Hindawi
Mobile Information Systems
Volume 2022, Article ID 1875815, 8 pages
https://doi.org/10.1155/2022/1875815

Research Article

Music Sense Analysis of Bel Canto Audio and Bel Canto Teaching Based on LSTM Mixed Model

Zhangcheng Tang

Hunan First Normal University, Changsha 410205, China

Correspondence should be addressed to Zhangcheng Tang; tzc8573@hnfnu.edu.cn

Received 21 April 2022; Revised 23 May 2022; Accepted 28 May 2022; Published 22 June 2022

Academic Editor: Le Sun

As we all know, as the soul of music artists, the cultivation of music sense is an indispensable and important part of Bel Canto teaching. Traditional music classroom education lags behind the development of the information age. According to the educational method of Bel Canto teaching, the recognition experiment of Bel Canto audio is carried out, which helps students analyze the content of music sense contained in music works, and teachers cultivate students' music sense and music theory knowledge according to effective results. In order to better let students appreciate the true meaning of music, it is necessary to add online tools to assist Bel Canto teaching. Traditional methods neither teach students in accordance with their aptitude from the actual situation, but use the rapidly developing computer technology to match resources, nor does it seriously cultivate students' ability to appreciate music and perceive emotions. Based on the above problems, this paper starts from the field of deep learning and plans to build a hybrid model related to LSTM. The results of this paper are as follows: (1) The CNN-LSTM model has the highest recognition rate curve, and the recognition rate of some emotions is over 90%; the loss rate tends to be stable at 200 iterations, and the convergence speed is rapid. (2) After preprocessing, the emotion recognition rate is higher, and the average accuracy of audio features extracted based on spectrogram + LLDs in emotion is about 0.7. (3) According to the actual scene application, the best effect of music sense cultivation is to use the model to assist classroom teaching, and the highest score can reach 8.8 points. In addition, the error between the emotional expression identified by the model and the original work is between 0 and 0.5 points, and the emotional expression effect is excellent. (4) The model can also recognize different kinds and times of emotion in 5-minute Bel Canto works. The above experimental results show that the model basically meets the requirements of the subject, and its performance is excellent, but the details need to be optimized.

1. Introduction

The rapid development of the Internet and the rapid innovation of technology promote the change in music education and open up new ideas for music teaching. In the past, Bel Canto teaching was mainly a traditional "face-to-face" classroom, which required teachers to teach students various skills without relying on any tools. This leads to great pressure on teachers, which easily neglects the cultivation of students' sense of music and wastes abundant teaching resources. Based on the wide application and excellent performance of deep learning in the audio field, this paper chooses to study multimodal data fusion and constructs a model with hybrid convolution neural network and cyclic neural network. On the one hand, students can correctly perceive the scene and emotion of music through professional and detailed model analysis. On the other hand, the model encourages students to sort out their self-awareness and can carry out a lot of training according to their own situation to cultivate their musical sense ability. After reducing teaching tasks, teachers can adjust the teaching content of Bel Canto according to the results of model test and find the correct mode of training different students. Combined with the existing Bel Canto teaching system, in order to get more reference ideas and improvement methods, we refer to a large number of relevant literature and materials for research. This paper provides reference methods and ideas for the following researchers as shown below.

By improving the traditional music teaching of skill training, it explores the cultivation of students' sense of

music and appreciation teaching [1]. Influenced by COVID-19, the teaching mode changed, and online music education was adopted to start the vocal music teaching mode [2]. Vocal music teaching in higher music education needs diversification and innovation and realizes teaching in the new period by using modern information technology [3]. Based on the open and cooperative teaching mode, this paper studies the cultivation of music sense in vocal music teaching and music education in colleges and universities [4]. According to the music teaching under the new situation, students' comprehensive music quality and sense of music are cultivated through music learning, and the basic music theory knowledge is increased [5]. This paper analyzes the importance of cultivating students' sense of music in music education and explores effective ways to enhance students' sense of music [6]. The understanding of music, the mastery of Bel Canto technology and the second creation of works, discusses the singing art of Bel Canto [7]. According to the popular trend of music, a prediction model combining long-term memory network with attention mechanism is designed [8]. LSTM and attention mechanism are integrated to classify music emotionally, which solves the problem that it is difficult to find preferred music [9]. According to the feature extraction of song audio emotion classification, CNN-LSTM model is constructed to identify music sense [10]. Based on the classification method of music content, music genres are classified by using long-term and short-term memory network [11]. Combining key object recognition and deep self-attention, Bi-LSTM model is constructed for emotion classification [12]. Introducing the penalty term, a stacked LSTM model embedded with self-attention mechanism is proposed for audio and video emotion recognition [13]. Using the LSTM model, a new multimodal fusion music emotion classification method based on audio and lyrics is proposed [14]. This paper analyzes the significance of music literacy to Bel Canto in the performance of vocal singers [15].

