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

A Visual Interaction Design Method for Music Based on Multi Audio Features

Fei Ma

School of Music and Dance, Zhengzhou University of Science and Technology, Zhengzhou, Henan 450000, China, yangmabenteng@163.com

Corresponding author: Fei Ma, yangmabenteng@163.com

Abstract. Music visualization is the transformation from music to images, which is a process presentation method and provides a brand-new interpretation and deduction way for music appreciation. If the music effect can be expressed in the form of images, and the audio-visual combination can be achieved, it will be more vivid and can better understand the artistic conception of the author. Because music itself has rich and subtle emotional information, it is difficult to mechanically transform it into vision by a single rule in the process of music visualization. In this article, a visual expression method of music based on multi-audio features and CAD is proposed. Firstly, multiple features of music are extracted, and then these features are comprehensively visualized, so that images can express more music information, thus improving the interactive experience of music appreciators. The results show that the proposed convolutional neural network (CNN) model achieves higher accuracy of original pitch than the traditional method. The intelligent recognition method of notes based on CAD uses hierarchical filtering and segmentation method for a musical melody to complete the operation, thus reducing the task of music feature and improving the efficiency of music feature recognition.

Keywords: Music Visualization; Multi-Audio Features; CAD; Human-Computer Interaction
DOI: https://doi.org/10.14733/cadaps.2024.S7.1-14

1 INTRODUCTION

Voice is the most common means of message transmission in people's daily life, which accounts for most of the total data. Visualizing the relevant features of voice can intuitively show the hidden attributes inside the voice. Data visualization method is an intuitive, simple and reasonable method to summarize and present data, and its main purpose is to convey information more clearly and efficiently by means of graphics or images. Baradaran [1] preprocesses EEG signals, including denoising, filtering, alignment, and other operations, to facilitate subsequent emotion recognition. Divide the preprocessed data into training and testing sets for the purpose of training

and testing the emotion recognition model. Build a customized 2D CNN model based on the needs of emotion recognition. The model can include convolution layer, pooling layer, full connection layer, etc., and is trained and optimized by Backpropagation. Train the constructed 2D CNN model using a training set to automatically recognize emotions in EEG signals. Use a test set to test the trained 2D CNN model to evaluate its accuracy and error rate in emotion recognition. Analyze and optimize the test results to further improve the accuracy and stability of emotion recognition. A reasonable visualization result can help users to analyze and interpret data, thus making the original complex data easy to understand and use. Bishop et al. [2] explored the importance of movement in communication and interaction, focusing on the body movement patterns in music duo performances. In music performances, physical movement is an important means of expressing emotions, establishing relationships, and conveying information. Especially when two performers perform together, their body movements are more collaborative and interactive. Through analysis and research, we can better understand the role and significance of body movement in music performance. In music duo performances, mobility is the key to achieving communication and interaction. Effective physical exercise can help performers better express emotions, establish relationships, and convey information. For example, by coordinating steps and movements, performers can establish a common emotional state, thereby creating a shared sense of rhythm in music. In addition, by observing and imitating the actions of the other party, performers can better understand their intentions and emotions, thereby establishing a closer cooperative relationship. In order to better transform the auditory image of music into visual image and serve the values and aesthetics of modern appreciators, the perceptual auditory standard of music is deeply integrated with emerging multimedia to a higher extent. The growht of visualization provides a new way for the expression of music, using specific rules to interpret and produce reproducible visual effects. Music visualization is the transformation from music to images, which is a process presentation method and provides a brand-new interpretation and deduction way for music appreciation. In the field of digital media design, the research on music visualization is helpful to build a bridge between audio-visual media and expand the expression of dynamic media in a cross-border way. By touching an artwork, visually impaired individuals can perceive its shape, texture, texture, and color. Cho's [3] research has shown that touch can help visually impaired individuals better understand and appreciate artistic works. For example, Monet's Impressionism painting "Water Lilies" has been proved to be able to perceive the subtle changes of light, shadow and color through touch. Hearing is also one of the important senses for visually impaired individuals to perceive and understand art. By observing the sound effects and sound effects of artistic works, visually impaired individuals can obtain important information about the work, such as emotions, themes, or storylines. At the same time, hearing can also help visually impaired individuals perceive and understand the colors of artistic works. The association between these colors and emotions can be conveyed and perceived through sound and tactile experiences.

Dotov et al. [4] explored the collective effects of music listening, with a particular focus on enhancing motor energy and visual social cues. In the collective effect of music listening, exercise energy is an important factor. Research has shown that music can stimulate listeners' desire for exercise, improve their athletic ability, and enhance their athletic energy by resonating with people who listen to music collectively. This effect is widely applied in many occasions, such as in sports competitions, dance performances, carnivals, and so on. In order to enhance the collective effect of music listening, some researchers have proposed methods using grooves and visual social cues. Groove is a music arrangement technique that can adjust elements such as rhythm, melody, and harmony to make music more suitable for specific occasions and audiences. Visual social cues refer to conveying information and resonating with others through specific body language, facial expressions, eye contact, and other means. After the music is composed, the interpretation methods such as playing musical instruments or singing human voices can convey it to the audience. In order to make the audience fully understand the emotions and thoughts expressed by music, the creators or communicators adopt different auxiliary expressions. If the music effect can be expressed in the form of images, and the audio-visual combination can be achieved, it will be more vivid and can better understand the artistic conception of the author. With the growht of

