Hindawi
Mobile Information Systems
Volume 2022, Article ID 7629367, 9 pages
https://doi.org/10.1155/2022/7629367

Research Article

Modeling Method of Intelligent Scoring for Solfeggio Training Based on Deep Learning

Ying Lu

School of Music, Baoji University of Arts and Sciences, Baoji 721013, Shaanxi, China

Correspondence should be addressed to Ying Lu; luying@bjwlxy.edu.cn

Received 16 August 2022; Revised 7 September 2022; Accepted 17 September 2022; Published 10 October 2022

Academic Editor: Muhammad Zakarya

In the solfeggio training, melody model singing is an important part of auditory training. Reasonable and effective model singing training is of great significance to train the singer's hearing. So, the level of model singing also reflects the level of the singer's solfeggio to a certain extent. The correct model singing score will help students to master their own real melody model singing level and help students to better improve their solfeggio skills. However, the melody model singing training scoring system that has been developed has low accuracy and needs to be improved. In order to improve the accuracy of the intelligent scoring results of melody model singing training, this paper used deep learning technology to study the intelligent scoring method of melody model singing training. This paper first extracted the features of songs, then collected a large number of song features, trained the existing feature data, and established an intelligent scoring model for melody model singing training. Then, the students' model songs and standard songs were used as two inputs, and the similarity of the characteristics of the two songs was analyzed; then, the score results of the model songs were obtained. The validity of this research should be verified by a comparative experiment. The research results showed that the melody model singing training intelligent scoring model, established by deep learning technology, has a higher accuracy rate, and the accuracy rate was increased by 6.82% compared with the original scoring model, indicating the research has practical significance.

1. Introduction

Model singing training can effectively improve the trainer's musical hearing ability and effectively improve the trainer's auditory discrimination ability and pitch perception ability, prompting them to continuously enhance their singing skills, and model singing training scoring is an important part of it. With the wide application of artificial intelligence and the popularization of the Internet and smartphones, the problem of intelligent scoring of melody model singing training is more and more worthy of in-depth study. At present, although there has been preliminary exploration on the intelligent scoring of melody model singing training, these scoring systems have low accuracy and have not achieved satisfactory results. In order to further improve the scoring accuracy of the melody model singing training intelligent scoring system, this study attempted to use deep learning technology to develop the melody model singing training intelligent scoring system.

At present, there are many research studies on solfeggio. Debevc et al. conducted an experiment to examine the effectiveness of an interactive mobile app mySolfeggio in solfeggio learning. Experimental results showed that students show higher scores in musical interval and rhythm accuracy when using the mobile app [1]. Based on dynamic teaching theory and teaching practice, Maofang constructed a dynamic teaching mode of solfeggio class to meet the needs of talent training in basic music education and improve the teaching effect of solfeggio class [2]. Aycan has developed a vocal training system based on bona exercises by examining vocal exercises adapted to rhythmic pronunciation exercises in the classroom. The research results showed that bona exercises play an important role in solfeggio training [3]. Ding conducted research on SIFT-based audio processing technology. The experimental results showed that the introduction of digital multimedia music production technology in music solfeggio teaching could not only make the teaching form vivid, standardize the teaching process, and

enrich the teaching content but also improve the quality of teaching and promote students' enthusiasm and interest in learning [4]. Ying analyzed and compared the effect of CAT technology in the auxiliary teaching system of solfeggio and ear training and the difference with the original traditional teaching. The results showed that more than 60% of the students are satisfied with the cat technology learning assistance system [5]. Although there are many studies on solfeggio training, there are few studies on melody model singing training in solfeggio training, and the research on intelligent scoring of melody model singing training has not yet achieved satisfactory results.

Due to their advantages in feature extraction, deep learning techniques are widely used. Geert investigated the use of deep learning for image classification, object detection, segmentation, registration, and other tasks, having brief overview of research in each application area [6]. Shen et al. introduced the fundamentals of deep learning methods and reviewed their success in image registration, anatomy, cellular structure detection, tissue segmentation, computer-aided disease diagnosis, and prognosis [7]. Kermany et al. has built a deep learning framework-based diagnostic tool to screen patients for common treatable blinding retinal diseases; the study of which can help speed up diagnosis and referral for these treatable diseases [8]. Rajkumar et al. proposed a representation of the patient's complete raw EHR records based on the Fast Healthcare Interoperability Resource (FHIR) format, and studies have demonstrated that deep learning methods using this representation can accurately predict multiple medical events from multiple centers [9]. Sinha et al. demonstrated that deep neural networks (DNNs) can be trained to solve the inverse problem in computational imaging, given a raw intensity image recorded at a certain distance, training a DNN to recover phase objects [10]. Although deep learning technology is widely used, no one has studied the application of deep learning technology in melody model singing training scoring.

In this paper, the Mel sound spectrum was used as the sound spectrum feature to extract song features; then, the existing feature data were trained to build a melody model singing intelligent scoring model based on the convolutional neural network. The feature similarity between them was used to score analog songs. In the experimental part, 10 subjects were selected to sing a Chinese song, an English song, a German song, and a pure music through a man-machine comparison experiment to verify the melody model using deep learning technology and the effectiveness of singing and training smart scoring models [11].

2. Intelligent Scoring of the Melody Model Singing Training Based on Deep Learning

Model singing mainly refers to imitating the rhythm or music sung. If people want to intelligently analyze the training of melody model singing, they must first extract the features of the model singing songs and then establish an intelligent scoring model for melody model singing based on deep learning. The task of the scoring model is to match the songs sung by the students with the standard songs, calculate the similarity between them, and then output the corresponding scores to achieve automatic scoring [12]. The basic process of melody model singing training scoring is shown in Figure 1.

