music shows its unique charm and vitality under the influence of the network. A music visualization system can have only one or more, but usually only one main function. For example, the music visualization system used in music performance occasions, game design, film and television and other entertainment places has an ornamental role. Under the general and specific optical environment, the physical vibration mechanism of the instrument is studied by using visualization technology, and its role is to observe. Music information visualization is an auxiliary expression in the field of music informatics and other disciplines. It has strong observation power. The visualization system of music information plays a prompt role in music teaching.

Image is an objective thing observed by a variety of observation systems in different ways, which can directly or indirectly affect people’s eyes, thus forming a material form of vision. Usually, the result of image acquisition is a series of energy. Therefore, images are usually represented by matrices or arrays. The coordinates of each element represent the location of the site, and the value of the element is a specific feature of the site attractions. Video is a series of digital images that change with time, and it also contains some sound, which is a very important information.

In most scenes, moving objects are the most noticeable to human eyes. In fact, the main features of image and video are rich original data, strong correlation between adjacent frames, and dynamic changes in time domain, which make it possible to detect, segment and recognize moving objects. This not only provided an important theoretical basis for content-based interactive services and image compression technology, but also provided a theoretical basis for some advanced applications of video technology.

The preprocessing of music video images improved the ability of visualization of regions of interest, and provided convenience for future work. Therefore, although the preprocessing method is simple, it is more important and directly affects the quality of music image in the next step. Music video image processing is a technology based on image features, which can be processed according to different application environments and targets. It is generally used to improve the visual quality of video images. At present, due to the limitations of video capture, presentation, video transmission and other technologies, the quality of music video images that users can see in real life would be reduced, and sometimes it would become unbearable. Real music images are often degraded due to some random errors, commonly referred to as noise. In the process of music video image acquisition, presentation, transmission and processing, there are noises, which can be content related or independent. The noise generated by the video image capture device can be divided into two categories: spatial and temporal. The spatial noise is caused by the change of the spatial pixel value of the image. In human visual features, and in plane images, the human eye is more sensitive to the fluctuation of noise. Time domain noise refers to the random error generated by the sensor of the same video image at different time points. In addition, due to the limitations of the music video display device itself, the accuracy of its expression would bring noise. Channel noise would be introduced in the transmission process, and error noise would also be caused when encoding music images.

4 Conclusions

Nowadays, people’s exploration and research on audio visualization is still growing. The way and scope of visualization are also changing, and the carrier of visualization is also constantly updating. Although the product design of virtual reality technology and visualization technology is not mature, and the combination of the two technologies is far from perfect, with the progress of technology, music visualization technology would become better and better. Therefore, this paper can try to understand the shortcomings of some modern visualization products, learn their advantages, and optimize them. Tunes play a great role in expressing emotions and influence. The melody of the music is also a distinctive feature of the music’s own emotional changes. It would not only affect the direction of the tone, but also the length of the tone. On this basis, this paper improved the design of audio visualization scheme through examples, but there are still some problems. For example, the improvement range of precision, the design of visualization scheme of music emotional theme, the inconvenience caused to the conversion of individual music libraries, and the oneness of visualization carrier are all future research directions.

About the Author

Xin Chen was born in Jingmen, Hubei, China, in 1996. Her master's degree from Guangxi Arts University, and her Doctor's degree from Sookmyung Women's University. Her research interest include Music and Dancology.
E-mail: cx20217653@163.com

Liqiong Duan was born in Zhangjiakou, Hebei, P.R. China, in 1982. She received a Bachelor’s Degree from Langfang Normal University, P.R. China. Now, she works in Admissions Office of Zhangjiakou Open University. Her research interest focuses on music education teaching.
E-mail: momo_918@163.com

Liu Xiaowei was born in Zaozhuang City, Shandong Province, China, in 1995. She received her bachelor's degree from Yangzhou University, her master's degree from Guangxi Arts University, and her doctoral degree from Sangmyung University. Her research interests include vocal music (bel canto).
E-mail: heidi9527@163.com

References

[1] Wu, Yongmeng. "Open symphony: Creative participation for audiences of live music performances." IEEE MultiMedia 24.1 (2017): 48-62.

[2] Khulusi, Richard. "A survey on visualizations for musical data." Computer Graphics Forum 39.6(2020): 82-110.

[3] Lima, Hugo B., Carlos GR Dos Santos, and Bianchi S. Meiguins. "A survey of music visualization techniques." ACM Computing Surveys (CSUR) 54.7 (2021): 1-29.

[4] De Prisco, Roberto. "Understanding the structure of musical compositions: Is

visualization an effective approach?." Information Visualization 16.2 (2017): 139-152.

[5] Jin, Yucheng. "Effects of personal characteristics in control-oriented user interfaces for music recommender systems." User Modeling and User-Adapted Interaction 30.2 (2020): 199-249.

[6] Birajdar, Gajanan K., and Mukesh D. Patil. "Speech and music classification using spectrogram based statistical descriptors and extreme learning machine." Multimedia tools and applications 78.11 (2019): 15141-15168.

[7] Wang, Bryan, and Yi-Hsuan Yang. "PerformanceNet: Score-to-audio music generation with multi-band convolutional residual network." Proceedings of the AAAI Conference on Artificial Intelligence 33.01(2019) :1174-1181.

[8] de la Fuente, Carlos. "Multimodal image and audio music transcription." International Journal of Multimedia Information Retrieval 11.1 (2022): 77-84.

[9] Herremans, Dorien, and Ching-Hua Chuan. "The emergence of deep learning: new opportunities for music and audio technologies." Neural Computing and Applications 32.4 (2020): 913-914.

[10] Pandeya, Yagya Raj, and Joonwhoan Lee. "Deep learning-based late fusion of multimodal information for emotion classification of music video." Multimedia Tools and Applications 80.2 (2021): 2887-2905.

[11] Liu, Caifeng. "Bottom-up broadcast neural network for music genre classification." Multimedia Tools and Applications 80.5 (2021): 7313-7331.

[12] Ripani, Giulia. "What do adults think of music across the lifespan?." Psychology of Music 49.6 (2021): 1701-1720.

[13] Way, Lyndon CS. "Populism in musical mash ups: recontextualising Brexit." Social Semiotics 31.3 (2021): 489-506.

[14] Craton, Lincoln G., Geoffrey P. Lantos, and Richard C. Leventhal. "Results may vary: Overcoming variability in consumer response to advertising music." Psychology & Marketing 34.1 (2017): 19-39.

[15] Murodova, Durdona. "Scientific And Theoretical Aspects of Musical Thinking." Zien Journal of Social Sciences and Humanities 1.1 (2021): 196-199.

[16] Shang, Lanyu. "CCMR: A Classic-enriched Connotation-aware Music Retrieval System on Social Media with Visual Inputs." Social Network Analysis and Mining 11.1 (2021): 1-14.

[17] Kirkman, Andrew, and Philip Weller. "Music and image/image and music: the creation and meaning of visual-aural force fields in the later Middle Ages." Early Music 45.1 (2017): 55-75.

[18] Briot, Jean-Pierre, and François Pachet. "Deep learning for music generation: challenges and directions." Neural Computing and Applications 32.4 (2020): 981-993.

[19] Polo, Antonio, and Xavier Sevillano. "Musical Vision: an interactive bio-inspired sonification tool to convert images into music." Journal on Multimodal User Interfaces 13.3 (2019): 231-243.

[20] Lee, J., and J. Nam. "Multi-Level and Multi-Scale Feature Aggregation Using Pre-trained Convolutional Neural Networks for Music Auto-tagging." IEEE Signal Processing Letters 24.8(2017):1208-1212.