Engineering Reports

WILEY

Engineering Reports Open Access

RESEARCH ARTICLE OPEN ACCESS

Application of Music Data Visualization Technology in Music Appreciation Teaching

Xiaowei Chen

College of Music, Xinjiang Normal University, Urumqi, China

Correspondence: Xiaowei Chen (cxiaowei1210@163.com)

Received: 23 December 2025 | **Revised:** 6 May 2026 | **Accepted:** 14 May 2026

Keywords: computer science | interactivity | music appreciation | music data | visualization

ABSTRACT

Traditional music appreciation teaching has limitations such as excessive reliance on oral explanations, passive acceptance by students, and difficulty in visually presenting the internal structure and emotional logic of music. The purpose of this article is to explore the integration application path of music data visualization technology and Deep Learning (DL) in teaching scenarios. By constructing an experimental system that simulates a real appreciation environment, this study utilizes deep learning algorithms to efficiently preprocess audio data, extract multidimensional features, and recognize rhythm information, achieving intuitive and visual expression of music data. The experiment selected three typical music samples: popular, classical, and folk for verification. The results indicate that this method has high accuracy in audio feature recognition and can generate visual representations that accurately reflect different musical styles, rhythm structures, and emotional features. The research conclusion shows that data visualization technology based on deep learning can significantly enhance the interactivity and participation of music appreciation teaching, assist learners in deeply understanding the inherent laws of music, and effectively stimulate aesthetic creativity imagination.

1 | Introduction

In the information age of the 21st century, music, as an important part of human culture, has undergone unprecedented changes in its educational methods. With the rapid development of science and technology, especially the continuous progress of computer technology and artificial intelligence technology, the processing, analysis and presentation of music data have undergone earth-shaking changes [1]. Music data visualization technology, as an important part of this transformation, is gradually being integrated into the field of music appreciation, offering fresh perspectives and immersive experiences that enrich how audiences engage with music [2]. At the same time, advances in Deep Learning (DL) and audio recognition technologies have opened up unprecedented opportunities for enhancing music appreciation through data-driven approaches. Music data visualization involves representing musical information—such as audio signals, score data, and structural features—in graphical, pictorial, or animated visual forms, enabling listeners to intuitively perceive, interpret, and emotionally connect with the underlying elements of a musical work [3]. The origin of this technology can be traced back to the early field of music signal processing. At that time, researchers mainly focused on how to transform music signals into visual waveforms or spectrograms for music analysis or teaching [4, 5]. However, with the continuous development of computer graphics and human-computer interaction technology, the forms and contents of music data visualization have become more diverse, from simple waveform diagrams and spectrum diagrams to today's dynamic music videos, three-dimensional music models, Virtual Reality (VR) music

experiences, and so on, greatly broadening the dimensions of music expression [6, 7].

The traditional way of music appreciation often depends on the oral description of the lecturer and the passive listening of the audience. This "one-to-many" communication mode can convey basic music knowledge to a certain extent, but it is obviously insufficient in stimulating the audience's aesthetic perception, deepening the understanding of the inherent logic of music and improving the overall music literacy [8]. The traditional way is difficult to visually present the internal structure, emotional expression and development of music works, and the audience often stays in the superficial feeling stage [9]. In addition, this kind of way lacks personalization and interactivity, and can't meet the differences in interest preferences, cognitive level and aesthetic needs of different audiences, resulting in uneven appreciation effects [10]. With the help of the construction of deep neural network, the computer can automatically learn the feature representation of music and realize the tasks of music style recognition, emotion analysis and even music generation. The breakthrough of these technologies not only provides a new perspective for music research, but also brings new possibilities for music appreciation. For example, the DL-based system can intelligently recommend music works or visually analyze the content according to the audience's auditory preferences and appreciation history. DL can also automatically analyze and interpret music works, and help the audience to understand the melody trend, rhythm and emotional changes in a more in-depth way in a visual form, thus enhancing the immersion of music appreciation.

Audio recognition technology can accurately identify notes, rhythm, melody, and other elements in music, and even capture the singer's emotional expression and deductive skills [11]. In the process of music appreciation, with the help of audio recognition technology, the audience can get an immediate analysis of the structure and performance characteristics of the listened works, so as to more keenly perceive the artistic tension of music. This technology can also provide an objective style and emotional label for music works, and help the audience to establish a deeper aesthetic cognition [12]. Music data visualization technology can transform abstract music information into intuitive visual images [13]. For example, through dynamic audio-visual video or three-dimensional music model, the audience can clearly "see" the unfolding process of music in time and space, and intuitively feel the interweaving of melody, harmony level, and rhythm [14]. Music data visualization can also be deeply integrated with DL technology to build an intelligent music appreciation assistant system [15]. By analyzing the audio data through DL algorithm, the system can automatically generate visual interpretations and emotional annotations for different music works, helping the audience to understand the composition intention, performance style, and emotional atmosphere more comprehensively [16].

Artificial intelligence and multimedia technology profoundly reshape the paradigm of modern education. The core of this paper is to explore the application potential of music data visualization technology in music appreciation teaching and build an intelligent bridge connecting hearing and visual cognition. The research will deeply analyze how this technology breaks through the limitations of traditional teaching and optimizes the learning effect from three aspects. On the one hand, students' rational understanding of complex music structure and logical laws is enhanced by visual atlas; on the other hand, students' emotional resonance and aesthetic experience are deepened by dynamic visual feedback; on the other hand, it can activate students' creative thinking and stimulate their aesthetic imagination in audio-visual interaction. This paper combines the powerful feature mining ability of deep learning with audio recognition technology and opens up a new path of cutting-edge technology application in digital education scenes through visual modeling. On the one hand, the visual atlas can help students understand the complicated music structure and logical laws more clearly and enhance their rational understanding; on the other hand, dynamic visual feedback is helpful to deepen students' emotional resonance and aesthetic experience; in addition, audio-visual interaction can also activate students' creative thinking and stimulate their aesthetic imagination.