computer-generated 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. Images are usually associated with emotions and are supported by relevant psychological theories. Because music itself has rich and subtle emotional information, in the process of music visualization, it is difficult to mechanically transform it into vision by a single rule. Nowadays, most of the music visualization methods regard each music element as a whole for unified display and fail to distinguish the main melody from the auxiliary melody, so as to highlight the main content of music. How to enhance the humanization of visual works through multi-channel mapping mode is the research purpose of this article. In view of this, this article proposes a visual expression method of music based on multi-audio features and CAD, which extracts the main theme of music and highlights it visually, thus improving the interactive experience of music appreciators.

Georges and Seckin [5] combine computer graphics and visualization technology to transform music data into intuitive graphical display. In the field of classical music, music Information visualization can help us better understand the composer's creative style and work structure. Network diagram is a common way of music Information visualization. It can convert melody, harmony, rhythm and other data in music works into nodes and edges, intuitively displaying the structure and relationships of music works. By adjusting the layout and node markers of the network graph, we can better observe patterns and trends in music works. Multidimensional zoom is also a common music Information visualization technology. By transforming music data into points in a multidimensional space, we can classify, cluster, and visualize music works. Multidimensional scaling can help us better understand the complex relationships and dynamic changes in music works. In the process of audio visualization, by using different shapes, sizes, colors and animations to express the characteristics of audio, visual psychology and auditory psychology can reach a unified state, so that audio can be visualized more appropriately. This article studies the application of CAD in music visual interactive design:

⊖ This article proposes a design scheme for a music visualization system and describes the process and processing mechanism of this system. The visualization effect can be basically consistent with the rhythm, style, emotions, and other factors of music, achieving good results.

⊖ This study is based on a music feature recognition model combined with CNN, and designs a comprehensive visualization method based on multiple audio features. Firstly, multiple music features are extracted, and then these features are comprehensively visualized to achieve more music information expression in the image.

The first section introduces the connotation of music visualization research and the significance of CAD in music visualization Interaction design; In the second section, a music visual Interaction design method based on multi audio feature recognition and CAD is proposed; The third section verified the performance of the modified system through simulation experiments; Finally, the main achievements and contributions of the article are summarized.

2 RELATED WORK

Music emotion divides the process of music emotion recognition into two stages: segmented level and two-stage learning. He and Ferguson [6] segmented the music samples into emotional categories, such as happiness, sadness, and excitement. In each emotion category, emotion analysis technology is used to identify music emotions and obtain the emotional scores for each emotion category. Using this model, emotion recognition can be performed on new music samples to obtain their emotion categories. The advantage of this method is that it can effectively improve the accuracy of emotion recognition. By using segmented levels and two-stage learning, emotional elements in music can be better captured and more accurately classified and recognized. In addition, this method can also provide personalized music recommendations based on different emotional categories to meet the different needs of users. In the field of psychology, background music has a potential impact on people's emotions and cognition. Klein et al. [7] analyzed that

specific types of music can affect people's emotions and perceptions, thereby affecting the evaluation of visual images. For example, solemn background music may increase people's perception of tragic images, while cheerful background music may make people more positive about joyful images. In the field of marketing, background music is also used as a marketing strategy to influence consumers' evaluation of products or services. For example, quiet background music may make consumers pay more attention to the details of the product, thereby enhancing their liking for the product. In the training process of GAN, the generator and discriminator are like a zero sum game, where every time one party wins, the other party loses. As the training progresses, the generator will gradually learn to generate more realistic data, while the discriminator will gradually learn to distinguish between real and generated data. GAN is widely used, such as image generation, image conversion, Inpainting, super-resolution, style conversion, art generation, etc. They have brought many innovative applications to computer vision, Natural language processing and other fields. Through adversarial training of these two networks, a generator can be obtained that can generate new data similar to real data. Lattner and Nistal [8] use generators to recover randomly disturbed music audio. Evaluate the quality of recovered music audio. It should be noted that GAN requires a large amount of training data and complex model settings when processing highly structured and complex data such as music audio. In addition, due to the instability and difficulty in training of the GAN model, it is necessary to carefully adjust and optimize the model parameters to obtain better results.