2. Theoretical Basis

2.1. Bel Canto Teaching Method. Bel Canto [16]: its literal meaning can be interpreted as "beautiful singing." It is a soft singing method originated in Europe in the 17th century, and it has influenced music singing all over the world in its continuous development and dissemination. The audience is usually infected by the singer's emotions and feelings and resonates, thus realizing the artistry of music works. Different from other singing methods, Bel Canto adopts the laryngeal low voice method, using true and falsetto resonance and mixed sound area skills. This also leads to Bel Canto that is very dependent on teachers' teaching and needs to pay attention to teaching skills and style so that students can learn the essence of singing correctly.

In Bel Canto teaching, we can divide the contents into two categories. First, it is the teaching of vocal music skills. Teachers need to teach students to sing with the correct way of breathing and ensure the circulation of breath in the body through scientific training. Students should also learn the skills of vocalization, adjust the timbre according to their own voice and timbre characteristics, and cultivate the ability of listening and distinguishing through a lot of ear training. The second point belongs to nontechnical vocal music teaching. In view of the emotional characteristics contained in Bel Canto works, teachers teach students to inherit innovative music styles, let students control and express subtle emotional changes, and improve the stage singing effect.

2.2. Musical Sense. Music sense [17]: it is a manifestation of music appreciation level and music accomplishment. In essence, music sense is the ability to perceive music emotion. For many people, it is not difficult to operate and play musical instruments mechanically or use skills to perform songs. However, model training cannot make people really and deeply feel the charm and emotion of music. If we want to resonate emotionally to generate and realize the fun of learning music, we need to pay attention to the cultivation of music sense. Therefore, compared with the traditional classroom, the original education method only aiming at skills is no longer applicable, and teachers need to adjust their educational policies and concepts and make changes in teaching programs. How to cultivate students' sense of music and make students deeply understand the emotional meaning behind music has become a new topic for teachers to think about. Similar to the cultivation of music sense is language sense, both of which need a lot of time and excellent music works as the foundation. Through long-term and persistent training, students can perceive the connotation of music expression. Teachers can cultivate students' music perception ability from the aspects of tone, style, timbre, range, strength change, emotion, and rhythm.

2.3. Audio Segmentation Method. Audio segmentation [18]: in this paper, using the vertical axis of audio, cutting Bel Canto audio works can assist the model to analyze music from different frequency bands. Segmentation rules:

$$\text{note} \in \begin{cases} \text{bass} & \text{pitch} < 46 \\ \text{middle} & 46 \le \text{pitch} < 70 \\ \text{treble} & \text{pitch} \ge 70 \end{cases}. \tag{1}$$

Usually, the audio level is different, so we can use this feature to segment the music. Generally speaking, music has two sound regions: high school and low school, and three groups of note sequences are obtained. Therefore, we will divide it into three segments of audio, which are low frequency, intermediate frequency, and high frequency. Then, using the statistical calculation method, three audio sequences are operated. The extracted note sequence information is as follows:

Average pitch:

$$\text{Range} = \frac{1}{n}\sum_{i=1}^{n} \text{pitch}_i. \tag{2}$$

Average sound intensity [19]:

$$\overline{\text{Intensity}} = \frac{1}{n}\sum_{i=1}^{n} \text{Intensity}_i. \tag{3}$$

Pitch trend:

$$DirPitch = \frac{\sum_{i=1}^{n} Interval_i * duration_i}{duration_i - duration_n}. \quad (4)$$

Phonetic standard aberration:

$$StaPitch = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(pitch_i - Range)^2}. \quad (5)$$

Intervals [20]:

$$Intercal_i = pitch_{i+1} - pitch_i. \quad (6)$$

Interval standard deviation:

$$StaInterval = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(|Interval_i| - |\overline{Intercal}|\right)^2}. \quad (7)$$

Mean value of interval absolute value:

$$SpanInterval = \frac{1}{n-1}\sum_{i=1}^{n}|Interval_i|. \quad (8)$$

2.4. Feature Classification Method. "Deep Learning" [21]: this concept is put forward mainly to study various performance achievements of artificial neural networks in the field of machine learning. This method is suitable for data fields such as voice, image, text, and video and has made great achievements in artificial intelligence research in recent years.

Method 1: convolution neural network. It can superimpose three layers of networks, namely, convolution layer, pooling layer, and full connection layer, to increase the depth and width of the network.

$$Y_k = f(W_k * X). \quad (9)$$

For formula (9), after the feature map passes through the kth convolution kernel, Y_k represents the kth output feature map. X represents the input feature, and W_k represents the kth convolution kernel. The symbol "$*$" represents a two-dimensional convolution operator.