2 | Literature Review

2.1 | Present Situation of Music Data Visualization Technology

Visualization of music data is a bridge between music and visual arts. Its history can be traced back to the early stage of music signal processing. Initially, researchers used simple waveforms and spectra to show the time and frequency characteristics of music [17]. These basic visualization tools, although rudimentary, provide an intuitive perspective for music analysis [7]. With the development of computer graphics, music visualization has gradually moved from static to dynamic, expanded from two-dimensional to three-dimensional, and even incorporated VR and Augmented Reality (AR) technologies, providing listeners with an immersive music experience.

In the field of music appreciation, the application of music data visualization technology is mainly reflected in three aspects: music ontology analysis, structure perception and aesthetic inspiration [18]. Displaying the timbre characteristics of different musical instruments through spectrograms is helpful for listeners to intuitively understand the composition and differences of sound spectrum; using dynamic graphics to present the ups and downs of melody lines can help listeners to grasp the structural context and development logic of music more clearly; in addition, with the help of a three-dimensional model to show the layout relationship of music works in time and space, it can stimulate the imagination of the audience.

2.2 | The Integration of Deep Learning and Music Appreciation

The rise of DL has brought innovative changes to music appreciation. By constructing the deep neural network, the computer can automatically learn and extract the high-order features of music. In the field of music appreciation, the application of DL is mainly reflected in personalized music recommendation, emotional analysis of works, and auxiliary aesthetic experience [19]. For example, the music recommendation system based on deep learning can intelligently recommend music works or visually

interpret the content according to the audience's listening history, preferences, and aesthetic tendencies [20]. In addition, with the help of audio analysis technology, the system can automatically analyze music works, identify the emotional trend, rhythm, and structural level, and present them in a visual way, helping listeners to perceive the internal logic of music more intuitively. DL can also be used to generate music fragments with specific styles, which not only enriches the appreciation materials but also provides listeners with a new perspective to understand different music languages and creative thinking.

2.3 | The Application of Audio Recognition Technology in Music Appreciation

Audio recognition technology, especially audio recognition based on DL, shows a broad prospect in the field of music appreciation. This technology can accurately identify the basic elements of music such as pitch, rhythm, and harmony, and even further analyze the emotional color and expression style conveyed by singers or performers.

In the process of music appreciation, audio recognition technology can analyze the works in real time and transform abstract elements such as melody trend, rhythm, and emotion fluctuation into intuitive visual information, helping the audience to perceive the structural level and artistic expression of music more clearly [21]. This real-time and dynamic analysis mechanism is of great significance to deepen the audience's understanding of the connotation of music and enhance aesthetic sensitivity. In the scene of work analysis and guided appreciation, this technology can also automatically generate music structural annotation, emotional label, or style classification, reduce the cost of manual analysis, and enhance the professionalism and consistency of appreciation content. By combining auditory experience with visual interpretation, audio recognition technology is becoming an important bridge between the public and music art.

2.4 | Integrated Technological Approaches for Enhanced Music Appreciation

By integrating music data visualization technology, DL and audio recognition technology, a more intelligent and efficient music appreciation support system can be constructed [22]. The system can analyze the audience's listening behavior, preference characteristics, and emotional feedback with the help of deep learning, so as to intelligently recommend music works that meet their aesthetic interests [23]. Audio recognition technology can analyze the played music in real time and accurately capture elements such as melody, rhythm, harmony, and emotional expression; combined with music data visualization technology, this abstract music information is transformed into dynamic graphics, color changes, or interactive maps, which provide intuitive and vivid aesthetic guidance for the audience.

Combined with VR technology, it can also create an immersive music appreciation environment. In the virtual space, the audience can not only see the structure and emotional flow of music, but also explore the details of the work through interactive operation and feel the artistic charm and inner mystery of music more deeply in the multi-sensory integration. This technology fusion expands the boundary of traditional appreciation and also provides an innovative path for public aesthetic education and digital humanistic practice.

3 | Application of Music Data Visualization Technology in Music Appreciation

The core value of music data visualization technology lies in its ability to transform abstract music data into intuitive visual information, thereby assisting learners in better comprehending the internal structure, emotional expression, and stylistic characteristics of music [24]. This transformation enriches pedagogical approaches and enhances students' capacity to perceive and analyze musical elements. In the context of music appreciation teaching, the application of this technology is primarily reflected in three aspects: first, visualization tools display basic elements such as notes, rhythm, and chords, enabling students to grasp music theory concepts more intuitively; second, presenting spectrum diagrams, waveform diagrams, and melody lines through visualization aids learners in deeply analyzing the composition and style of musical works, fostering a deeper understanding of their artistic connotation; third, these visual tools stimulate creative inspiration, guide students to explore diverse musical styles and expressions, and cultivate their aesthetic perception and innovative thinking.

In music appreciation, the feature detection method based on DL can automatically learn the complex characteristics of music and provide more accurate support for personalized guide and intelligent analysis [25]. For example, by transforming music data into point cloud data, PointNet can learn the spatial distribution of music data and its feature relationship and provide powerful feature representation for music classification, emotion analysis, style recognition, and other tasks. This method can adapt to different types of music and their styles and provides technical support for personalized music appreciation experience and intelligent work analysis (see Figure 1).

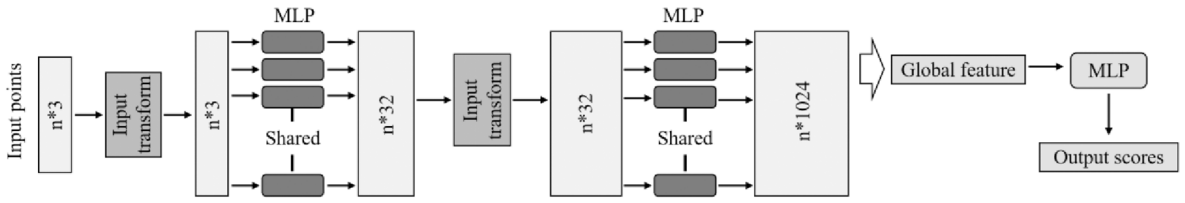

FIGURE 1 | Network structure of PointNet.