Liu et al. [9] conducted an automatic classification of various music genres by combining emotional and intelligent algorithms. It converts musical works into digital form and extracts musical features such as melody, harmony, rhythm, etc. Enhance the preprocessed data, such as increasing the amount of data, generating virtual data, etc., to improve the accuracy and stability of classification. Select important features from preprocessed data to reduce feature dimensions and improve classification efficiency. Use sentiment analysis algorithms to classify music works into emotions, such as happiness, sadness, anger, etc. Use intelligent classification algorithms to classify music works, such as support vector machines, decision trees, neural networks, etc. Integrate the results of emotional analysis and intelligent classification to obtain more accurate classification results. Evaluate and optimize the classifier to improve its accuracy and stability. In short, the automatic classification method of multiple music genres that combines emotion and intelligent algorithms can classify music works and obtain more accurate classification results, while also providing new technologies and tools for the music field. The digital representation of musical works refers to the transformation of musical works into digital form for processing and analysis on a computer. Lopes and Tenreiro [10] analyzed in digital representation that musical works can be represented as a series of data structures such as numbers, vectors, matrices, etc. This facilitates various mathematical and statistical analyses. Multidimensional scaling technology is a commonly used data dimensionality reduction technology, which can transform high-dimensional data into low-dimensional data while preserving the characteristics of the original data as much as possible. In music Information visualization, multidimensional scaling technology can be used to transform music works into 2D or 3D graphics for intuitive analysis and observation. For example, multidimensional scaling technology can be used to convert melody, harmony, rhythm, and other data of music works into multidimensional vectors, and then project them onto a two-dimensional plane to form a two-dimensional graph. By observing this figure, we can find patterns and trends in music works, such as the trend of melody, changes in harmony, etc. Melchiorre et al. [11] analyzed that EmoMTB is an intelligent audiovisual interface primarily used for music discovery and recommendation. It uses sentiment analysis technology to perceive emotional elements in music, providing users with more personalized and accurate music recommendation services. Through emotional analysis of music, EmoMTB can perceive emotional elements such as happiness, sadness, excitement, etc. in music, thereby providing users with more accurate music recommendations. Based on users' emotional needs and preferences, EmoMTB can recommend music that matches their emotional elements to meet their personalized music needs. EmoMTB can analyze users' emotional states, identify their emotions, such as happiness, sadness, anger, etc., and provide users with more personalized music recommendation

services. EmoMTB can understand users' music needs and preferences through interaction with users, such as voice interaction, gesture interaction, etc., in order to provide more accurate music recommendation services. Pan et al. [12] analyzed the impact of audiovisual integration on musical emotions. Research has shown that music can trigger emotional responses in people, and this response can manifest as behavioral and physiological changes. For example, certain types of high-energy music can stimulate people's vitality, which can be reflected by measuring physiological indicators such as heart rate, blood pressure, and skin resistance. In terms of CAD human-computer interaction audiovisual, there may be some studies exploring the impact of this technology on music emotions, but I cannot provide specific behavioral and physiological evidence. If you are interested in specific research or applications, we suggest that you further investigate relevant literature or consult with experts in the relevant field. In the implementation process, music data needs to be preprocessed into a format suitable for neural network input, usually represented in time series or frequency domain. Then we design a neural network model suitable for dealing with music multi notes, such as Recurrent neural network (RNN) or Convolutional neural network (CNN) and learn the patterns and rules in music data by training the model. Ultimately, the trained neural network model can be used for intelligent fusion of music multiple notes, generating new note sequences similar to the original music style based on the input music data. This method has broad application prospects in fields such as music generation, music recommendation, and music understanding.

Tian [13] uses feedforward neural networks to train and test data to achieve intelligent fusion of multiple notes. In specific implementation, deep learning technologies such as Convolutional neural network and Recurrent neural network can be used to intelligently fuse multiple musical notes. By adjusting the number of layers, nodes, Activation function and other parameters of the neural network, different intelligent fusion effects can be achieved. For example, Convolutional neural network can be used to extract features of melody, harmony, rhythm, etc. of music works, and input feature vectors into Recurrent neural network for sequence modeling, to achieve intelligent fusion of multiple musical notes. In the training process, the Cross-entropy Loss function can be used to optimize, and the Backpropagation can be used to update the weights and offsets of the neural network. Xu et al. [14] analyzed a dual mode emotion recognition algorithm that combines audio signal and speech context mixed features. The algorithm achieves emotion classification by combining early fusion methods and attention mechanisms. Specifically, the model uses bidirectional LSTM to model audio and text inputs separately, and then performs attention calculations on the vectors at each token position to achieve feature fusion. This method not only considers the different features of audio and text, but also effectively combines them through attention mechanisms to improve the accuracy of emotion recognition. In addition, this method also uses text tokens to find the most relevant features to audio, which is why this method is called learning alignment. Through this method, text and audio features can be aligned for better emotional analysis. Overall, this bimodal emotion recognition algorithm combines early fusion methods and attention mechanisms to achieve the fusion of mixed features of audio signals and speech context, thereby improving the accuracy of emotion recognition. Reinforcement learning is a kind of technology to learn the best behavior strategy through agent interaction with the environment. In music coordination, an intelligent agent can be an automatic melody coordination system, the environment can be a melody sequence, and the best behavior strategy is a melody transformation sequence that can make two melody sequences coordinate with each other. Zeng and Lau [15] achieved automatic melody coordination by exploring the structured reinforcement of melody sequences through learning. Represent a melody sequence as a state in the state space. The state can include the notes of the melody and their timestamps, as well as the sequence of previous transitions. Define a series of possible actions, such as adding, deleting, moving a note in a melody, or changing the volume of a note. Train the agent repeatedly until it can coordinate two melody sequences or reach a certain number of training times. This method can be applied to various types and styles of melody sequences and can be improved by adjusting parameters and model architecture. At the same time, this method can also be extended to other types of sound and music sequences and can even be applied to other art forms and interactive media. Zhang

