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

Virtual Reality-Based Interactive Experience of Music Creation Using a Diffusion Model

Xiyang Chen[1] and Yifei Huang[2]

[1]School of Arts, Guangxi Minzu Normal University, Chongzuo, Guangxi 532200, China, 2020T1319@pwu.edu.ph
[2]School of Science, Guilin University of Aerospace Technology, Guilin 541004, China, yifeihyf@163.com

Corresponding author: Yifei Huang, yifeihyf@163.com

Abstract. As information technology rapidly evolves, computer-aided design (CAD) and virtual reality (VR) technologies are progressively infiltrating the realm of music creation, significantly diversifying the techniques and approaches used and thereby boosting its societal relevance and value. In this article, the possibility of combining the Diffusion model, CAD technology, and VR technology to realize a new interactive experience in music creation is deeply discussed. The diffusion model, with its powerful generating ability, can generate audio clips with specific styles and characteristics, which provides continuous inspiration for music creation. CAD technology enables music creators to intuitively edit scores and design chords and rhythms, greatly improving music creation's efficiency and accuracy. VR technology enables users to interact with musical instruments and creative tools in virtual space in real-time so as to experience the fun of music creation more deeply. The experimental results show that this method can not only effectively inspire the creators and improve the efficiency and quality of creation but also bring users more rich and diverse music creation experiences.

Keywords: Diffusion Model; Music Creation; CAD; VR; Interactive Experience
DOI: https://doi.org/10.14733/cadaps.2024.S28.252-265

1 INTRODUCTION

Driven by emerging technologies such as 5G, artificial intelligence (AI), and big data, VR technology—a prominent representative of modern information technology—has demonstrated vast potential for application across multiple professional industries, including medicine, gaming, aviation, and education. The existing optical music recognition (OMR) technology typically adopts an end-to-end approach, which involves processing the entire staff image at once and identifying the music symbols within it. The shape characteristics determine the pitch and length of a note, while the height directly reflects the specific position of the note on the staff. In existing methods, the shape of notes and their vertical position in the staff limit the complete parsing and deep application of music information. Traditional methods simplify the processing flow and do not require precise alignment of each note and its corresponding label, which helps to create and retrieve rich music corpora [1]. Through this approach, the model can not only accurately identify each note in the score but also gain a deeper understanding of their specific positions and interrelationships within the music structure. The application in these fields not only greatly promoted the innovation and development of the industry but also brought unprecedented immersive experiences to people. Technological progress is developing at an unprecedented pace and has had a profound impact on various industries, including the music industry. Andrea and Zahra [2] proposed a new method based on deep learning and convolutional neural networks (CNN), which focuses on identifying the position of notes in the staff. With the rapid development of Internet technology, people feel the arrival of a new era - the new media era. The forms of computer music creation and traditional music creation are interdependent. The efficient and convenient use of computer music software by people has also formed a certain degree of impact on the traditional form of music creation. In the process of rapid development of new media information technology, computers, mobile phones, etc., have become indispensable tools in people's lives, affecting people's lives. The emergence of computer music production software has given music creators more choices in their music creation methods. Baro et al. [3] analyzed the subcultural phenomena contained in virtual music characters in the context of new media communication from a cultural perspective, focusing on virtual music characters. The combination of music and computer technology has revolutionized the creation and presentation of music works. This gives music more possibilities in terms of creation, presentation, and other aspects. Analyze the differences and characteristics between traditional media music production and new media music production through changes in communication structure. It uses the "sunshine model" to analyze the transmission channels of virtual music roles in the Internet era. Secondly, the author starts from the perspective of communication studies and studies the changes in the ways virtual music characters are created, disseminated, presented, and accepted in the context of new media. Finally, the author analyzed the economic benefits brought by virtual music characters and proposed some possible suggestions for the future development of virtual music characters.

