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

Human-computer Interaction Based Music Emotion Visualization System and User Experience Assessment

Ruidi He[1], Miaoping Geng[2] and Jia Guo[3]

[1]The Conservatory of Music, Hebei Institute of Communications, Shijiazhuang, Hebei 056000, China, heruidi@hebic.edu.cn
[2]The Conservatory of Music, Hebei Institute of Communications, Shijiazhuang, Hebei 056000, China, gengmp@hebic.edu.cn
[3]The Conservatory of Music, Hebei Institute of Communications, Shijiazhuang, Hebei 056000, China, goojiia@gmail.com

Corresponding author: Miaoping Geng, gengmp@hebic.edu.cn

Abstract. The perceptual and auditory standard of music is deeply integrated with the emerging multimedia to a higher degree, thus forming the music visualization. It is a process presentation method, which provides a brand-new way of interpretation and deduction for music appreciation. In this article, the application of computer aided design (CAD) in music emotion visualization system is studied, and a mapping model between music characteristics and emotion for digital music emotion recognition is constructed by combining with convolutional neural network (CNN). Combined with CAD technology, the structural music features are extracted and calculated, and the main melody and auxiliary melody of music are obtained. Then, based on the separated main melody and auxiliary melody, comprehensive visualization design is carried out to realize the visualization method of highlighting the main melody. In the experimental part, the performance of music emotion recognition algorithm is tested and the user experience is assessed. The results show that the simulation accuracy and user interaction experience of this system have achieved good results, which can improve the interaction between CAD design and viewing of music emotion visualization. Compared with the recurrent neural network (RNN), support vector machine (SVM) and other emotion recognition models, this model has a higher recognition rate of music emotion, which is of great significance to the research of music emotion visualization system.

Keywords: Human-Computer Interaction; CAD; Music Visualization; Emotion Recognition
DOI: https://doi.org/10.14733/cadaps.2024.S7.133-147

1 INTRODUCTION

The perceptual and auditory standard of music is deeply integrated with the emerging multimedia to a higher degree, thus forming music visualization, that is, the transformation of music into

images. Multi-layer dynamic GCN is a data structure used to process large-scale graphic data. It can combine the information of nodes and edges to better capture the complex structure of the graph. Bao et al. [1] recalibrated the convolutional network model based on style, which is a kind of deep learning model. It combines style transfer technology and Convolutional neural network, thus realizing automatic style conversion of images. By utilizing multi-layer dynamic GCN, the complex structure of EEG signals can be better captured, thereby improving the accuracy of emotion recognition. And style-based recalibration of CNN can provide more powerful image processing capabilities, thereby better processing EEG signals. The combination of the two can achieve more accurate and efficient emotion recognition based on EEG. It is a process presentation method, which provides a brand-new way of interpretation and deduction for music appreciation. In order to realize music visualization based on emotion, it is often needed to mark the emotion of music works. Audio text emotion recognition requires processing multiple types of data, such as speech, text, and emotion images. In order to better understand emotions, these multimodal data can be fused. In engineering mathematics, probability graph models, deep learning models, etc. can be used to fuse these multimodal data, thereby improving the accuracy of emotion recognition. Cai et al. [2] performed emotion recognition on audio texts based on an improved neural network. In audio text emotion recognition, the training dataset may not be sufficient, which can lead to overfitting of the model. In order to solve this problem, we can use the Transfer learning technology, take the model that has been trained on other data sets as the pre training model, and then use its feature extraction ability for audio text emotion recognition task. In engineering mathematics, Backpropagation and least square method can be used to optimize the parameters of the model, so as to improve the accuracy of emotion recognition. If you use manual methods to mark the emotion of massive music, it will not only be a huge workload, but also be inefficient. Therefore, it is an inevitable choice to study music emotion recognition technology and realize automatic emotion annotation of music works. In the process of music visualization, how to enhance the human-computer interaction (HCI) experience of visual works through multi-channel mapping mode is the research purpose of this article. With the rapid development of artificial intelligence and computer technology, human-computer interaction methods are also constantly evolving. From traditional graphical user interfaces (GUIs) to current audiovisual user interfaces (AVUIs), designers and developers have continuously explored and innovated in creating more intuitive, vivid, and three-dimensional interactive experiences. Correia et al. [3] introduces the application of AVUI in the field of human-computer interaction and related technology transactions. At present, the main technical means for implementing AVUI include speech recognition, image recognition, virtual reality, augmented reality, and so on. For example, through speech recognition technology, the system can recognize users' voice commands and achieve voice interaction; Through image recognition technology, the system can recognize user behavior and actions, achieving more intuitive interactive operations. On the basis of GUI, AVUI adds various interactive methods such as visual and auditory, making human-computer interaction more natural and intuitive. For example, through visual design and animation effects, users can have a clearer understanding of interface operations; Through audio prompts and voice interaction, users can complete operations more conveniently. The human-computer interaction technology for museums and exhibitions based on gaze relies on computer vision technology, which captures the user's gaze direction and gaze movement through a camera, thereby identifying the user's points of interest and interaction behavior. Dondi and Porta [4] analyzed human-computer interaction in museums and exhibitions based on gaze. Through the human-computer interaction interface, users can interact with exhibition content in a more natural and intuitive way. Big data and AI technology are used to analyze and understand users' staring behavior and interaction behavior, so as to provide more personalized and intelligent services. Intelligently guide users to visit exhibitions by capturing their gaze direction and eye movement. Gazing based human-computer interaction in museums and exhibitions can also be used to create more interactive experiences, such as controlling multimedia content or interactive scenes in exhibitions through the user's gaze direction and gaze movement.

