www.cad-journal.net
Computer-Aided Design AND APPLICATIONS

Computer Aided Music Visualization Teaching Model Based on Emotional Feature Detection

Yu Deng[1], Jie He[2] and Linyuan Hu[3]

[1]Faculty of Education and Arts, College of Arts and Science of Hubei Normal University, Huangshi, Hubei 435109, China, dengyu503@126.com
[2]Shool of Computer Science and Technology,Wuhan University of Bioengineering, Wuhan, Hubei 430415, China, hhjjkingshejie@sina.com
[3]School of Music, Wuhan University of Bioengineering, Wuhan, Hubei 430415, China, lyhu@whsw.edu.cn

Corresponding author: Linyuan Hu, lyhu@whsw.edu.cn

Abstract. Computer-aided instruction (CAI) technology has an important influence in the stage of music education and teaching, no matter in educational concept, educational content or curriculum form. With people paying more and more attention to music education, the application of CAI in music teaching has become a hot spot in today's education. This article presents a computer-aided visual music instructional model based on emotional feature detection. In this model, the graph-convolution network (GCN) is used to identify the emotional characteristics of music, and the emotional information in music is extracted and transformed into a visual form to help students better understand and feel the emotional expression in music. The application of this model can transform the audio characteristics of music works into visual images through image processing technology, so that students can better understand the emotional characteristics of music works. Moreover, the flexibility and operability of CAI system also provide more possibilities and creativity for music teaching. The model proposed in this article has a wide application prospect and can provide an effective auxiliary tool for music teaching and music emotion research.

Keywords: Emotional Recognition; Computer-Aided Instruction; Musical Visualization
DOI: https://doi.org/10.14733/cadaps.2024.S10.46-60

1 INTRODUCTION

Music is a borderless art form, which can express people's feelings and thoughts through special languages and symbols. In music teaching, how to make students better understand and feel the emotional characteristics in music works is very important. Bahari [1] analyzed the application of computer-aided nonlinear dynamic methods in second language teaching in blended and distance

learning. The nonlinear dynamic method uses complex system theory to explore the dynamics of language development in a nonlinear manner. This method can demonstrate how language evolves in social contexts by simulating the interactive behavior of language users. In second language teaching, computer-aided nonlinear dynamic methods can provide an interactive environment, allowing students to learn and understand language through participation and practice. For example, by simulating language interaction in real contexts, students can engage in dialogue, writing, and other language activities in a virtual environment, thereby improving their listening, speaking, reading, and writing skills. In addition, computer-aided nonlinear dynamic methods can also evaluate students' language proficiency and development trends through data analysis. By tracking and recording students' language behavior in a virtual environment, teachers can obtain real-time feedback on students' language abilities for personalized teaching guidance. In summary, the application of computer-aided nonlinear dynamic methods in second language teaching in blended and distance learning can provide students with a more interactive, practical, and personalized learning environment, while improving teachers' teaching efficiency and students' learning outcomes. However, traditional music teaching methods often focus on theoretical knowledge and skill training, while ignoring the identification and understanding of emotional characteristics.

Music is an art form that can touch human emotions. In music teaching, it is very important to understand the emotional characteristics in music works. Bogach et al. [2] used a technique called "transfer learning" to further improve the performance of the model. Transfer learning involves utilizing pre trained models that have already been trained on other datasets, and then fine-tuning them on the new dataset. The advantage of this method is that it can utilize the common features already learned by the pre trained model, which can help the new model better process new image data. Firstly, it is necessary to extract features related to rhythm from the audio. These features may include pitch, energy, duration, etc. Pitch can reflect the rhythm of music, and changes in energy can also provide clues to rhythm. The information of duration (such as the length of syllables) can also be used for modeling rhythm. In this step, you can use deep learning models (such as recurrent neural networks, RNNs) or other machine learning models to learn both acoustic and language features. RNN models are particularly suitable for processing temporal data, such as voice signals or text, as they can capture time dependencies in audio signals. However, traditional music teaching methods often focus on theoretical knowledge and skill training, while ignoring the identification and understanding of emotional characteristics. In recent years, with the continuous growth of computer technology, computer-assisted music teaching has become a new research direction. One of the main learning objectives of the music theory course is to cultivate students' ability to analyze music forms. During this process, the charts in parentheses are a commonly used tool that can help students better understand the structure and elements of music while reading and analyzing music scores. CAI has an important influence in the stage of music education and teaching, whether in educational concept, educational content or curriculum form. With people paying more and more attention to music education, the application of CAI in music teaching has become a hot spot in today's education. Human emotions can be conveyed to others through body language, facial expressions, tone of voice, and other means. At the same time, people can also understand their emotions through self-perception, such as realizing whether their facial expressions are smiling or depressed, or hearing their voice tremble. In psychology, the relationship between self-awareness and emotional communication is considered a feedback mechanism that can help people better understand and manage their emotions. For example, when a person realizes that their body language and facial expressions are conveying angry emotions, they can try adjusting their posture and expressions to reduce the transmission of angry emotions and better control their emotions. Chaturvedi et al. [3] analyzed the signal classification results under the feature parameter music emotion model. A visual emotional framework was constructed by classifying human emotions and music types.