[16] selects important features from multiple extracted features to reduce feature dimensions and improve classification efficiency. Integrate the selected important features together to obtain more accurate emotional recognition results. Train emotion recognition models using fused features, such as support vector machines, decision trees, neural networks, etc. Evaluate the trained emotion recognition model to determine its accuracy and stability. Optimize the emotion recognition model to improve its accuracy and stability. In music emotion recognition methods based on multi feature fusion, feature extraction and fusion are two key steps. By fusing different musical features together, it is possible to better capture the emotional information of musical works, thereby improving the accuracy and stability of emotional recognition. In addition, using multi feature fusion can also reduce the dimensionality of features and improve classification efficiency. In summary, music emotion recognition methods based on multi feature fusion can improve the accuracy and stability of music emotion recognition by fusing multiple different music features together, and also provide new technologies and tools for the music field.

3 METHODOLOGY

3.1 Audio Information Extraction

Data visualization method is a process of presenting abstract data in the form of intuitive graphics or images, and visualized data can convey the information in the data more clearly and effectively. In order to improve the classification accuracy of audio classification system and ensure the availability of data sets, it is needed to preprocess, extract and select the features that are beneficial to the classification accuracy. The visual method of audio data can be used to display the auditory content visually. Although the current research on audio visualization has played a role in many scenes, there are some problems in the related research and application, such as inaccurate extraction of some sound features, subjective results of data visualization, and inability to balance the presentation form and content in the visualization process.

After the music is composed, the interpretation methods such as playing musical instruments or singing human voices can convey it to the audience. According to the characteristics of sound data and the actual training results, the training methods and parameters are adjusted until a higher recognition rate is obtained. Finally, the trained network parameters are applied to the training of two types of neural networks in other categories until all the categories are combined in pairs. The music information extraction process of the music information visualization system is shown in Figure 1.

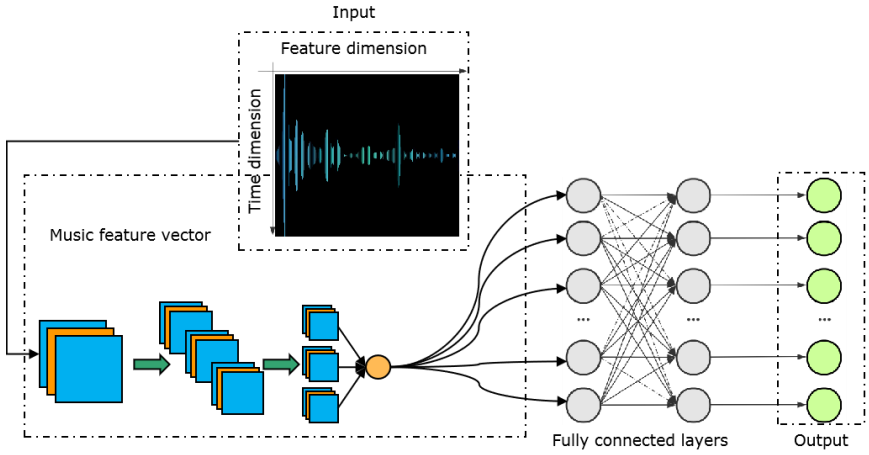

Figure 1: Music information extraction process.

Before feature detection, one-step silent frame discrimination is needed to reduce unnecessary interference and calculation. If these frames are allowed to enter the next set of frames to extract features, it will not only increase the calculation amount, but also make the extracted features biased, which will seriously lead to the failure of the experiment. In order to make the audience fully understand the emotions and thoughts expressed by music, the creators or communicators adopt different auxiliary expressions. In order to simplify the extraction and analysis operation, it is needed to frame and window a large segment of audio data, and then calculate its short-term characteristics frame by frame by formula. After preprocessing the musical samples, it is needed to extract as many attributes from the musical samples as possible that can reflect the task according to the specific identification and classification task, then train and model the extracted feature quantity, and use the established extraction model to identify and classify the test samples of different music to get the prediction results.

A piece of music signal can be divided into several frames:

$$N_{\text{frames}} = \left[\frac{N_x - N_0}{N_f - N_0} \right] \tag{1}$$

Where N_x is the total length of the signal, N_0 is the length of interframe overlap, and N_f is the length of one frame.

After framing, windowing is required. The purpose of windowing is to gradually change the amplitude of a frame signal to zero at the beginning and end, which can improve the resolution of the spectrum. When the window moves continuously and smoothly on the original musical sound signal, it completes the operation of framing and windowing at the same time. The framing can be continuous or discrete. Different window functions will also have different effects on the results of framing and windowing.

Window functions commonly used in music include rectangular window, Hanning window, etc:

Rectangular window:

$$w(n) = 1, 0 \le n \le N-1 \tag{2}$$

Hanning window:

$$w(n) = 0.5\left(1 - \cos\left(2\pi \frac{n}{N-1}\right)\right), 0 \le n \le N-1 \tag{3}$$

Among them, N is the frame length of the music signal, and different window functions will have different influences on the analysis of the characteristic parameters of the music signal. The choice of window functions should be based on the characteristics of the music signal parameters, which has achieved better extraction of the essential characteristics of music.

$$\Delta f = \frac{1}{NT_s} \tag{4}$$

It can be seen that when the sampling period T_s is constant, the frequency resolution Δf decreases with the increase of the window size N, that is to say, the frequency resolution increases correspondingly while the time resolution decreases.