Bidirectional Convolutional Recursive Sparse Network (BCRSN) is an effective music emotion recognition model. This model combines the advantages of Convolutional neural network (CNN) and recurrent neural network (RNN), and adopts a sparse connection strategy to achieve better performance and lower number of parameters. In the BCRSN model, each time step is associated with a feature vector, which is calculated using a recurrent neural network. Recurrent neural networks can remember previous inputs and use them for future calculations, which enables the BCRSN model to better process time series data. Dong et al. [5] also adopted a sparse connection strategy through the BCRSN model, which introduces some sparsity constraints in the network, making the model more robust and more generalized. In general, BCRSN model is an effective music emotion recognition model, which can process time series data, combines the advantages of Convolutional neural network and recurrent neural network, and adopts a sparse connection strategy to achieve better performance and lower number of parameters. 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 addition to the inherent subjective influence of music emotion recognition itself and the performance of existing automatic recognition algorithms, the characteristics as model input greatly affect the accuracy of emotion recognition model. The appearance of computer has greatly changed people's understanding of the world, which not only makes people communicate more conveniently, but also makes people enter a virtual world constructed by themselves. In this article, the application of CAD in music emotion visualization system is studied, and the mapping model between music characteristics and emotion for digital music emotion recognition is constructed with CNN.

Music is an important ideology that expresses people's thoughts and feelings through auditory forms and reflects social real life. Computer music is an art form combining information technology with music theory. It expresses human subjective feelings through multimedia technology. However, no matter how different the physical properties are, the emotional information contained in the same music works is the same. As the highest abstract form of music content and the information form directly understood and communicated by people, emotional semantics occupies an important position in the whole computer music field. The complexity of musical emotion puts forward new requirements for pattern recognition, artificial emotion, fuzzy classification and other fields, thus promoting the research of related theories and the development of technology, and making computer intelligence closer to real intelligence at this stage. Based on the concept of HCI, this article studies the music emotion recognition algorithm based on CNN. Its main innovations are as follows:

⊖ In the research process of music emotion visualization system, a rule-based music emotion cognition method is designed, which uses production rules and uncertain reasoning technology to analyze music emotion.

(2) On the visualization method of highlighting the main melody, this article studies the method of extracting and calculating the structural music features with CAD technology to get the main melody and auxiliary melody, and then carries out comprehensive visualization design based on the separated main melody to realize the visualization method of highlighting the main melody.

⊛ In the automatic classification of music, this article proposes an automatic classification of music based on the fractal dimension characteristics of music. Firstly, the fractal dimension of the whole music is calculated, and then the music feature classification index is constructed to classify different music styles, and different image elements are used to correspond to different styles of music.

In this article, the construction idea of music emotion visualization system model is introduced from two aspects: music emotion feature extraction and classification and music emotion visualization modeling. Then the simulation test proves the effectiveness of the algorithm in the HCI design of music visualization system; Finally, the work and limitations of this article are summarized, and the next research direction of music emotion visualization system is pointed out.

2 RELATED WORK

The layout of human-computer interaction interface of electronic music products based on ERP technology was optimized. Han [6] optimized the interface layout and improved the user experience and efficiency through in-depth analysis of user behavior and needs. By analyzing users' operational behavior and demand data, we can understand their usage habits and preferences, thereby determining the optimization direction of interface layout. Based on the data analysis results, optimize the interface layout, such as placing the most commonly used functions in the most prominent position, optimizing the visual effects of buttons and icons, and adding guidance prompts, in order to improve user efficiency and experience. Design personalized interface layouts based on user preferences and needs, such as different themes, colors, fonts, etc., to meet the needs and preferences of different users. By increasing the interaction between users and the interface, such as voice interaction, Gesture recognition and other ways, the user experience and interaction can be improved. By optimizing feedback mechanisms, such as adding animation effects and voice prompts, users can better understand the operation results and status, improving their user experience and efficiency. Personalized push is the recommendation of personalized content through algorithms based on user behavior and interests. Human computer interaction refers to the interaction between humans and computers. Liang and Willemsen [7] achieve interactive communication between humans and computers through technologies such as interactive interfaces, speech recognition, and image recognition. In music exploration, human-computer interaction can realize the interaction between users and computers through interactive interface, voice recognition, image recognition and other technologies. Users can explore music in a variety of ways, such as voice input, Gesture recognition, expression recognition and other ways to interact with computers, so as to better experience music. By combining personalized push and human-computer interaction, music exploration can be better promoted. For example, through personalized push, users can recommend their favorite music works. Users can explore the details and emotions of music through human-computer interaction, such as voice recognition, Gesture recognition and other ways to interact with computers to better experience music. Intelligent interactive music information research based on visualization technology is a research field that utilizes visualization technology, artificial intelligence technology, and interaction technology to achieve intelligent interaction and visual presentation of music information. Liao [8] utilizes visualization technology to present various attributes of music information (such as pitch, rhythm, harmony, etc.) in the form of images, charts, or dynamic visual effects, in order for users to better understand and analyze music information. It utilizes artificial intelligence technology to intelligently analyze music information, including sequence analysis, harmony analysis, style analysis, etc., in order to better understand the emotions, emotions, and structures of music. By utilizing interactive technology, users are allowed to explore music information in an interactive manner. For example, through an interactive interface, users can explore different elements of music, such as melody, harmony, rhythm, etc., and visually see different aspects of music through visualization technology. Apply the above technologies to practical scenarios, such as music education, music therapy, music emotion recognition, etc., to achieve better music experience and transmission of music information. In summary, intelligent interactive music information research based on visualization technology is a research field that combines visualization technology, artificial intelligence technology, and interaction technology, aiming to achieve efficient, intuitive, and intelligent presentation and exploration of music information, thereby providing users with better music experiences and services.