The automated emotional computing of biological signal analysis and deep learning methods is a complex but increasingly important field. This field mainly covers the use of various biological signals, such as speech, facial expressions, physiological signals, etc., for emotion recognition and

computation through deep learning algorithms. Firstly, it is necessary to collect a large amount of labeled data, including signals and corresponding emotional labels. This type of data typically needs to be collected from human subjects, each of whom generates signals in different emotional states. Filippini et al. [4] preprocessed the collected signals, including noise cancellation, feature extraction, and other steps. Select the most emotionally relevant features from the preprocessed signal. This can be achieved through methods such as mutual information and chi square testing. Select and train deep learning models. Use a portion of data that has not been trained to validate and test the performance of the model. This can help understand whether the model can accurately recognize emotions. Optimize and adjust the model based on the results of validation and testing. This may include changing the architecture of the model, adjusting hyperparameters, and increasing data augmentation. Emotional characteristics refer to the emotional elements expressed in music works, such as happiness, sadness and tension. In music teaching, understanding and identifying the emotional characteristics of music works will help students better understand and feel music works, and further deepen their ability to understand and feel music. However, traditional music teaching methods often only focus on theoretical knowledge and skill training, while ignoring the identification and understanding of emotional characteristics. This makes it difficult for students to understand the emotional characteristics in music works. Images are usually associated with emotions and emotions, and are supported by relevant psychological theories. With the growth of computer generation technology, music visualization and image style transfer have become an important part of computer information interaction, providing more personalized and customized services and works in the experience economy. Computer image processing technology has also been widely used in music visualization teaching. Music visualization teaching refers to transforming music works into visual images through image processing technology, so that students can better understand and feel music works. The purpose of this article is to construct a music visual instructional model based on CAI system by using emotional feature detection technology to improve students' ability to recognize and understand emotional features of music works. This model can not only help students better understand and feel music works, but also provide more flexible and efficient teaching means and methods for music teachers, and promote the continuous growth of music teaching.

Hai [5] conducted speech recognition tracking for computer-aware interaction. The speech recognition system converts recorded speech into text and sends the recognition results to a network assisted learning platform. The network assisted learning platform analyzes speech recognition texts through natural language processing technology, detects language errors, and provides corrective suggestions. At the same time, the system can also provide real-time feedback and evaluation based on students' oral expression, helping them understand their learning progress and shortcomings. Students revise and practice themselves based on systematic feedback and suggestions, continuously improving their English oral expression skills. The system can track students' learning progress and provide regular evaluation reports to help students understand their learning status and goal achievement, while providing teaching references and suggestions for teachers. In short, the computer-assisted teaching mode for intelligent English oral learning based on speech recognition and network assistance can provide students with personalized, efficient, and convenient English oral learning experiences, help improve students' oral expression ability, and provide teaching aids and data support for teachers. While enjoying music, it is an inevitable trend to obtain information such as music image and artistic conception through imagination and then visualize it. Visual music provides people with more colorful information, at the same time, people can easily accept this art form, and unconsciously form their own different musical art views on different music. In the past, the music that people understood was just a feast for the ears, especially for some people, it was just a complex and changeable signal set. The main goal of this article is to explore the relationship and expression between music characteristics and image expression, so as to deeply study the intelligent visual design of music information. The significance of the research lies in:

(1) By using emotional feature detection technology and computer image processing technology, a computer-aided music visualization instructional model based on emotional feature

detection is constructed, aiming at improving students' ability to recognize and understand emotional features of music works.

(2) This model is based on CAI system, and it can make use of the technical advantages of computer to realize the efficient teaching of music visualization. By transforming musical works into visual forms, students can understand and feel the emotional characteristics of musical works more intuitively.

The structure of this article is as follows: The first section introduces the background and significance of computer-aided music visualization instructional model based on emotional feature detection. The second section introduces the related research of emotional feature detection and computer-aided music visualization teaching in detail. The third section describes in detail the construction stage of computer-aided music visualization instructional model based on emotional feature detection. The fourth part verifies the application effect of emotional visualization model. The fifth section summarizes the research results of this article and the future research direction.

2 RELATED WORK

Han et al. [6] analyzed the change detection in drone images used for monitoring road construction progress. During the training phase, it is first necessary to collect a set of multi temporal images, which may include images of the construction area at different time points. Then, a technique called "image differentiation" is used to extract changing regions from these images. Specifically, this involves calculating the difference between every two images to generate a differential image that primarily highlights the areas where changes occur between the two images. Next, this differential image can be used to train a deep learning model. The specific model type can be selected based on specific requirements and dataset characteristics, such as convolutional neural networks (CNN). During the training process, the purpose of the model is to learn how to identify construction areas from differential images. To achieve this, a technique called "label propagation" can be used, which allows the model to predict the boundaries of the construction area on previously unseen images. Iyer et al. [7] constructed CNN and LSTM models for identifying emotions from EEG signals. The CNN model can be used to extract local features from EEG signals. The LSTM model can be used to capture time dependencies in signals. Train CNN and LSTM models using training sets. During the training process, ensemble learning methods can be used to fuse the prediction results of multiple models to improve recognition accuracy. Evaluate the trained model using a test set to obtain indicators such as accuracy, recall, F1 score, etc.

