Development and Implementation of an Embedded Systems-Based Artificial Intelligence-Driven Music Teaching System for Vocational Colleges

Fan He[1]

[1]School of Art, Xinxiang Institute of Engineering, Xinxiang 453700, China

Corresponding author: Fan He, hefan198382@126.com

Abstract. When the continuous advancement for the brand-new courses is undergoing, teaching for music courses, from the outdated closed teaching to open teaching. From the traditional pure manual teaching method to the use of artificial intelligence technology to assist teaching. Based on artificial intelligence technology, this paper builds and practices the music teaching system in vocational colleges, and draws the following conclusions: The bottleneck of the music teaching system in vocational colleges is the network bandwidth； In terms of music pitch extraction, the improved algorithm proposed in this paper has greatly improved the period, robustness and stability of the extracted signal compared with the traditional cepstrum method, and the periodic detection effect of pitch is obvious；Using virtual reality technology to create a virtual music classroom algorithm, by comparing with the other three algorithms, it is good for it；Compared with the P.563 method and the FDGSVM-based method under the ITU-T P. Supplement-23 Database, the music quality evaluation method studied in this paper improves the correlation coefficient with subjective evaluation and reduces the mean square error; Compared with the method based on FDGSVM, it has increased by 78.59%, but in terms of comprehensive performance, the improved method still has advantages and can effectively evaluate the quality of music.

Keywords: music teaching; Embedded Systems-Based artificial intelligence technology; virtual reality technology; cepstrum method; music quality evaluation; Context of Digital Art
DOI: https://doi.org/10.14733/cadaps.2024.S8.180-195

1 INTRODUCTION

Virtual music teaching is to use virtual reality technology in music teaching to create an immersive virtual music teaching environment for learners, so that learners can see and hear music through visual, auditory, tactile and other senses, and even simulate with virtual reality technology[16]. The objects that come out interact with each other, creating a feeling of swimming in the virtual music world[1]. The products of virtual reality technology can be used as a new tool for learning music, so that on the one hand, learners can obtain experiences and feelings that cannot or are difficult to obtain in the real world from the virtual imaginary space, and on the other hand, it also helps In the era of rapid development of information technology, music teachers and music educators are changing traditional music teaching thinking and enriching music teaching models[12]. Virtual reality technology will help teachers to teach more quickly and effectively[4].Virtual music teaching provides a new platform and space for the innovation of music teaching mode, and realizes the great possibility of teaching mode innovation[3]. The advantages of virtual reality technology are very beneficial for music teachers to build a new and rich teaching environment and means[22]. This interactive and intuitive learning environment is conducive to stimulating learners' interest and enthusiasm in learning music, thereby improving the quality and effect of teaching and learning[9].The music teaching system is driven by artificial intelligence technology[21].

2 LITERATURE REVIEW

2.1 Artificial Intelligence Technology and Music Teaching

It was first put forward by Aaron Lacier[13]. It is the world computer simulation system that integrates it real. Now, virtual reality has been successfully used in military aircraft, navigation, medical training, entertainment and education[11]. It has become an indispensable technology for innovative production, design, R&D and education[23].Three characteristics for interacting with each other, namely immersion, imagination and interactivity[10].

(1) Immersion. Immersion means that learners have a sense of immersion in the virtual music environment created by virtual reality technology, feel that they are part of this environment, are surrounded by the virtual music world, and are more focused during the experience[18]. Immersive music teaching can help learners transform the original music learning process into an external form[15]. For example, in a virtual music environment, the learner appropriately reflects his perception, understanding and imagination of music, and the teacher will listen to the music from the learner[7]. The positioning of eyes during music, the movement of hands, the movement of footsteps and other external behaviors, to judge the learner's sense of participation in the virtual music environment, the attention to music, the continuity of music learning, etc[19]. The situation of knowledge acquisition or the efficiency of learning can be better monitored[8].Embedded Systems-Based Artificial Intelligence (AI)-driven music teaching system tailored for vocational colleges focusing on digital art disciplines. The proposed system will leverage cutting-edge AI technologies to provide interactive and personalized music education experiences for students.

(2) Imaginative. Imagination means that the use of virtual reality technology can combine learners' rational cognition of music with perceptual cognition, and stimulate learners' musical imagination[14]. For example, in the teaching process of foreign music appreciation or ethnic minority music appreciation, learners can be brought into a specific virtual music environment, and they can experience the relationship between music through the era characteristics of the foreign music works to be appreciated or the singing and dancing scenes of ethnic minorities[6]. Differential characteristics. Learners can also imagine that they have come to a foreign country, wandering among the music styles of different ethnic groups and music genres in different regions, so as to further help learners to quickly understand the background of music creation and the style of music[20].

(3) Interactivity. Interaction means that learners do not need to use tools such as mouse and keyboard to interact naturally with the virtual music environment, and at the same time get feedback from the music environment[2]. Devices such as virtual headsets and virtual joysticks are used as

interactive tools. It is an immersive VR music maze exploration teaching experiment. The music maze exploration scene is a three-dimensional maze, which makes full use of the immersive, imaginative and interactive characteristics of music virtual reality technology, so that learners have the opportunity to Practice and feel confident in training hearing skills in 3D environments[5]. The mobile virtual reality helmet worn by the learner will block the learner's vision and hearing, and the screens of the left and right eyes will display virtual images, creating a three-dimensional sense in the learner's mind. be strengthened.