The analog sound signal in audio signal is a continuous quantity, which is composed of many sine waves with different vibration amplitudes and frequencies. In order to facilitate the computer to process and analyze the audio files, it is needed to sample and quantize the analog signals in the audio first.

3.2 Visual Design of Music

The graphic elements of audio features are not only line type and color, but also the size of shape. If you want to show exaggerated features or emphasize a certain feature, you can use large-area or small and dense graphics. Small and dense graphic combinations can be used when the music rhythm is fast, and large and full graphics can be used when the music characteristics do not change obviously with time. In the research of many fields, such as data mining, it is usually needed to collect a large quantity of data that can represent the characteristics of the research object. Multivariable and large sample data sets not only enrich the information, but also increase the task and workload of the research, and the correlation between the data will also increase the complexity of the research. It is needed to extract and select the features that are beneficial to the classification accuracy.

Users can fully understand the information that data can convey by observing the visualization content. When deleting the data in the data set, important information may be lost in the original data set, resulting in a huge error in the final result. Therefore, it is needed to study the corresponding data selection algorithm to reduce the amount of data in the data set, while retaining as much useful information as possible in the original data set to complete the optimization of the original data set. The CNN model of music feature recognition is shown in Figure 2.

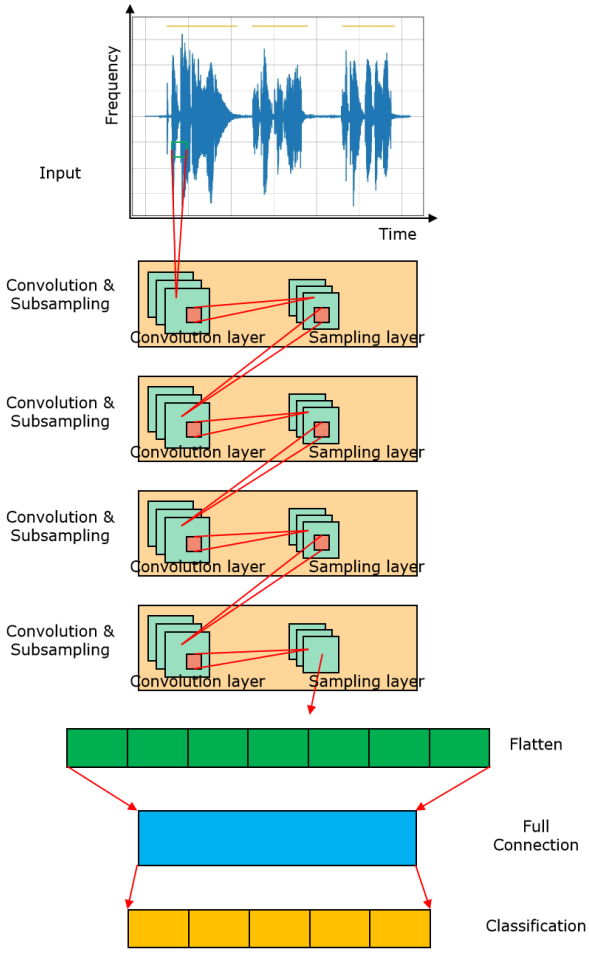

Figure 2: CNN model of music feature recognition.

Because there may be high similarity between different parameters in the data set, and different data parameters have different degrees of importance in solving the target problem, how to find out which data parameters are redundant and which data parameters play a decisive role in solving the target problem. The visual method of audio data can be used to display the auditory content visually. Although the current research on audio visualization has played a role in many scenes, there are some problems in the related research and application, such as inaccurate extraction of some sound features, subjective results of data visualization, and inability to balance the presentation form and content in the visualization process.

The CNN model is defined as:

$$s(i,j)=(X*W)(i,j)+b=\sum_{k=1}^{n_{in}}(X_k*W_k)(i,j)+b \tag{5}$$

Both voice and music signals belong to unsteady signals, but they can be considered to be steady in a short time range. In the short-term analysis of audio signals, audio signal segments are generally framed first, and the frame length of framing can not be too short or too long. In order to simplify the extraction and analysis operation, it is needed to frame and window a large segment of audio data, and then calculate its short-term characteristics frame by frame by formula. Hash codes can be defined as:

$$q=\phi\left(W_H^T h_{end}+v_H\right) \tag{6}$$

$W_H \in R^{H\times K}$ represents the weight of the hash layer, $v_H \in R^{K\times 1}$ represents the offset, and $\phi(\cdot)$ represents the tanh function.

Then, the output value of CNN output layer node is used as the input of HMM, and the posterior probability of HMM state is calculated by Softmax regression model. The output vector set at time t is O_n, and the distribution expression of output state is:

$$y_n(s)=P(s|O_n)=\frac{\exp[a_n(s)]}{\sum_s \exp[a_n(s)]} \tag{7}$$

Where $a_n(s)$ is the activation probability of the output layer state s, that is, the output value.