Optimize and adjust the model based on the evaluation results. Possible optimization methods include adjusting network structure, adding hidden layers, and changing activation functions. Applying the trained model to practical human emotion recognition tasks can enable real-time emotion classification of new EEG signals. It should be noted that EEG signals have high noise levels and individual differences, so appropriate preprocessing and feature extraction methods need to be taken when conducting emotion recognition based on EEG. In addition, to improve recognition accuracy, ensemble learning methods can be used to fuse the prediction results of multiple models. It is currently a difficult topic to abstract and expand the development and creation of computer music. Maba [8] has conducted background development on the topic of computer-aided music. It utilizes software programs to assist in the creative achievements of information technology. Its research topic focuses on the development of computer-aided music, especially the use of software programs to assist music creativity. This may include using algorithms and artificial intelligence technologies to generate new music works, or using music analysis tools to explore and understand the structure and features of music works. Through such research and development, computers can become powerful tools for music creators, providing new creative possibilities, and opportunities for deeper understanding and exploration of music. At the same time, it can also help music learners better understand music and improve their skills and creativity through practice and feedback. Mustaqeem [9] solves the transformation analysis of local emotional features with hierarchical correlation. By exploring the spatiotemporal learning of

the features of the correlation function, it constructs and analyzes the global information spatiotemporal clue probability weights of the loss function. This model contains multiple hierarchical structures, and each layer of ConvLSTM network can perform time series analysis on input features and extract local features at different time scales. These local features are then extracted and classified through the fully connected layer (FC) for global features. Pei and Wang [10] conducted a theoretical analysis of image data state development parameters for computer teaching visual training. By assisting in visual information transformation of music information, audio editing of teaching and training results can be better carried out. Process and analyze image data through visual training, while transforming music information into visual information for better audio editing and teaching training. Among them, computer teaching refers to the process of imparting knowledge and skills through computer technology and related tools. Visual training is the process and analysis of image data through specific image processing algorithms and techniques to achieve a certain goal. For example, visual training can be used to recognize objects in images, extract image features, and so on.

Through the auxiliary transformation of visual information, audio editing of teaching and training results can be better carried out. For example, visual information can be corresponded and matched with audio information to achieve more accurate audio editing and processing. Peng et al. [11] constructed a human centered holistic approach to examine second language learners' autonomous participation in mobile learning activities. It requires comprehensive consideration from learners' needs and goals, skills and levels, learning methods and preferences, as well as autonomy and self-management. In order to provide them with a more suitable, flexible, and personalized mobile learning experience. In the CALL method, computers play an important role. Computers can provide various forms of auxiliary tools, such as speech recognition, natural language processing, grammar checking, and vocabulary translation, to help learners improve their language skills. In addition, computers can provide feedback based on learners' performance and progress to help them improve their learning methods. Informal mobile language learning is a form of CALL method that utilizes the characteristics of mobile devices to enable learners to engage in language learning at anytime and anywhere. This learning method is very flexible and can meet the different needs of learners. For example, learners can improve their listening skills by listening to foreign language songs on their way to and from work, or communicate with locals through translation applications while traveling. Qiu et al. [12] analyzed an unsupervised music latent representation learning method based on deep 3D convolutional denoising autoencoder. 3D-DCDAE is an unsupervised music latent representation learning method based on deep 3D convolutional denoising automatic encoder. This method utilizes a three-dimensional convolutional neural network to convolve music signals, capturing temporal information in the music signal, and automatically encodes and decodes it through a denoising autoencoder to achieve unsupervised learning of potential musical representations. Using a three-dimensional convolutional neural network to perform convolution operations on music signals and capture temporal information in the music signal. Using a denoising autoencoder to automatically encode and decode the convolutional music signal to remove noise and extract potential musical representations. By minimizing reconstruction errors, potential representations of music can be learned. Use clustering algorithms to cluster and analyze the learned potential representations to discover similarities and patterns in music. Quan [13] analyzed the development of a computer-aided classroom teaching system based on machine learning prediction and artificial intelligence KNN algorithm. It has designed and developed a user-friendly interface that allows teachers and students to use the system. Integrate the KNN model into the system to achieve prediction and feedback functions. Test the system in a real environment, collect feedback, and make necessary optimizations. Deploy the system to a server or cloud, allowing teachers and students to access it anytime, anywhere.

Regularly update models and systems to adapt to new requirements and data. This is a complex project that requires knowledge and skills in various fields such as computer science, data science, and education. During the development process, many other factors need to be considered, such as system scalability, security, privacy protection, etc. Shalini et al. [14]

constructed a music recommendation system based on facial emotions using computer vision and machine learning techniques. It combines technology from computer vision, machine learning, and music recommendation systems. Next, I will explain in detail the working principle of this system. Firstly, the core of this system is facial emotion recognition. Through computer vision technology, the system can analyze facial expressions and identify various emotions. This is mainly achieved through the use of deep learning models, such as convolutional neural networks (CNN). The model abstracts facial features from a large number of facial expression images and learns to associate them with specific emotions. Then, this system will recommend music based on the recognized emotions. The main technology used here is machine learning. A possible approach is to establish a music library where each piece of music is associated with one or more emotions. When the system recognizes a user's emotion, it searches for music that matches that emotion and then recommends it to the user. In addition, this system may also consider other factors, such as user music preferences and history, to provide more personalized recommendations. This can be achieved through other machine learning algorithms, such as collaborative filtering or content filtering. Suzuki et al. [15] constructed an emotion estimation model based on the eeg/hrv index using feature extraction and selection algorithms. This model requires extracting features from EEG and HRV signals. There may be many extracted features, but not all of them are useful. To improve the performance of the model, various feature selection methods can be used to select the most important features. Choose a suitable machine learning algorithm to construct an emotion estimation model for the model. The goal of this model is to predict or classify emotions based on the selected features. Use a portion of data (usually a part of the training set) for cross validation and parameter optimization. Then use another portion of the data (usually the test set) to evaluate the performance of the model. Evaluate the performance of the model based on evaluation indicators such as accuracy, recall, F1 score, etc. Further analysis can also be conducted, such as exploring the predictive ability of the model for different emotional types, or analyzing the performance of the model for different individuals or groups. The above steps can be adjusted based on specific research objectives and datasets. If the data volume is large, a deep learning model can be considered. If feature selection is complex, integration methods can be used. The visual psychological imagery triggered by music does not refer to specific and vivid visual images or images. It refers to a more abstract visual experience, such as the emotions conveyed by music, the image of a musician or instrument. Even the visual understanding of the scenes or stories described by music. This experience may include elements such as color, shape, motion, and space, as well as their changes and combinations in time and space [16]. Yuan [17] used CAD technology to enhance his thinking ability in the field of static multimedia teaching. When describing the motion of a parabola, CAD software can be used to simulate the motion process of the parabola, and the shape, velocity, angle, etc. of the parabola can be adjusted and demonstrated, enabling students to intuitively understand the motion law of the parabola. The use of CAD technology to enhance thinking ability in the field of static multimedia teaching is a very promising direction. However, this also requires teachers to make corresponding adjustments and improvements in teaching methods and technology applications. In addition, the application of CAD technology in multimedia teaching requires detailed design and implementation based on specific teaching content and objectives.