As an important carrier of human emotion, music has gradually emerged with new functions and forms with the development of the mobile Internet. These highly active users have gathered together to form a unique NetEase Cloud Music community. We have gathered a large number of loyal users through innovative designs, high-quality music reviews, and a distinct sharing mechanism. It is suggested that community operators should encourage users to have real names, expand the song copyright database, and community organizers should strengthen organizational management. On the one hand, exploring the impact of interactions among members on virtual communities at a micro level reveals that interactions within a community can enable individuals to gain emotional energy. On the other hand, from a macro perspective, it has been found that online socializing can have an impact on real-life interactions. NetEase Cloud Music, as a rising star in the music app market, clearly proposed the concept of "music socialization" at the beginning of its launch. Based on relevant survey data, Cai et al. [4] depicted portraits of members of the NetEase Cloud Music Community. It discovered that virtual identity and diverse roles are two major characteristics of members of the NetEase Cloud Music community. It creates a personalized spiritual space for individuals, expands diverse ways of social interaction, and builds a communication bridge for the offline extension of online relationships. A detailed explanation of the four elements and interactive process of the interactive ceremony has restored the interactive landscape of the NetEase Cloud Music community. Once again, study the interactive results. Exploring specific solutions to problems from a detailed perspective, in order to provide reference for the orderly operation of virtual communities in music apps. In addition, this model can also be used for music evaluation, helping listeners or music professionals to more accurately understand and feel the emotional connotations of music works. Experimental results show that compared with single mode, the overall recognition accuracy using text information is improved by 6.70%, while the accuracy of emotion recognition using voice information is increased by 13.85%. In the field of music creation, Calilhana [5] draws inspiration from the multi-rhythm and improvisation elements of Ogong music, injecting new vitality into modern music works. This not only helps composers better understand the internal structure of music but also provides them with new creative ideas. This method is not only applicable to Ibo music but also to various music genres around the world, as music itself is a borderless art. By applying psychoacoustic methods such as ski hill and loop diagrams, we can conduct more in-depth analysis and visualization of the beat and rhythm of music. Meanwhile, the elements of improvisation can also provide inspiration for modern music creation, allowing music to flow more naturally and smoothly. Especially, visualizing the theory of rhythm and showcasing the evolution of Western music theory through ski hill and loop diagrams. This enables music creation to transcend the boundaries of culture and theory and absorb the essence of different music genres. In addition, this article uses psychoacoustic methods to reveal how Ibo music can be accurately preserved for future generations and included in academic discussions. This research method can also be applied to modern music creation, helping composers to have a deeper understanding of emotional expression in music and the perceptual experience of the audience. In the field of art, especially in music creation, the integration of VR technology also indicates the arrival of a brand-new creative era. Music, as an important part of human culture, has always been an important carrier of emotional expression, aesthetic pursuit and creative exploration. In today's society, with the increasingly obvious trend of cultural diversity and individualization, music creation also presents a diversified development trend. No matter the creative idea, creative method or creative form, it is necessary to innovate with the times to meet the music aesthetic needs of the masses in the new era.

This kind of innovation not only requires music creators to have profound artistic accomplishment and creative ability but also requires them to master advanced technical means and combine science and technology with art to create more contemporary and innovative music works. However, traditional music creation methods are often limited by space and technology, which makes it difficult for creators to give full play to their creativity and inspiration. Through VR technology, the creator can be in a realistic virtual music creation environment, play and debug with virtual instruments in real time, and feel the realism and immersion of music creation.

The innovations are as follows:

(1) In this article, the combination of Diffusion model, CAD technology, and VR technology is proposed and applied to the field of music creation. This cross-disciplinary technology fusion breaks the limitations of traditional music creation and brings a new perspective and possibility for music creation.

(2) By combining VR technology, this article provides an immersive creative experience for music creators. Creators can interact with musical instruments and creative tools in real time in the virtual environment. This brand-new interactive way can inspire the creators, improve creative efficiency, and make the music creation process more intuitive and interesting.

(3) This article uses the Diffusion model to generate audio clips with specific styles and characteristics, which provides a steady stream of inspiration for music creation. The powerful generating ability of the diffusion model makes music creation more flexible and diverse and allows it to meet the various needs of creators.

At the beginning of this article, the background and significance of the research are deeply analyzed. On this solid basis, we further tap the broad application potential of the Diffusion model, CAD technology, and VR technology in the field of music creation. Subsequently, we elaborate on how to build an interactive experience of music creation based on these technologies and the related algorithm flow. In order to verify the actual efficiency and feasibility of this comprehensive method, we carefully planned and implemented a series of experiments. In conclusion, this article systematically summarizes the main research achievements and innovations of the interactive experience of the diffusion model, CAD technology, and VR technology in music creation. Furthermore, we look forward to future research directions and to putting forward targeted suggestions in order to promote innovation in the field of music creation.