The structure comprises an input layer, a transformation network, feature extraction layers, and a max pooling layer. The input layer receives music point cloud data, while the transformation network aligns the feature space. Convolutional layers are used to extract local features, and the max pooling layer aggregates global information. This architecture ensures a robust representation of the spatial distribution characteristics of music.

In this study, the Convolutional Neural Network (CNN) algorithm is selected to process music data. The music data used in this study was sourced from publicly available datasets and authorized collections. The data acquisition process complied with copyright regulations and was conducted solely for academic research purposes. The samples cover three music genres—pop, classical, and folk—with a uniform sampling rate of 44.1 kHz and stored in WAV format. All data processing was carried out in a simulated environment to ensure experimental controllability. In order to input music data into the CNN model effectively, a series of preprocessing work is carried out at first. The core step of this preprocessing process is to convert the original audio signal into a visual representation that can reflect the frequency and time characteristics of audio; that is, spectrogram or Mel spectrogram. This transformation makes music data appear in the form of images. Next, the label of music rhythm is specially coded and converted. In this study, the method of single heat coding is adopted, and each rhythm type is transformed into a unique vector consisting of 0 and 1.

After the above pretreatment, a CNN model suitable for music data processing is constructed. This model is designed to receive preprocessed spectrogram or Mel spectrogram as input, and learn the deep features in music data through its internal convolution layer, pooling layer and full connection layer. A typical CNN architecture comprises various convolution layers, pooling layers, and fully connected layers. Although CNN were originally designed for image processing, music spectrograms are structurally analogous to images, as both contain spatially localized features. As a result, CNN are capable of effectively extracting both temporal and spectral characteristics from the spectrograms. The extraction process of MFCC is designed to emulate the perceptual characteristics of the human auditory system. The audio is first divided into frames and windowed, followed by the application of a Mel filter bank, and finally transformed using the discrete cosine transform. This process converts the audio signal into a feature vector that reflects timbral energy, facilitating subsequent visual mapping. Given a set of N notes, one can formulate a maximum likelihood model as follows:

$$\left(\hat{f}_0^1, \ldots, \hat{f}_0^N\right) = \arg \max_{f_0^1, \ldots, f_0^N \in F} p\left(O \middle| f_0^1, \ldots, f_0^N\right) \tag{1}$$

In this article, the time-domain discrete signal is segmented into overlapping frames using a framing technique, where O represents the observation frequency spectrum, $f_0^1, \ldots, f_0^N$ denotes the fundamental frequency in N logarithmic scales, and F indicates the potential frequency range of the fundamental frequency.

$$X_{\mathrm{STFT}}(k, n) = \sum_{m=0}^{N-1} x(n-m) w(m) e^{-j2k\pi m/N} \tag{2}$$

In this context, k denotes frequency coordinates, n signifies the center of the short-time Fourier transform window, and $w(m)$ refers to the hamming window. The prominent $X_{\mathrm{STFT}}(k, n)$ is transformed into a 12-dimensional vector, $p(k)$, where each dimension signifies the intensity of a semitone, following the mapping formula:

$$p(k) = \left[12 \log_2 \left(\frac{k}{N \cdot f_{\mathrm{sr}}} \cdot f_{\mathrm{ref}}\right)\right] \bmod 12 \tag{3}$$

here f_{ref} represents the reference frequency and f_{sr} the sampling rate. The frequency values of the points corresponding to each sound level are summed to yield the contour characteristic components for each sound level in each time segment, using the following formula:

$$PCP(p) = \sum_{k=p(k)=p} |X(K)|^2 \qquad p = 1, 2, 3, \ldots, 11 \tag{4}$$

In this article, the tone level is represented by a 12-dimensional vector that captures the relative intensity of notes across 12-tone intervals in the chromatic scale. At the model's output, a softmax layer is employed to generate the probability of each rhythm type.

MFCC (Mel frequency cepstral coefficients) coefficient is a common feature detection method in music signal processing. It can effectively reflect the spectrum characteristics of audio signals and is especially suitable for music identification and classification tasks [26]. In music appreciation, the combination of MFCC extraction and visualization technology can provide listeners with more intuitive music analysis tools [27].

Figure 2 shows the extraction process of linear MFCC coefficients of audio short frames. Through this process, the original audio signal is transformed into a series of MFCC coefficients, which effectively reflect the energy distribution characteristics of sound in different frequency bands. By presenting these coefficients in visual forms such as charts or images, the audience can intuitively observe the changes in the spectral characteristics of music in different time periods, so as to more clearly perceive its rhythm, timbre texture, and emotional ups and downs. This data-based visual guide enhances the depth and dimension of music appreciation and also builds a bridge to the inner structure and artistic expression of music for non-professional listeners.

The process consists of pre-emphasis, framing, windowing, fast Fourier transform, Mel filtering, and discrete cosine transform. The resulting coefficient matrix reflects the energy distribution of the signal across different frequency bands. In music appreciation, the visualization technology of MFCC coefficient has been widely used. For example, by showing the MFCC coefficient maps of different musical instruments, the audience can intuitively observe the unique characteristics of various musical instruments in the spectrum energy distribution. In addition, by comparing the MFCC visual maps of different music styles, the audience can clearly perceive the differences in rhythm density, frequency domain energy center of gravity, and dynamic changes [28]. MFCC visualization can also be combined with real-time audio analysis to provide instant auditory–visual feedback for listeners in digital guided tours or interactive exhibitions, so that they can "see" the structure and emotional flow of music synchronously during listening.