2.1. Feature Extraction. For the two songs that need to be matched, feature extraction is to extract some salient features from them. The song melody features mainly include four feature vectors of pitch, length, speed, and dynamics [13, 14].

2.1.1. Pitch. The physical characteristic that corresponds to the pitch is the frequency of vibration. The higher the vibration frequency, the higher the sound and vice versa. F is defined as a function for evaluating the pitch attribute of a song. Assuming that there are $m\,(m \geq 1)$ tracks in the current song, the pitch features of each track are extracted for correlation extraction, and the pitch feature vector x_f is determined according to the relevant attributes of the main track. Setting the pitch feature value of the song is to match the song to $p_i\,(x_f)$ and $i = [1, 2, ..., n]$, then:

$$f\left(x_f\right) = Max\left(p_i\right). \tag{1}$$

2.1.2. Sound Length. The sound length represents the duration of the note. In the identification of the pitch length, the pitch attribute is mainly determined according to the length of its duration. In order to measure the sound length characteristics of the song, the value of the sound length evaluation function g for:

$$g\left(x_g\right) = \begin{cases} long, x_g \geq x_{switch}, \\ short, x_g < x_{switch}. \end{cases} \tag{2}$$

According to formula (2), the length and duration of each note can be determined, so that the properties of each note length of the whole song can be determined accordingly.

2.1.3. Speed. Tempo refers to the speed of the music beat. The beat characteristics of each piece are certain, and the tempo characteristics can be directly obtained.

2.1.4. Strength. The information that reflects the strength of the note is mainly contained in the data bytes in the note-off and note-on events. The strength is reflected in the numerical range of 0–127.

In this paper, the Mel spectrum is selected as the feature of the spectrum, and the extraction process of the Mel spectrum is shown in Figure 2 [15].

First, it is necessary to perform a short-time Fourier transform on the sound signal of the music. Then, the Mel scale is used to transform the frequency on the amplitude spectrum. Then, the amplitude is converted by the Mel filter and the result is the Mel spectrum representation of each frame; then, the corresponding Mel spectrum is obtained by

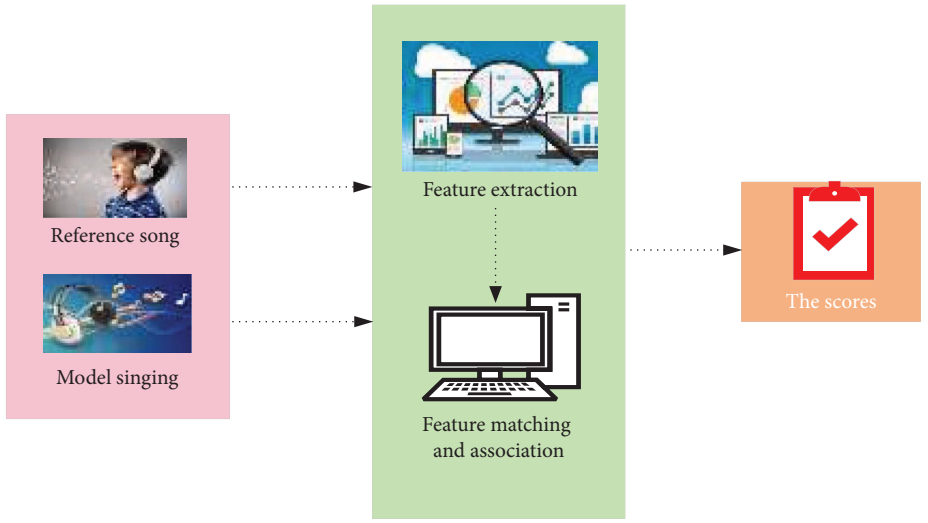

Figure 1: Melody singing training basic process of scoring.

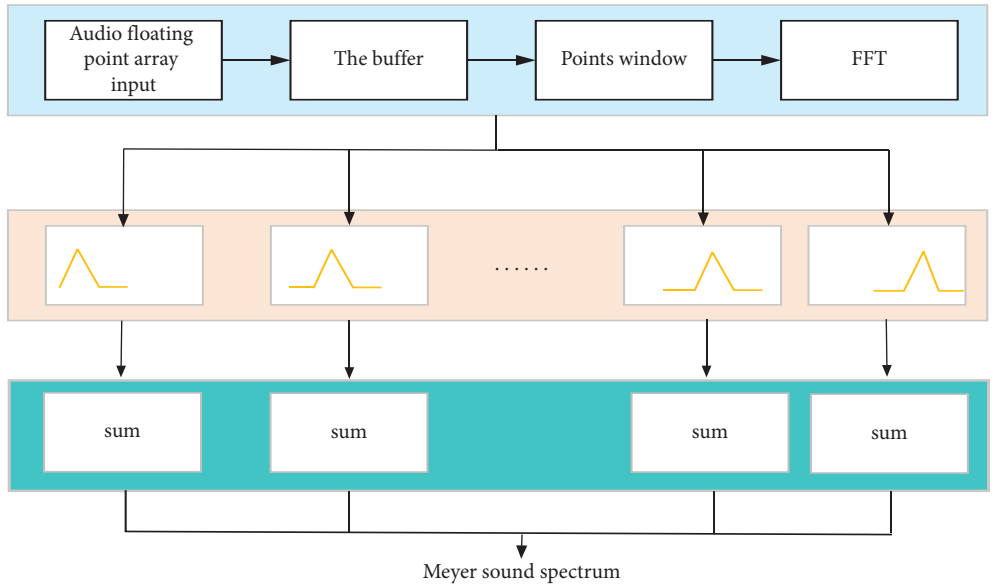

Figure 2: Song feature extraction process.

piecing together the spectrum within the analysis window length.