Method 2: cyclic neural network. LSTM [22]: it is called "long short-term memory." Because of the special network structure of this method, it is convenient to deal with serialized feature data. LSTM is extremely suitable for serialized information such as text and audio. In addition, because LSTM has a special gate structure, compared with the traditional basic RNN, LSTM can solve the problem of long-term dependence.

$$s_t = f(Ux_t + ws_{t-1}), \quad (10)$$

$$o_t = \text{soft}\max(Vs_t), \quad (11)$$

wherein the input vector of the tth time node is x_t; the hidden state of the node is s_t. And the function f is a nonlinear function. U and w represent the weights of the input layer and the hidden layer. It is worth noting that s_o is usually initialized to 0, and V represents the weight of the hidden layer versus the output layer.

Method 3: support vector machine. Its abbreviation is "SVM," which was put forward in 1996. This method can deal with binary classification problem without the limitation of sample number.

(1) Linearly separable SVM:

$$w \cdot x + b = 0, \quad (12)$$

$$f(x) = \text{sign}(w \cdot x + b). \quad (13)$$

Calculate the distance of support vector:

$$\text{margin} = \frac{2}{\|w\|}. \quad (14)$$

Constrained optimization problem:

$$\min_{w,b} \frac{1}{2}\|w\|^2, \quad (15)$$

$$s.t. \quad y_i(w \cdot x_i + b) - 1 \geq 0, \quad i = 1, 2, \ldots, N. \quad (16)$$

(2) Linearly indivisible SVM:
Change constraints:

$$s.t. \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, N. \quad (17)$$

Corresponding questions:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i. \quad (18)$$

(3) Nonlinearity [23]:

$$K(x, z) = \phi(x) \cdot \phi(z), \quad (19)$$

where w and b represent normal vector and load distance, respectively. The greater the margin value, the higher the possibility of correct classification. By introducing the Lagrange operator, the original optimization problem can be simplified and the dual problem can be calculated. ξ_i stands for the relaxation variable and C for the penalty function. ϕ represents spatial transformation; introducing kernel function can solve the problem of increasing the cost of inner product calculation.

3. LSTM Hybrid Model Based on Music Sense Analysis

3.1. Audio Preprocessing. Before feature extraction of the Bel Canto audio signal, the first step of processing audio emotion data is preprocessing. This is because the collected audio is easily affected by some factors, such as incomplete

and missing data, too much noise in the data, or too much mute to cause interference. Therefore, it is necessary to enhance the audio signal and supplement the missing data before extracting features, discard useless data, standardize data, and remove all audio impurities.

(1) Frame processing by audio segmentation method.

(2) Reduce the influence of data leakage by adding windowing function. According to the signal type and target, choose the appropriate function properly.

$$S_w(n) = s(n) \times w(n). \tag{20}$$

Rectangular window:

$$w(n) = 1. \tag{21}$$

Hanning window:

$$w(n) = 0.5 - 0.5\cos\left(\frac{2\pi n}{N}\right). \tag{22}$$

Hamming window:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right). \tag{23}$$

(3) Endpoint detection can accurately control the sound segment and silent segment and reduce the computational complexity of the model.

(4) Noise reduction: approximate estimation of speech signal:

$$|\widehat{X}(k,l)|^2 = G(k,l)|Y(k,l)|^2. \tag{24}$$

Calculate the gain function:

$$G(k,l) = \frac{\xi(k,l)}{1+\xi(k,l)}\exp\left(\frac{1}{2}\int_{v(k,l)}^{\infty}\frac{e^{-t}}{t}dt\right)\gamma(k,l), \tag{25}$$

$$v(k,l) = \frac{\gamma(k,l)\xi(k,l)}{1+\xi(k,l)}. \tag{26}$$

Calculate each signal-to-noise ratio:

$$\gamma(k,l) = \frac{|Y(k,l)|^2}{\delta_d^2(k,l)}. \tag{27}$$

(5) Feature selection and dimension reduction.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}. \tag{28}$$

(6) Reduce data differences and normalize them.