2 RELATED WORK

Dai and Wu [6] explored the effectiveness of feedback from peers and/or ASRs in mobile-assisted music creation and pronunciation learning. The research methods not only include a pre-test, post-test, and delayed post-test of music composition and pronunciation but also include student perception questionnaires and in-depth interviews. In the perception questionnaire, the three groups of students did not show any obvious preference or bias towards any method, indicating that each method has its unique value and advantages. The second group only relies on peer feedback (Co no ASR group); The third group consists of contract partners and feedback from ASR (Co ASR group). However, students in the Co no ASR and Co ASR groups performed better in music expression and pronunciation clarity. Students generally believe that real-time feedback from peers can provide more direct and targeted advice, while immediate feedback from ASR can help them quickly correct pronunciation errors. This may be related to real-time communication and interaction between peers, as well as the immediate feedback provided by ASR. They discovered some common and unique technical, social/psychological, and educational inspirations in music creation and pronunciation learning. In exploring methods for visualizing acoustic music graphics, Danchenko et al. [7] delved into the unique challenges in the field of electronic music. The particularity of this musical form requires us to rethink the role of notation in creation and presentation. Not only to understand and analyze the structure of electronic music but also to inspire and drive innovation in music creation. Especially its hidden sound sources and lack of traditional written notation characteristics. However, this does not mean that notation becomes irrelevant in music creation and presentation. On the contrary, it provides us with a brand new creative space and expression methods. In terms of research methods, they adopt a combination of structure and function to analyze the characteristics of sound objects deeply. And understand their impact on the overall work by comparing different electronic music synthesis methods. They analyzed the graphic fixing functions in audio generation software, which allow music creators to operate and edit sound in an intuitive way. This approach shares similarities with the creative process of folk songs, both based on intuitive auditory understanding and expression. Music emotion recognition not only plays an important role in the field of music information retrieval but also provides a unique perspective and tool for music creation. Dong et al. [8] proposed a new model called Bidirectional Convolutional Recursive Sparse Network (BCRSN) based on convolutional neural networks and recurrent neural networks. For music creation, this technology that can continuously predict the emotions of audio files has enormous potential. Emotion recognition technology can assist composers in gaining a deeper understanding of the emotional connotations of music, thereby creating more infectious works. By adjusting musical elements to match specific emotional states, composers can achieve more precise emotional control during the creative process. This model adaptively learns the sequence information of significantly influential features (SII-ASF) from the two-dimensional time-frequency representation (i.e., spectrogram) of music audio signals, thereby capturing the emotional dynamics that change over time in music. By adjusting the components on the WIM model, creators can change the properties of music in real-time, such as pitch, volume, timbre, etc. Through this technology, creators can freely change the form and structure of music, creating richer and more diverse musical effects. This intuitive operating method can greatly reduce the threshold for music creation, allowing more people to participate in music creation. By introducing these design indicators, creators can analyze and evaluate their music works more systematically, thus making wiser creative decisions. In terms of design indicators, although there are differences between music creation and product design, some concepts can also be borrowed. Analogy weight to the complexity or depth of music, and centre of mass to the core theme or emotional expression of music. Free-form mesh deformation technology is not only applicable to product design but can also be introduced into music creation [9].

He [10] proposed an interactive visual teaching model for music creation based on an intelligent classroom environment. It establishes a music cognitive map, guides the integration of music concepts and creation, and evaluates learners' transfer and application abilities in music creation. In this design, they fully utilized intelligent cloud service technology to collect, transmit, and analyze teaching data of music theory concept maps, thereby developing a personalized set of music creation learning resources. In order to better adapt to the intelligent classroom environment, they adopted various encoding types of music creation interaction analysis systems and optimized the interaction between people and technology. Especially with the addition of specific code specifically designed to analyze various interactions during the music creation process. In the process of music creation, students can gain a deeper understanding of music theory knowledge through visual concept maps. Combining teaching examples of music courses can also stimulate students' musical creative potential and promote the development of their high-level musical thinking abilities. In music education, the integration of information technology should be regarded as the core principle to promote the development of students' music knowledge and experience. Composing, as one of the pinnacle manifestations of musical creativity, with the support of computer technology and software applications, can help students better understand this highly demanding process and give them the possibility to achieve music creation. In addition, the application of information technology in music education has made some abstract and unimaginable music concepts intuitive and easy to understand. These projects not only provide rich music creation tools but also encourage students to explore the boundaries of music creation through collaboration and innovation. Whether it is harmony, counterpoint, or other aspects of music theory, through computer-aided learning, students can quickly grasp this basic knowledge and apply it to practical music creation [11].

With the continuous progress of technology, the application of computer technology in the field of teaching is becoming increasingly widespread, especially in music education, its potential is immeasurable. Pei and Wang [12] designed and developed an advanced digital music course teaching system based on computer-aided teaching technology. This provides music creators with an intuitive and precise audio editing environment. This system transforms complex music theories and skill methods into vivid image data information, making the music creation and learning process more intuitive and easy to understand. By utilizing advanced audio processing software and synthesizers, creators can create richer and more unique music effects, expanding the expressive boundaries of music. For example, during the composition process, creators can use this system to intuitively explore the development of music themes, the construction of structures, and post-production, greatly improving the efficiency and quality of creation. The practice has proven that the digital music curriculum teaching system not only provides visual information and parameter support for music teaching and training. By utilizing the ease of collecting teaching data in the context of the Internet of Things, Yang [13] analyzed and understood the literacy and skills required for music creation. It proposes the concept of personalized recommendation based on deep learning algorithms and applies it to the field of music creation. Through the analysis of educational data mining-related theories, we can discover that information on music genres, styles, and the biographical backgrounds of musicians in history courses has important enlightening effects on music creation. In the experimental results, they found that deep learning algorithms performed excellently in terms of recommendation accuracy. These recommendations can help students understand the musical characteristics of different historical periods, stimulate their creative inspiration, and integrate these elements into their actual creations. Especially when facing music datasets, algorithms can adaptively adjust the difficulty of recommended music materials to meet the creative needs of different students.