2.2. Scoring Model. Currently, the most widely used deep learning model is the convolutional neural network (CNN) [16]. A convolutional neural network is a special kind of the multilayer perceptron or feedforward neural network. The feature of local connection and weight sharing, in which a large number of neurons are organized in a certain way and respond to overlapping areas in the field of vision. A standard convolutional neural network generally consists of an input layer, alternating convolutional layers (also called detection layers) and pooling layers (also called down-sampling layers), a fully connected layer, and an output layer. The convolution kernel generally needs to be trained, but it can sometimes be fixed [17]. The convolutional neural network model is shown in Figure 3.

This paper attempted to use a convolutional neural network to intelligently score students' melody model singing training. The scoring model first collected a large number of song features, trained the existing feature data through a neural network, and built a melody model singing intelligent scoring model based on the convolutional neural network. The student model songs and standard songs were taken as two inputs and sent to the same neural network structure for processing, and two corresponding high-level representations were obtained. These two high-level representations were calculated by the similarity calculation of the upper layer of structured network, and the obtained similarity features were adjusted by a fully connected layer, so as to obtain the predicted score of the analog song in the actual

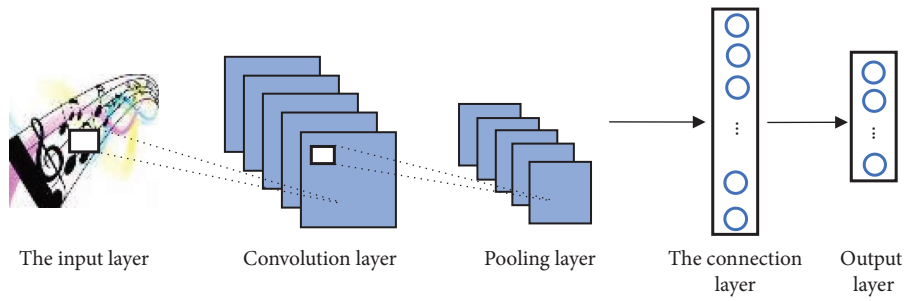

Figure 3: Convolution neural network model.

situation and output it. As the score of students' melody model singing, automatic scoring was realized [18].

2.2.1. Convolutional Layer. In the convolution layer, there are usually multiple learnable convolution kernels. The output feature graph of the previous layer can be obtained by convolution operation with the convolution kernel, that is, dot product between the input item and the convolution kernel. Each output feature map may be a combination of values convolved with multiple input feature maps [19]. The calculation of the output value a_j^l of the jth unit of the convolutional layer L is as follows:

$$a_j^l = f\left(b_j^l + \sum_{i \epsilon M_j^l} a_j^{l-1} * k_{ij}^l\right). \tag{3}$$

2.2.2. Pooling Layer. Pooling layers usually appear after convolutional layers, the two alternate with each other, and each convolutional layer has a one-to-one correspondence with a pooling layer [20]. The calculation formula of the activation value a_j^l in the pooling layer L is as follows:

$$a_j^l = f\left(b_j^l + \beta_j^l down\left(a_j^{l-1}, M^l\right)\right). \tag{4}$$

Among them, down (.) indicates the pooling function, b_j^l is the bias, β_j^l is the multiplier residual, and M^l indicates that the size of the pooling box used in the first layer is $M^l * M^l$.

2.2.3. Fully Connected Layer. The raw output of layer L is as follows:

$$z^{(l)} = \omega^{(l)} \bullet a^{(l-1)} + b^{(l)}. \tag{5}$$

The output of this layer after activation by function f is as follows:

$$a^{(l-1)} = f\left(z^{(l)}\right). \tag{6}$$

The original output error δ of the $L-1$ layer can be inversely obtained from the output error of the L layer as follows:

$$\delta^{(l-1)} = \left(\left(\omega^{(l)}\right)^T \delta_{fc}^{(l)}\right) \bullet f'\left(z^{(l-1)}\right). \tag{7}$$

The gradient of the parameter can be calculated from the pre-activation output error of this layer and the post-activation output of the previous layer:

$$\begin{aligned} \nabla_{W(l)} Loss(W, b; x, y) &= \delta^{(l)}\left(a^{(l-1)}\right)^T, \\ \nabla_{b(l)} Loss(W, b; x, y) &= \delta^{(l)}. \end{aligned} \tag{8}$$

2.2.4. Pooling Layer Backpropagation. For the max-pooling layer (max-pooling), due to the pooling process, only the max value element is sampled in each group of data blocks (block) of the output matrix, so the gradient of each data block, the max value element is 1, and the other elements do not participate in error propagation and the gradient is 0. In order to allow the error δ of the maximum pooling layer to be correctly transmitted back, during max-pooling processing, the position index of the max value of each data block at each depth of the input data needs to be saved after sampling and reused directly during backpropagation [21]. The backpropagation of the error by the pooling layer can be expressed as follows:

$$\delta^{(l-1)} = upsample\left(\delta^{(l)} \bullet f'\left(z^{(l-1)}\right)\right). \tag{9}$$