2.2 3D Reconstruction Technology of Virtual Reality

A lot of researches have been done on the problem of weak texture in 3D reconstructed images. Zhang Yifei et al. proposed a matching algorithm fused with image segmentation to solve the problem of inaccurate matching of weak texture regions in stereo matching. Xu Xuesong et al. proposed a stereo matching algorithm using convolutional neural network (CNN) to solve weak texture or highlight regions. Liu Yifan et al. proposed a 3D reconstruction algorithm for binocular stereo vision based on iterative closest point (ICP) and motion recovery structure (SFM), which overcomes the shortcomings of the two algorithms to the greatest extent. Aiming at the problem of robot vision autonomous navigation and obstacle avoidance in weak texture scenes, Lin Yimin et al. proposed a solution based on laser point projector. Feng Hui proposed a weak texture 3D image detection algorithm based on Harris corner detection, which can improve the quality of image processing. Chen Jia et al. introduced the application of deep learning in 3D reconstruction of single image objects and the research status of traditional methods in detail. Wang Fang et al. proposed a 3D reconstruction algorithm for weak texture images that can be used in a frame-missing environment, which can improve the algorithm's image texture recognition and repair and reconstruction capabilities.In this paper, virtual reality technology is used to realize the real instrument teaching scene in a virtual environment, such as drums, bass, guitar and other instruments. In order to achieve interactive learning with students.In the context of virtual reality (VR) technology, digital arts have been utilized to create realistic instrument teaching scenes in virtual environments. This enables interactive learning experiences for students, allowing them to learn and practice playing instruments such as drums, bass, and guitar in a simulated and immersive environment

2.3 Music Quality Evaluation Algorithm

At present, many domestic scholars have begun to study music quality evaluation from different aspects. Among them, Liang Weiqian and others can be regarded as the first group of researchers in China to start to study music quality evaluation. They mainly use music recognition as the basis to evaluate music quality by judging the accuracy and fluency of music signals. Then Zhao Heming and Qi Jianyu put forward a new idea. They used the HMM method to evaluate the quality of music from two aspects of tone and phoneme. In the 21st century, Wang Jing et al. began to study the music quality evaluation without reference, and they proposed an objective evaluation method of narrowband music quality. The method first extracts PLP features from pure music, and then inputs them into Gaussian mixture model GMM for training to obtain the GMM model of pure music. Secondly, the music activity detection VAD is used to divide the music signals with transmission impairment into three categories, and the feature differences between the characteristics of each type of music signal and the pure music Gaussian mixture model are calculated separately, and then the support vector regression method SVR is used as the music feature. Compared with the existing P.563 method, this method reduces the mean square error and improves the correlation coefficient7. After further research, they then published an article on Fuzzy Gaussian Mixture Models and Fuzzy Neural Networks for Reference-Free Music Quality Assessment Methods. First, the pure music is divided into clear, voiced and silent, and the perceptually weighted linear prediction feature parameters are extracted from each category to establish the corresponding FGMM reference model. Then, the PLP features of distorted music are also divided into three categories, the consistency parameters of each category of distortion features and the corresponding FGMM reference model are calculated, and finally the obtained consistency parameters are mapped to the subjective MOS score using the fuzzy neural network model. The improved method is obviously better than the previous music quality evaluation method based on Gaussian mixture model, and the performance is improved. Subsequently, another domestic author proposed a new method for music quality evaluation based on output-based multilingual music samples using Fuzzy Directed Graph Support Vector Machine (FDGSVM). The method organizes multiple fuzzy support vector machines that can be classified into two types into a directed graph structure with a unique root node, and obtains a multi-class classifier FDGSVM; extracts the Mel cepstral coefficients of the music signal to be tested and uses it as a feature vector , and then map the feature vector to the nonlinearly divided subjective mean opinion score MOS interval through FDGSVM, and the mapped value is the objective evaluation result of the output music quality. In practical applications, the sample points in the training process will affect the final result, some of which are important and some can be ignored. In the music quality evaluation method based on FDGSVM, each input MFCC vector is not simply classified as a certain MOS level, but the weight is assigned according to its importance. The experimental simulation results verify the effectiveness of the method.In this paper, the accuracy of the music teaching quality evaluation system is judged by calculating the root mean square error value RMSE of the objective evaluation results and subjective evaluation results of music teaching quality; the correlation is described by Karl-Pearson correlation coefficient. Sex is mainly expressed according to this indicator.

3 METHODOLOGY

The artificial intelligence technology-driven music teaching system in vocational colleges as shown in Figure 1.

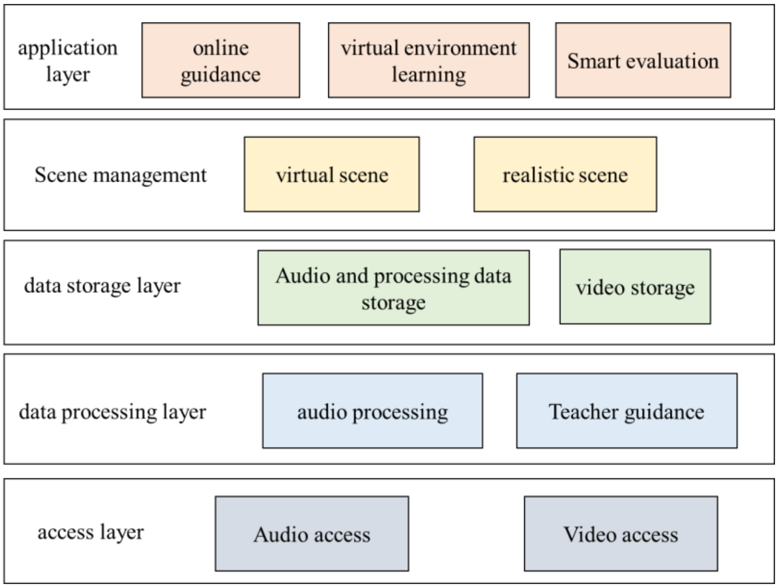


Figure 1: Vocational College Music Teaching System.

(1) Access layer: The access layer mainly includes audio access and video access. Audio access realizes the input of sound through a voice capture device (such as a microphone); video access realizes the input of video through a video capture device (such as a camera). (2) Data processing layer: including audio processing system and teacher guidance system. The audio processing system realizes the extraction of sound feature data and the comparison of two sets of pitch data. (3) Data storage layer: realize the storage of audio and audio feature data and video storage. The storage interface adopts the commonly used ODBC and JDBC data access methods, and the audio and video file interface provides file access services through the file IO system that has been encapsulated by the Java platform. (4) Scene management layer: including virtual scene and real scene. Virtual scenes are virtual learning environments and characters created by software. The real scene is realized by accessing the real video. (5) Application layer: realize the intelligent evaluation of conventional online guided learning, virtual environment learning and autonomous learning.