With the development of the mobile Internet, listening to music is no longer a lonely personal behaviour. Users share their favourite songs in the WeChat circle of friends. Alternatively, one can tell their own story in specific song comment sections and socialize with music as the theme gradually becomes normalized. Based on the Collins Interactive Ritual Chain theory, this study investigates the virtual community of NetEase Cloud Music. Due to the birth of this theory in the traditional social era, Collins emphasized the importance of being present in person, but the development of technology has made being present in person no longer a necessary condition for interaction. Taking Netease Cloud Music Virtual Community as an example, this paper explains the four elements of interaction ceremony, interaction process, and interaction results in detail in order to enrich the connotation of this theory in the mobile Internet era. Case study method: By comparing the virtual community operations of current music apps, the author chose NetEase Cloud Music, which is representative of this field, as the research object. NetEase Cloud Music has accumulated a large number of users in the short term. Compared to other similar music applications, NetEase Cloud Music has a clearer social attribute. The high-quality output content within the community has formed a unique culture of the NetEase Cloud Music community, and therefore, research on it is of representative significance [14]. It not only greatly enriches teaching methods, but also shows great potential in the field of music creation. Zhang and Yi [15] have deeply integrated computer automatic matching technology into music teaching and proposed a new music teaching model, aiming to stimulate students' potential for music creation. The research results indicate that the application of computer technology in music teaching not only effectively assists music teaching but also improves students' understanding and mastery of basic music knowledge. Especially with the introduction of computer automatic matching technology not only supports the automatic generation of random teaching content but also provides powerful tools for cultivating students' musical thinking and creative abilities. Students can use computer-generated music materials to create, combine their learned knowledge with practical applications, and form their own music works. Students can use a computer automatic matching system to randomly obtain music materials, harmony patterns, melody structures, and other elements and then apply the music theory and skills they have learned to combine and create.

To sum up, the Diffusion model, CAD technology and VR technology have broad application prospects in the field of music creation. This study combines the three to explore a new way of music creation and experience. Through simulation experiments, this article will verify the feasibility of this combination and provide more scientific and innovative ideas and methods for future music creation.

3 THE APPLICATION OF DIFFUSION MODEL, CAD TECHNOLOGY, AND VR IN MUSIC CREATION

The diffusion model has unique advantages in music creation; its stability and reliability support large-scale complex data processing, especially when processing music flow field data information. Through the process of diffusion and denoising, the model can learn the inherent laws of music data and generate music fragments with specific styles. In terms of transfer learning, the Diffusion model shows strong applicability and generalization ability, which can easily adapt to data with different downsampling ratios, realize cross-task knowledge transfer, and provide rich creative tools and inspiration for music creators.

VR technology brings innovation and convenience to music creation, especially immersive experience design, which greatly enriches creative forms. CAD technology plays a key role in music creation with its accurate and intuitive characteristics, which improve creative efficiency and enrich the performance of works. VR technology allows the creator to be in a realistic virtual music environment, interact with virtual instruments in real time, feel the charm of music more intuitively, and stimulate creative inspiration (Figure 1). The combination of the two brings unprecedented possibilities for music creation.

Figure 1: Various manifestations of VR in music creation.

Creators can use VR technology to build a virtual concert hall or stage to simulate different sound effects and performance scenes. By wearing VR equipment, the audience can feel the live atmosphere of music and have a closer emotional connection with the creators. In addition, VR technology can also be used in the field of music education and training to help students better understand and master music knowledge by simulating the real performance environment and music creation process.

4 MUSIC CREATION INTEGRATING DIFFUSION MODEL, CAD TECHNOLOGY, AND VR

4.1 Diffusion Audio Generation Model

Traditional music creation often relies on the personal experience and inspiration of creators, while the Diffusion model provides a data-driven approach. These generated audio clips not only have a similar style to the original music but can also exhibit different characteristics and variations to a certain extent. This data-driven approach can help creators break traditional creative frameworks and explore new music styles. By studying and analyzing a large amount of music data, models can discover patterns and patterns in music, providing creators with new sources of inspiration. Due to the fact that the Diffusion model learns statistical features of music data during training, it is able to generate diverse audio clips. This diverse generation of results can stimulate the imagination and creativity of creators, bringing more possibilities to music creation. Figure 2 shows the Diffusion model structure of this article.

