Hindawi
Scientific Programming
Volume 2021, Article ID 3807666, 10 pages
https://doi.org/10.1155/2021/3807666

*Research Article*

# Research on Music Emotion Recognition Model of Deep Learning Based on Musical Stage Effect

**Cuiqing Huang and Qiang Zhang**

*School of Music and Dance, China-ASEAN College of Arts, Chengdu University, Chengdu, Sichuan 610106, China*

Correspondence should be addressed to Qiang Zhang; zhangqiang01@cdu.edu.cn

Received 10 August 2021; Revised 17 September 2021; Accepted 30 September 2021; Published 26 October 2021

Academic Editor: Bai Yuan Ding

The change of life style of the times has also prompted the reform of many art forms (including musicals). Nowadays, the audience can not only enjoy the wonderful performances of offline musicals but also feel the charm of musicals online. However, how to bring the emotional integrity of musicals to the audience is a technical problem. In this paper, the deep learning music emotion recognition model based on musical stage effect is studied. Firstly, there is little difference between the emotional results identified by the CRNN model test and the actual feelings of people, and the coincidence degree of emotional responses is as high as 95.68%. Secondly, the final recognition rate of the model is 98.33%, and the final average accuracy rate is as high as 93.22%. Finally, compared with other methods on CASIA emotion set, the CRNN-AttGRU has only 71.77% and 71.60% of WAR and UAR, and only this model has the highest recognition degree. This model also needs to update iteration and use other learning methods to learn at different levels so as to make this model widely used and bring more perfect enjoyment to the audience.

## 1. Introduction

With the development of the times and technology, people can easily get digital music, drama, film, and television on mobile phones, iPad, computers, and other electronic devices, while tapes, CDs, records, videos, and so on gradually disappear into people's daily life. The stage appeal of musicals is very strong, but simply the video recording technology is poor, and the audience cannot be there. It is difficult for the stage effect to have an effect on the audience online. If the musical is to be moved online well, so that the audience can be infected by the musical without being on the scene, it is necessary to study how to completely restore the musical scene. This paper is one of the most important step, using the existing technology to make a musical emotional recognition model, which can perfectly identify the emotions contained in the musical. Literature [1] used the LEO model to process tree structure format with Gabor feature representation in order to recognize human facial emotion. Literature [2] proposed a layered coding cascade optimization model for the facial expression recognition system, which optimizes direct similarity and Pareto-based function optimization. Literature [3] easily created an entropy-based maximum emotion recognition model using individual average differences of emotion signals. Literature [4] proved that small-world network is the most suitable model to capture the cognitive basis of facial emotions. Reference [5] designed an improved wave physics model based on depth wave field inference in speech emotion recognition. Literature [6] used the Gaussian mixture model fitting method to design neutral profile dictionary to solve the baseline problem. Literature [7] collected emotional physiological data sets under four induced emotions, and the group-based IRS model improved the performance of emotion recognition. Literature [8] introduced traditional acoustic features and new vector functions to represent speech signals abstractly. In Literature [9], cyclic neural network is effectively replaced by convolution network and self-attention, which is close to transformer performance. Literature [10] proposed a new multidimensional cyclic convolution algorithm to achieve the least number of multiplications in theory. Based on CRT and Vinograd's minimum multiple complexity theorem, a simplified cyclic convolution formula is obtained in literature [11]. Literature [12] used the music emotion recognition model based on deep learning to solve the problem of low accuracy of emotion recognition. Literature [13] computed audio function ideas, capturing music form, texture, and expression elements to advance music emotion recognition. Literature [14] constructed a balanced music video emotional data set and integrated multimodal transport information based on deep learning. Literature [15] created a model based on deep neural network to identify and classify music genres and Chinese traditional musical instruments. In order to learn to restore the important human emotion nodes in music, this paper focuses on designing the emotional recognition model of musical based onstage effect, introduces the related theoretical basis, respectively, and then designs a hybrid model based on CNN and RNN according to several points that need to be discussed emphatically in emotional recognition (signal preprocessing, emotional data set, recognition algorithm, and evaluation). Then, the model is simulated and tested.