3.1 Audio Processing

The audio processing system includes two parts: feature extraction module and feature processing module. The feature extraction module includes a pitch frequency recognizer and a pitch length acquirer, as shown in Figure 2

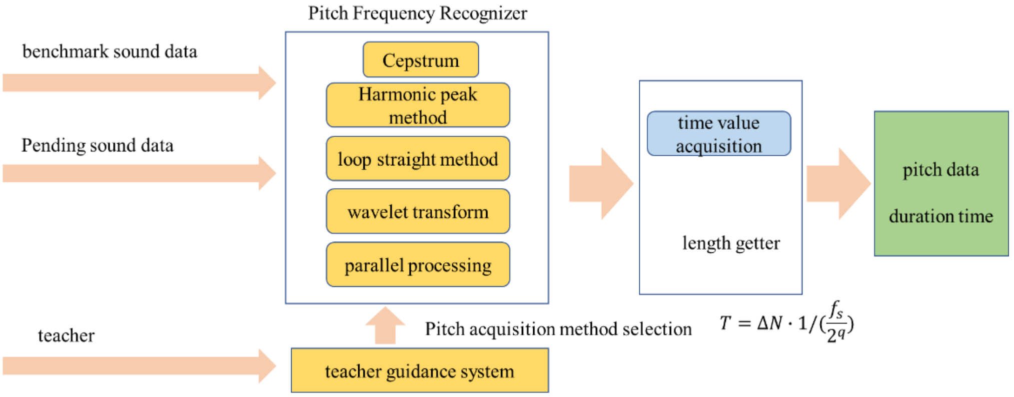


Figure 2: Feature Extraction Module.

The fundamental frequency identification methods built in the fundamental frequency identifier include: cepstrum method, harmonic peak method, cyclic straight method, wavelet transform method and parallel processing method. According to the method of obtaining the pitch selected by the teacher in the teacher guidance system, the pitch frequency data features are extracted.

(1) Improved cepstrum method to extract musical features

In a pure environment, the cepstrum method has a better detection effect, but in a noisy environment, it is more difficult to detect the pitch of the music signal, and the cepstrum method is easily affected by environmental factors such as formants, and it is difficult to extract the pitch period normally. The cepstrum algorithm process is: the original music signal is divided into frames by windowing, the least squares method is used to remove the trend term, the multi-window spectrum estimation spectrum is subtracted to remove the noise, the linear prediction method is used to remove the formant influence, the cepstrum method is used to measure the pitch and the median filter is smoothed. Get the pitch period.

The improved part of this algorithm is "multi-window spectral estimation spectral subtraction denoising", the noise in the signal is inevitable in the signal detection, and reducing the interference of the noise to the greatest extent becomes a key step in the signal detection. The improved algorithm adopts spectral subtraction. The basic idea of spectral subtraction is to analyze the noise components in the speech by using the silent segment, and then subtract the noise-containing speech to obtain pure speech.

If y(t) is a noise signal, n(t) is a noise signal, and s(t) is a pure speech signal, there are:

$$y(t)=s(t)+n(t) \tag{1}$$

Using Y(w) .s(w).N(w) to represent the Fourier transform of y(t), s(t), and n(t) respectively, we can get:

$$Y(\omega)=S(\omega)+N(\omega) \tag{2}$$

Compared with the conventional spectral subtraction method, the improved spectral subtraction method uses several orthogonal data windows to calculate the direct spectrum separately, and further obtains the spectral estimation through mean processing, and obtains the appropriate estimated variance. It is defined as:

$$S_k^{mt}(\omega)=\frac{1}{L}\sum_{k=0}^{L-1}S_k^{mt}(\omega) \tag{3}$$

Smt is the spectrum of the data window, and the data window is L:

$$S_k^{mt}(\omega)=\left|\sum_{k=0}^{L-1}a_k(n)\ x(n)e^{-jnw}\right|^2 \tag{4}$$

In the formula: N is the sequence length; is the kth data window; x(n) is the data sequence, and the different data windows are orthogonal.

$$\begin{cases}\sum a_k(n)a_j(n)=1,\ k=j\\ \sum a_k(n)a_j(n)=0,\ k\neq j\end{cases} \tag{5}$$

The pitch getter uses the algorithm:

$$T=\Delta N\cdot 1/(f_s/2^q) \tag{6}$$

Extract the time value of the sound length, where △N is the number of samples between the two note endpoints, q is the wavelet decomposition scale, and it is the initial sampling frequency of the signal.

The feature processing module includes a frequency pitch converter, a pitch comparator and a duration comparator. The way the frequency pitch converter converts is:

$$X=\left[12\times\lg(y/27.5)\right]/\lg 2+1 \tag{7}$$

Among them, y is the fundamental frequency, and X is the corresponding one. Form a practical and reasonable real form, so as to bring a real teaching experience. Intangible music can also be made into images that learners can touch and control. Transform invisible music into tangible scores or light-sensitive images that can represent music, and present them in front of sound fun virtual scene.

3.2 Music Teaching Under Virtual Reality

In music teaching, teachers can use virtual reality technology to combine the content of music textbooks to transform the scenes that learners cannot experience in time or have not experienced in the original abstract theory or practical teaching into virtual scenes. Form a practical and reasonable real form, so as to bring a real teaching experience. Intangible music can also be made into images that learners can touch and control. Transform invisible music into tangible scores or light-sensitive images that can represent music, and present them in front of each learner. The learner will see the red light-sensing image used to represent the melody feature of exciting music; the yellow light-sensing image used to represent the melody feature of soothing music; like the green light-sensing image used to represent the melody feature of jumping music picture. The light-sensing images of different colors will flow with the music playing, and learners can accurately capture the images of different colors to clarify the characteristics of the music in the process of listening to the music, so as to achieve the training of improving the listening ability of music in the virtual music scene. Three-dimensional reconstruction technology is used to realize the real instrument teaching scene in a virtual environment, such as drums, bass, guitars and other instruments.