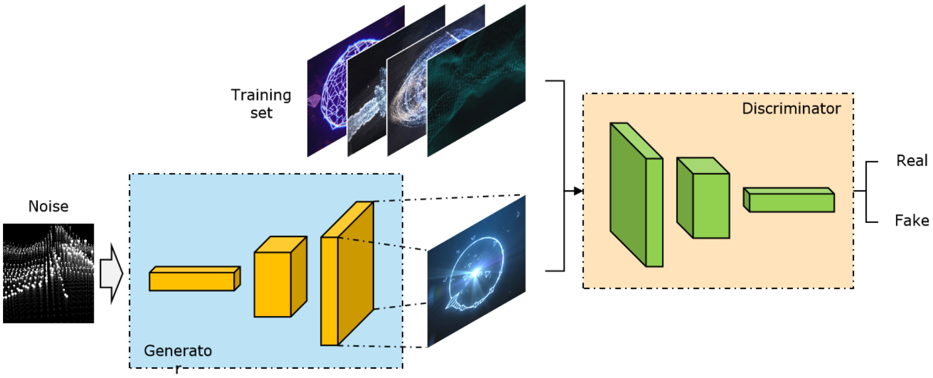

Figure 2: Diffusion model.

After that, more audio features are extracted based on the Mel spectrum, which enriches model input and improves performance. Finally, the features are standardized to eliminate the differences in dimension and distribution, which helps the model to learn the internal laws and deep features of data more easily.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}t} + \in} \hat{m}t \tag{1}$$

Where θ_t is the parameter vector at the time step t, η is the learning rate, $\hat{v}t$ and $\hat{m}t$ are the first-order moment estimation (average) and the second-order moment estimation (decentralized variance) of the gradient, respectively, and $\in$ is a small positive number to prevent the error of dividing by zero. The updated formulas of first-order moment estimation and second-order moment estimation are as follows:

$$mt = \beta_1 m_{t-1} + 1 - \beta_1 \ g_t \tag{2}$$

$$vt = \beta_2 v_{t-1} + \ 1 - \beta_2 \ \ g_t^2 \tag{3}$$

$$\hat{m}t = \frac{mt}{1-\beta_1^t} \tag{4}$$

$$\hat{v}t = \frac{vt}{1-\beta_2^t} \tag{5}$$

Where g_t is the gradient at time step t, and β_1 and β_2 are superparameters, usually with values of 0.9 and 0.999. $\hat{m}t$ and $\hat{v}t$ are deviation corrections mt vt to ensure that there will be no deviation due to initialization values at the initial stage of training.

Training is not one-off but requires constant iteration, during which independent verification data sets are regularly used to evaluate the model performance. The purpose of this is to ensure that the model not only performs well on the training data but also has good generalization ability on the unknown data, so as to avoid the occurrence of over-fitting. The whole training process will continue until the model reaches the preset number of training rounds or meets other preset stopping conditions.

The unique Diffusion process of diffusion model plays a core role in the application of music generation. In this process, random noise is skillfully added to the original music data, and then the required music samples are gradually constructed from the noise through a specific algorithm. This diffusion method not only increases the diversity of data but also makes the generated music more creative and changeable. The diffusion model is learned through a fixed process, which gradually and hierarchically extracts the internal structure and melody of music from noise. Furthermore, these potential music variables are treated in the same way as the original music data in the learning process, which ensures that the generated music is consistent with the original data in style, melody, and structure, and at the same time injects new elements and possibilities.

In the forward process, Gaussian noise is gradually added to the image to get a time-frequency image with noise. Where the distribution at time X_t is equal to the distribution at time X_{t-1} plus the noise of Gaussian distribution:

$$X_t = \sqrt{\alpha_t} X_{t-1} + \sqrt{1-\alpha_t} Z_1 \tag{6}$$

Where X_t represents the image at t a time, α_t is the attenuation value of noise, and Z represents Gaussian noise. After repeated iterations, it can be distributed at any t time of the initial state X_0:

$$X_t = \sqrt{\bar{\alpha}_t} X_0 + \sqrt{1-\bar{\alpha}_t} Z_t \tag{7}$$

Where $\bar{\alpha}_t$ represents the factorial attenuation value of noise at t the moment.

In the reverse process, the original time-frequency map is gradually restored from the noise. According to the Bayesian formula:

$$q\ X_{t-1} \middle| X_t\ = q\ X_t \middle| X_{t-1}\ \frac{q\ X_{t-1}}{q\ X_t} \tag{8}$$

$q\ X_{t-1} | X_t$ is the probability distribution of X_{t-1} when X_t is given. Because $q\ X_{t-1}$ distribution can't be obtained directly, the conditional probability distribution $q\ X_{t-1} | X_t, X_0$ is used for the approximate solution:

$$q\ \ X_{t-1}\left|X_t, X_0\ \ = q\ \ X_t\right|X_{t-1}, X_0\ \ \frac{q\ \ X_{t-1}\left|X_0\right.}{q\ \ X_t\left|X_0\right.} \tag{9}$$

When X_0 is known, the distribution of X_{t-1} time can be found:

$$X_{t-1} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(X_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\xi_\theta\ \ X_t, t\right) + \sigma_t Z \tag{10}$$

After repeated iterations, the distribution of X_0 time can be predicted finally. σ_t represents the predicted noise variance, and the noise predicted by t the time model is expressed as:

$$\beta_t = -1\alpha_t, \xi_\theta\ \ X_t, t \tag{11}$$

Audio generation is a creative process, that begins with a series of detailed condition setting and sampling strategy selection. Before the audio is generated, this article sets specific conditions for the Diffusion model, such as the music style we want, the unique rhythm, or a specific melody. These conditions, like the artist's creative inspiration, provide clear guidance for the model so that the generated audio clips can meet our specific needs.