## 2. Theoretical Basis

*2.1. Emotional Description Model under Stage Effect.* The charm of musicals is always infectious, which most people cannot stop. Different from the performance of simple music and drama dance, musical is a mixture of music elements, dance elements, and drama elements and gradually develops and changes into a unique artistic expression in the torrent of years. It brings wonderful enjoyment in visual, auditory, and other senses. The stage effect cooperates with the music to immerse the audience in the performance of the plot. In this study, we want to establish an emotion recognition model [16]. We want to build an emotion recognition model [16] to define and classify various emotions.

(1) A large number of scholars [17] have studied the definition of relatively basic emotion [18] as shown in Table 1.

Emotion is a reflection of the relationship between objective things and subject needs. However, emotion is an extremely complex psychological process, which is a complex body with multidimensions, multiforms, and multifunctions. Each individual is influenced by his own environment, social environment, and his own experience and cognition, and everyone has different definitions of emotion. Furthermore, human research on psychology is not thorough, and there are many blank fields. Therefore, there are too many reasons for the differences in definitions of different scholars, involving a wide range of fields, which need to be systematically studied and discussed by later generations.

(2) Because human emotions are complex and changeable, it is difficult to express them with simple basic emotional definitions. Therefore, in this study, we tend to combine a more comprehensive and complex two-dimensional emotion description model with a simple emotion definition as shown in Figure 1.

Table 1: Definition of emotion by scholars.

| Scholar | Emotion |
|---|---|
| Mowrer | Pain, joy |
| Panksepp | Expectation, fear, anger, panic |
| James | Fear, sadness, love, anger |
| Pultchick | Acceptance, anger, expectation, disgust, joy |
| Gray | Fear, sadness, surprise |
| Tomkins | Anger, fear, anxiety, joy |
| Ekman, Freesen | Anger, interest, contempt, disgust, pain |
| Weiner, Graham | Fear, joy, shame, surprise |
| Frijida | Anger, disgust, fear, joy, sadness, surprise |

*2.2. Deep Learning Method.* Machine learning [19] has a method called deep learning [20]. This method is very mature in the field of speech recognition, so we can use it to recognize musical emotion.

*2.2.1. Deep Confidence Networks.* Deep confidence network (DBN) [21] is a probability generation model [22], and it is also a special neural network [23]. It consists of a variety of “constrained Boltzmann machines” as shown in Figure 2.

*2.2.2. Parameter Pretraining.* Constrained Boltzmann machine [24] is widely used. Set *n, m, v*, and *h* as visible layer nodes, hidden layer nodes, visible units, and hidden units, respectively. The energy of the system [25] is defined as follows:

$$E(v,h \mid \theta) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} v_i W_{ij} h_j. \tag{1}$$

When the model state is constant, the joint probability distribution is as follows:

$$P(v,h \mid \theta) = \frac{e^{-E(v,h|\theta)}}{Z(\theta)},$$
$$Z(\theta) = \sum_{v,h} e^{-E(v,h|\theta)}. \tag{2}$$

The activation probability of hidden unit is as follows:

$$P\left(h_j = 1, \mid v, \theta\right) = \sigma\left(b_j + \sum_{j} v_i W_{ij}\right). \tag{3}$$

Visible cell activation probability is as follows:

$$P\left(v_i = 1, \mid h, \theta\right) = \sigma\left(a_i + \sum_{i} h_j W_{ij}\right). \tag{4}$$

The sigmoid activation function is as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \tag{5}$$

*2.2.3. Parameter Tuning.* The output of the activation function hidden layer node is as follows:

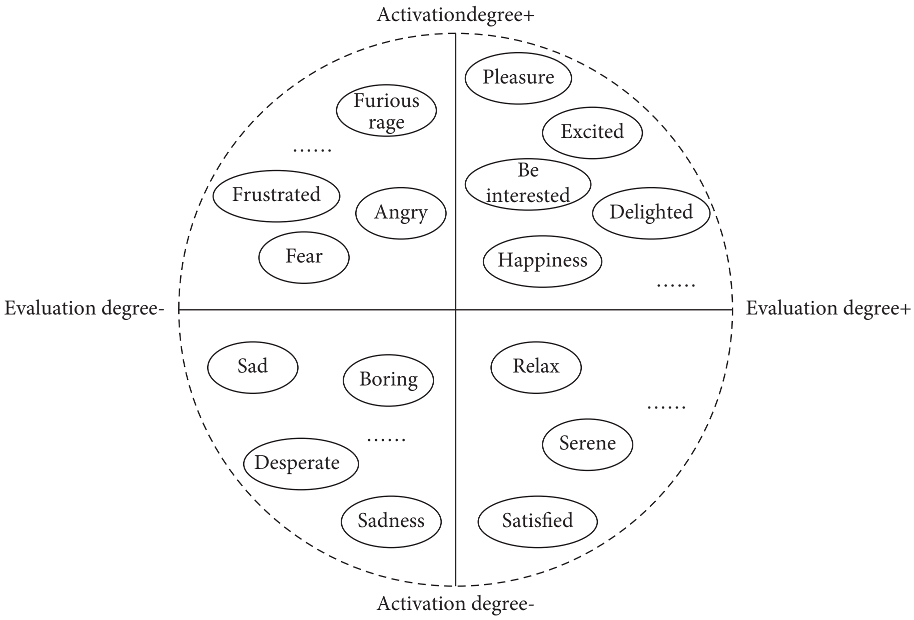

Figure 1: Two-dimensional (activation and evaluation) emotion model.

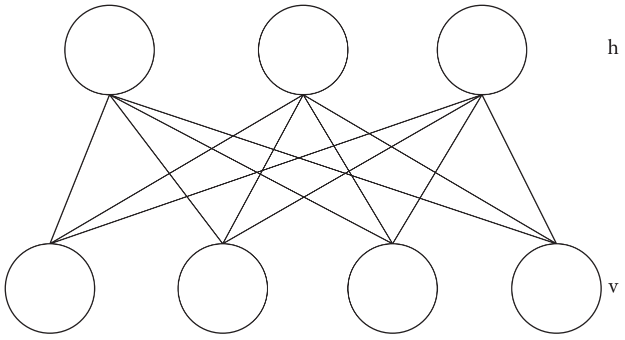

Figure 2: Constrained Boltzmann machine.

$$\begin{aligned} v^{k+1} &= \omega^k h^k + b^k, \\ h^k &= \sigma\left(v^k\right). \end{aligned} \tag{6}$$

Softmax function is as follows:

$$p_s = \frac{\exp\left(v_s^{N+1}\right)}{\sum_j \exp\left(v_s^{N+1}\right)}. \tag{7}$$

When $d_s = 0$, the cross entropy function is as follows:

$$L = -\sum_s d_s \log\, p_s. \tag{8}$$

*2.2.4. CNN.* "CNN" is the abbreviation of convolution neural network. CNN offers supervised and unsupervised learning. It is usually used for processes corresponding to natural access and language. Generally, three-dimensional CNN has two operations: convolution and pooling. Important formulas of convolution layer are as follows:

$$\begin{aligned} Z^{l+1}(i,j) &= \left[Z^l \otimes \omega^{l+1}\right](i,j) + b, \\ Z^{l+1}(i,j) &= \sum_{k=1}^{K_l} \sum_{x=1}^{f} \sum_{y=1}^{f} \left[Z_k^l\left(s_0 i + x, s_0 j + y\right)\omega_k^{l+1}(x,y)\right] + b, \\ L_{l+1} &= \frac{L_l + 2p - f}{s_0} + 1\,(i,j) \in \{0,1,\dots L_{l+1}\}, \\ Z^{l+1} &= \sum_{k=1}^{K_l} \sum_{i=1}^{L} \sum_{j=1}^{L} \left(Z_{i,j,k}^l \omega_k^{l+1}\right) + b = \omega_{l+1}^T Z_{l+1} + b, \\ L^{l+1} &= L. \end{aligned} \tag{9}$$

Excitation function to help express complex characteristics is as follows:

$$A_{i,j,k}^l = f\left(Z_{i,j,k}^l\right). \tag{10}$$

Manifestations of *Lp* pooling are as follows:

$$A_k^l(i,j) = \left[\sum_{x=1}^{f} \sum_{y=1}^{f} A_k^l\left(s_0 i + x, s_0 j + y\right)^p\right]^{1/p}. \tag{11}$$

Hybrid pooled linear combination is as follows:

$$A_k^l = \lambda L_1\left(A_k^l\right) + L_\infty\left(A_k^l\right), \lambda \in [0,1]. \tag{12}$$

*2.2.5. RNN.* "RNN" is short for cyclic neural network. The information of the sequence can be better handled. The details are as follows:

$$\begin{aligned} O_t &= g\left(V \cdot S_t\right), \\ S_t &= f\left(U \cdot X_t + W \cdot S_{t-1}\right). \end{aligned} \tag{13}$$

Simply RNN is shown in Figure 3.