For time-varying audio signals, the signal segments within a certain duration range are considered as steady-state signals, while signals beyond this range are still not considered as steady-state signals. Therefore, in the process of framing, a partial overlap is usually set between two adjacent frames. In order to effectively convey the information in the data, visualization needs to ensure that the artistic form of presentation is fully considered on the basis of presenting accurate data characteristics, so as to lead users to deeply observe the obscure and complicated data characteristics and associations in the data set.

$$S=\sum_{n-1}^{m=0} x*(n-m)*w(m) \tag{8}$$

$$a_k=\begin{cases}1, D_{DTW}(v_i,v_k)<y_i(v_i)\\ 0, other\end{cases} \tag{9}$$

At this time, $A(G)$ is a symmetric matrix, which is also a musical note similarity matrix.

$$PCP(p,m)=\sum_{g(k)=p}|STFT(k,m)| \tag{10}$$

4 RESULT ANALYSIS AND DISCUSSION

In this article, the deep belief network (DBN) recognition method is designed as the comparison method of the experiment, and the same music melody is segmented and recognized according to three methods. The experimental data set consists of 100 tone value sequence files and their tone sequences plus different levels of noise, and each tone value sequence is a simple spectrum sequence selected from a song. The result of sound station estimation in the presence of fundamental frequency is shown in Figure 3.

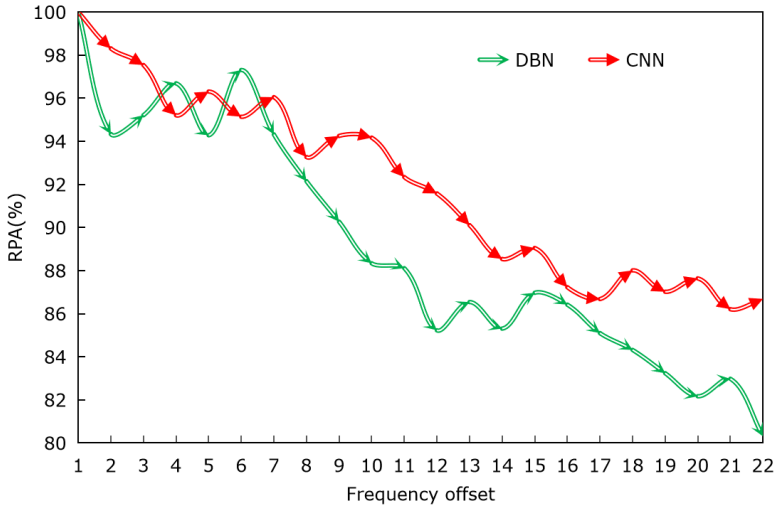

Figure 3: Fundamental or harmonic offset.

The proposed CNN method obtains higher original pitch accuracy than the traditional method. Model training is particularly important in music recognition, which requires users to provide a large quantity of original databases. Therefore, before starting the research on musical tone recognition, we should first establish a musical tone database. The results of recall of speech frames with different classifiers are shown in Figure 4.

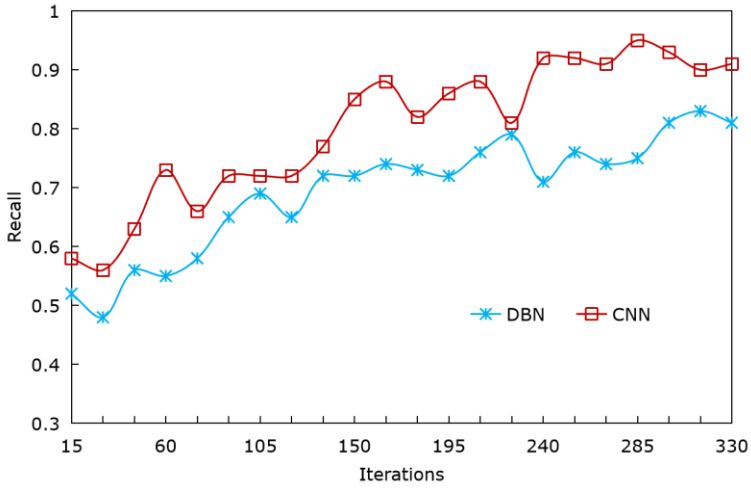

Figure 4: Recall rate of speech frames.

CNN has continuous-time nonlinear dynamics, large-scale parallel distributed processing, global function of network, high robustness and learning and association ability.

Two-dimensional music visualization needs low data dimension and is relatively simple to realize, while three-dimensional music visualization is suitable for constructing complex objects because of its strong expressive force and is mainly used in occasions with high immersion requirements. In two-dimensional visualization, static images show the correlation of parameters of two dimensions of audio, or mine some structural and text features that appear, and there are also dynamic representations of images changing with time. However, three-dimensional transcends the dimension limitation of plane space and can superimpose time flow on two dimensions other than time in two-dimensional space. Figure 5 shows the original pitch accuracy of this method on different databases.