2.2.5. Convolutional Layer Backpropagation. When strides is 1, on an input channel, the convolutional layer error δ transfer of a filter is expressed as follows:

$$\delta_{i,j}^{(l-1)} = \frac{\partial Loss^{(l)}}{\partial a_{i,j}^{(l-1)}} \frac{\partial a_{i,j}^{(l-1)}}{\partial z_{i,j}^{(l-1)}}. \tag{10}$$

Among them,

$$\frac{\partial Loss^{(l)}}{\partial a_{i,j}^{(l-1)}} = \delta^{(l)} * rot180^o\left(W^{(l)}\right). \tag{11}$$

Expanding to item-by-item accumulation:

$$\frac{\partial Loss^{(l)}}{\partial a_{i,j}^{(l-1)}} = \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} \omega_{r,c}^{(l)} \delta_{i-r,j-c}^{(l)}. \tag{12}$$

The data block elements of δ correspond to the elements of filter W in reverse order and then calculate the dot product.

$$\frac{\partial a_{i,j}^{(l-1)}}{\partial z_{i,j}^{(l-1)}} = f'\left(z_{i,j}^{(l-1)}\right). \tag{13}$$

Putting the two parts together, then

$$\delta_{i,j}^{(l-1)} = \sum_{r=0}^{R-1}\sum_{c=0}^{C-1} \omega_{r,c}^{(l)} \delta_{i-r,j-c}^{(l)} f'\left(z_{i,j}^{(l-1)}\right), \tag{14}$$

or written in a cross-correlation form:

$$\delta^{(l-1)} = \delta^{(l)} * rot180^{o}\left(W^{(l)} \bullet f'\left(z^{(l-1)}\right)\right). \tag{15}$$

After processing formula (15), we obtain as follows:

$$\frac{\partial Loss}{\partial \omega^{(l)}} = a^{(l-1)} * \delta^{(l)}. \tag{16}$$

Expanding to item-by-item accumulation:

$$\frac{\partial Loss}{\partial \omega^{(l)}} = \sum_{r=0}^{R-1}\sum_{c=0}^{C-1} \delta_{r,c}^{(l)} a_{i+r,j+c}^{(l-1)}. \tag{17}$$

Calculating the weight parameter gradient of the filter as follows:

$$\frac{\partial Loss}{\partial \omega^{(l)}} = \frac{\partial Loss}{\partial z^{(l)}} \frac{\partial z^{(l)}}{\partial \omega^{(l)}}. \tag{18}$$

Calculating the bias gradient of the filter:

There are convolution kernels with D filters and there are also D bias item gradients. The bias item gradient of this convolution kernel in order d is the sum of the elements of the network layer error tensor δ on the component matrix of d:

$$\frac{\partial Loss^{(l)}}{\partial b_d^{(l)}} = \frac{\partial Loss^{(l)}}{\partial z_d^{(l)}} \frac{\partial z_d^{(l)}}{\partial b_d^{(l)}} = \sum_i \sum_j \delta_{d,i,j}^{(l)}. \tag{19}$$

3. Intelligent Scoring Experiment of Melody Model Singing Training

A man-machine comparison experiment was carried out to test the scoring effect of the melody model singing intelligent scoring system after using the deep learning technology [22]. 10 subjects were selected to sing a Chinese song, an English song, a German song, and pure music. The human graders were professional music teachers. The machine grading system included the automatic grading model established in this paper and two other automatic grading models [23]. The automatic grading model established in this paper was recorded as Model 1, and the other two automatic grading models were recorded as Model 2, 3. To set the full score of 10 points, the subjects' model singing results were scored; then, the difference between the human-computer scores was calculated, and the data were analyzed.

3.1. Chinese Songs. Manual scoring and three scoring models were used to score the Chinese song model singing results of the subjects, and data analysis was performed on the scoring results, as shown in Figure 4.

Because the scoring algorithms of the 3 scoring systems were different, the scoring results were different. From the scoring results, it could be seen that the scoring results of Model 1 and Model 3 were lower than those of the manual raters, and the scoring results of Model 2 were higher than those of the manual raters. In the manual scoring results, 8 people were higher than the passing line, with an average score of 7.4 points, and the pass rate was 80%. The scoring results of Model 1 were 0.4–0.8 points lower than the manual scoring results, and 6 people were higher than the passing line, with an average score of 6.79 points, and the passing rate was 60%. The scoring results of Model 2 are 0.7–1.4 points higher than the manual scoring results, with an average score of 8.46 points. 10 people were above the pass line, and the pass rate was 100%. The scoring results of Model 3 were 0.9–1.3 points lower than the manual scoring results and 6 people were higher than the passing line, with an average score of 6.25, and the passing rate was 60%. Among them, the average disparity between the Model 1 marking consequence and the factitious marking consequence was 0.61 points, the average disparity between the Model 2 marking consequence and the factitious marking consequence was 1 point, and the disparity difference between the Model 3 marking consequence and the factitious marking consequence was 1.15 points. The comparison of the results showed that the automatic scoring model established in this paper was more accurate in the scoring results of Chinese songs.

3.2. English Songs. It is necessary to use manual scoring and three scoring models to score the English song model singing results of the subjects and to perform data analysis on the scoring results, as shown in Figure 5.