This section only discusses the instrument feature matching algorithm, that is, the improved Harris-scale-invariant feature transform matching algorithm. The Harris-Scale Invariant Feature Transform (SIFT) algorithm has good real-time performance, but the accuracy is average. In order to improve the accuracy of the algorithm, the extracted feature corners are first purified by the entropy method; then the Euclidean distance is changed to the Mahalanobis distance, and the random extraction-consistency algorithm is used to replace the threshold method in the screening process.

(1) Feature corner extraction

First, the Harris algorithm is used to extract feature corners from the preprocessed grayscale image, which can be expressed as:

$$M(x,y) = \mathrm{Det}\left[H(x,y)\right] - k \cdot \left\{\mathrm{Tr}[H(x,y)]\right\}^2 \tag{8}$$

In the formula, (x, y) is the pixel coordinates, M(x, y) is the corner response function, H is the Hessian matrix, Tr(H) is the trace of the matrix H, and Det(H) is the determinant of the matrix H. value, k is a constant, and the experiment takes = 0.04, and the local maximum point of the response function of the corner point is the corner point. Generally speaking, the entropy of an image is proportional to the size of the information contained in the image. Therefore, the feature corner points extracted by the original Harris-SIFT algorithm are purified by the entropy method, that is, the feature corner points with small entropy value are deleted, so as to improve the matching accuracy of the algorithm, reduce the time used for matching, and improve the real-time performance of the algorithm.

(2) Generate SIFT feature descriptor

The Harris algorithm can obtain the position information of the feature corners and accurately locate all the feature corners. However, when the SIFT algorithm matches the image, it is necessary to obtain the orientation information of the feature corners. When obtaining the direction information of the feature corners, we must first calculate all the gradient directions and modulus values of the image pixels, which can be expressed as:

$$m(x,y) = \left\{\left[L(x+1,y) - L(x,y-1)\right]^2 + \left[L(x,y+1) - L(x,y-1)\right]^2\right\}^{\frac{1}{2}} \tag{9}$$

$$\theta(x,y)=\arctan\left\{\left[L(x,y+1)-L(x,y-1)\right]/\left[L(x+1,y)-L(x-1,y)\right]\right\} \tag{10}$$

where m(x, y) is the modulus value of the gradient at (.x, y), $\theta(x, y)$ is the direction of the gradient at (x, y), and L(x, y) is the key point at all scales. According to the sampling principle of 3σ (σ is the scale space factor), the gradient modulus value is weighted by a Gaussian distribution of 1.5σ, the radius of the neighborhood window is set to 3 × 1.5σ, and the gradient and direction of the corner points in the neighborhood are counted through the gradient histogram. , so as to generate the required 128-dimensional feature descriptor.

(3) Feature matching

The Mahalanobis distance is used to replace the Euclidean distance of the original Harris-SIFT algorithm to determine the similarity between two feature vectors. Compared with Euclidean distance, the one between two vectors can be expressed as:

$$D_{XY}=\sqrt{(X-Y)^{T}S^{-1}(X-Y)} \tag{11}$$

$$S=E\left\{\left[X-E(X)\right]\left[Y-E(Y)\right]\right\} \tag{12}$$

In the formula, X and Y are two sets of vectors, S is the covariance matrix of the two sets of vectors, S-1 is the inverse covariance matrix, and E is the mean. It can be found that compared with the Euclidean distance, the Mahalanobis distance has an additional inverse covariance matrix, which can measure the distance relationship more accurately. When the covariance matrix is the identity matrix, it evolves into the Euclidean distance. Although the amount of computation is increased, Mahalanobis distance eliminates it of matching accuracy is higher, and the overall effect is better than Euclidean distance. In order to further improve the matching accuracy of the algorithm, it is necessary to eliminate the wrong matching. Therefore, the threshold method is improved to the random sampling consistency (RANSAC) algorithm, which can accurately estimate the model parameters in the data with many external points.

3.3 Music Quality Evaluation System

The digital processing of music sound signal is to serve the human hearing, and the human auditory system has its unique perception characteristics to the sound signal. Therefore, in order to accurately predict the true MOS score, the objective evaluation method of music quality needs to select speech features that can fully characterize the speech characteristics and human perception characteristics as parameters. Since the output-based voice quality objective evaluation method has no reference signal, this paper selects the parameters of the stack auto-encoder to train the stack-type auto-encoder by selecting a database containing multiple impairment types of voices, so that the SAE can extract the essential features that can better characterize the voice quality. During the training process, a single auto-encoder is trained in turn. When the training of one auto-encoder is completed, the next auto-encoder takes the output of the hidden layer of the first auto-encoder as input, and then repeats the previous step until all AEs are Training is complete. The essential features of speech are extracted by the stack auto-encoder, and then the features are mapped to the subjective MOS scores through the BP neural network. The overall framework of testing and training of the objective evaluation method of speech quality based on the stack auto-encoder is shown in Figure 3:

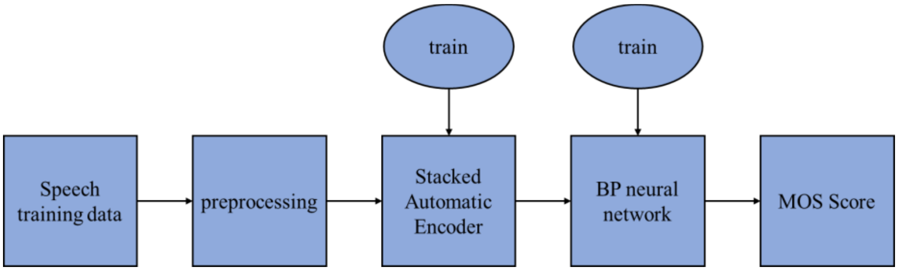


Figure 3: Frame Diagram of Music Quality Evaluation System.