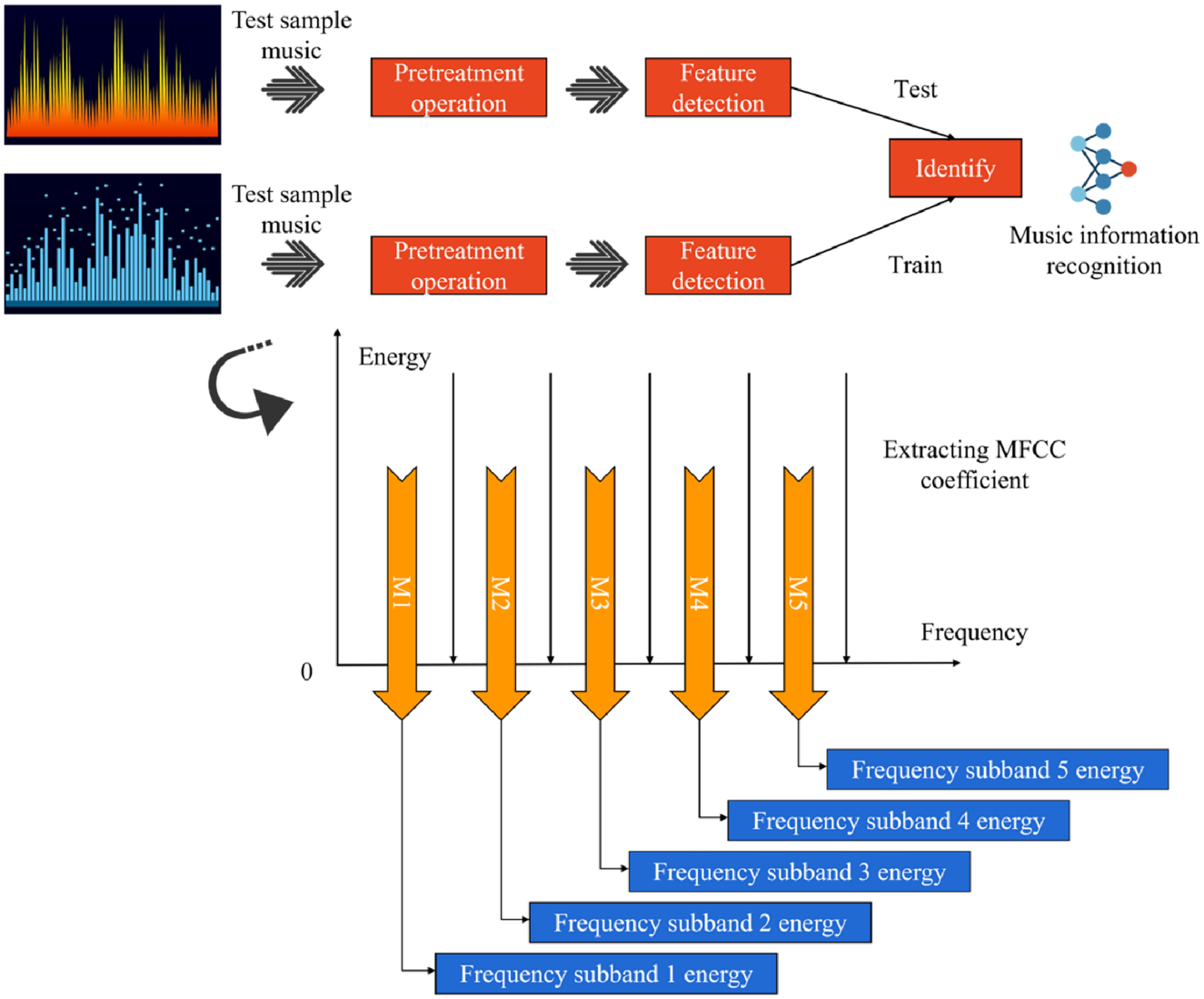

FIGURE 2 | Linear MFCC coefficient extraction of audio short-time frames.

In audio signal processing, analyzing the frequency spectrum is essential for extracting tonal features. Among these, the chromaticity vector and mode features are particularly effective in describing musical modes. We employ the Constant Q Transform for time-frequency domain conversion and derive chromaticity characteristics in the frequency domain under this transform. Additionally, we design a filter bank following an $\log_2$ change rule, where each filter's center frequency is determined as follows:

$$f_c\left(k_{lf}\right) = f_{\min} \cdot 2^{\frac{k_{lf}}{\beta}} \tag{5}$$

In this study, the minimum center frequency of the filter bank is set at 220 Hz, denoted as $f_{\min}$. The filter index number is represented by k_{lf}, while β signifies the number of filter banks per octave. To measure the interchangeable energy between signals $x(t)$, $y(t-\beta)$ and determine their similarity, the cross-correlation function is employed. The function is defined as follows:

$$\gamma_{xy}(\alpha,\beta) = \int_{-\infty}^{+\infty} [x(t) \cdot y(\alpha \cdot (t-\beta)) \cdot dt] \tag{6}$$

When $\alpha = 1$, β is utilized to remove the time displacement.

With the representation function Φ defining $R^{H\times W\times D} \to R^d$ and a specified target code $\varphi_o \in R^d$, the objective of the music data visualization technique is to identify an image $X \in R^{H\times W\times D}$:

$$\min_{X\in R^{H\times W\times D}} R_\alpha(X) + R_{TV\beta}(X) + Cl\left(\Phi(X), \Phi_o\right) \tag{7}$$

The loss function l quantifies the discrepancy between the desired coding and the coding of the produced image, while the regularization term $R_\alpha + R_{TV^\beta} : R^{H\times W\times D} \to R_+$ encapsulates the prior knowledge about natural images. C denotes the balance between the natural image prior and the target loss. In generating an image through maximum activation, the neural network identifies regions on the map where the target neuron is active and refines these areas via optimization to enhance activation. This iterative process continues until the image converges. The loss function is formulated as:

$$l\left(\Phi(X), \Phi_o\right) = -\frac{1}{Z}\left\langle \Phi(X), \Gamma\left(\Phi_o\right)\right\rangle \tag{8}$$