There were 6 people who are above the pass line in the manual scoring results, with an average score of 6.62 points, and a pass rate of 60%. The scoring results of Model 1 were 0.5–0.8 points lower than the manual scoring results, and 6 people were higher than the passing line, with an average score of 6 points, and the passing rate was 60%. The scoring results of Model 2 were 0.9–1.5 points higher than the manual scoring results, and 10 people were higher than the passing line, with an average score of 7.8 points, and the passing rate was 100%. The scoring results of Model 3 were 0.8–1.2 points lower than the manual scoring results, and 4 people were higher than the passing line, with an average score of 5.58, and the passing rate was 40%. Among them, the average disparity between the Model 1 marking consequence and the factitious marking consequence was 0.62 points, the average disparity between the Model 2 marking consequence and the factitious marking consequence was 1.18 point, and the disparity difference between the Model 3 marking consequence and the factitious marking consequence was 1.04 points. The comparison of the results showed that the automatic scoring model established in this paper was more accurate in the scoring results of English songs.

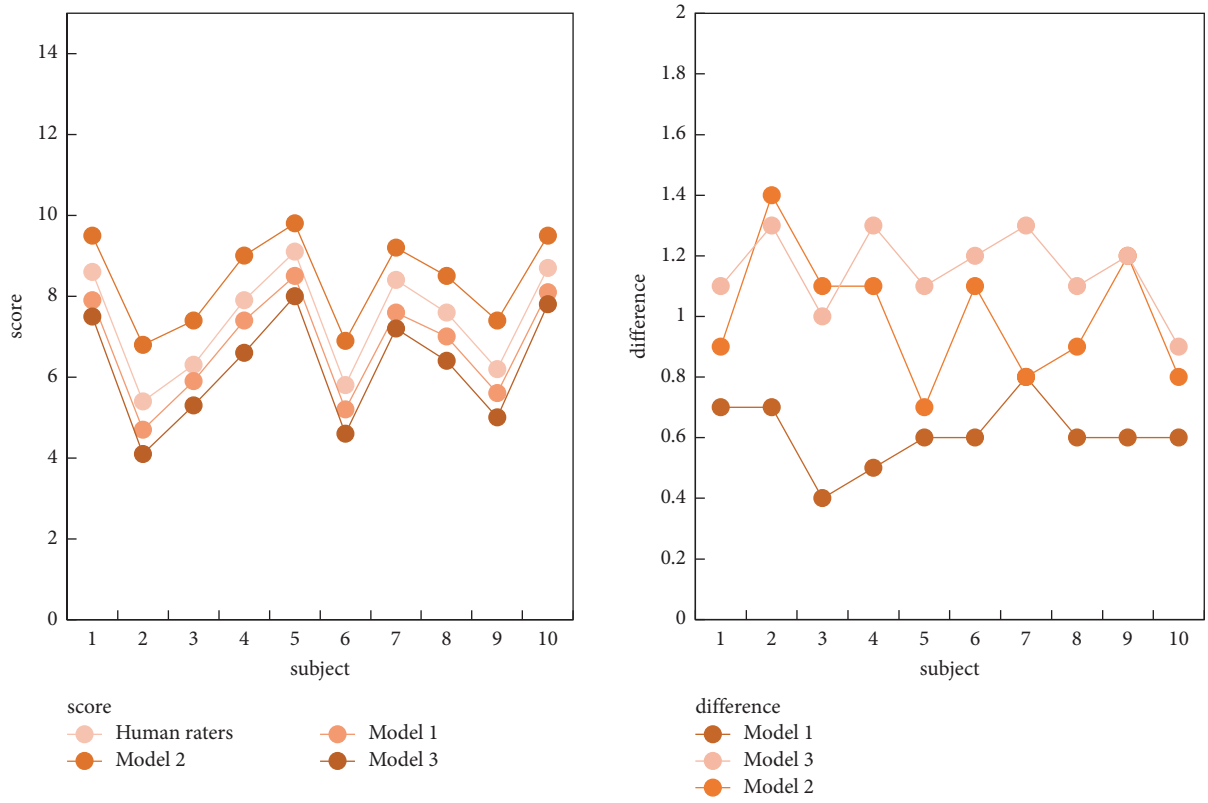

Figure 4: The result of the Chinese song singing rating.