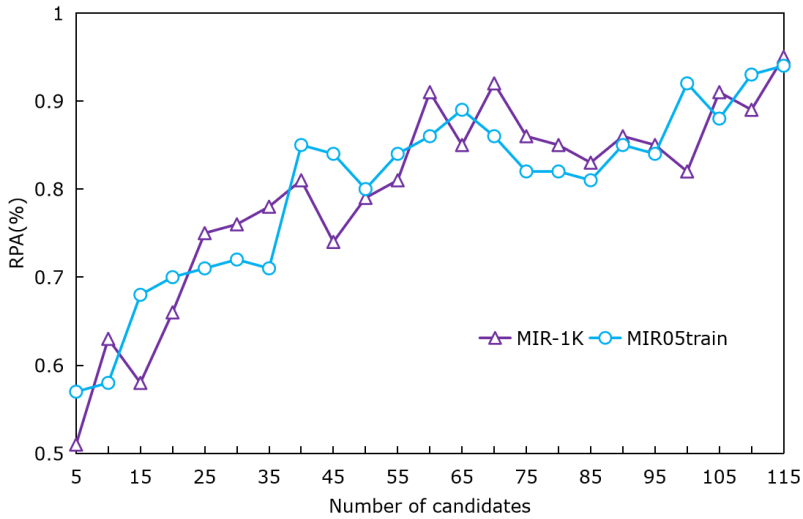

Figure 5: Original pitch accuracy.

When the learning rate is too large, the weight correction of the system may not be adjusted to the best, and the system will be unstable; When the learning rate of the system is too small, the training cycle will be long and the learning rate of the system will be too slow, thus affecting the system's ability.

In the audio feature detection, 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. The error of different algorithms is shown in Figure 6.

CNN shows great flexibility and adaptability to the processing of nonlinear signals and systems whose data cannot be described by rules and formulas. For the audio classification system, the more feature parameters, the more detailed the description of the target audio, and the more profound the impact on the classification system. In this large quantity of features, there must be some features that have little or even negative effects on the classification results of the system, and too many feature parameters will increase the calculation time of the system and reduce the operation efficiency of the system. The experimental results of different algorithms to identify time are shown in Figure 7.

In many cases, convertible data itself is not the primary problem in the works, but through the transmission of these data, the interesting points of some data streams can be better understood and experienced by people or perceived in a new way. The advantage of this method is that it is simple in feature detection, and the visualization effect is richer, and it can be displayed visually with more representative features. This also makes listening to music and watching music achieve

a good combination, so that the audience has a better understanding of the feelings expressed by music.

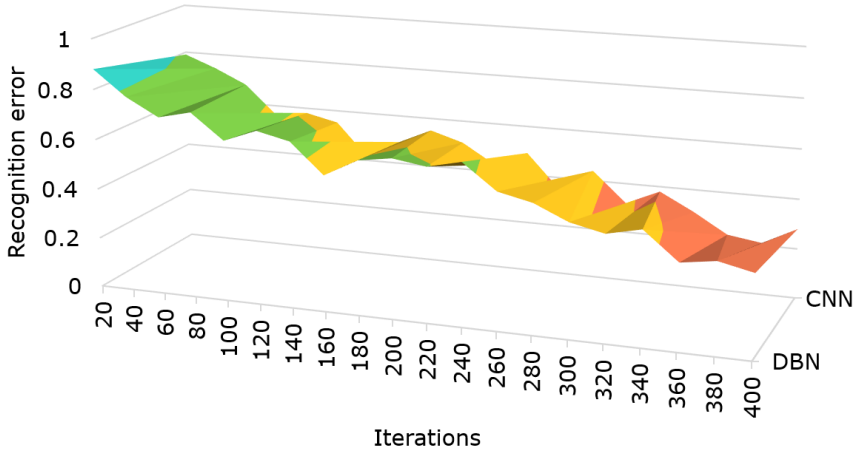

Figure 6: Error situation of different algorithms.

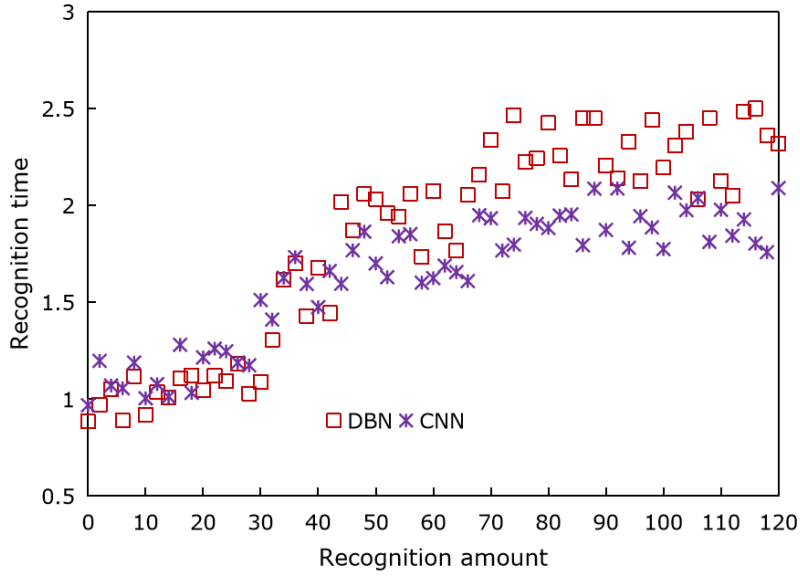

Figure 7: Experimental results of identification time.