Min-max method [24]:

$$X_{\text{normalization}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{29}$$

Z-score [25]:

$$x_{\text{normalization}} = \frac{x - \mu}{x_{\max} - \delta}. \tag{30}$$

3.2. Audio Feature Extraction. In the past, the traditional audio emotion data sets were mostly pure music fragments or pure human voice fragments, both of which had short time and single sound composition. Bel Canto audio is mostly stored in the form of digital music, and the time is usually less than 5 minutes, which is longer than the traditional audio test time, and the emotional expression is different in different periods. In addition, Bel Canto audio is not a single component and usually includes two kinds of sounds: music part and human voice part. However, the Bel Canto audio studied in this research is quite different from the audio studied in the past. Therefore, in the process of audio feature extraction, we need to perform fine-grained audio segmentation and voice separation operations. Using these two methods, we can solve the two problems of large feature dimension and complex composition. This is all due to the long singing time of Bel Canto. In addition, the spectrogram can be used to output emotional features. In this paper, some feature parameters are extracted as shown in Table 1.

Among them, HSFs are statistical features based on LLDs. The extraction of MFCC is closely related to LLDs. The spectrum is obtained by FFT, and the final M-dimensional MFCC coefficients are obtained by transformation formula, triangular filter, logarithmic operation, and DCT transform. Besides, some other features can be extracted from audio signals such as zero-crossing rate, spectrum centroid, spectrum bandwidth, spectrum attenuation, spectrum flux, and chromaticity characteristics.

$$\text{Mel}(f) = 25951g\left(1 + \frac{f}{700}\right). \tag{31}$$

3.3. Construction of Model. In order to comprehensively consider the performance of frequency spectrum and time sequence characteristics in the emotional classification of Bel Canto Audio, this paper combines CNN and LSTM neural networks to construct a hybrid emotional classification model for emotional data of Bel Canto Audio. This model absorbs the essence of Bel Canto teaching and can be applied to the analysis of music sense. The network frame structure of the model is clear, and the layout is reasonable as shown in Figure 1.

The processed audio features are used as the network input part of the whole model. Then, the model can be divided into two main parts for a series of operations. The first main part is composed of spectrogram and CNN-LSTM. The first agent can output a set of serialized feature vectors and then add an attention mechanism to the output. Considering that the feature classification ability of a single spectrogram is not enough to meet the experimental requirements, the second main part of the model fuses LLDs features and DNN in the network. LLDs evolved into HSFs by statistical combination, and DNN was used to reduce the dimension. The second agent hopes that with the help of these three components, the classification performance can be improved to make up for the expression of emotional information. Finally, the feature vectors obtained by the two

Table 1: Audio characteristics.

Characteristics of audio frequency	Classification of features
Maximum, mean, variance	Advanced statistical features (HSFs)
MFCC, zero crossing rate, spectral centroid, spectral bandwidth, spectral attenuation, spectral flux, chroma characteristics	Low-level descriptive features (LLDs)

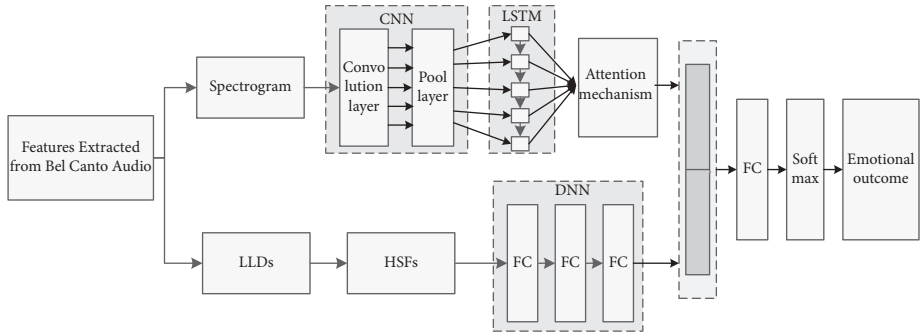

Figure 1: Framework based on CNN-LSTM hybrid model.

agents are vertically spliced, and the audio fusion features are input into the Softmax layer. After classification processing, the emotional classification results are obtained.

4. Experimental Analysis

4.1. Experimental Preparation. The LSTM hybrid model we build needs to be trained and tested on a professional server platform. Therefore, before the simulation experiment, we publicly show the specific detailed parameters of some hardware and software information as shown in Table 2.

4.2. Emotional Data Set. The purpose of our model is to accurately identify emotional states in Bel Canto audio and to help teachers and students develop their sense of music. This is undoubtedly a great success for Bel Canto teaching. In the construction of this data set, the description of emotion needs to be recognized by a wide range of experts and scholars. Therefore, we choose the discrete emotion theory as the foundation and collect the commonly used discrete emotion data sets such as CASIA, EMODB, and IEMOCAP and establish a high-quality, high-resolution, and high-success rate emotion data set suitable for this study. According to many scholars' different definitions of basic emotions, we choose the most widely used emotion classification method. For Bel Canto teaching, cultivating students' musical sense is equivalent to letting students comprehend rich levels of emotion in music. When students can properly perceive and analyze the delicate emotions in different periods of music works, students have the ability of music sense. In order to correctly identify emotional categories, it is necessary to construct emotional data sets of Bel Canto audio. This method, proposed by Ekman, Friesen, and Ellsworth, divides emotion into six concepts. They are happy, surprised, angry, disgusted, afraid, and sad. Our experiment will identify Bel Canto audio clips according to these six emotions.