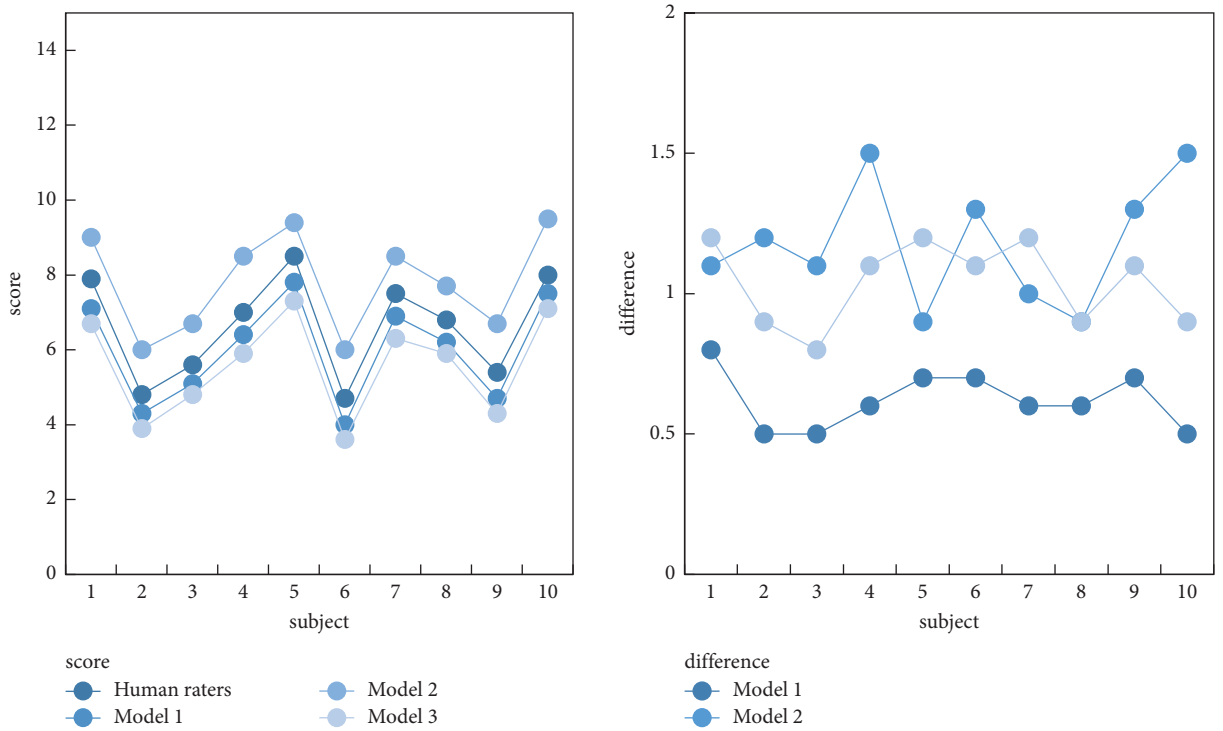

Figure 5: The result of the English song singing rating.

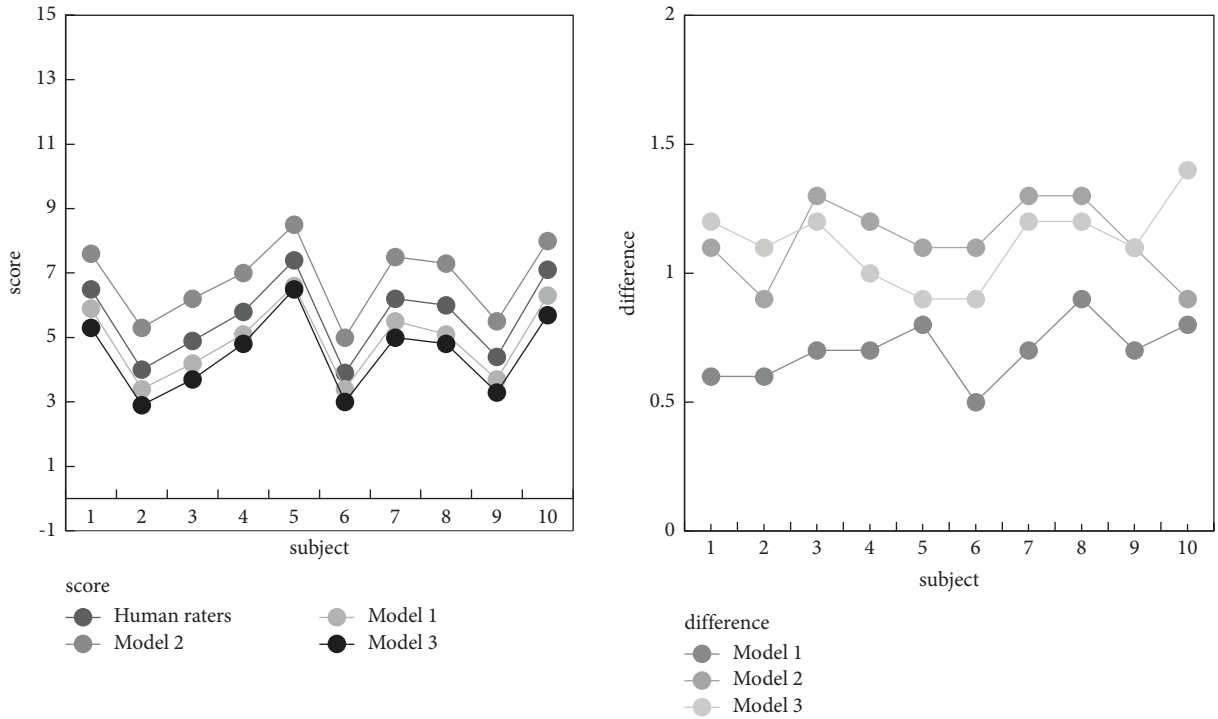

Figure 6: The result of the German song singing rating.

3.3. German Songs. Manual scoring and three scoring models were used to score the German song model singing results of the subjects, and data analysis was performed on the scoring results, as shown in Figure 6.

In the manual scoring results, 5 people were higher than the passing line, with an average score of 5.62, and the passing rate was 50%. The scoring results of Model 1 were 0.5–0.9 points lower than the manual scoring results, and 2 people were higher than the passing line, with an average score of 4.92 points, and the passing rate was 20%. The scoring results of Model 2 were 0.9–1.3 points higher than the manual scoring results, and 7 people were higher than the passing line, with an average score of 6.79 points, and the passing rate was 70%. The scoring results of Model 3 were 0.9–1.4 points lower than the manual scoring results, and 1 person was higher than the passing line, with an average score of 4.5 points, and the passing rate was 10%. Among them, the average disparity between the model 1 marking consequence and the factitious marking consequence was 0.7 points, the average disparity between the model 2 marking consequence and the factitious marking consequence was 1.13 point, and the disparity difference between the model 3 marking consequence and the factitious marking consequence was 1.12 points. The comparison of the results showed that the automatic scoring model established in this paper was more accurate in the scoring results of German songs.