## 3. Research on Emotion Recognition

*3.1. Music Signal Preprocessing.* When we want to identify the influence of music in musicals on the audience, the preprocessing of music signals is the first step of all work. We try our best to extract many features of music, then identify their categories, and build models to identify different music. This is an extremely important step, and all subsequent work is based on this step. Music signals will be uniformly converted into good formats, which is convenient for data management and high-quality operation. The flowchart is shown in Figure 4.

Preprocessing is to read audio information for feature extraction. We use SciPY speech processing tool and Python for speech information reading and feature extraction, librosa for speech processing, and MATLAB for pre-filtering. Advantages can eliminate the influence of aliasing, high frequency, and other factors brought by equipment on the quality of voice signals, ensure that the processed music signals are more uniform and smooth, provide high-quality parameters, and improve the processing quality.

The preprocessing parameter data are shown in Table 2.

The parameter settings in Table 2 are basically reasonable. The input of the model is the digital amplitude mel spectrum, and the audio frequency of 22 s is input, and the fast Fourier transform STFT is performed (FFT is the content of decomposing the whole time domain into countless small processes with equal length). Hop size is the overlapping area between two windows. Mel filter is used to obtain the logarithmic amplitude mel spectrum; the preprocessing process is carried out in Librosa, and the output size is array. Zero padding is the last step in the preprocessing process. By adding 75 frames on the time axis to inject more data with the same information, the input signal has better frequency resolution.

*3.2. Music Emotional Data Set.* For many years, researchers have been studying databases that contain many kinds of human emotions, and these databases contain various forms of data. Music emotion database can collect signals through audience's reaction to different music materials. In this paper, we quote EMO-DB, CASIA, SAVEE, DEAP, and MAHNOB-HCI.

*3.3. Emotion Recognition Algorithm*

(1) Bayesian network is shown in Figure 5.

(2) Hidden Markov model (HMM) is shown in Figure 6.

(3) The Gaussian mixture model (GMM) is shown by the following formula:

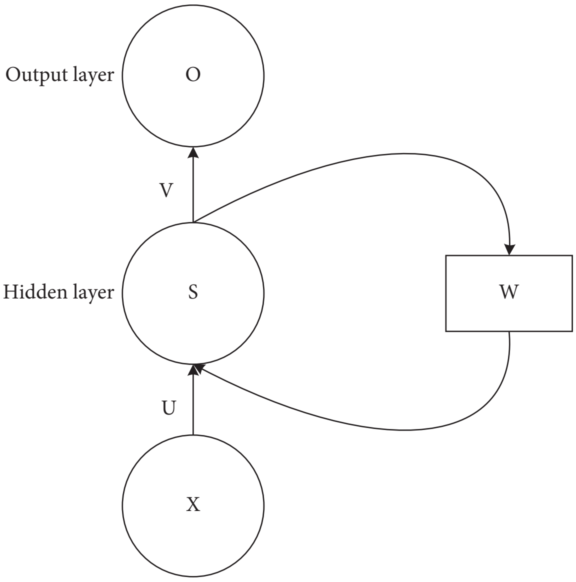

FIGURE 3: Simple cyclic neural network.

$$P(X \mid \theta) = \sum_{k=1}^{M} \pi_k N\left(X; \mu_k; \sum_k\right). \tag{14}$$

*3.4. Evaluation Aspects.* When we apply the musical emotion recognition model, we will find that there will be an important problem in the process of construction: overfitting and under-fitting. In order to solve this problem well, we choose to evaluate the model again and again and use various evaluation indexes to test it. In this study, the K-fold cross-validation method and UA, WA, SD, SI, Precision, Recall, and other evaluation indicators were used. Here are some of the formulas:

$$UA = \frac{TP + TN}{TP + FP + TN + FN}, \tag{15}$$

$$WA = \frac{\sum_i^n a_i}{n}, \tag{16}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{17}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{18}$$

Formula 19 is used to evaluate the accuracy of the model. TP means positive case prediction is positive case. FN means positive case prediction is negative case. FP means negative case is predicted as positive case. TN means negative case predicts negative case.