Emotional visualization based on visual and scene changes is a method of presenting emotional trends through images, charts, and other means. This method can use emotion analysis technology to process video, image, text and other data, and convert them into visual forms such as emotion Run chart or curve chart. In emotion visualization based on visual and scene changes, Lin et al. [9] preprocessed the input data. For example, preprocessing images, including image enhancement, denoising, edge detection, and other operations, for subsequent emotional analysis. Then, sentiment analysis algorithms are used to classify the preprocessed data into emotional polarities, such as using deep learning models to classify images into positive, negative, or neutral

emotional polarities. Next, visualize the classification results of emotional polarity, such as using different colors or shapes to represent different emotional polarities, such as red representing positive emotions, blue representing negative emotions, and green representing neutral emotions. At the same time, you can use graphs or Line chart to show the changes of emotional trends, so as to better understand the changes of emotional trends. The deep learning of intelligent human-computer interaction music is a Applied science, which uses deep learning technology to study human-computer interaction of music. Lv et al. [10] utilized deep learning techniques to identify music styles, such as distinguishing between different styles such as classical, pop, and rock music, in order to better understand and analyze the attributes of music. Using deep learning techniques to identify emotions in music, such as distinguishing different emotions such as happiness, sadness, excitement, and calmness, in order to better understand the emotions and emotions of music. By utilizing deep learning technology, music therapy assistance tools can be developed, such as analyzing elements such as rhythm, melody, and harmony of music, to help therapists better understand and analyze patients' music therapy needs. In short, the in-depth learning of intelligent human-computer interaction music is a Applied science. It utilizes deep learning technology to study and solve human-computer interaction problems in music. To enhance people's understanding and analytical abilities in music, and to provide better services for music related fields. The application of computer-aided music education technology can transform students from passive learning to active exploration. This has a great promoting effect on their innovative and exploratory spirit, and also allows them to fully utilize their talents, thereby promoting the development of the music industry. At the same time, using computers for music teaching not only has a large class capacity, but also is vivid and vivid, which can stimulate students' learning enthusiasm and improve classroom teaching efficiency. Maba analyzed [11] computer-assisted music education and music creativity. Music creativity is very important in music education. Music is an art form that expresses emotions and creativity, and its value needs to be reflected through creation and performance. Computer assisted music education can provide more possibilities for Musical composition. For example, using computers to teach music theory can make abstract theories more intuitive, which is conducive to revealing key points and breaking through difficulties.

Through computer-assisted music education, students can better understand and master music knowledge, while also being able to control elements such as rhythm, melody, and timbre. It makes music teaching and Musical composition more convenient, and it can also let music teachers and trendy musicians generate inspiration and let them fully play their imagination. It is an effective way of human-computer interaction to enhance the self-disclosure of users by visualizing the common activities and dialogue atmosphere of Chatbot. Chatbot can show friendly and cordial attitude through language and expression, so as to establish a good dialogue atmosphere and make users more willing to share their information. Provide personalized questions based on the characteristics and interests of users, guiding them to think more deeply and express their ideas. Visualize the common activities that users participate in through images, videos, and other means, such as listening to music or watching movies together, to enhance users' sense of participation and interactivity. When users disclose themselves, Chatbot can provide positive feedback and encouragement. Pujiarti et al. [12] enhances users' self-disclosure by visualizing the common activities and dialogue atmosphere of Chatbot. This can be achieved by establishing a friendly dialogue atmosphere, providing personalized questioning, visualizing joint activities, and providing feedback and encouragement. This human-computer interaction method can enhance users' sense of participation and interactivity, thereby better meeting their needs and expectations. Tang et al. [13] analyzed an intelligent shadow play system with multidimensional interactive perception. It combines computer technology, artificial intelligence, and traditional shadow puppetry art, providing users with a brand-new interactive experience. This system can perceive various information such as user actions, speech, and vision through various sensors and machine learning algorithms, thereby better understanding the user's intentions and needs. The system analyzes and processes user interaction behavior through intelligent algorithms, which can automatically adjust and optimize the effect of shadow puppetry performances, providing a more

high-quality interaction experience. This system combines traditional shadow puppetry art with modern technology, not only retaining the unique charm of traditional shadow puppetry art, but also bringing users a brand-new interactive experience. In terms of human-computer interaction, the system emphasizes the user experience and the naturalness of interaction. Through multidimensional interaction perception and intelligent control technology, users can interact with shadow puppetry performances in a more intuitive and natural way. For example, users can control the movements and performances of shadow puppets through gestures, speech, or body movements, making human-computer interaction smoother and more natural. Computer graphics visualization plays an important role in the research of controlled users in human-computer interaction. Wang et al. [14] used visualization technology to present information such as user behavior, actions, and emotions in the form of images, animations, etc., in order to better understand the user's operational intentions and experience. By tracking and recording user actions, they can be presented in the form of images or animations, thereby better understanding the user's operational process and experience. By analyzing users' emotional data, they can be presented in the form of images or animations to better understand their emotional state and experience. By collecting and analyzing user feedback information, it is possible to present it in the form of images or animations, thereby better understanding the user's feedback content and experience. The bimodal emotion recognition model of Southern Min songs is an informatics model, which is used to identify the emotions in Southern Min songs. This model combines two different modal data, audio and text, to obtain more accurate and comprehensive emotion recognition results. When building this model, Xiang et al. [15] needs to collect and process a large amount of Southern Min song data, including audio files and lyrics text. Firstly, preprocess the audio file, including audio feature extraction, pre-emphasis, and noise reduction, for subsequent emotion recognition. Next, emotional analysis is conducted on the lyrics text to extract emotional vocabulary and polarity information. When constructing a bimodal emotion recognition model, it is necessary to combine audio features and text emotion information, and generate a comprehensive feature vector through feature interaction and fusion strategies. Common fusion strategies include weighted fusion, overlay fusion, and deep fusion. The social emotional Sensory design of Affective computing based on neural divergent experience is the design of Affective computing and emotional interaction through neural network and artificial intelligence technology. This design aims to simulate the process of human emotions and cognition, in order to better understand the needs and emotional states of users, and thus provide more personalized and humanized services. Social affective Sensory design is an important aspect of Affective computing, which focuses on the performance and perception of human emotions in social interaction. This design identifies the user's emotional state and interaction behavior by analyzing their emotional signals, such as facial expressions, voice intonation, body language, etc. Then, according to the user's emotional state and behavior, social emotional Sensory design can adjust the interaction mode and content to provide more personalized and humanized services. Zolyomi and Snyder [16] calculate and analyze users' emotional state and behavior through Affective computing algorithm to understand users' needs and preferences. Design interaction methods and content based on the emotional state and behavior of users to provide more personalized and humanized services.