3.4. Pure Music. It is necessary to use manual scoring and three scoring models to score the pure music song model singing results of the subjects and perform data analysis on the scoring results, as shown in Figure 7.

In the manual scoring results, 8 people were higher than the passing line, with an average score of 7.22, and the passing rate was 80%. The scoring results of Model 1 were 0.4–0.8 points lower than the manual scoring results, and 6 people were higher than the passing line, with an average score of 6.64 points, and the passing rate was 60%. The scoring results of Model 2 were 0.9–1.4 points higher than the manual scoring results, and 10 people were higher than the passing line, with an average score of 8.4 points, and the passing rate was 100%. The scoring results of Model 3 were 0.8–1.3 points lower than the manual scoring results, and 6 people were higher than the passing line, with an average score of 6.19 points, and the passing rate was 60%. Among them, the average disparity between the Model 1 marking consequence and the factitious marking consequence was 0.58 points, the average disparity between the Model 2 marking consequence and the factitious marking consequence was 1.18 point, and the disparity difference between the model 3 marking consequence and the factitious marking consequence was 1.03 points. The comparison of the results showed that the automatic scoring model established in this paper had more accurate scoring results in pure music.

3.5. Comprehensive Analysis. The data of the four evaluation results are summarized, and the intelligent scoring effect of the three systems is comprehensively analyzed. The data are shown in Table 1.

The average difference between the evaluation results of the four songs of Model 1 and the manual evaluation results was 0.63 points, and the accuracy rate was 90.66%. The accuracy rate of this scoring model was relatively high. The

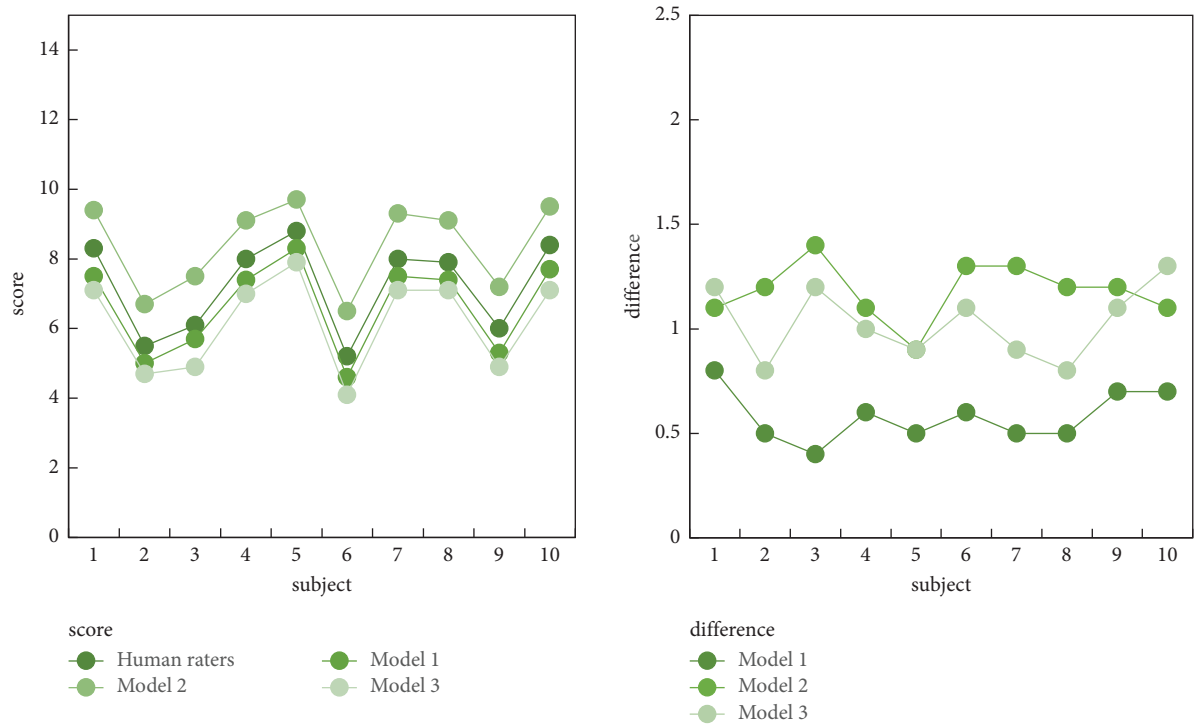

Figure 7: The result of the pure music song rating.

Table 1: Summary of evaluation system data.

Model	The average difference	Accuracy (%)
Model 1	0.63	90.66
Model 2	1.12	82.9
Model 3	1.09	83.84

average difference between the evaluation results of the four songs of Model 2 and the manual evaluation results was 1.12 points, and the accuracy rate was 82.9%. The average difference between the evaluation results of the four songs of Model 3 and the manual evaluation results was 1.09 points, and the accuracy rate was 83.84%. Comparing the data, it could be seen that compared with Model 2 and Model 3, the melody model singing intelligent scoring result of Model 1 had a higher accuracy and was closer to the real scoring result. Compared with the original melody model singing intelligent scoring system, the accuracy of the automatic scoring model established in this paper was increased by 6.82%, indicating that the research had certain feasibility.