3 CONSTRUCTION OF COMPUTER AIDED VISUAL MUSIC INSTRUCTIONAL MODEL

3.1 Music Emotion Recognition Model

To divide music into segments, we first divide the music into small segments with unit time, and then connect the small segments that can express similar emotional types to form music segments. In this article, we choose the bar length of the music itself as the unit time, and then judge whether the bars can be connected according to the similarity of the characteristics of adjacent bars. The GCN model of musical emotion recognition is shown in Figure 1.

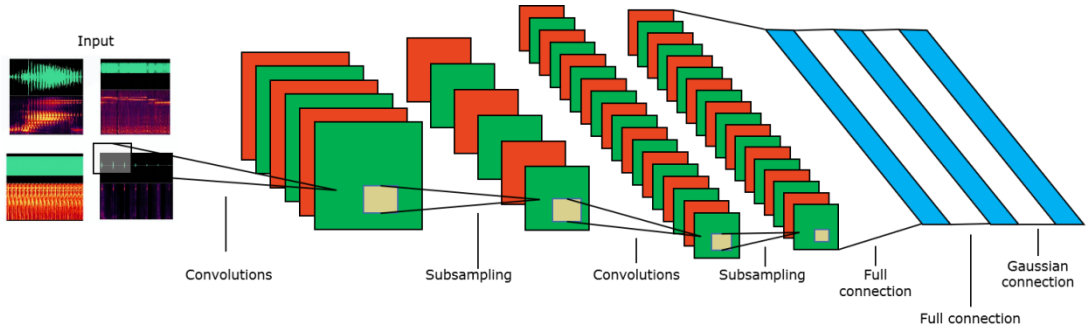

Figure 1: GCN model of music emotion recognition.

This system is designed as a real-time interactive music visualization system based on emotion recognition, which uses an optional virtual character to express emotional actions. The system needs to reduce the system load and perform a music analysis and visual performance in real time at the same time. The music features extracted from the music input of the live performance microphone and electronic organ are processed and filtered by relevant information, then sent to the music animation script and engine, and finally the music emotion is expressed by virtual characters. Assuming that an L-level GCN works on a dependency graph $g=(v,\varepsilon)$, where v,ε is a node set and an edge set respectively, then that output of the k lay node i is expressed as:

$$h_i^{(k)}=\rho\left(\sum_{j=1}^{q}A_{ij}W^{(k)}h_j^{(k-1)}+b^{(k)}\right) \tag{1}$$

Where, $h_j^{(k-1)}$ represents the output representation of node j at the $k-1$ layer GCN, $W^{(k)}$ represents the weight matrix, $b^{(k)}$ represents the partial differential vector, $\rho(\cdot)$ represents the activation function RELU, and A represents the adjacency matrix of the dependency in the dependency tree.

The probability occupation of emotional analysis is described as:

$$P(i_j|k,\theta)=\frac{\exp(x_j(k,\theta))}{\sum_{1\le i\le|X|}\exp(x_j(k,\theta))} \tag{2}$$

Where $x_j(k,\theta)$ is the average pool result; Parameter set θ corresponds to class j; The class space is represented as X. In order to minimize the negative logarithmic probability, the random gradient descent method is used. The loss function uses the cross-entropy function:

$$loss=-\sum_i\sum_j y_i^j\log\hat{y}_i^j+\lambda\|\theta\|^2 \tag{3}$$

Where y is the expected value, $\hat{y}$ is the predicted value, λ is L2 regularization, and θ is the parameter set of the neural network.

The cognitive consciousness of music performance can take many forms. When hearing a special melody about happiness, the character may feel happy, or when receiving a series of specific chords that feel horrible, the character may enter a state of fear. If the emotions of adjacent sections are similar, then their differences in characteristics are within a range to some extent. The feature selected in this article is the feature vector composed of the pitch, length and

intensity of notes. If their characteristic difference is less than a certain threshold range, then the two sections are connected.

It is assumed that all the training sample data can be fitted linearly without error, and the distance from the sample point to the hyperplane is less than or equal to ε as a penalty function. Because the function f is unknown, we can only use the linear regression function $f(x)=w\cdot x+b$ to fit the sample data according to the collected samples, and get:

$$\begin{aligned} &y_i-(w\cdot x_i+b)\le\varepsilon \\ &(w\cdot x_i+b)-y_i\le\varepsilon, \qquad i=1,2,\cdots,n \end{aligned} \tag{4}$$

SVM classification algorithm can transform the problem of finding hyperplane solution into a solution:

$$\begin{cases} \min \|w\|^2/2 \\ s.t. \quad y_i(w\cdot x+b)\ge 1 \end{cases} \tag{5}$$

The corresponding prediction function is:

$$f(x)=\operatorname{sgn}\left(\sum_{i=1}^{n}a_i y_i\langle x_i\cdot x\rangle+b\right) \tag{6}$$

Because different features have different ability to distinguish emotions, and the feature dimension is too high, it is easy to be over-fitted. Therefore, in order to shorten the modeling time and improve the progress of the model, it is necessary to screen out the key features in the input features. In this article, the proportional selection operator based on relative fitness is used for selection.