3 CONSTRUCTION OF MUSIC EMOTION VISUALIZATION SYSTEM MODEL

3.1 Music Emotion Feature Extraction and Classification

The artistic creation of music visualization is not a simple computer 3D animation, but the creation of real 3D images in an immersive virtual environment. Immersion refers to the immersive feeling that the audience has from beginning to end when enjoying the music visual works. It is a high-level experience centered on the participants. From the perspective of pure technology, music visualization also belongs to the research category of virtual reality technology. In the information age when people's needs are getting higher and higher, the simple plane and 3D graphics can no longer meet people's needs, which also gives birth to virtual reality application technology. Music

visualization is inseparable from emotion, and images directly touch human perception, so images become the core element of immersion. The essence of music visualization is from sound to graphics. In the process of visualization, it is generally difficult to extract natural audio features, and the data is not accurate enough. Therefore, the extraction of melody features of multi-track MIDI has become the first choice. Music visualization, as a new visualization technology, has had an important influence in various popular media playback software. This article proposes a music emotion recognition model based on CNN, as shown in Figure 1.

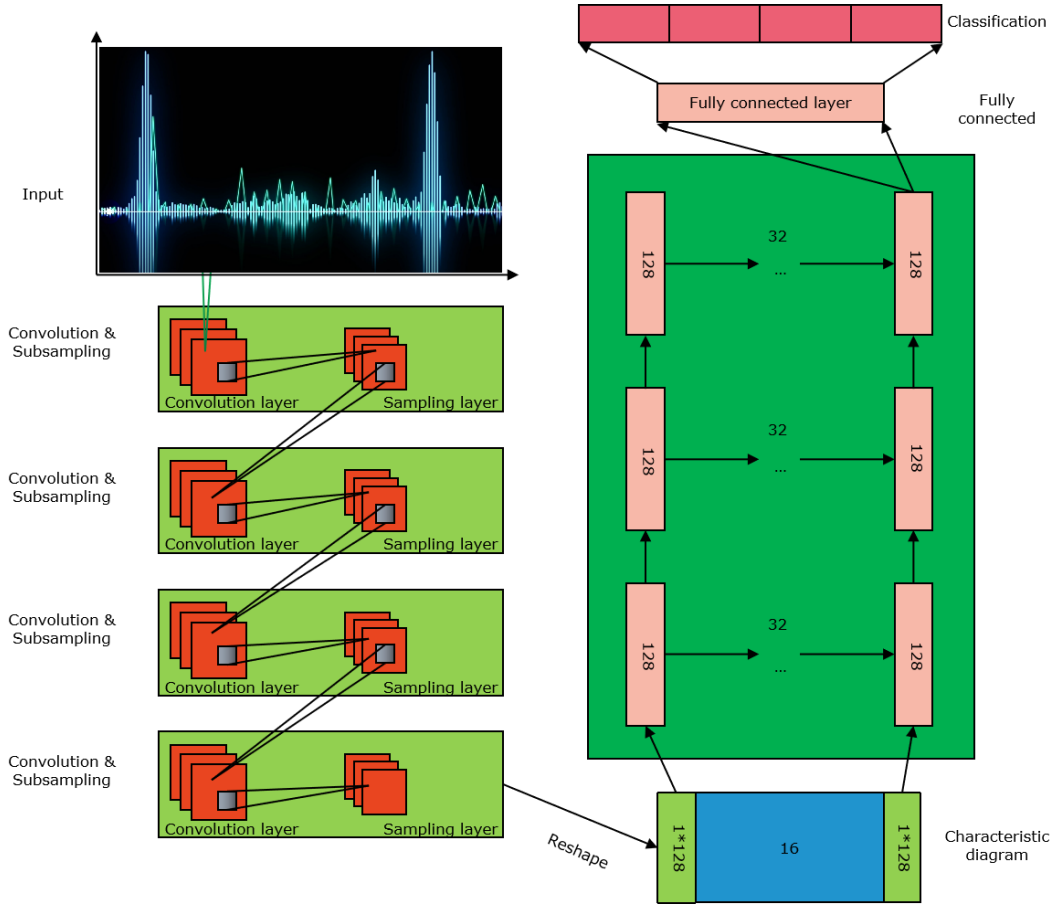

Figure 1: CNN music emotion recognition model.

Before feature extraction, one-step silent frame discrimination is needed to reduce unnecessary interference and calculation. If these frames are allowed to enter the frame set where we want to extract features next, it will not only increase the calculation amount of research, but also make the features extracted by us deviate, which will seriously lead to the failure of the experiment. In the process of audio feature extraction, the first step is to extract the features of the main and auxiliary tracks, and to find out the special audio features, so as to distinguish the auxiliary tracks from the main tracks more accurately. Dig out the information about MIDI file in the description part of audio track information, and then use this information to get rid of the interfering audio tracks. In order to make the changes of two adjacent frames smoother, it is needed to adopt the method of partial overlap between frames, that is, the last frame repeats the length of the previous frame by nearly half.

The persistence of images enables visual works to maintain their real-time nature when users interact, and creates a virtual environment by means of computer system for processing and calculation. Through a certain preprocessing process, audio signals can be sorted into data forms that can be processed by algorithms and computers. This process is generally divided into the following steps: audio signal filtering processing, pre-emphasis processing, framing and windowing processing, and mute frame discrimination processing. The short-term average energy $E(i)$ of the music signal in the i-th frame can be obtained by one of the following three algorithms:

$$E(i)=\sum_{n=1}^{N}\left|X_i(n)\right| \tag{1}$$

$$E(i)=\sum_{n=1}^{N}X_i^2(n) \tag{2}$$

$$E(i)=\sum_{n=1}^{N}\log X_i^2(n) \tag{3}$$

Among them, N is the frame length, and $X_i(n)$ is an amplitude energy of the audio information at the n point. The spectrogram itself originally covers the spectrum of all sound signals, which is a dynamic spectrum. The generated fast Fourier transform is as follows:

$$X(n,k)=\sum_{M=0}^{N-1}X_n(m)e^{-j\frac{2\pi km}{N}} \tag{4}$$

Among them, $X_n(m)$ is the n-th frame signal of the framed audio. $0\le k\le N-1$, then $\left|X(n,k)\right|$ is the short-term amplitude spectrum estimate of $X(n)$, and the spectral energy density function $p(n,k)$ at m is:

$$p(n,k)=\left|X(n,k)\right|^2 \tag{5}$$

When the frame length is too long, the frequency resolution will be very small, that is, the resolution will be very high. The disadvantage of this is that the number of frames we get on the time axis is too small to meet the requirements of the application. On the contrary, when the window length is too short, the frequency resolution will be very large, that is, the resolution will be too low, and too many frames will be obtained, which is too complicated to calculate. Even if the calculation amount is not considered, when the frame length is too small, this function is equivalent to a high-pass filter, which makes many low-frequency signals ignored, resulting in abnormal feature transformation such as short-term average energy, forming an atypical feature function.

3.2 Visual Modeling of Musical Emotion

The main melody of a piece of music, that is, the main track of an old file, usually has a good pitch continuity, because the main melody needs to be played continuously. If the pitch fluctuates greatly, it will give people an abrupt feeling, while the accompaniment track does not need to be played continuously, so a large fluctuation will not affect the musical sense. When designing music scores, the main melody usually has good continuity. After determining the category of the sound to be measured, it is needed to collect the signals of these sounds, process the collected signals, extract the features of the sound signals, and add the extracted features to the template database, so that the sound can be category-matched. The model of music signal recognition system is shown in Figure 2.

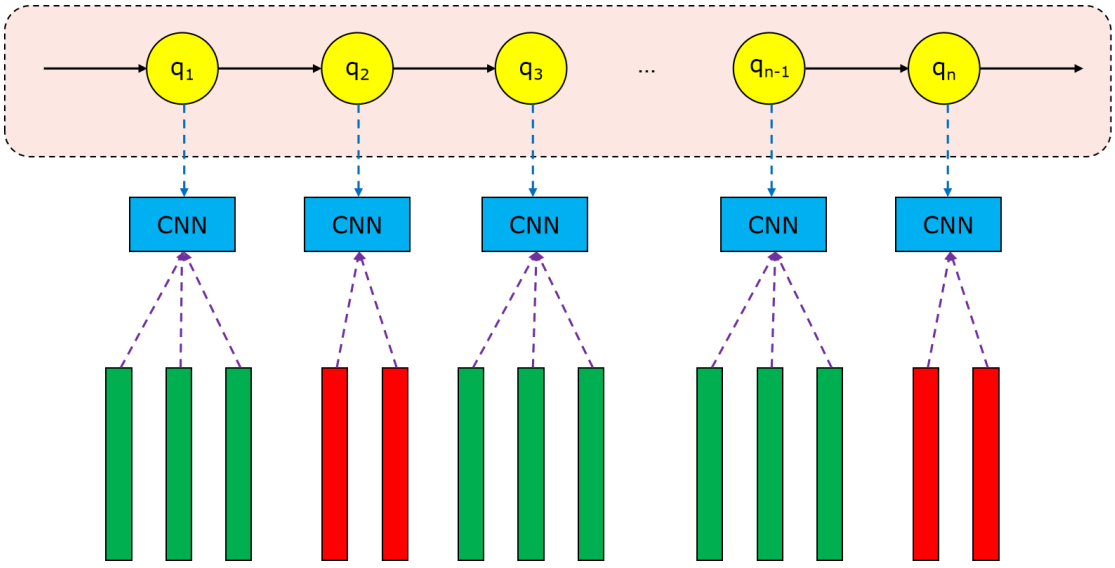

Figure 2: Music signal recognition system model.

The traditional recognition program can only provide users with a recognition result after the recognition, and the recognition rate is only the statistical value after the recognition. In this article, the introduction of credibility gives users the ability to actively grasp the recognition results in the recognition process. By calculating the reliability rate, users can get more satisfactory recognition results.

When analyzing an audio track, its corresponding note queue is empty. When encountering an instruction to open a note, the system will parse the instruction, get the pitch, strength, duration and start time of the note, and add the note to the queue. When you encounter an instruction to close a note, find the note from the note queue and fill in its end time. In this way, after analyzing the track, the queue contains information such as the pitch, intensity, duration and start-stop time of all the notes in the track, which is ready for the division of music sections below.

Time-frequency domain transform adopts constant Q transform, and the solution of chroma features is deduced in frequency domain under constant Q transform, and a filter bank with $\log_2$ change rule is designed, in which the central frequency of each filter is:

$$f_c\left(k_{lf}\right)= f_{\min}\cdot 2^{\frac{k_{lf}}{\beta}} \tag{6}$$

Where $f_{\min}$ is the minimum center frequency of the filter bank, this article selects 220 Hz; k_{lf} is the filter index number. β is the number of filter banks/levels per octave. The cross-correlation function can calculate the exchangeable energy between the signals $x(t), y(t-\beta)$, and quantify this exchangeable energy, so as to obtain the similarity between the two signals. The formula of the cross-correlation function is as follows:

$$\gamma_{xy}(\alpha,\beta)= \int_{-\infty}^{+\infty}\left[x(t)\cdot y(\alpha\cdot(t-\beta))\cdot dt\right] \tag{7}$$

In which $\alpha = 1$ and β are used to eliminate the time offset.

For MIDI music, many features are easy to extract, but this music format is also flawed in feature extraction, that is loudness. Because the loudness in MIDI can only be estimated or summed from other values (such as speed), such data is not accurate enough, because the speed of notes is not linearly related to the loudness felt by people. For symbolic music data, such as piano music, that is, piano music that can be recorded to a computer, symbolic information includes the onset, pitch and speed of notes. It is easy to extract these features from music in MIDI format. But in MIDI format, loudness can only be estimated from the speed value. However, such an estimate is inaccurate, because the speed of tapping is not linearly related to the perceived loudness. He pointed out that the current method of imitating loudness is to sum all the notes played in a given time.