Table 2: Software and hardware platforms.

Type name	Supreme law
Open source software library	TsorsFlow 1.13.1
CPU	E5-2650V4**2
Memory	16G_DDR4_2400**8
Graphic card model	NVIDIA GTX1080Ti
Operating system	Ubuntu 16.04 LTS
CUDA version	3584
CuDNN version	Giant network
Solid-state drive	480G_SSD_6G**2
Mechanical hard disk	4T_7200_6G**2
Numpy	Handheld intelligence

4.3. Test Model

4.3.1. Model Performance Comparison. In the simulation test of this section, we test the performance of the LSTM hybrid model constructed in this paper.

(1) Recognition Rate of Different Models. Comparing the recognition rate of the four models on six emotions, the recognition rate curve of the mixed CNN-LSTM model has obviously increased. Among them, the highest recognition rate of happiness is 61%, which is the lowest among the six emotions, and the relevant recognition work needs to be strengthened; the recognition rate of surprise emotion is 88%; the recognition rates of anger, sadness, and fear are all over 90%. The recognition rates of LSTM and CNN models are similar, but there is little difference, and the recognition effect of music is not good enough. Although the emotion recognition performance of the 3DACRNN model is not better than that of the CNN-LSTM model, the recognition performance of the 3DACRNN model is slightly improved compared with other single models, LSTM and CNN models, as shown in Figure 2.

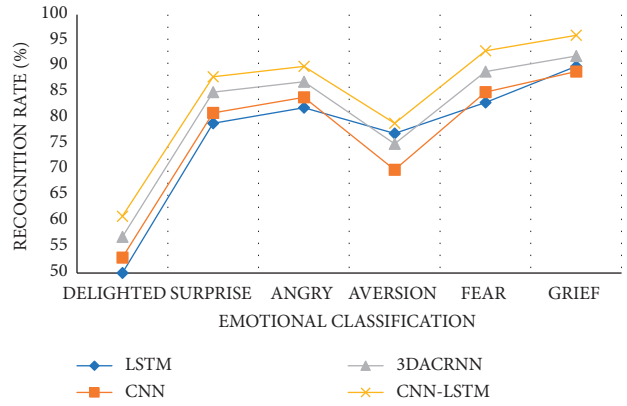

Figure 2: Recognition rate of different models.

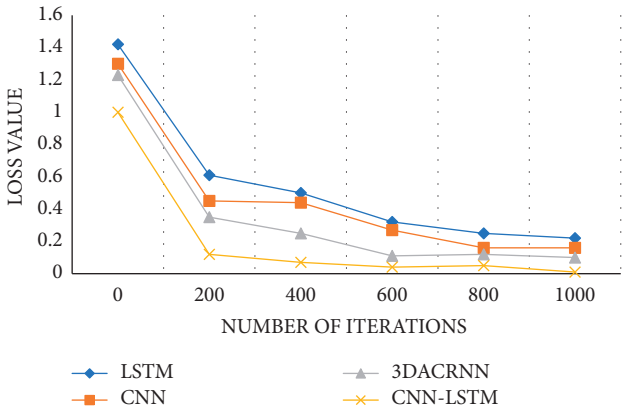

Figure 3: Loss rate of different models.

(2) Comparison Of Loss Rates. The change of loss function of the test model can intuitively reflect the convergence speed of the model. The loss values of the four models are compared under different iteration times. It can be clearly seen that the CNN-LSTM model converges rapidly on the data set and tends to be stable at the 200th iteration, and the loss function value is 0.01 at the 1000th iteration, and the subsequent loss value approaches 0 wirelessly. However, when the 3DACRNN model is close to 600 iterations, the loss function value is close to 0.1. After 1000 iterations of LSTM and CNN, the loss value is close to 0.2 and stops convergence. To sum up, the CNN-LSTM model has the fastest training speed, and its convergence speed is better than the other three models as shown in Figure 3.

4.3.2. Emotion Recognition Experiment. In this experiment, we mainly aim at different preprocessing methods and different audio features to explore the model for music emotional classification recognition accuracy.

(1) Preprocessing Audio Segmentation. A 1-minute Bel Canto audio is divided into three parts: original segment, pure background music, and vocal singing. The proposed method analyzes the classification accuracy of the six emotions in this performance. We can find that the original fragment has not been preprocessed, and the recognition rate of six emotions is low. After pretreatment, the recognition rate of pure background music and vocal singing has obviously increased as shown in Figure 4.