It is also important to interact with virtual characters in a natural musical expression. Therefore, this article should ensure that once the system designed in this way is running, users can interact with the program only by providing music input mode of controller or microphone. Microphones and controllers should be placed in specific locations so that users can easily observe the reaction of virtual characters when performing in real time. The threshold of similarity of feature vectors is used to judge whether the similarity of two bars is enough to connect them to form a music segment. If the similarity of feature vectors of two bars is less than this value, then they are connected, otherwise, they are regarded as the dividing point of the music segment and belong to the two music segments before and after.

The function of music perception filter layer is two-sided. On the one hand, the system needs to analyze the complex music data input stream in order to extract meaningful music features to prepare for the related work in the future. On the other hand, in order to simulate the emotional response of human beings to music, the system needs to adopt the previous research methods on music emotion to organize these characteristics. The minimum threshold of segment length indicates how many bars are connected together in a segment. If the quantity of bars in a segment is less than this value, then the next bar belongs to the segment. Otherwise, start a new passage from the next section. The maximum threshold of segment length indicates how many bars can be connected together at most in a segment. If the quantity of bars in a segment is less than this value, then continue to judge whether the next bar belongs to the segment, otherwise, start a new segment from the next bar.

Usually, the calculation method of the emotional tendency of the whole music is to accumulate the emotional tendencies of all the melodies, and the emotional tendency of the melodies is determined by the sum of the emotional tendencies of the characteristic words, so that the overall tendency of the document can be measured by gradually accumulating. The emotional tendency of characteristic words is:

$$tendency=\frac{1}{n}\sum_{i=1}^{n}sim\left(word,seed_{1i}\right)-\frac{1}{m}\sum_{i=1}^{n}sim\left(word,seed_{2j}\right) \tag{7}$$

It is found that when $a_i \neq 0$, the corresponding vector X_i is the support vector. Therefore, the decision function is rewritten as:

$$f\left(x\right)=\operatorname{sgn}\left(\sum_{i=1}^{M}a_i y_i K\left(x,x_i\right)+b\right) \tag{8}$$

Where M represents the quantity of support vectors.

The calculation of similarity between two words is transformed into the calculation of similarity between two concepts. Similarly, the similarity calculation between concepts will be transformed into the similarity calculation between sememes:

$$Sim\left(p_1,p_2\right)=\frac{a}{d+a} \tag{9}$$

Where p_1 is the original meaning 1; p_2 is original meaning 2; d is the path length of p_1, p_2 in the semantic tree system; a is an adjustable parameter.

3.2 Visual Modeling of Musical Emotion

In this article, a music visualization system is established around music emotion, which can be used for music analysis and music visualization. Through the real-time voice and processing input of live performances, through processing and analysis, it is finally mapped to virtual characters. The system takes a microphone and an electronic organ as input parameters, extracts relevant music features through processing, then carries out emotion detection processing, and expresses emotional factors in real time with changes in facial expressions and actions of virtual characters and changes in virtual scenes, and finally displays them on a big screen. The structure of music emotion visualization model is shown in Figure 2.

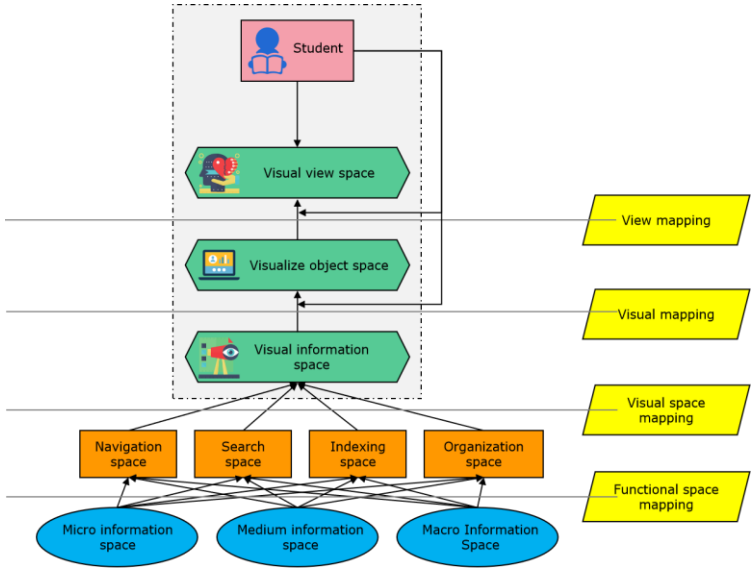

Figure 2: Visual model of music emotion.

Emotional description can be expressed by discrete words, and there are always clear boundaries and differences between these words. However, people's emotional changes are gradual, smooth and transitional, and there is no clear dividing point to distinguish happiness, calmness, depression and pain in people's emotions. Therefore, the corresponding relationship between music works and emotions is not fixed and clear, but loose and vague.