4.2 VR Immersive Music Creation Experience

In the VR music creation environment, creators can break free from the constraints of traditional creative tools and interact with music in a more natural and intuitive way. This interactive approach not only reduces learning costs but also allows creators to experience the changes and charm of music more intuitively. The VR music creation environment has the ability to provide real-time feedback on audio effects. Creators can immediately hear the sound effects of their works during performance or creation, including timbre, volume, rhythm, etc. It can directly operate virtual instruments, such as pianos, guitars, drums, etc., for performance, debugging, and creation. This real-time audio feedback mechanism greatly improves the efficiency and accuracy of creation. In addition, VR environments also support rich audio processing functions such as reverberation, equalization, compression, etc., helping creators create more layered and dynamic music works. In addition, VR technology also supports online collaboration among multiple people, allowing creators to share their creative space in real-time with others and co-create music works. In virtual space, creators can try various unprecedented music forms and styles, such as music performances that combine virtual reality and augmented reality technology, interactive music games, etc. The introduction of a VR music creation environment not only provides creators with a more intuitive and interesting creative experience but also expands the possibilities of music creation. These innovative music forms not only attract more audience and player attention but also bring new growth points to the music industry.

After the creation, the creator can choose to export the final music score and audio file. These files can be saved and shared in various formats for use on different platforms and devices. In this way, creators can easily share their works with others and get more feedback and recognition.

5 RESULT ANALYSIS AND DISCUSSION

5.1 Experimental Design

By combining VR technology with music creation, this article provides a brand-new, efficient, and creative creation platform for creators. On this platform, creators can feel the change and charm of music more intuitively and create it in a more natural and interesting way. Furthermore, through the close integration with the CAD system and Diffusion model, we ensure the smoothness and consistency of the whole creative process. Experiments are carried out in this section to verify the superiority of this method.

Experimental goal: to verify the superiority of the Diffusion audio generation model in music creation; Test the role of CAD score editing and optimization tools in improving the efficiency and accuracy of music creation; Evaluate the influence of VR immersive music creation experience on the creative process and quality of works.

Experimental scheme:

Data preparation: diverse music data sets are collected for training the Diffusion model and testing the CAD and VR systems.

System construction: integrate the Diffusion model, CAD technology and VR technology to build a complete music creation platform.

User recruitment: Recruit creators with different music creation backgrounds and skill levels to participate in the experiment.

Experimental task: Participants are required to use our music creation platform to complete the creation of an original music work.

Data collection and analysis: record participants' behaviour data, work quality and their feedback during the creative process.

5.2 Experimental Process

Diffusion model training and testing: use the collected music data set to train the Diffusion model, and conduct a series of tests after the training to verify its ability to generate audio clips and style diversity. The test results are shown in Figure 3.

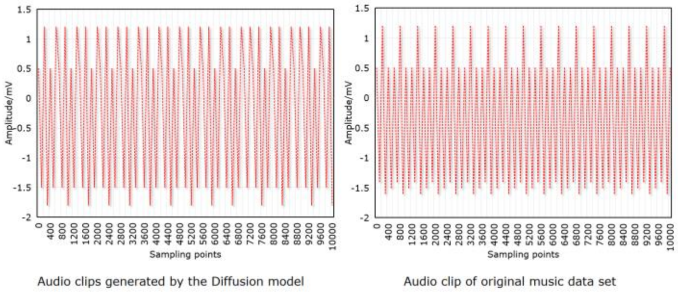

Figure 3: Generating audio clips.

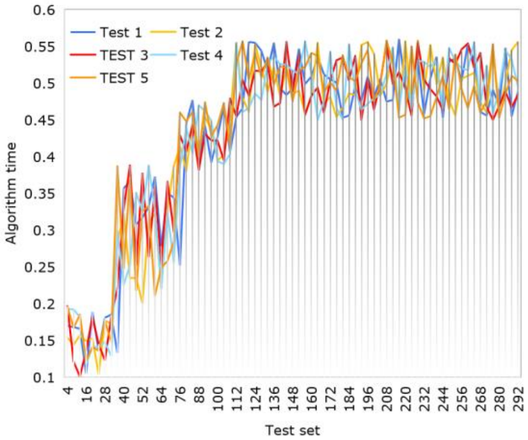

Figure 4: Creative efficiency.