$\Gamma\left(\Phi_o\right)$ signifies the choice and modification of the most prominent visual element in the target code. When $\Gamma\left(\Phi_o\right) = e_i$, indicating that only a single neuron's value is maintained while the rest

are zero, the gradient of the produced image primarily originates from the targeted neuron. The coding inversion loss function is defined as follows:

$$l\left(\Phi(X), \Phi_o\right) = \left|\Phi(X) - \Phi_o\right| \tag{9}$$

Φ_o is the feature map derived from the target image via network mapping, denoted as $\Phi\left(X_o\right)$. To emphasize the information of the target image within a specific region of the visual target code Φ_o, apply a mask M. The mask M retains the region of interest in the target code Φ_o while suppressing other areas, resulting in the following loss function:

$$l\left(\Phi(X), \Phi_o\right) = \frac{\left\|\left(\Phi(X) - \Phi_o\right)\Theta M\right\|}{\Phi_o \Theta M} \tag{10}$$

By integrating music data visualization technology, feature detection method based on DL and MFCC coefficient extraction, a more intelligent and efficient music appreciation support platform can be constructed. The platform uses deep learning to automatically extract the style, emotion, and structural characteristics of music works, and combines with the visual analysis of MFCC coefficient to provide listeners with personalized recommendation content that suits their interest preferences and aesthetic level.

The platform also has the function of real-time audio analysis and interactive feedback: when users listen to or interact with music, the system can instantly generate visual information such as frequency spectrum, melody line, or emotional trajectory to help them perceive rhythm, timbre change, and emotional evolution more intuitively.

On this platform, listeners can choose their own content according to their own interests and appreciation experience, and explore the structural level, style characteristics, and cultural background of different music works with the help of visual tools. The platform generates portraits of users' auditory behaviors and preferences through data analysis, which not only optimizes the follow-up recommendation strategy but also provides insights for curators, tour guides, or aesthetic education workers.

4 | Verification of Visualization Effect of Music Data

4.1 | Experimental Results

In order to verify the actual effect and feasibility of music data visualization technology based on DL, a comprehensive simulation environment is built in this study, and the goal is to assess the performance and potential of this technology by simulating the real music data visualization process. The model development was carried out in three stages: training, validation, and testing. The dataset was split using a 7:2:1 ratio. During training, the weights were optimized using the backpropagation algorithm, with cross-entropy loss serving as the loss function. The Adam optimizer was employed, with an initial learning rate set to 0.001. The model was trained for 100 epochs, and the validation set was used to monitor overfitting. In the testing stage, an independent test set was used to evaluate the model's generalization capability.

The experimental design covers music data preprocessing, feature detection, rhythm information recognition, and also involves the correlation of visual design parameters, the establishment of parametric models, and the final simulation and assessment of visual effects.

At the beginning of the experiment, three representative music signals were simulated, corresponding to pop music, classical music, and national music respectively, as shown in Figures 3–5. These waveforms show the time-domain characteristics of different music types, which provide a data basis for subsequent feature detection and recognition. By simulating the music data in the real world, the performance of technology in practical application can be assessed more accurately.

In the preprocessing stage, the advanced DL is used to extract the depth features of the music signal. These network structures can automatically learn complex patterns in music data, such as melody, rhythm, and harmony features. In particular, the network structure such as PointNet is introduced to deal with the

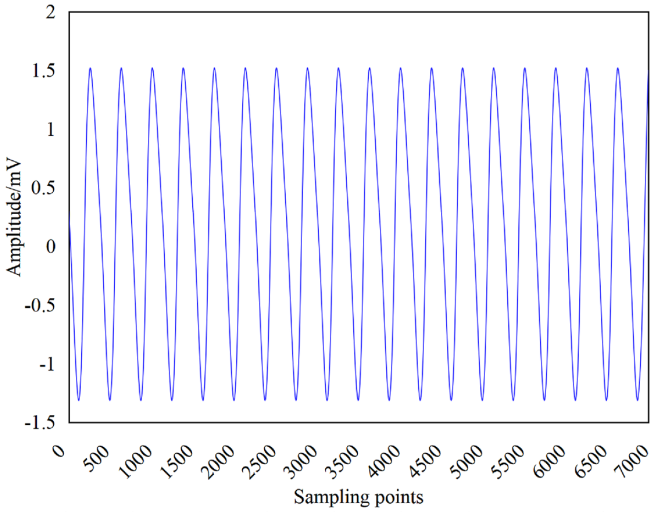

FIGURE 3 | Waveform of a pop music signal.

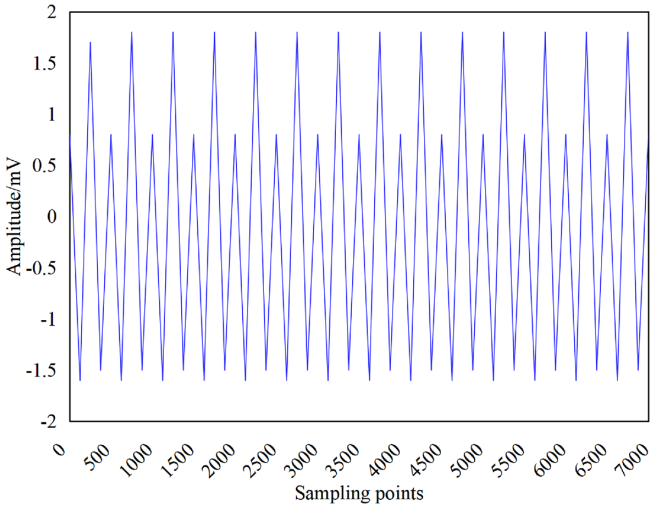

FIGURE 4 | Waveform of a classical music signal.