In the music perception filtering layer, the system extracts important music features by analyzing the input audio information, and then sends the results of this layer as input to the role and scene cognition layer. Finally, these extracted music features are sent to the feature recorder for the next stage of processing, so as to meet the requirements of complex real-time audio analysis. The expression of musical emotion is expressed by the melody composed of notes connected according to certain rules. It is assumed that all the training sample data can be fitted linearly without error, and the distance from the sample point to the hyperplane is less than or equal to ε as a penalty function. Because the function f is unknown, we can only use the linear regression function $f(x)=w\cdot x+b$ to fit the sample data according to the collected samples, and get:

$$\begin{aligned} &y_i-(w\cdot x_i+b)\le\varepsilon \\ &(w\cdot x_i+b)-y_i\le\varepsilon, \qquad i=1,2,\cdots,n \end{aligned} \tag{11}$$

SVM classification algorithm can transform the problem of finding hyperplane solution into a solution:

$$\begin{cases} \min \|w\|^2/2 \\ s.t. \quad y_i(w\cdot x+b)\ge 1 \end{cases} \tag{12}$$

The corresponding prediction function is:

$$f(x)=\operatorname{sgn}\left(\sum_{i=1}^{n}a_i y_i\langle x_i\cdot x\rangle+b\right) \tag{13}$$

In the cognitive layer, the information sent by the perceptual layer needs to be received and analyzed to simulate the internal state of the virtual character. The feature organizer contains the information of these live music performances, the importance data of each singing note, the information about singing rhythm and the chord data played by users. Then this article needs to simulate a cognitive consciousness of perceptual data, because the role needs to have an interesting internal state expressed through actions.

4 RESULT ANALYSIS AND DISCUSSION

Human emotions are always changing, just like the sea, sometimes quiet and sometimes surging. When listening to music, this kind of emotion will be more athletic, just like the sea water pushed by the wind, which fluctuates with the speed of the wind. Under the rendering of music, the computer seems to be a free and easy painter, describing to the audience how it feels after hearing music, and gradually accumulating a repainted background image with background music emotions over time, thus forming synaesthesia of music hearing and vision, and visually expressing it through intelligent design. The level, rhythm and speed of music are constantly changing, and it is these changes that reflect the changes of emotions. Therefore, in the stage of music appreciation, emotions are constantly changing and moving with the movement of music, forming a curve of ups and downs.

Using GCN as the emotion recognition model, the algorithm is tested by MSE and RMSE, and its formula is as follows:

$$\square\ MSE = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - y'_k\right)^2 \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\left|y_k - y'_k\right|\right)^2} \tag{11}$$

Where y_k is the actual value and y'_k is the output value. The MSE results of different algorithms are shown in Figure 3. RMSE results of different algorithms are shown in Figure 4.

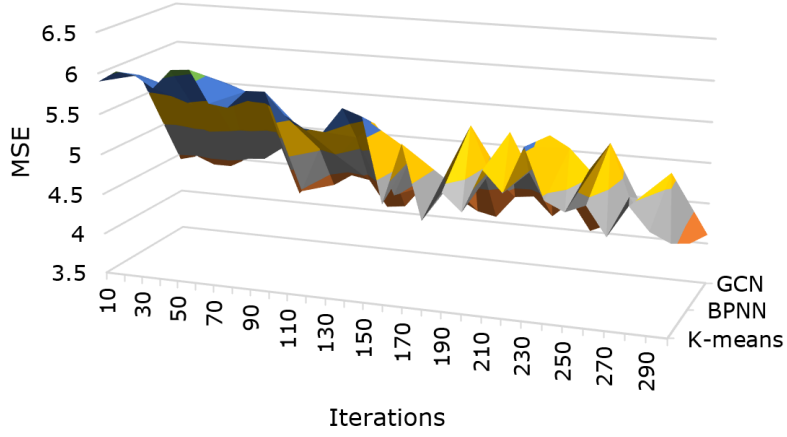

Figure 3: MSE results.

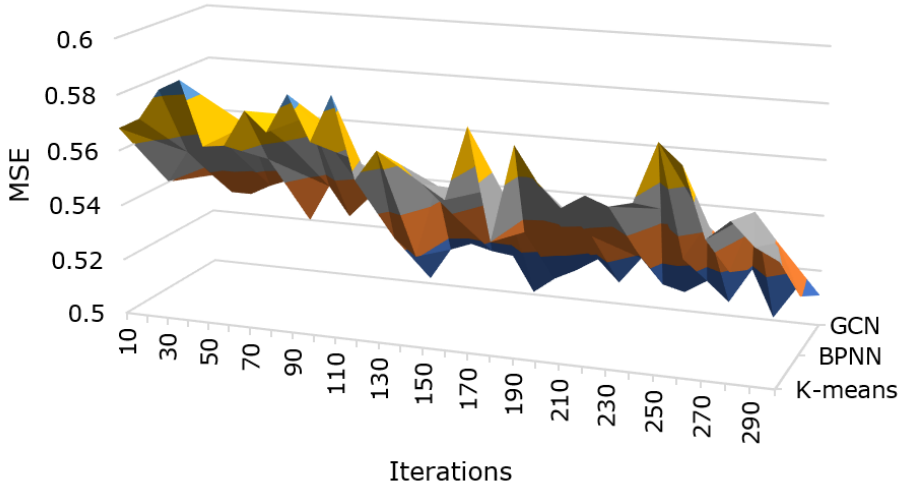

Figure 4: RMSE results.

The error of this algorithm is small. Similar to speaking, the ascending melody will give people positive emotional feelings, even solemn and passionate emotional feelings, while the descending music can express sad and sad emotions, while the calm and comfortable emotions can be expressed by parallel melodies. Because it is not only the pitch that affects the directionality of melody, the duration of notes will also affect the directionality of melody.