(2) Different Audio Characteristics. Support vector machine (SVM) is used as a classification method to recognize the emotion of a Bel Canto audio for single LLDs, HSFs features, and spectrogram+LLDs audio features proposed in this paper. We can find that a single feature performs poorly in the classification of happy and angry emotions, and the classification accuracy of other emotions can reach more than 0.6. The average accuracy of single LLDs and HSFs is about 0.61. Under this method, the average classification accuracy of six emotions is about 0.7, and the overall value is much higher than the performance of the other two characteristics as shown in Figure 5.

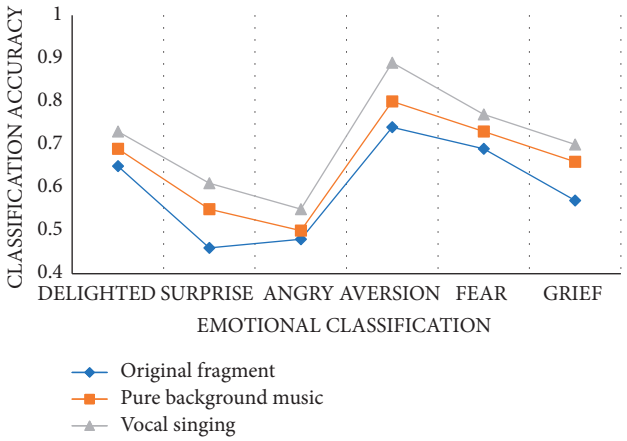

Figure 4: Emotion recognition after audio segmentation.

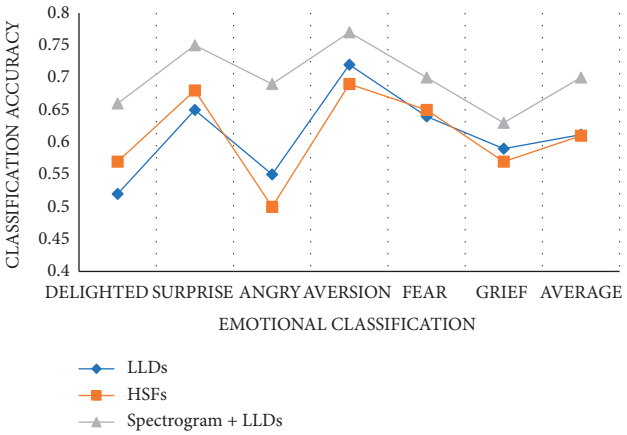

Figure 5: Accuracy of different audio features.

4.3.3. Practical Teaching Application

(1) Mixed Teaching Mode. We compare traditional Bel Canto teaching, pure online teaching, and mixed model assisted teaching. Six study groups were set up, and each study group was divided into six students, who learned a

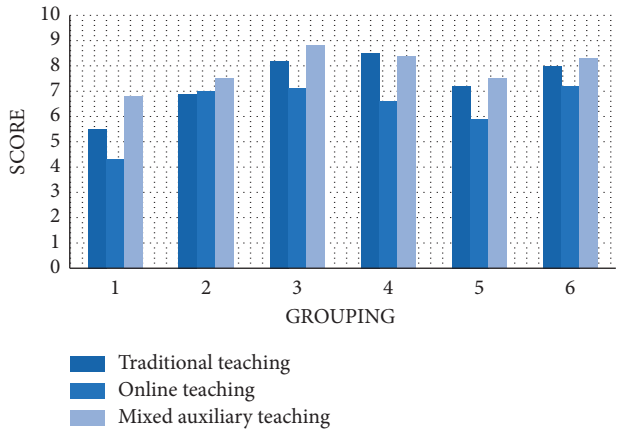

Figure 6: Music sense cultivation in different modes.

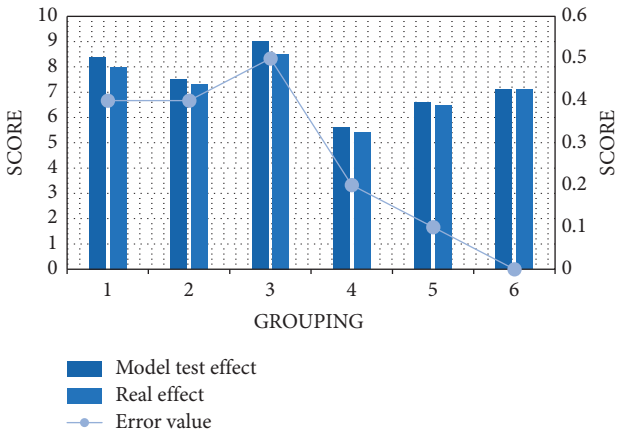

Figure 7: Comparison of real effects.