Calculate the sum of fitness sizes of all individuals in the population F according to the following formula:

$$F = \sum_{i=1}^{N} f(i) \tag{8}$$

Where N is the population size and f_i is the fitness function value of the i th individual. Then calculate the relative fitness of each individual in the population, and the formula for calculating the relative fitness P_i of the i -th individual is:

$$P_i = \frac{f(i)}{F} \tag{9}$$

Finally, the roulette operation is simulated for multiple rounds of selection. The larger the P_i, the larger the area occupied by the individual in the roulette, so the greater the probability and times of being selected.

4 MODEL TESTING AND RESULT ANALYSIS

A variety of features are extracted, a feature set is constructed and the extraction results are displayed graphically. The comprehensive visual design based on different features makes the display effect no longer based on a single element, which makes the audience realize the relationship between the internal characteristics of audio and the music itself, and also gives them multiple visual experiences. Pattern recognition needs to describe and extract some features of the object as the basis for pattern recognition. Music that people can hear can't be directly recognized by computers. Music must be turned into a music file that computers can read, and then music can be analyzed and features can be extracted through a series of algorithms. In order to make the results more convincing, the network convergence trend diagram is derived during the solution process, as shown in Figure 3.

As can be seen from the figure, after about 19 iterations, the output error of the algorithm has converged to a certain extent. The reason why CNN can be used in music emotion analysis is that the characteristics of music are not only time domain characteristics, but also frequency domain characteristics. The speech signal can be divided into frames, each frame is Fourier transformed into a spectrogram, and then the spectrogram is connected in the time dimension to form an idiom spectrogram.

120 typical classical music in wav format, 150 hip-hop music and 115 country music were selected as the music library of the experiment, with a total of 385 pieces of music with a sampling rate of 44100Hz and eight bits of storage. Through experiments, the accuracy of emotion mapping

of the music emotion visualization system constructed in this article is shown in Figure 4. The comparison of emotion mapping accuracy of different methods is shown in Figure 5.

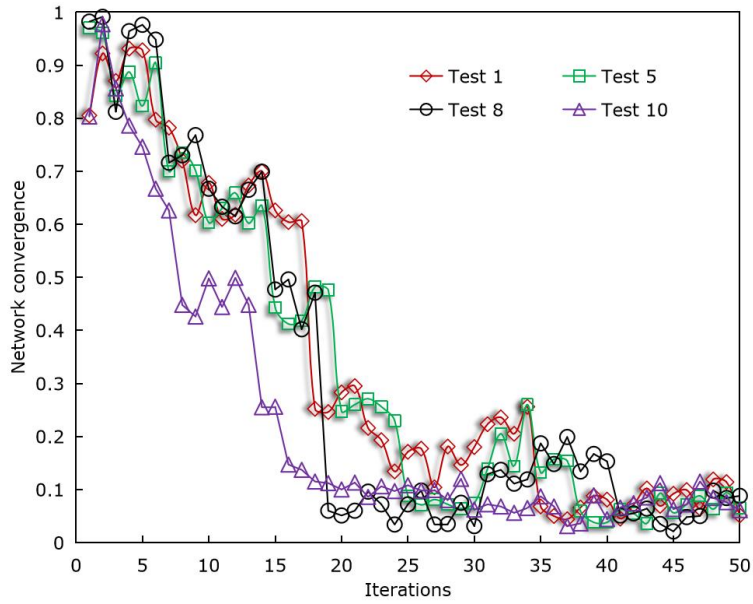

Figure 3: Network convergence trend diagram.

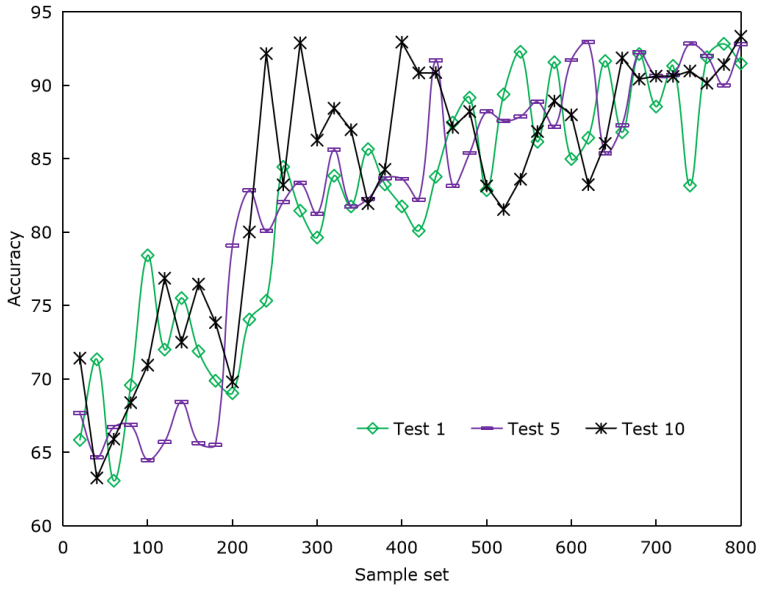

Figure 4: Emotion mapping accuracy of CNN method.

The results show that the emotion mapping accuracy of the music emotion visualization system based on CNN model combined with CAD technology is high, which can reach more than 90%. However, the emotion mapping accuracy of the music emotion visualization system constructed by RNN method is about 75%. The accuracy of sentiment mapping of SVM method is only 60%.