For each pattern, the category of its neighboring patterns will affect its category. Therefore, if the quantity of patterns in a certain category is very large, it will affect the classification of all patterns, that is to say, the categories of all patterns will tend to the category with the largest number, resulting in a wrong classification result. The first training recognition rate fitting curve and the final training recognition rate fitting curve of the model are shown in Figures 5 and 6.

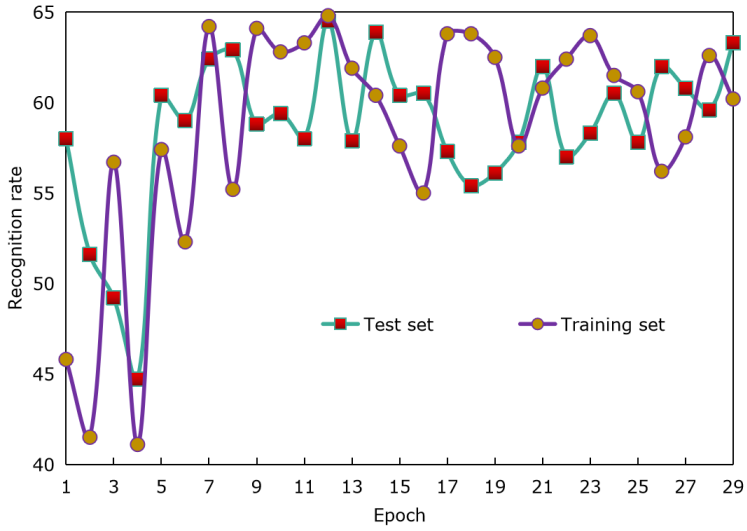

Figure 5: First training recognition rate curve.

Figure 6: Recognition rate curve for the 20th training.

When the quantity of iterations of the model increases, the size of the training set will also decrease, thus improving the fitting degree of the model. After preliminary training, the prediction accuracy of this method has reached 65.5%. In 20 training stages, the final model test accuracy reached 94.8%.

In the cognitive layer, the internal state of the role is defined. Then this internal state is visualized by the animation generated automatically in the expression layer. In order to achieve this visual effect of music, the system needs to create a complex enough role to generate behavior, which is considered by users and observers to be a timely response to music input. In the classification cognitive model, the music feature vector obtained from the music feature model is used as input, and the feature vector is mapped to the music emotion model by using the pattern recognition algorithm provided by the classification cognitive model to get the category of music. Process animation means that the motion or deformation of an object in animation is described by a process. The simplest process animation is to use a mathematical model to control the geometric shape and movement of objects, such as the movement of water waves with the wind. The more complicated ones include deformation, elasticity theory, dynamics, collision detection and so on. The animation technology based on physical model considers the genus of objects in the real world.

For the music emotion recognition system, when the quantity of feature parameters is large, its description of the target sound will be more detailed and its classification system will be greatly affected. Among a large quantity of music features, some will have little or even negative classification effect on the system, and if there are too many feature parameters, the calculation time of the system will be increased, thus reducing the operating efficiency of the system. The results of the identification time of different algorithms are given in Figure 7.

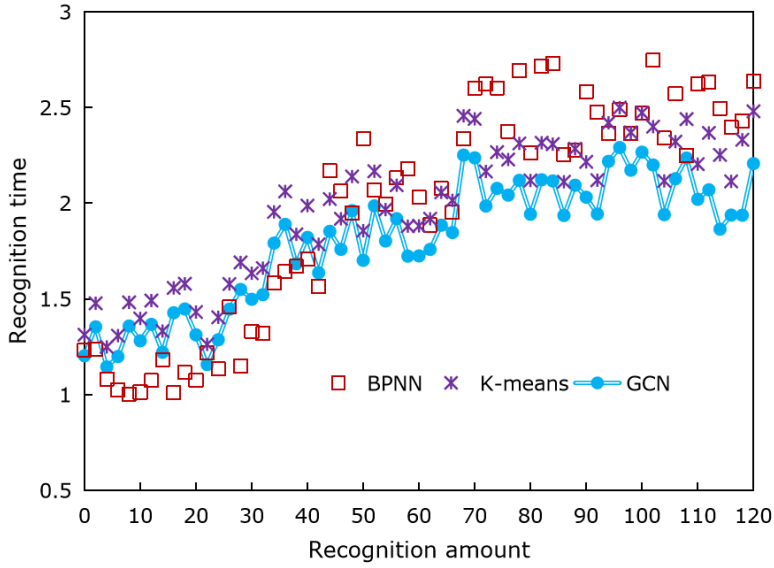

Figure 7: Experimental results of identification time.

The recognition time of different musical emotional features is also different. The method proposed in this paper has a shorter recognition time, because it blocks the notes according to the changes of their physical characteristics, thus efficiently extracting the emotional features of the notes. In contrast, the other two traditional methods are inefficient in extracting emotional features, resulting in longer recognition time. Different combinations of musical features will correspond to an infinite quantity of emotional feelings, and there may only be subtle changes and differences between these emotional feelings. Therefore, this should also be taken into account in emotional analysis. In the stage of emotional cognition, we should grasp the characteristics of music as a whole in order to get the most complete emotional experience. This method can not only identify

emotions in music, but also provide a deeper understanding, such as identifying the influence of different musical instruments or music types on emotions.