3-minute Bel Canto audio together. In order to quantify the cultivation of students' sense of music, we use 10 scores to score the results. We can find that the third group scored the highest, reaching 8.8 points, by using the model to assist the cultivation of music sense in classroom teaching. After adding the model, these six groups can achieve the traditional teaching effect, even better. However, the effect of Bel Canto teaching by using the online model alone is the worst, with the lowest score of 4.3 and the highest score of 7.2. This shows that Bel Canto teaching without practical classroom training is not feasible as shown in Figure 6.

(2) Manual Evaluation of Subjective Emotion. There are 6 groups in this test group, which test the musical sense of 6 different Bel Canto works. Three relevant experts were invited to score the real actual effect, and a 10-point scoring method was adopted. It mainly tests the consistency between the model and the real Bel Canto emotion, in order to verify the effect of the model on music analysis. We can see that there is not much difference between the test effect of 6 works and the emotional effect of real works. The overall error values are between 0 and 0.5. This shows that the emotional expression recognized by the model is very consistent with the expression of the original work as shown in Figure 7.

(3) Emotion Recognition in Different Periods. According to a 5-minute Bel Canto audio, the recognition of music sense is carried out, and how many different emotions are recognized in this period of time is obtained. With 1 minute as the limit, they were divided into 5 groups, and the types and times of musical emotions measured in each group were different. According to the recognition situation, we can determine the music style, emotional change state, rhythm speed, and other conditions of this Bel Canto work, which can help teachers and students to learn Bel Canto comprehensively and cultivate music sense better. The first two minutes of this audio are mainly happy and surprised, and then, the atmosphere gradually becomes sad, depressed, and filled with sad emotions with time. Happy and surprised emotions were detected 6 times; anger was detected 3 times;

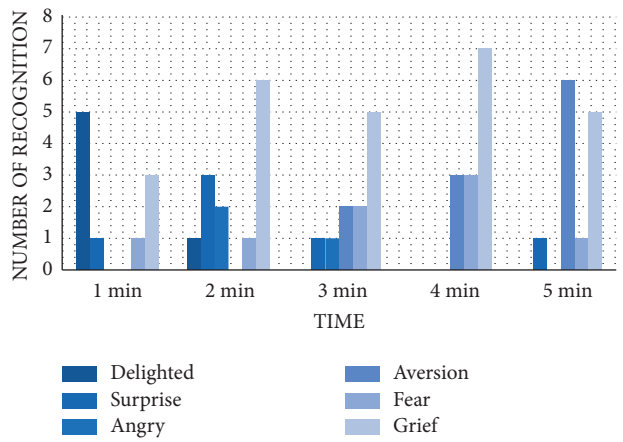

Figure 8: Model recognition performance in different periods.

disgust was detected 11 times; there are 8 times of fear; sadness is the most, up to 26 times as shown in Figure 8.

5. Conclusion

For the music sense analysis of Bel Canto specialty, there are few studies at home and abroad. This paper focuses on the cultivation of musical sense in Bel Canto teaching and discusses the different emotions contained in Bel Canto works. By means of audio feature classification and extraction method, the fusion emotion information mode cuts out fine-grained audio data to extract different features, and classifies the emotion for model output. The experiment avoids the classification defects caused by the direct dimension reduction method and makes good use of the correlation among various modes to obtain the best performance. Both the effect of music emotion recognition and the performance of classification accuracy have been improved to a certain extent. During this period, in order to improve the defects of single feature and single network classification, the paper uses the existing music recognition technology to transform the model based on LSTM and CNN. The research results of this paper show that this model combines the emotional characteristics and music style of Bel Canto and follows the existing results to modify the experiment based on this preference. Our hybrid

model makes up for the defect that it is difficult to extract audio temporal features in a single mode, and the calculation speed is faster.

Although the model constructed in this paper performs well in Bel Canto audio recognition, it proves that the research in this paper is effective. However, for some performance of the model and the application of actual scenarios, there are still some problems that need further research. In the future research process, we can test and explore from the following angles: adjusting the hybrid network model and optimizing the parameters to improve the classification accuracy, expand recognizable emotional categories for music sense, and make emotional classification more subtle, which is convenient for Bel Canto teaching. In the stage of Bel Canto audio feature extraction, all features are considered comprehensively, and various features are introduced to discuss, and the word vector method is combined to improve the classification performance, increase the amount of data taken in the experiment to make the experimental results more comprehensive and universal, and improve the training algorithm of the model to reduce the computational complexity and time; when the model recognizes Bel Canto audio, it is easily affected by network jamming and delay, which leads to deviation of sound accuracy, and more experiments need to be carried out on noise reduction.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest regarding this work.