CAD score editing and optimization test: participants are required to use CAD tools to edit and optimize the initial score, and record the number of operations, time and the quality of the final score in the editing process. The scoring results of creative efficiency and music quality are shown in Figure 4 and Figure 5.

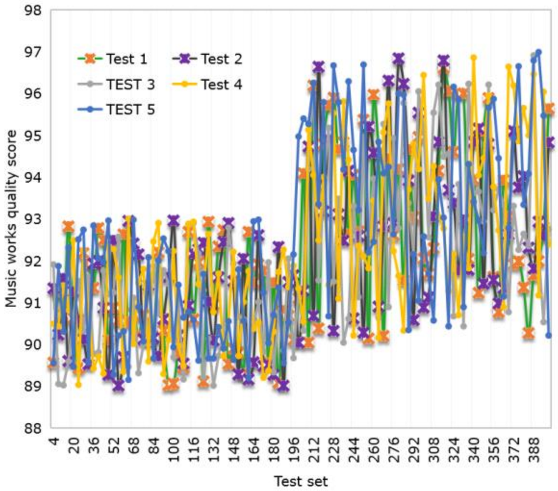

Figure 5: Quality rating of music works.

VR immersive music creation experience: let participants enter the VR environment to create music, observe their interaction with virtual instruments, and the quality of the created music works. The survey results of VR immersive music creation experience satisfaction are shown in Table 1 and Figure 6.

Participant number	*Interaction time with virtual instruments (minutes)*	*Number of music works created*	*Music score (out of 10)*	*Experience satisfaction (out of 10)*
001	35	3	9.0	9.5
002	20	2	8.0	8.5
003	45	1	7.5	8.0
004	30	2	8.5	9.0
005	25	3	9.2	9.5
006	40	1	7.8	8.0
007	30	2	8.7	9.2
008	20	2	8.2	8.7
009	45	3	9.5	10.0
010	35	1	7.6	8.0
011	25	2	8.3	8.8
012	40	2	9.1	9.6

Table 1: Interaction of participants.

Data synchronization and output test: verify the data synchronization performance among CAD, VR and Diffusion, and the output quality of the final music score and audio file. The test results of data synchronization and output performance are shown in Figure 7.

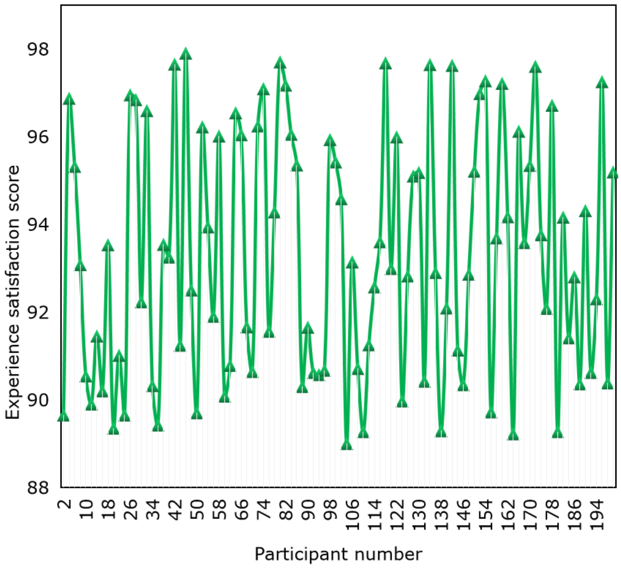

Figure 6: VR Immersive music creation experience satisfaction survey.

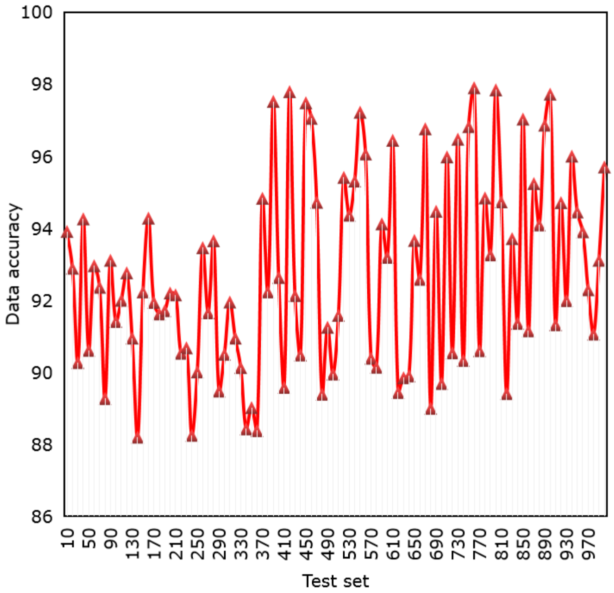

Figure 7: Data synchronization and output performance testing.