The performance for music is mainly judged by two indicators: accuracy and correlation. The accuracy is judged by error value RMSE of the objective evaluation results and subjective evaluation results of music teaching quality; the correlation is described by Karl-Pearson correlation coefficient, and the correlation of the objective evaluation system of music teaching quality It is mainly expressed according to this indicator.The accuracy of the objective evaluation will be often measured by the mean square error value index. The mean square error (RMSE) is also called the standard error. It is the difference between the output score of the music teaching quality objective evaluation system and the subjective MOS score value. Take the square root of the result of the ratio of the sum of squares to the number of tests N.The root of the objective evaluation one will be to use computer simulation system to it . and the one absolute prediction error is the difference between the result value output by the computer and the true MOS score, which can be calculated by formula (13):

$$P_{error}(i) = MOS_S(i) - MOS_O(i) \tag{13}$$

The root mean square error RMSE is calculated as shown in formula (14):

$$RMSE=\sqrt{\left(\frac{1}{N-1}\sum_{i=1}^{N} P_{error}(i)^2\right)} \tag{14}$$

Among them, N represents the total number of test corpus, and the reason why N-1 is used in formula (14) is to make the root mean square error an unbiased estimate.The correlation index between subjective and objective evaluations is analyzed below. Karl-Pearson correlation coefficient is usually used to represent the correlation between the music teaching quality output score and the subjective real value, that is, the linear correlation between them. Therefore, the size of the correlation coefficient R is used in this paper to indicate the strength of the correlation between the objective evaluation result and the subjective real value. The larger the value of R, the stronger the correlation between the objective evaluation result value and the subjective MOS real value. sex is weaker. The calculation of R is shown in formula (15):

$$R=\frac{\sum_{i=1}^{N}\left((MOS_o(i)-\overline{MOS_O})(MOS_S(i)-\overline{MOS_S})\right)}{\sqrt{\sum_{i=1}^{N}\left((MOS_o(i)-\overline{MOS_O})^2\sum_{i=1}^{N}(MOS_S(i)-\overline{MOS_S})^2\right)}} \tag{15}$$

The two performance indicators for judging the quality of the objective evaluation system of music teaching quality have been analyzed before. Equations (16) and (17) are the performance difference parameters of the music teaching quality evaluation system:

$$\Delta_{R} = \frac{R_{proposed} - R_{compare}}{1 - R_{compare}} \times 100\% \tag{16}$$

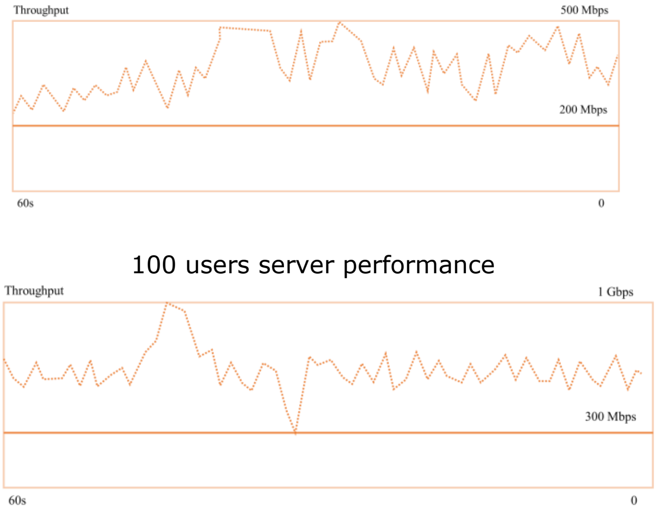
$$\nabla_{RMSE} = \frac{RMSE_{compare} - RMSE_{proposed}}{RMSE_{compare}} * 100\% \tag{17}$$

Among them, it is the growth rate of the correlation coefficient of the research method in the text compared with the comparison method, it is the decrease rate of the mean square error of the method in the text compared with the comparison method, the subscript "proposed" indicates the method studied in the paper, and "compare" means to compare methods, namely P.563 method and music in line with FDGSVM. the value of is positive, which proves that the method in this paper improves the correlation with subjective quality evaluation compared with the comparison method, and the value is negative, indicating that the one reduces the correlation with subjective quality evaluation.

4 RESULT ANALYSIS AND DISCUSSION

4.1 Performance Test of Music Teaching System

Due to limited VR equipment, this test uses PC VR player instead of VR equipment to play. Using 25 VM it runs 4 player processes, such server performance is shown in Figure 4(a), the one will be 47%, and the memory consumption is 15%. Bandwidth consumption is about 41.5%. Each VM virtual machine runs 8 VR player processes, such server performance is shown in Figure 4(b). The one is 79%, such memory will be 15%, bandwidth begins to bottleneck, which is close to 1Gbps. , jitter occurs.



100 users server performance

200 users server performance

Figure 4: Performance Test Results of Music Teaching System.

Therefore, the bottleneck of the music teaching system is the network bandwidth. If you need to continue to increase the pressure, you can deploy the cluster server.

4.2 Experimental Results of Improved Cepstrum Extraction of Musical Features

The pure speech "music teaching" superimposes Gaussian noise. After MATLAB simulation, it can be analyzed that the spectral subtraction method using multi-window spectral estimation can filter out signal noise. Compared to it, it can ensure filtering of noise signals while ensuring The integrity of the original signal, the result analysis in Figure 5.

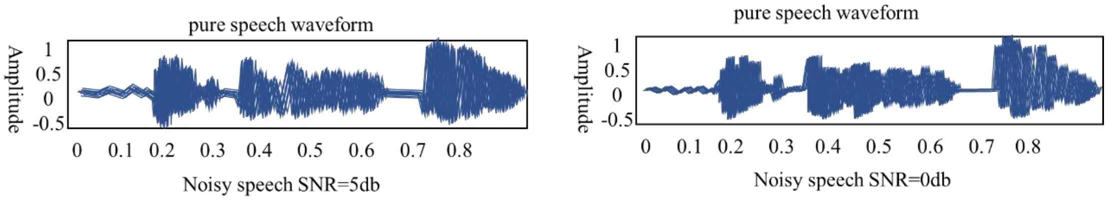


Figure 5: Comparison of the Processing of Speech With Noise "Music Teaching".