References

[1] Z. Zhang, "On the teaching analysis of students' music sense cultivation and appreciation in junior high school music teaching," *Northern Music*, vol. 40, no. 5, p. 2, 2020.

[2] S. Yang, L. Wang, and K. Weihua, "Exploration and thinking on online vocal music teaching in normal universities during the epidemic in novel coronavirus pneumonia," *Educational Research*, vol. 3, no. 7, pp. 129–131, 2020.

[3] W. Zhang, "Diversified innovation and research of vocal music teaching in higher music education in the new period," *Northern Music*, vol. 40, no. 4, p. 2, 2020.

[4] X. Guo, "Practical research on vocal music teaching mode in colleges and universities based on open cooperation--comment on research on vocal music teaching and music education in colleges and universities," *Food Science and Technology*, vol. 46, no. 2, p. 2, 2021.

[5] J. Fan, "Analysis on the cultivation of junior high school music sense and appreciation teaching under the new situation," *Reading, Writing and Computing*, vol. 000, no. 003, p. 74, 2019.

[6] M. Zeng, "Discussion on the importance of music sense cultivation in middle school music education," *Middle School Curriculum Counseling: Teaching Research*, vol. 10, no. 002, pp. 144-145, 2016.

[7] L. Kong, "Analysis of the music singing art of Bel Canto," *Northern Music*, vol. 38, no. 14, p. 1, 2018.

[8] Z. Wang, C. Ye, and W. Wang, "Prediction of music pop trend based on LSTM-Att method," *Computer Technology and Development*, vol. 30, no. 9, p. 6, 2020.

[9] F. Pengyu, P. Chen, and J. Shen, "Recommended method of music classification integrating LSTM and attention mechanism," *Computer Science and Application*, vol. 10, no. 12, p. 11, 2020.

[10] C. Chen, "Song audio emotion classification based on CNN-LSTM," *Communications Technology*, vol. 52, no. 5, p. 5, 2019.

[11] L. He and B. Yuan, "Classification of music genres by using long-term and short-term memory network," *Computer Technology and Development*, vol. 29, no. 11, p. 5, 2019.

[12] L. Lei, X. Wu, and J. Liu, "Bi-LSTM affective analysis model integrating key object recognition and deep self-attention," *Small Microcomputer System*, vol. 42, no. 3, p. 6, 2021.

[13] T. Liu, L. Zhang, and W. Yu, "Audio and video emotion recognition based on embedded attention mechanism level LSTM," *Progress in Laser and Optoelectronics*, vol. 58, 2021.

[14] K. Chen and L. Han, "Research on music emotion classification based on audio and lyrics," *Electronic Measurement Technology*, vol. 41, no. 22, p. 6, 2018.

[15] Y. Xie, "Analysis of the importance of music literacy to Bel Canto in vocal singing," *Northern Music*, vol. 40, no. 11, p. 2, 2020.

[16] D. Lu, "Analysis of the influence of Bel Canto color on Bel Canto music works," *Xijiangyue*, vol. 000, no. 017, 252 pages, 2013.

[17] A. Yu, "Some thoughts on singing Chinese songs with reference to Bel Canto," *Northern Music*, vol. 39, no. 10, p. 2, 2019.

[18] X. Chang, "Analysis of the training of "killfully" using breath in vocal music teaching," pp. 166-167, 2021.

[19] W. Liu, "Research on the application of nationalization of Bel Canto in vocal music teaching," *International Education Forum*, vol. 2, no. 10, p. 188, 2020.

[20] X. Yang, "Research on introducing pop music into junior high school music teaching," *Music Time and Space*, vol. 503, no. 24, pp. 189-190, 2015.

[21] Y. Chen, "Application of new media technology in music teaching in junior high school," *Middle School Curriculum Resources*, vol. 139, no. 1, pp. 35-36, 2019.

[22] S. Zhou, "Training method of intonation in junior high school music course," *Huaxia Teacher*, vol. 81, no. 21, p. 78, 2017.

[23] Z. Fan, "Application of mixed teaching mode in vocal music teaching in colors and universities in post-epidemic period," *Art Evaluation*, no. 18, pp. 73–75, 2020.

[24] S. Liu, "Analysis of online teaching effect of vocal music course in local comprehensive normal collections," *Northern Music*, vol. 407, no. 23, pp. 208–210, 2020.

[25] A. Nie, "Research on innovative mode of online vocal music teaching in colleges and universities," *Daguan (Forum)*, vol. 228, no. 10, pp. 121-122, 2020.