5 CONCLUSIONS

Visual music provides people with more colorful information, at the same time, people can easily accept this art form, and unconsciously form their own different musical art views on different music. This method divides notes into blocks based on changes in their physical characteristics, effectively extracting emotional features of notes, and has a shorter recognition time. In contrast, the other two traditional methods have lower efficiency in extracting emotional features, resulting in longer recognition time. In this article, a computer-aided visual instructional model of music based on emotional feature detection is proposed. GCN is used to identify the emotional features of music, extract the emotional information in music and transform it into a visual form to help students better understand and feel the emotional expression in music. The results show that the prediction accuracy of this method reaches 65.5% after preliminary training. In 20 training stages, the final model test accuracy reached 94.8%. This model can not only help students better understand and feel music works, but also provide more flexible and efficient teaching means and methods for music teachers, and promote the continuous growth of music teaching.

If the algorithms and functions to be realized are complex enough, such as 3D virtual reality technology, it is necessary to form a network with multiple computers for cluster parallel processing, and different computers will handle different functions to realize the overall function in the shortest possible time.

Yu Deng, https://orcid.org/0009-0006-9974-3745
Jie He, https://orcid.org/0009-0009-1202-7407
Linyuan Hu, https://orcid.org/0009-0000-5526-4108

REFERENCES

[1] Bahari, A.: CANDA: Computer-assisted nonlinear dynamic approach for the L2 teaching in blended and distance learning, Interactive Learning Environments, 31(2), 2023, 752-778. https://doi.org/10.1080/10494820.2020.1805774
[2] Bogach, N.; Boitsova, E.; Chernonog, S.; Lamtev, A.; Lesnichaya, M.; Lezhenin, I.; Blake, J.: Speech processing for language learning: A practical approach to computer-assisted pronunciation teaching, Electronics, 10(3), 2021, 235. https://doi.org/10.3390/electronics10030235
[3] Chaturvedi, V.; Kaur, A.-B.; Varshney, V.; Garg, A.; Chhabra, G.-S.; Kumar, M.: Music mood and human emotion recognition based on physiological signals: a systematic review, Multimedia Systems, 28(1), 2022, 21-44. https://doi.org/10.1007/s00530-021-00786-6
[4] Filippini, C.; Di Crosta, A.; Palumbo, R.; Perpetuini, D.; Cardone, D.; Ceccato, I.; Merla, A.: Automated affective computing based on bio-signals analysis and deep learning approach, Sensors, 22(5), 2022, 1789. https://doi.org/10.3390/s22051789
[5] Hai, Y.: Computer-aided teaching mode of oral English intelligent learning based on speech recognition and network assistance, Journal of Intelligent and Fuzzy Systems, 39(4), 2020, 5749-5760. https://doi.org/10.3233/JIFS-189052
[6] Han, D.; Lee, S.-B.; Song, M.; Cho, J.-S.: Change detection in unmanned aerial vehicle images for progress monitoring of road construction, Buildings, 11(4), 2021, 150. https://doi.org/10.3390/buildings11040150
[7] Iyer, A.; Das, S.-S.; Teotia, R.; Maheshwari, S.; Sharma, R.-R.: CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings, Multimedia Tools and Applications, 82(4), 2023, 4883-4896. https://doi.org/10.1007/s11042-022-12310-7

[8] Maba, A.: Computer-aided music education and musical creativity, Journal of Human Sciences, 17(3), 2020, 822-830. https://doi.org/10.14687/jhs.v17i3.5908
[9] Mustaqeem, K.-S.: CLSTM: Deep feature-based speech emotion recognition using the hierarchical ConvLSTM network, Mathematics, 8(12), 2020, 2133. https://doi.org/10.3390/math8122133
[10] Pei, Z.; Wang, Y.: Analysis of computer aided teaching management system for music appreciation course based on network resources, Computer-Aided Design and Applications, 19(1), 2021, 1-11. https://doi.org/10.14733/cadaps.2022.S1.1-11
[11] Peng, H.; Jager, S.; Lowie, W.: A person-centred approach to L2 learners' informal mobile language learning, Computer Assisted Language Learning, 35(9), 2022, 2148-2169. https://doi.org/10.1080/09588221.2020.1868532
[12] Qiu, L.; Li, S.; Sung, Y.: 3D-DCDAE: Unsupervised music latent representations learning method based on a deep 3d convolutional denoising autoencoder for music genre classification, Mathematics, 9(18), 2021, 2274. https://doi.org/10.3390/math9182274
[13] Quan, Y.: Development of computer aided classroom teaching system based on machine learning prediction and artificial intelligence KNN algorithm, Journal of Intelligent & Fuzzy Systems, 39(2), 2020, 1879-1890. https://doi.org/10.3233/JIFS-179959
[14] Shalini, S.-K.; Jaichandran, R.; Leelavathy, S.; Raviraghul, R.; Ranjitha, J.; Saravanakumar, N.: Facial emotion based music recommendation system using computer vision and machine learning techiniques, Turkish Journal of Computer and Mathematics Education, 12(1), 2021, 912-917. https://doi.org/10.17762/turcomat.v12i2.1101
[15] Suzuki, K.; Laohakangvalvit, T.; Matsubara, R.; Sugaya, M.: Constructing an emotion estimation model based on eeg/hrv indexes using feature extraction and feature selection algorithms, Sensors, 21(9), 2021, 2910. https://doi.org/10.3390/s21092910
[16] Taruffi, L.; Küssner, M.-B.: A review of music-evoked visual mental imagery: Conceptual issues, relation to emotion, and functional outcome, Psychomusicology: Music, Mind, and Brain, 29(2-3), 2019, 62-74. https://doi.org/10.1037/pmu0000226
[17] Yuan, Y.: Design and realization of computer aided music teaching system based on interactive mode, Computer-Aided Design and Applications, 19(S2), 2020, 92-101. https://doi.org/10.14733/cadaps.2021.S2.92-101