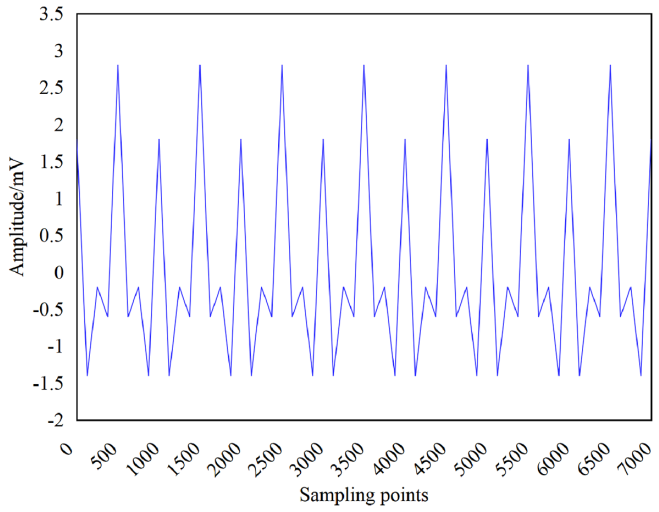

FIGURE 5 | Waveform of a folk music signal.

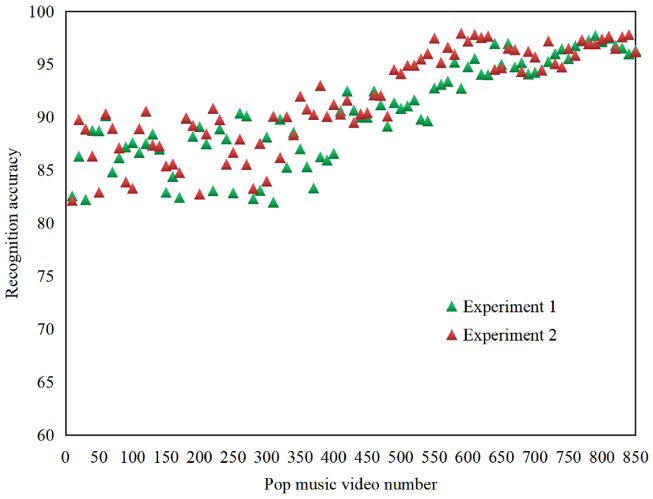

FIGURE 6 | Rhythm recognition accuracy in popular music videos.

point cloud features in music signals, which further improves the accuracy and robustness of feature detection.

The identified music rhythm information is then mapped to visual design parameters, such as color, shape, motion trajectory, etc., thus establishing a parametric model. This mapping process fully considers the natural relationship between music and vision. For example, a fast rhythm may correspond to bright colors and fast movements, while a slow rhythm may correspond to soft colors and slow changes. With the help of fine parameter design, the harmony between music and visual effect is ensured.

The core of the experiment is to assess the accuracy of audio feature recognition. As shown in Figures 6–8, the recognition accuracy in pop music, classical music, and national music data is shown respectively. The results show that this method has achieved excellent accuracy in pop music, reaching over 97%. In classical music, although the rhythm is more complex and changeable, this method still maintains an accuracy of about 94%, showing a strong generalization ability. In folk music, due to the diversity and regionality of rhythm features, the accuracy rate is slightly reduced to 92%, but it is still at a high level, which proves the stability and applicability of the method.

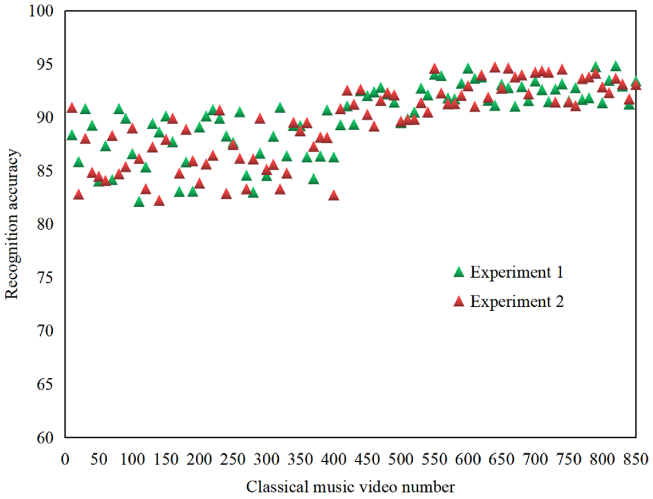

FIGURE 7 | Rhythm recognition accuracy in classical music videos.

Based on the identified music rhythm information, the parametric model is driven for dynamic design, and the visual effects of different types of music are generated, as shown in Figure 9. These visual works intuitively show the rhythm structure of music, and also convey the emotion and atmosphere of music with the help of color, shape, and dynamic changes. By comparing the visual effects of different types of music, it is not difficult to see the close relationship between music style and visual design, which further verifies the effectiveness of this method.

To assess the practical value of the visualization, a subjective evaluation was conducted with 50 participants, who were divided into two groups: professional musicians and general listeners. The evaluation tasks included rhythm recognition, emotional perception, and structural understanding. Results showed that the visualization group achieved significantly higher accuracy in rhythm

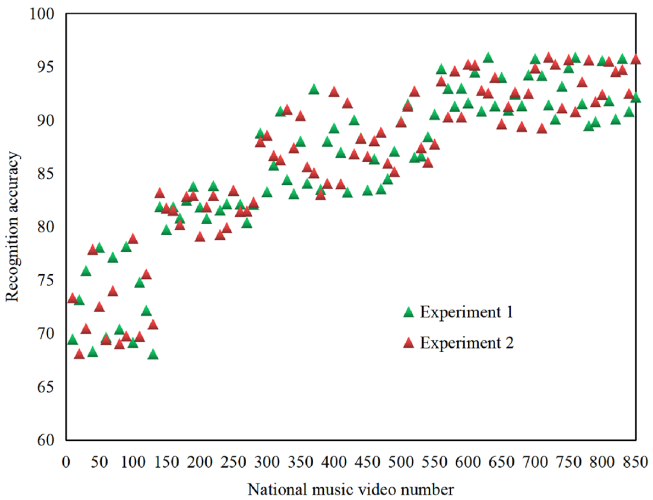

FIGURE 8 | Rhythm recognition accuracy in national music videos.