The pitch detection results of the traditional method and the pitch detection using the improved in Figure 6. the cepstrum method improved algorithm the period, robustness and stationarity of the extracted signal, and the period detection effect of the pitch is obvious.

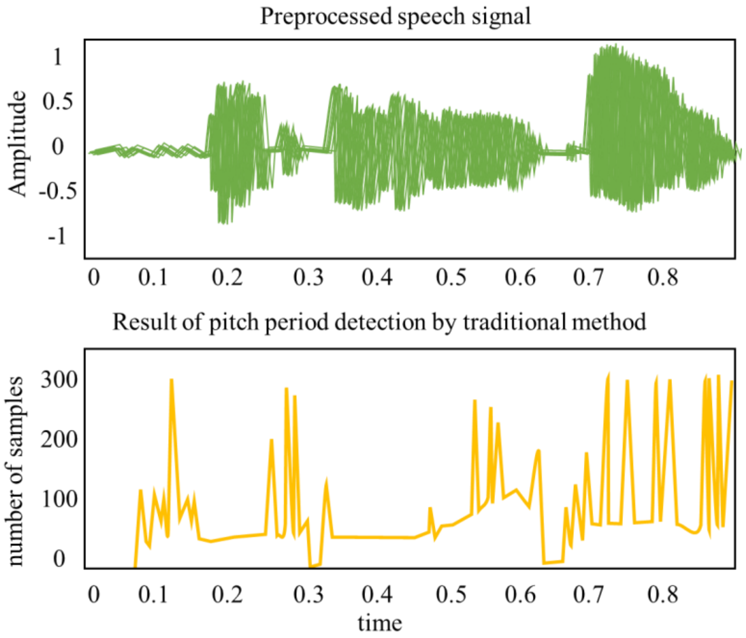


Figure 6: Comparison of the Improved Cepstrum Pitch Period Detection Results.

4.3 Comparison Test of Virtual Reality Scene Reconstruction Algorithm

Figure 7 is a graph of the experimental results of the four matching algorithms after processing with the image grayscale compensation algorithm. The reason is that the algorithm uses the entropy method to screen the feature corners extracted by the Harris-SIFT algorithm, which reduces the workload of the subsequent matching process, and uses the Mahalanobis distance instead of the Euclidean distance. Compared with the original Harris-SIFT algorithm, the matching accuracy is improved by 1.25 percentage points, and the time taken is reduced by 8.38%.

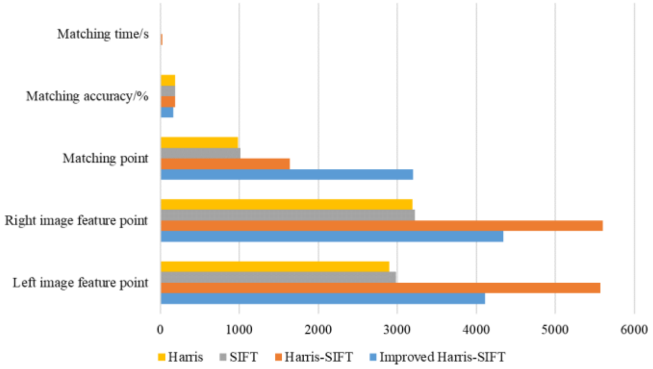


Figure 7: Accuracy and Running Time of the Four Algorithms.

4.4 Test Results of Music Quality Evaluation System

Music system studied when such number of units in the first AE hidden layer in the stacked automatic encoder is determined, how to select some units for such second AE is very important. The units selected is different, and corresponding experiments are carried out to compare and analyze.

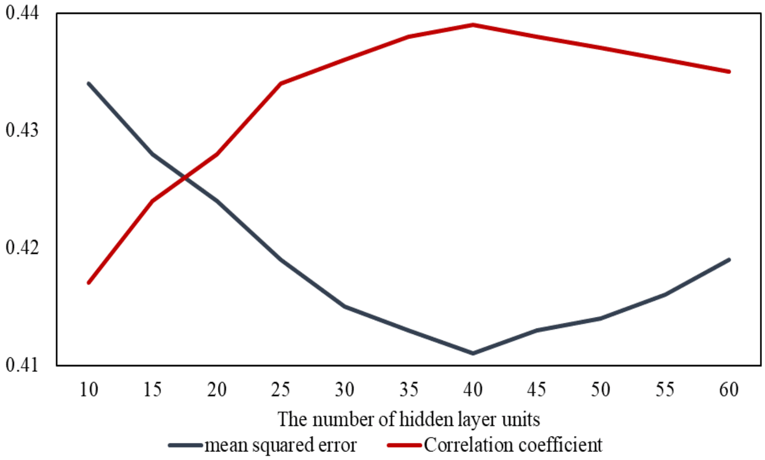
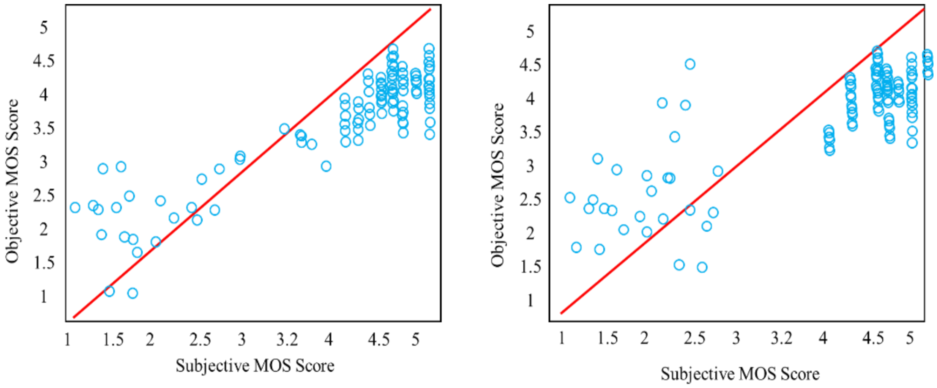


Figure 8: The Relationship Between the Number of Hidden Layer Units and the Performance of the Music Quality Evaluation System.