5 CONCLUSIONS

Data visualization method is an intuitive, simple and reasonable method to summarize and present data, and its main purpose is to convey information more clearly and efficiently by means of graphics or images. In the process of music visualization, by using different shapes, sizes, colors and animations to express the characteristics of audio, visual psychology and auditory psychology can reach a unified state, so that audio can be visualized more appropriately. In this article, a visual expression method of music based on multi-audio features and CAD is proposed, and the

main melody is extracted and highlighted visually. The results show that this method is simple in feature detection, and the visualization effect is more abundant, and it can be displayed visually with more representative features. In many cases, convertible data itself is not the primary problem in the works, but through the transmission of these data, the interesting points of some data streams can be better understood and experienced by people, or perceived in a new way. This method has the advantages of easy implementation, simple application and fast implementation, and can better express the theme information and emotion of music, which is helpful to improve the human-computer interaction experience of music visualization. If the algorithms and functions to be realized are complex enough, such as 3D virtual reality technology, it is needed 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.

Fei Ma, https://orcid.org/0009-0007-3233-2260

REFERENCES

[1] Baradaran, F.; Farzan, A.; Danishvar, S.; Sheykhivand, S.: Customized 2D CNN model for the automatic emotion recognition based on EEG signals, Electronics, 12(10), 2232. https://doi.org/10.3390/electronics12102232

[2] Bishop, L.; Cancino, C.-C.; Goebl, W.: Moving to communicate, moving to interact: Patterns of body motion in musical duo performance, Music Perception: An Interdisciplinary Journal, 37(1), 2019, 1-25. https://doi.org/10.1525/mp.2019.37.1.1

[3] Cho, J.-D.: A study of multi-sensory experience and color recognition in visual arts appreciation of people with visual impairment, Electronics, 10(4), 2021, 470. https://doi.org/10.3390/electronics10040470

[4] Dotov, D.; Bosnyak, D.; Trainor, L.-J.: Collective music listening: movement energy is enhanced by groove and visual social cues, Quarterly Journal of Experimental Psychology, 74(6), 2021, 1037-1053. https://doi.org/10.1177/1747021821991793

[5] Georges, P.; Seckin, A.: Music information visualization and classical composers discovery: an application of network graphs, multidimensional scaling, and support vector machines, Scientometrics, 127(5), 2022, 2277-2311. https://doi.org/10.1007/s11192-022-04331-8

[6] He, N.; Ferguson, S.: Music emotion recognition based on segment-level two-stage learning, International Journal of Multimedia Information Retrieval, 11(3), 2022, 383-394. https://doi.org/10.1007/s13735-022-00230-z

[7] Klein, K.; Melnyk, V.; Voelckner, F.: Effects of background music on evaluations of visual images, Psychology & Marketing, 38(12), 2021, 2240-2246. https://doi.org/10.1002/mar.21588

[8] Lattner, S.; Nistal, J.: Stochastic restoration of heavily compressed musical audio using generative adversarial networks, Electronics, 10(11), 2021, 1349. https://doi.org/10.3390/electronics10111349.

[9] Liu, J.: An automatic classification method for multiple music genres by integrating emotions and intelligent algorithms, Applied Artificial Intelligence, 37(1), 2023, 2211458. https://doi.org/10.1080/08839514.2023.2211458

[10] Lopes, A.-M.; Tenreiro, M.-J.-A.: On the complexity analysis and visualization of musical information, Entropy, 21(7), 2019, 669. https://doi.org/10.3390/e21070669

[11] Melchiorre, A.-B.; Penz, D.; Ganhör, C.; Lesota, O.; Fragoso, V.; Fritzl, F.; Schedl, M.: Emotion-aware music tower blocks (EmoMTB): an intelligent audiovisual interface for music discovery and recommendation, International Journal of Multimedia Information Retrieval, 12(1), 2023, 13. https://doi.org/10.1007/s13735-023-00275-8

[12] Pan, F.; Zhang, L.; Ou, Y.; Zhang, X.: The audio-visual integration effect on music emotion: Behavioral and physiological evidence, PloS One, 14(5), 2019, e0217040. https://doi.org/10.1371/journal.pone.0217040

[13] Tian, Y.: Multi-note intelligent fusion method of music based on artificial neural network, International Journal of Arts and Technology, 13(1), 2021, 1-17. https://doi.org/10.1504/IJART.2021.115763
[14] Xu, Y.; Su, H.; Ma, G.; Liu, X.: A novel dual-modal emotion recognition algorithm with fusing hybrid features of audio signal and speech context, Complex & Intelligent Systems, 9(1), 2023, 951-963. https://doi.org/10.1007/s40747-022-00841-3
[15] Zeng, T.; Lau, F.-C.: Automatic melody harmonization via reinforcement learning by exploring structured representations for melody sequences, Electronics, 10(20), 2021, 2469. https://doi.org/10.3390/electronics10202469
[16] Zhang, Y.: Music emotion recognition method based on multi feature fusion, International Journal of Arts and Technology, 14(1), 2022, 10-23. https://doi.org/10.1504/IJART.2022.122447