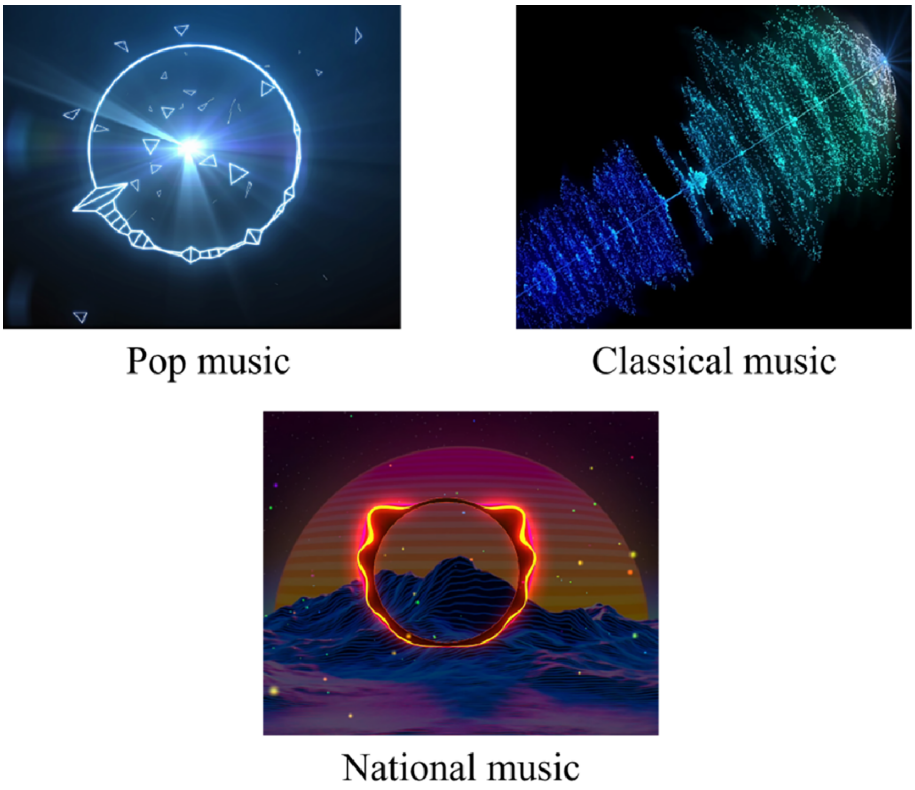

FIGURE 9 | Visualization of different types of music.

recognition compared to the audio-only group. General listeners reported that the visual aid notably lowered the barrier to comprehension and enhanced their sense of immersion. Professional musicians found the visualization tools helpful in analyzing complex musical structures.

4.2 | Discussion

It is found that the music data visualization technology based on DL is highly reliable in feature extraction and recognition. This technology realizes the information conversion from auditory perception to visual cognition and largely solves the problem that abstract concepts are difficult to concretize in traditional music appreciation teaching. Experiments show that the recognition accuracy of this technology in the field of pop music is over 97%, and it also maintains high robustness in classical and folk music. This result shows the generalization ability of the model to deal with complex structures with different musical styles. Subjective evaluation feedback reveals important pedagogical significance: dynamic visual feedback mechanism reduces students' cognitive load by using double coding effect, which enables learners to intuitively capture the logic of ups and downs and rhythms of music emotions and realize the cognitive mode change from passive listening to active and deep participation. This discovery verifies the original intention of technology-enabled aesthetic education, that is, to enhance the interactivity of music education by means of digital means.

This study also exposed some challenges worthy of discussion. In the aspect of folk music rhythm recognition, although the accuracy is high, it is lower than that of pop music. This is mainly because folk music has the characteristics of regional non-standardized rhythm and improvisation, which requires higher adaptability to deep neural networks based on structured data training. Although the current visualization scheme realizes basic emotional mapping, there is still room for improvement in artistic expression diversity and creative generation.

Follow-up research can be extended to multi-dimensional music feature fusion visualization, such as incorporating deep information such as harmony density and musical form structure into design parameters to achieve a more comprehensive and three-dimensional music expression. Combining VR, AR, and other emerging immersive technologies, it is expected to create a high-fidelity virtual music appreciation environment. We can also introduce an adaptive recommendation algorithm and learning analysis technology, develop a personalized music appreciation assistant system, and provide customized visual analysis according to different learners' cognitive level and aesthetic preference. This is helpful to improve the theoretical system of music data visualization and provide a more operational solution for aesthetic education practice under the intelligent background.

5 | Conclusions

This study focuses on the application of music data visualization technology in music appreciation teaching, analyzing its core principles, key technologies, and potential within educational contexts. Research indicates that this technology serves as a vital bridge connecting auditory art with visual perception, significantly enhancing students' understanding of musical structure, rhythm, and emotional expression while deepening their aesthetic sensitivity and artistic perception.

By constructing a simulation environment, this study replicates the visualization process of music data in real-world teaching scenarios. DL technology is employed to preprocess data, extract features, and identify rhythmic information. Experimental results demonstrate high recognition accuracy across pop, classical, and folk music genres, verifying the method's effectiveness and robust generalization capabilities. Based on these recognized rhythmic and structural elements, visual presentations tailored to different musical styles were successfully generated.

Consequently, this study offers a feasible pathway and practical foundation for integrating music data visualization into music appreciation instruction. Future research will further explore multi-dimensional visual expressions of musical features and incorporate emerging technologies such as VR and AR to create more immersive and personalized learning experiences for students.

Author Contributions

Xiaowei Chen: conceptualization, methodology, software, data curation, validation, formal analysis, writing – original draft, writing – review and editing, resources, investigation.

Funding

The author has nothing to report.

Ethics Statement

The user evaluation component of this study has received ethical approval. All participants provided informed consent prior to the experiment and were clearly informed that the data would be used solely for academic research purposes. The collection of music data complied with copyright regulations and posed no risk of privacy infringement.

Consent

The author has nothing to report.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

The dataset generated in this study is available from the corresponding author upon reasonable request. Reasonable requests refer to those intended for non-commercial academic research purposes and require the signing of a data usage agreement.