Figure 8 shows the corresponding R and RMSE values of the music quality evaluation system under different numbers of hidden layer units. The hidden layer units is small, R is proportional to the hidden layer single element, and RMSE is inversely proportional to the hidden layer single element. When they are 40, both reach the extreme point. With the increase of single elements in the hidden layer, the system performance deteriorates because the hidden layer is too dense or too sparse to effectively characterize the speech quality. Therefore, the stack encoder in this paper is 40.Input the essential features of music quality extracted by the stacked autoencoder into the trained BP neural network get MOS score. Figure 9(a)~(d) are the corresponding subjective evaluation values of the music quality evaluation method studied in this chapter, the music quality evaluation method based on FDGSVM, the P.563 music quality evaluation method and the music one. Scatter plot of objective evaluation values of test samples. The circle points in the figure indicate that the abscissa is the subjective MOS score, the ordinate is the objective prediction score, and the one where such prediction score and the subjective MOS score are completely consistent.



(a) The Music Quality Evaluation Method of this Paper (b) The FDGSVM Music Quality Evaluation Method

Figure 9: Comparison Between the Method in this Paper and Various Music Quality Evaluation Methods.

The quality of the music quality evaluation system can be judged according to the correlation between the output result and the subjective evaluation. The higher the correlation, the more ideal the music quality evaluation method is. The diagonal line in the scatterplot represents the ideal state in which the subjective score and the objective evaluation score are completely consistent. The closer the distribution of data points is to the diagonal line, the closer the evaluation result is to the subjective evaluation. Obviously, the method combining SAE and BP neural network (Figure9(a)) has a great improvement in correlation compared with the method based only on BP neural network (Figure9(d)). There are relatively outstanding performances in the music quality evaluation. In addition, the test data score distribution of the method studied in this paper (Figure9(a)) is closer to that of the existing ITU-T P.563 method (Figure9(c)) and the FDGSVM-based method (Figure 9(b)) The diagonal line shows that the method studied in this paper is more excellent. Figure 10 compares the correlation coefficient, mean square error and evaluation time of a single corpus of the three methods tested under the ITU-T P. Supplement-23 Database.

According to Figure 10, compared with the P.563 music quality evaluation the quality evaluation method studied in this paper is relatively reduced by 68.86%, and the correlation coefficient (R) is relatively increased by 76.93%; The music quality evaluation method of FDGSVM, the mean square error (RMSE) of the music quality evaluation method studied in this paper is relatively reduced by 43.24%, and the correlation coefficient (R) is relatively increased by 42.09%. Comparing the three methods to test the evaluation time of a single corpus, P.563 takes the longest time, and FDGSVM takes the shortest time.

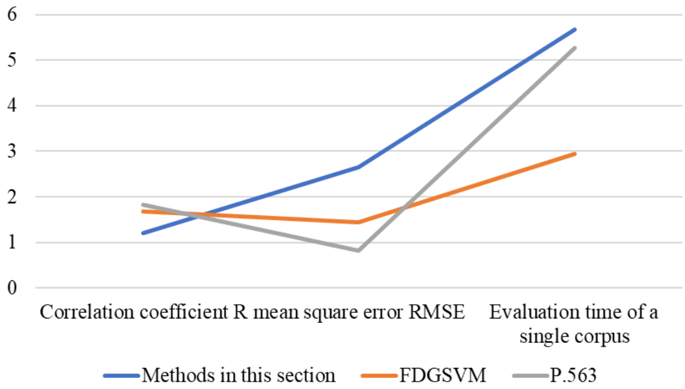


Figure 10: Performance Comparison of Three Methods in ITU-TP.Supplement-23 Database.

Compared with FDGSVM, the evaluation time of the method in this paper increases by 78.59%. The simulation results show that the music quality evaluation method studied in this paper improves the correlation coefficient with subjective evaluation and reduces the mean square error compared with the P.563 method and the method based on FDGSVM under the ITU-T P. Supplement-23 Database. Time increases compared to FDGSVM-based methods 78.59%, but in terms of comprehensive performance, the improved method still has advantages, which can effectively evaluate the quality of music.

5 CONCLUSION

The music quality evaluation method studied in this paper improves the correlation coefficient with subjective evaluation and reduces the mean square error; Compared with the method based on FDGSVM, it has increased by 78.59%, but in terms of comprehensive performance, the improved method still has advantages and can effectively evaluate the quality of music.Based on artificial intelligence technology, this paper builds and practices the music of vocational , draws the following conclusions:(1) The bottleneck for music teaching system in vocational colleges is the network bandwidth. If you need to continue to increase the pressure, you can deploy the cluster server.(2) In terms of music pitch for the improved one proposed it has greatly improved the period, robustness and stability of the extracted signal the cepstrum , and periodic detection effect of pitch is obvious.(3) Using technology for music classroom algorithm, by comparing with the other three algorithms,it is there.(4) Compared with the P.563 method and the FDGSVM-based method under the ITU-T P. Supplement-23 Database, the music quality evaluation method studied in this paper improves the correlation coefficient with subjective evaluation and reduces the mean square error; Compared with the method based on FDGSVM, it has increased by 78.59%, but in terms of comprehensive performance, the improved method still has advantages and can effectively evaluate the quality of music.