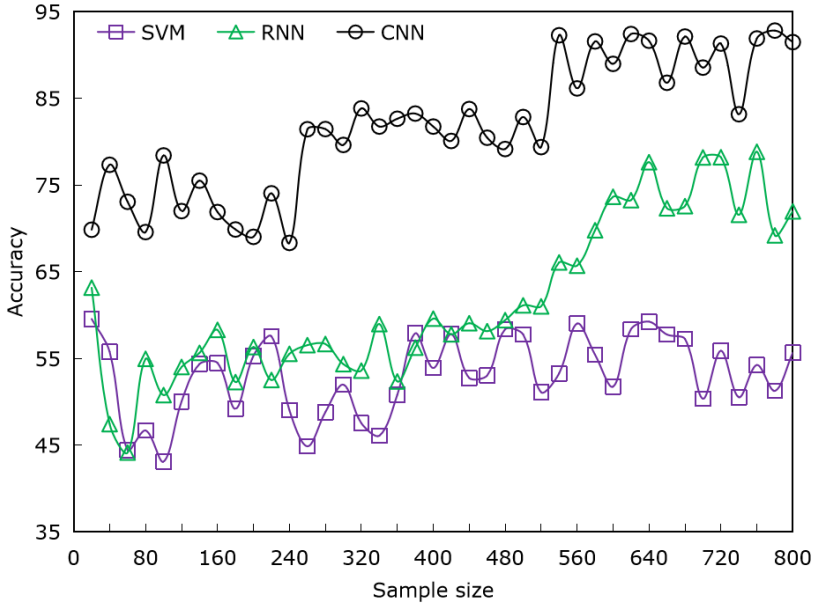

Figure 5: Comparison of emotion mapping accuracy.

A spectrogram can contain all the audio information, plus CNN's super automatic feature extraction ability for pictures, so it can be used as the original input of CNN, thus connecting music signals with CNN. On the Matlab platform, the efficiency of different methods of music emotion recognition model is tested, and the recognition efficiency is assessed by feature dimension reduction time, as shown in Table 1.

Music type	***Training sample***		***Test sample***	
	RNN	***CNN***	***RNN***	***CNN***
Classical music	7.85	6.04	8.22	5.64
Hip hop	8.36	5.48	6.65	4.72
Country music	7.69	6.55	7.81	3.53

Table 1: Dimension reduction time of music emotion recognition model.

Transfer learning in music emotion recognition has high generalization ability, jumps out of the local optimal solution better, and converges to a higher accuracy rate with fewer iterations. The accuracy of music emotion recognition by CNN and RNN is shown in Table 2.

Music type	***Training sample***		***Test sample***	
	RNN	***CNN***	***RNN***	***CNN***
Classical music	87.41%	92.48%	84.88%	92.46%
Hip hop	84.39%	95.26%	86.32%	95.27%
Country music	85.34%	95.53%	87.17%	95.99%

Table 2: Music emotion recognition accuracy.

Figure 6 shows the subjective score given by the observer on the visual effect of music emotion.

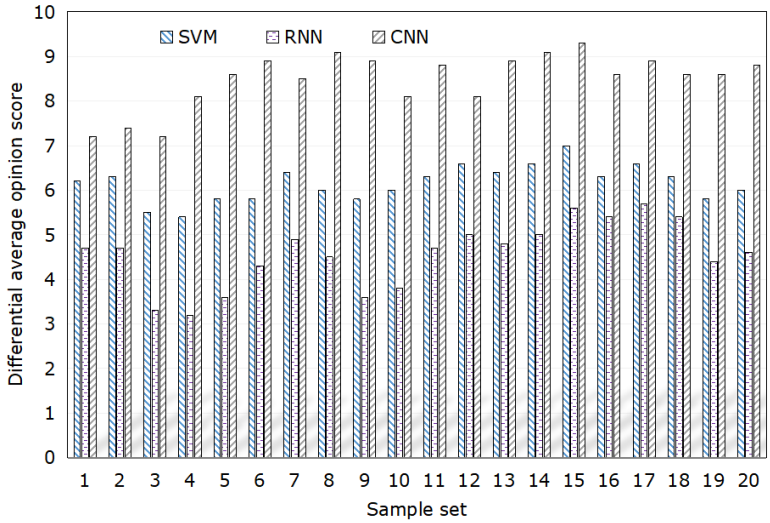

Figure 6: Subjective assessment of visual effect of music emotion given by observers.

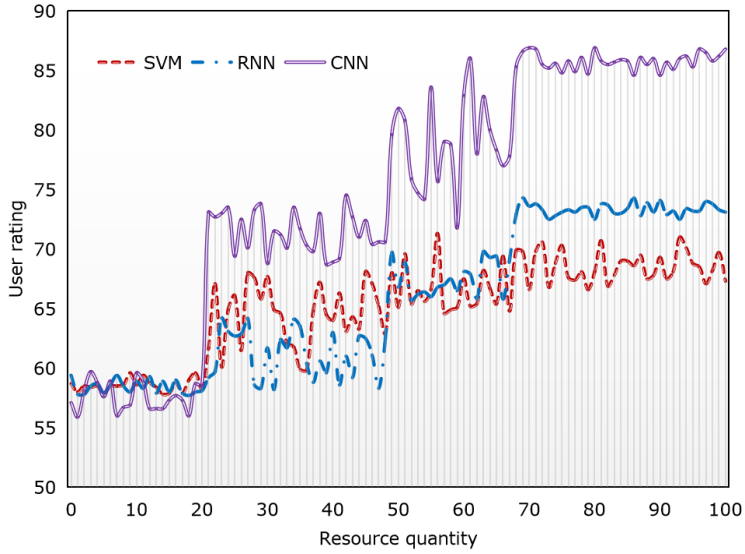

Figure 7: Viewer's rating on HCI experience of visual music exchange.

The music visualization method in this article has a higher score, so it can be considered that the music emotion visualization method combining CNN and CAD has achieved a better user experience. The system needs to analyze the complex music data input stream in order to extract meaningful music features to prepare for the following related work. In order to ensure that it can simulate the requirements of human emotional response to music, the system needs to adopt the previous research methods on musical emotion to organize these characteristics.

In order to make the characters get useful information in the audio information stream, the system needs to turn this stream into a simple format. Therefore, this article adds this extra music perception filter layer to the system framework. In this layer, the system extracts important musical 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. Because the concept of interactive design has been widely recognized and used by designers, the visual effect of music emotion has also joined the tide of development. Figure 7 shows the audience's assessment of the HCI experience of visual music.