6 CONCLUSIONS

This study combines the Diffusion model, CAD technology, and VR technology to construct an innovative music creation platform. The Diffusion model provides unlimited possibilities for music creation with its powerful generative ability. Through the application of this comprehensive technology, the aim is to explore new horizons in music creation and provide creators with unprecedented creative experiences. The following is an in-depth analysis of the expansion of this music creation platform. By using CAD tools to edit and optimize scores, creators can quickly adjust parameters such as notes, rhythm, and timbre, significantly improving creative efficiency. CAD technology optimizes the process of music creation with its precision and editability. The combination of the Diffusion model, CAD technology and VR technology not only reflects the progressiveness of technology but also shows innovative thinking in the field of music creation. VR technology has brought a brand new immersive experience to music creation. Creators can interact with various virtual instruments in real time in a virtual environment and experience the charm of music creation. This immersive experience not only inspires the creator's creative inspiration but also makes the music creation process more interesting and challenging. VR technology, on the other hand, allows creators to immerse themselves in the charm of music through its immersive experience. The combination of the three makes the music creation process more efficient, intuitive, and interesting.

7 ACKNOWLEDGEMENT

The 2022 Humanities and Social Sciences Project of the "Training Plan for Thousand Young and Middle-aged Backbone Teachers in Guangxi Higher Education Institutions", Exploration and Research on the Inheritance of Music Culture in Guangxi Zhuang GE XU under the Background of Rural Revitalization Strategy, Project Number: 2022QGRW056.

Xiyang Chen, https://orcid.org/0009-0005-5388-4115
Yifei Huang, https://orcid.org/0009-0009-0778-2176

REFERENCES

[1] Alfaro, C.-M.; Valero, M.-J.-J.: Exploiting the two-dimensional nature of agnostic music notation for neural optical music recognition, Applied Sciences, 11(8), 2021, 3621. https://doi.org/10.3390/app11083621

[2] Andrea, P.; Zahra, A.: Music note position recognition in optical music recognition using convolutional neural network, International Journal of Arts and Technology, 13(1), 2021, 45-60. https://doi.org/10.1504/IJART.2021.115764

[3] Baro, A.; Riba, P.; Calvo, Z.-J.; Fornés, A.: From optical music recognition to handwritten music recognition: a baseline, Pattern Recognition Letters, 123(5), 2019, 1-8. https://doi.org/10.1016/j.patrec.2019.02.029

[4] 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

[5] Calilhanna, A.: Ogene Bunch music analyzed through the visualization and sonification of beat-class theory with ski-hill and cyclic graphs, The Journal of the Acoustical Society of America, 148(4), 2020, 2697-2697. https://doi.org/10.1121/1.5147469

[6] Dai, Y.; Wu, Z.: Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study, Computer Assisted Language Learning, 36(5-6), 2021, 861-884. https://doi.org/10.1080/09588221.2021.1952272

[7] Danchenko, N.: Acousmatic music: principles of graphic visualization, Bulletin of Kyiv National University of Culture and Arts Series in Musical Art, 3(1), 2020, 37-47. https://doi.org/10.31866/2616-7581.3.1.2020.204337

[8] 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

[9] Evans, G.; Hoover, M.; Winer, E.: Development of a 3D conceptual design environment using a head-mounted display VR system, Journal of Software Engineering and Applications, 13(10), 2020, 258-277. https://doi.org/10.4236/jsea.2020.1310017

[10] He, G.: Schema interaction visual teaching based on smart classroom environment in art course, International Journal of Emerging Technologies in Learning (IJET), 15(17), 2020, 252. https://doi.org/10.3991/ijet.v15i17.16441

[11] Maba, A.: Computer-aided music education and musical creativity, Journal of Human Sciences, 17(3), 2020, 822-830. https://doi.org/10.14687/jhs.v17i3.5908

[12] Pei, Z.; Wang, Y.: Analysis of computer-aided teaching management system for music appreciation course based on network resources, Computer-Aided Design and Applications, 19(S1), 2021, 1-11. https://doi.org/10.14733/cadaps.2022.S1.1-11

[13] Yang, Z.: Data analysis and personalized recommendation of Western music history information using deep learning under Internet of Things, Plos One, 17(1), 2022, e0262697. https://doi.org/10.1371/journal.pone.0262697

[14] Yu, X.; Ma, N.; Zheng, L.; Wang, L.; Wang, K.: Developments and applications of artificial intelligence in music education, Technologies, 11(2), 2023, 42. https://doi.org/10.3390/technologies11020042

[15] Zhang, Y.; Yi, D.: A new music teaching mode based on computer automatic matching technology, International Journal of Emerging Technologies in Learning (IJET), 16(16), 2021, 117-130. https://doi.org/10.3991/IJET.V16I16.24895