4. Experimental Discussion

This paper used deep learning technology to study the intelligent scoring modeling method of solfeggio training in ear melody model singing training. The research results were as follows:

(1) It was necessary to extract the features of the melody model singing according to the Mel sound spectrum and then build a melody model singing intelligent scoring model based on the convolutional neural network. It could predict the score of the analog song and output it, realize automatic scoring, and enhance the scoring effect.

(2) A man-machine comparison experiment was carried out. The human graders are professional music teachers, and the machine grading system includes the automatic grading model established in this article and the other two automatic grading models to test the scoring effect of the melody-model singing intelligent scoring system after using the deep learning technology. The experimental results showed that the accuracy of the automatic scoring model established in this paper was 90.66%, and the accuracy of the other two automatic scoring models was 82.9% and 83.84% when evaluating the results of Chinese songs, English songs, German songs, and pure music, indicating that the automatic scoring models established in this paper had a higher accuracy, and the accuracy was increased by 6.82%, which was closer to the manual scoring results.

5. Conclusion

Based on deep learning technology, this paper constructs an intelligent scoring model for melody model singing training. First, the Mel map is used as the sound spectrum feature to extract song features; then, a large number of song features are collected, and the feature data are trained to build a scoring model. By comparing the feature similarity between the model song and the standard song, the score of the model song can be obtained. The man-machine comparison

experiment shows that the melody model singing training scoring model established by deep learning technology has a higher accuracy. Compared with the original melody model singing intelligent scoring system, the accuracy of the automatic scoring model established in this paper is increased by 6.82%. It shows that the research study has certain feasibility.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by Education Department of Shaanxi Provincial Government (Program No.2019JK0021).

References

[1] M. Debevc, J. Weiss, A. Šorgo, and I. Kozuh, "Solfeggio learning and the influence of a mobile application based on visual, auditory and tactile modalities," *British Journal of Educational Technology*, vol. 51, no. 1, pp. 177–193, 2020.

[2] L. Maofang, "The construction of dynamic teaching mode in solfeggio teaching in normal Universities," *Journal of Wuyi University*, vol. 26, no. 3, pp. 89–94, 2019.

[3] K. Aycan, "Using bona adaptation to improve accent defects as a voice training method," *European Journal of Educational Research*, vol. 17, no. 69, pp. 113–134, 2017.

[4] Q. Ding, "Research on SIFT-based audio processing technology in music teaching//CIPAE 2021: 2021 2nd international conference on computers," *Information Processing and Advanced Education*, vol. 52, no. 10, pp. 13–17, 2021.

[5] L. Ying, "Regular application of hidden melody multi-voice solfeggio in teaching practice," *The Guide of Science & Education*, vol. 17, no. 6, pp. 132–137, 2019.

[6] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 5, pp. 60–88, 2017.

[7] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017.

[8] D. S. Kermany, M. Goldbaum, W. Cai et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018.

[9] A. Rajkomar, E. Oren, K. Chen et al., "Scalable and accurate deep learning with electronic health records," *Npj Digital Medicine*, vol. 1, no. 1, pp. 18–21, 2018.

[10] A. T. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, pp. 1117–1144, 2017.

[11] Y. Li, Y. Zuo, H. Song, and Z. Lv, "Deep learning in security of internet of things," *IEEE Internet of Things Journal*, vol. 2021, no. 99, Article ID 3106898, 2021.

[12] E. A. Petersen, T. Colinot, F. Silva, and V. Turcotte, "The bassoon tonehole lattice: links between the open and closed holes and the radiated sound spectrum," *Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 398–409, 2021.

[13] S. Amiriparian, N. Cummins, and S. Ottl, "Sentiment analysis using image-based deep spectrum features//international conference on affective computing & intelligent interaction workshops & demos," *IEEE Computer Society*, vol. 2017, no. 8, 29 pages, 2017.

[14] Q. Fu, D. Lv, Y. Zhang et al., "Research on crane sound clustering of MFCC based on HHT," *Journal of Physics: Conference Series*, vol. 1693, no. 1, pp. 012134–13124, 2020.

[15] Y. K. Lin, M. C. Su, and Y. Z. Hsieh, "The application and improvement of deep neural networks in environmental sound recognition," *Applied Sciences*, vol. 10, no. 17, pp. 5965–5969, 2020.

[16] L. Wu, Q. Zhang, C. H. Chen, K. Guo, and D. Wang, "Deep learning techniques for community detection in social networks," *IEEE Access*, vol. 8, pp. 96016–96026, May 2020.

[17] C. Hu, Z. Yi, and M. K. Kalra, "Low-dose CT with a residual encoder-decoder convolutional neural network (RED-CNN)," *IEEE Transactions on Medical Imaging*, vol. 36, no. 99, pp. 2524–2535, 2017.

[18] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information Sciences*, vol. 405, no. 11, pp. 81–90, 2017.

[19] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, pp. 818–824, 2017.

[20] C. Yuan, X. Li, and Q. Wu, "Fingerprint liveness detection from different fingerprint materials using convolutional neural network and principal component analysis," *Computers, Materials & Continua*, vol. 53, no. 4, pp. 357–371, 2017.

[21] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, no. 11, pp. 120–147, 2018.

[22] Z. Lv, A. K. Singh, and J. Li, "Deep learning for security problems in 5G heterogeneous networks," *IEEE Network*, vol. 35, no. 2, pp. 67–73, 2021.

[23] R. Hou, C. Chen, and M. Shah, "Tube convolutional neural network (T-CNN) for action detection in videos," *IEEE Computer Society*, vol. 2017, no. 6, 5832 pages, 2017.