Fan He, https://orcid.org/0009-0003-3281-8712

REFERENCES

[1] Ahmad, T.; Zhang, D.; Huang, C.: et al. Artificial Intelligence in Sustainable Energy Industry: Status Quo, Challenges and Opportunities, Journal of Cleaner Production, 289, 2021, 125834. https://doi.org/10.1016/j.jclepro.2021.125834

[2] Bag, S.; Pretorius, J.H.C.; Gupta, S.: et al. Role of Institutional Pressures and Resources in the Adoption of Big Data Analytics Powered Artificial Intelligence, Sustainable Manufacturing Practices and Circular Economy Capabilities, Technological Forecasting and Social Change, 163, 2021, 120420. https://doi.org/10.1016/j.techfore.2020.120420

[3] Bag, S; Pretorius, J. H. C.; Gupta, S.: et al. Role of Institutional Pressures and Resources in the Adoption of big data Analytics Powered Artificial Intelligence, Sustainable Manufacturing Practices and Circular Economy Capabilities, Technological Forecasting and Social Change, 163, 2021, 120420. https://doi.org/10.1016/j.techfore.2020.120420

[4] Cai, Y.; Ramis Ferrer, B.; Luis Martinez Lastra, J.: Building University-Industry Co-Innovation Networks in Transnational Innovation Ecosystems: Towards a Transdisciplinary Approach of Integrating Social Sciences and Artificial Intelligence, Sustainability, 11(17), 2019, 4633. https://doi.org/10.3390/su11174633

[5] Chen, B.; Liu, Y.; Zheng, J.: Using Data Mining Approach for Student Satisfaction With Teaching Quality in High Vocation Education, Frontiers in Psychology, 2021, 12. https://doi.org/10.3389/fpsyg.2021.746558

[6] Chen, L.; Wang, P.; Dong, H.: et al. An Artificial Intelligence Based Data-Driven Approach for Design Ideation, Journal of Visual Communication and Image Representation, 61, 2019, 10-22. https://doi.org/10.1016/j.jvcir.2019.02.009

[7] Cope, B.; Kalantzis, M.; Searsmith, D.: Artificial Intelligence for Education: Knowledge and its Assessment in AI-Enabled Learning Ecologies, Educational Philosophy and Theory, 53(12), 2021, 1229-1245. https://doi.org/10.1080/00131857.2020.1728732

[8] Cukurova, M.; Kent, C.; Luckin, R.: Artificial Intelligence and Multimodal Data in the Service of Human Decision - Making: A Case Study in Debate Tutoring, British Journal of Educational Technology, 50(6), 2019, 3032-3046. https://doi.org/10.1111/bjet.12829

[9] Demchenko, I.; Maksymchuk, B.; Bilan, V.: et al. Training Future Physical Education Teachers for Professional Activities Under the Conditions of Inclusive Education, Brain. Broad Research in Artificial Intelligence and Neuroscience, 12(3), 2021, 191-213. https://doi.org/10.18662/brain/12.3/227

[10] Dermody, G.; Fritz, R.: A Conceptual Framework for Clinicians Working With Artificial Intelligence and Health - Assistive Smart Homes, Nursing Inquiry, 26(1), 2019, e12267. https://doi.org/10.1111/nin.12267

[11] Elmousalami, H.H.: Artificial Intelligence and Parametric Construction Cost Estimate Modeling: State-of-the-Art review, Journal of Construction Engineering and Management, 146(1), 2020, 03119008. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001678

[12] Felten, E.W.; Raj, M.; Seamans, R.: The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization, 2019 NYU Stern School of Business.

[13] Inkster, B.; Sarda S.; Subramanian, V.: An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study, Jmir Mhealth and Uhealth, 6(11), 2018, e12106. https://doi.org/10.2196/12106

[14] Jin, R.; Zou, P.X.W.; Piroozfar, P.: et al. A Science Mapping Approach Based Review of Construction Safety Research, Safety Science, 113, 2019, 285-297. https://doi.org/10.1016/j.ssci.2018.12.006

[15] Lu, Q.; Parlikad, A.K.; Woodall, P.: et al. Developing a Digital Twin at Building and City Levels: A Case Study of West Cambridge Campus, Journal of Management in Engineering-ASCE, 2020, 36(3). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000763

[16] Pedro, F.; Subosa, M.; Rivas, A.: et al. Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development, 2019.

[17] Sacks, R.; Girolami, M.; Brilakis, I.: Building Information Modelling, Artificial Intelligence and construction Tech, Developments in the Built Environment, 4, 2020, 100011. https://doi.org/10.1016/j.dibe.2020.100011

[18] Selamat, A.; Alias, R.A.; Hikmi, S.N.: et al. Higher Education 4.0: Current Status and Readiness in Meeting the Fourth Industrial Revolution Challenges, Redesigning Higher Education Towards industry, 4, 2017, 23-24.

[19] Shen, J.; Zhang, C.J.P.; Jiang, B.: et al. Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review, JMIR Medical Informatics, 7(3), 2019, e10010. https://doi.org/10.2196/10010

[20] Yang, R.: Artificial Intelligence-Based Strategies for Improving the Teaching Effect of Art Major Courses in Colleges, International Journal of Emerging Technologies in Learning, 15(22), 2020, 146-160. https://doi.org/10.3991/ijet.v15i22.18199

[21] Zhang, C.; Lu, Y.: Study on Artificial Intelligence: The State of the Art and Future Prospects, Journal of Industrial Information Integration, 23, 2021, 100224. https://doi.org/10.1016/j.jii.2021.100224

[22] Zhang, Y.; Xiong, F.; Xie, Y.: et al. The Impact of Artificial Intelligence and Blockchain on the Accounting Profession, IEEE Access, 8, 2020, 110461-110477. https://doi.org/10.1109/ACCESS.2020.3000505

[23] Zhao, Y.; Li, T.; Zhang, X.; et al.: Artificial Intelligence-Based Fault Detection and Diagnosis Methods for Building Energy Systems: Advantages, Challenges and the Future, Renewable and Sustainable Energy Reviews, 109, 2019, 85-101. https://doi.org/10.1016/j.rser.2019.04.021