Based on the above results, the simulation accuracy and user interaction experience of this system have achieved good results, which can enhance the interaction between CAD design and viewing of music emotion visualization. The music visualization system in this article needs to encode the characteristic information of live music input, so this article chooses the tonic as the reference note. All the later perceived notes will be encoded using this reference element, because they are all related to the tonic of the modal scale. During the live music performance, the cognitive processing of music input allows the animated characters to maintain a changing emotional state. Music animation expression layer uses animation to visualize character behavior, and this emotional state can then be conveyed to the audience through the expression layer.

5 CONCLUSIONS

Music is an important ideology that expresses people's thoughts and feelings through auditory forms and reflects social real life. 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 order to realize music visualization based on emotion, it is often needed to mark the emotion of music works. If you use manual methods to mark the emotion of massive music, it will not only be a huge workload, but also be inefficient. In this article, the application of CAD in music emotion visualization system is studied, and a method of music emotion visualization based on multi-features is proposed, which can selectively extract multi-features from music in music library and build a feature set. Then, according to the feature set, the multi-feature visual design of different types of music is realized. The results show that the simulation accuracy and user interaction experience of this system have achieved good results, which can improve the interaction between CAD design and viewing of music emotion visualization. Music appears as a whole, so calculating the fractal dimension of the whole music not only simplifies the calculation steps and complexity, but also makes the calculation results more prominent.

Music feature knowledge is subjective, complex and structural. In this article, it is inevitable that there are some defects in expressing music knowledge only by production. The application of various knowledge expressions will help to improve the performance of expert system.

6 ACKNOWLEDGEMENT

This work was supported by Key Project of Humanities and Social Sciences Research in Higher Education Institutions in Hebei Province: Study and Research on the Application of Ethnic Music Elements in the New Era of Music Creation from "Jingxing Lahua" (No. SD2022102).

Ruidi He, https://orcid.org/0009-0002-6468-7051
Miaoping Geng, https://orcid.org/0009-0003-1102-7997
Jia Guo, https://orcid.org/0009-0000-5030-6356

REFERENCES

[1] Bao, G.; Yang, K.; Tong, L.; Shu, J.; Zhang, R.; Wang, L.; Yan, B.; Zeng, Y.: Linking multi-layer dynamical GCN with style-based recalibration CNN for Eeg-based emotion recognition, Front Neurorobot, 2(24), 2022, 16:834952. https://doi.org/10.3389/fnbot.2022.834952

[2] Cai, L.; Hu, Y.; Dong, J.: Audio-textual emotion recognition based on improved neural networks, Mathematical Problems in Engineering, 2019(6), 2019, 1-9. https://doi.org/10.1155/2019/2593036
[3] Correia, N.; Tanaka, A.: From GUI to AVUI: situating audiovisual user interfaces within human-computer interaction and related fields, EAI Endorsed Transactions on Creative Technologies, 8(27), 2021, 1-9. http://dx.doi.org/10.4108/eai.12-5-2021.169913
[4] Dondi, P.; Porta, M.: Gaze-based human–computer interaction for museums and exhibitions: technologies, Applications and Future Perspectives, Electronics, 12(14), 2023, 3064. https://doi.org/10.3390/electronics12143064
[5] Dong, Y.; Yang, X.; Zhao, X.: Bidirectional convolutional recurrent sparse network (bcrsn): an efficient model for music emotion recognition, IEEE Transactions on Multimedia, 21(12), 2019, 3150-3163. https://doi.org/10.1109/TMM.2019.2918739
[6] Han, J.: Research on layout optimization of human-computer interaction interface of electronic music products based on ERP technology, International Journal of Product Development, 27(1-2), 2023, 126-139. https://doi.org/10.1504/IJPD.2023.129315
[7] Liang, Y.; Willemsen, M.-C.: Promoting music exploration through personalized nudging in a genre exploration recommender, International Journal of Human–Computer Interaction, 39(7), 2023, 1495-1518. https://doi.org/10.1080/10447318.2022.2108060
[8] Liao, N.-J.: Research on intelligent interactive music information based on visualization technology, Journal of Intelligent Systems 31(1), 2022, 289-297. https://doi.org/10.1515/jisys-2022-0016
[9] Lin, W.-Q.; Chao, L.; Zhang, Y.-J.: "Emotion visualization system based on physiological signals combined with the picture and scene, Information Visualization, 21(4), 2022, 393-404. https://doi.org/10.1177/14738716221109146
[10] Lv, Z.; Poiesi, F.; Dong, Q.; Lloret, J.; Song, H.: Deep learning for intelligent human–computer interaction, Applied Sciences, 12(22), 2022, 11457. https://doi.org/10.3390/app122211457
[11] Maba, A.: Computer-aided music education and musical creativity, Journal of Human Sciences, 2020, 17(3), 822-830. https://doi.org/10.14687/jhs.v17i3.5908
[12] Pujiarti, R.-N.; Lee, B.; Yi, M.-Y.: Enhancing user's self-disclosure through chatbot's co-activity and conversation atmosphere visualization, International Journal of Human–Computer Interaction, 38(18-20), 2022, 1891-1908. https://doi.org/10.1080/10447318.2022.2116414
[13] Tang, Z.; Hu, Y.; Weng, W.; Zhang, L.; Zhang, L.; Ying, J.: An intelligent shadow play system with multi-dimensional interactive perception, International Journal of Human–Computer Interaction, 39(6), 2023, 1314-1326. https://doi.org/10.1080/10447318.2022.2062839
[14] Wang, Z.; Ritchie, J.; Zhou, J.; Chevalier, F.; Bach, B.: Data comics for reporting controlled user studies in human-computer interaction, IEEE Transactions on Visualization and Computer Graphics, 27(2), 2020, 967-977. https://doi.org/10.1109/TVCG.2020.3030433
[15] Xiang, Z.; Dong, X.; Li, Y.: Bimodal emotion recognition model for minnan songs, Information (Switzerland), 11(3), 2020, 145. https://doi.org/10.3390/info11030145
[16] Zolyomi, A.; Snyder, J.: Social-emotional-sensory design map for affective computing informed by neurodivergent experiences, Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 2021, 1-37. https://doi.org/10.1145/3449151