References

1. L. Li and Z. Han, "Design and Innovation of Audio IoT Technology Using Music Teaching Intelligent Mode," *Neural Computing and Applications* 35, no. 6 (2023): 4383–4396.

2. S. M. Van Bonn, J. S. Grajek, and S. S. S. P. M. R. Rettschlag, "Interactive Electronic Visualization Formats in Student Teaching," *HNO* 72, no. 5 (2024): 341–349.

3. C. J. Okere, G. Su, X. Gu, B. Han, and C. Tan, "An Integrated Numerical Visualization Teaching Approach for an Undergraduate Course, Flow in Porous Media: An Attempt Toward Sustainable Engineering Education," *Computer Applications in Engineering Education* 29, no. 6 (2021): 1836–1856.

4. P. Georges and A. Seckin, "Music Information Visualization and Classical Composers Discovery: An Application of Network Graphs, Multidimensional Scaling, and Support Vector Machines," *Scientometrics* 127, no. 5 (2022): 2277–2311.

5. H. B. Lima, C. G. R. D. Santos, and B. S. Meiguins, "A Survey of Music Visualization Techniques," *ACM Computing Surveys* 54, no. 7 (2021): 1–29.

6. A. Calilhanna, "Ogene Bunch Music Analyzed Through the Visualization and Sonification of Beat-Class Theory With Ski-Hill and Cyclic Graphs," *Journal of the Acoustical Society of America* 148, no. 4 (2020): 2697–2697.

7. E. Ghaleb, M. Popa, and S. Asteriadis, "Metric Learning-Based Multimodal Audio-Visual Emotion Recognition," *IEEE Multimedia* 27, no. 1 (2019): 37–48.

8. J. Nam, K. Choi, J. Lee, et al., "Deep Learning for Audio-Based Music Classification and Tagging: Teaching Computers to Distinguish Rock From Bach," *IEEE Signal Processing Magazine* 36, no. 1 (2018): 41–51.

9. S. X. Li, "Optimization of Music Teaching in Colleges and Universities Based on Multimedia Technology," *Solid State Technology* 63 (2020): 2891–2902.

10. Y. Huang, "Research on Interactive Creation of Computer-Aided Music Teaching," *Solid State Technology* 64 (2021): 577–587.

11. D. Li, "Practical Application and Case Analysis of Computer Image Vision Technology in Music Education and Teaching," *Journal of Cases on Information Technology* 26, no. 1 (2024): 1–16.

12. C. Tang and J. Zhang, "An Intelligent Deep Learning-Enabled Recommendation Algorithm for Teaching Music Students," *Soft Computing* 26, no. 20 (2022): 10591–10598.

13. J. Sun, "Research on Resource Allocation of Vocal Music Teaching System Based on Mobile Edge Computing," *Computer Communications* 160 (2020): 342–350.

14. K. Yuan, "Research on Music Teaching Systems Assisted by Artificial Intelligence," *International Journal of e-Collaboration (IJeC)* 20, no. 1 (2024): 1–17.

15. Y. Zhang, "Enlightenment on Vocal Music Classroom Teaching From the Perspective of Neuroscience," *Neuro Quantology* 16, no. 6 (2018): 132–137.

16. L. Fan, "Audio Example Recognition and Retrieval Based on Geometric Incremental Learning Support Vector Machine System," *IEEE Access* 8 (2020): 78630–78638.

17. J. Xu, Z. Liu, J. Jiang, and Y. Dou, "High Performance Robust Audio Event Recognition System Based on FPGA Platform," *Cognitive Systems Research* 50 (2018): 196–205.

18. L. Liu, W. Li, X. Wu, and B. X. Zhou, "Infant Cry Language Analysis and Recognition: An Experimental Approach," *IEEE/CAA Journal of Automatica Sinica* 6, no. 03 (2019): 173–183.

19. W. Nie, M. Ren, J. Nie, et al., "C-GCN: Correlation Based Graph Convolutional Network for Audio-Video Emotion Recognition," *IEEE Transactions on Multimedia* 23 (2020): 3793–3804.

20. R. Panda, R. Malheiro, and R. P. Paiva, "Audio Features for Music Emotion Recognition: A Survey," *IEEE Transactions on Affective Computing* 14, no. 1 (2020): 68–88.

21. L. Xia, G. Chen, X. Xu, J. Cui, and Y. Gao, "Audiovisual Speech Recognition: A Review and Forecast," *International Journal of Advanced Robotic Systems* 17, no. 6 (2020): 302–306.

22. M. S. Hossain and G. Muhammad, "An Audio-Visual Emotion Recognition System Using Deep Learning Fusion for a Cognitive Wireless Framework," *IEEE Wireless Communications* 26, no. 3 (2019): 62–68.

23. S. Ayadi and Z. Lachiri, "Deep Neural Network Architectures for Audio Emotion Recognition Performed on Song and Speech Modalities," *International Journal of Speech Technology* 26, no. 4 (2023): 1165–1181.

24. S. H. Sayed, H. E. EIDeeb, and S. A. Taie, "Bimodal Variational Autoencoder for Audiovisual Speech Recognition," *Machine Learning* 112, no. 4 (2021): 1201–1226.

25. R. Craciunescu, "Non-Audio-Video Gesture Recognition Systems," *Wireless Personal Communications* 110, no. 2 (2020): 815–827.

26. X. Wang, J. Mi, B. Li, Y. Zhao, and J. Meng, "CATNet: Cross-Modal Fusion for Audio–Visual Speech Recognition," *Pattern Recognition Letters* 178 (2024): 216–222.

27. M. Hao, W. H. Cao, Z. T. Liu, M. Wu, and P. Xiao, "Visual-Audio Emotion Recognition Based on Multi-Task and Ensemble Learning With Multiple Features," *Neurocomputing* 391 (2020): 42–51.

28. A. Kashevnik, I. Lashkov, A. Axyonov, et al., "Multimodal corpus design for audio-visual speech recognition in vehicle cabin," *IEEE Access* 9 (2021): 34986–35003.