*3.5. Emotion Recognition Based on Hybrid Model.* Convolution neural network and cyclic neural network can not meet the research needs. Because both theories and methods have their own advantages and disadvantages, we

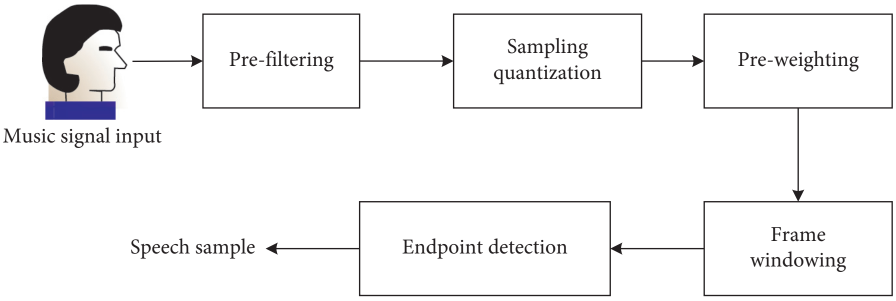

FIGURE 4: The flowchart of feature extraction.

TABLE 2: Preprocessing parameters of music signal.

| Duration | $n$ |
|---|---|
| FFT frame size | 512 |
| Hop size | 255 |
| Number of mel filters | 96 |
| Mel-time matrix size | $1 * 95 * 1300$ |
| Extended mel-time matrix | $1 * 95 * 1356$ |
| Extended boundary number | 75 |

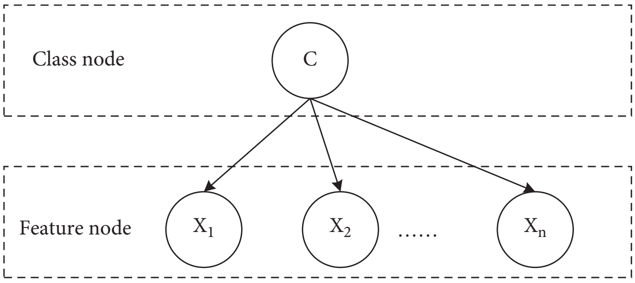

FIGURE 5: Basic structure of Bayesian network.

take the essence and discard the dross, so we improve these two kinds of networks and combine them into a more complex neural network to meet the needs of this study, that is, convolutional recurrent neural network (CRNN). We have modified many excellent performances for this model. It can have all kinds of advantages of CNN and RNN at the same time.

After the improvement of this model, the operation of variable length input can be carried out, and the interference of filling values to model data can be avoided. It can also ensure that the accuracy of the model will not be lost.

$$S_{\text{conv}} = \text{Conv}(S) \cdot \text{Mask}(S). \tag{19}$$

LSTM is an extension of RNN. It is a long-term and short-term memory network. We add this extension to the CRNN model, which can solve the problem of processing sequence change data. Then, we transport the data to the Softmax loss and quantification loss layer for processing and finally the optimization goal. The CRNN model is shown in Figure 7.

## 4. Music Emotion Recognition Model Test

In this experiment, the music emotion model is tested. The main tasks of the test model are as follows: accept music data samples of musicals for certain processing; after analyzing the sample, speculate and identify what emotions the stage effect will cause the audience (such as happiness, sadness, sadness, and excitement), the emotion description model should be applied here, and finally the audience's emotions should be collected, and the accuracy of the opinion collection should be compared with the results of our experimental model test, so as to evaluate whether the established CRNN model is available. If there is too much difference between the two results, it means that our music emotion recognition model is unqualified, and more modifications and adjustments are needed to carry out technical innovation. If the difference accuracy meets the requirements, it proves that the deep learning music emotion recognition model based on musical stage effect can be used.

*4.1. Experimental Environment.* In the process of music performance, various unpredictable external interference factors affect the audience, so it is necessary to deal with the noise. We chose to conduct the experiment in a theater where the surrounding environment is quiet and there is no performance task for the time being. In this experiment, we invited 30 volunteers (native speakers of Chinese) who were in good mental state and did not stay up late, drink alcohol, or get sleepy to watch a Chinese musical together. While watching the experiment, the volunteers were unable to do anything else. After watching the musical, the emotional feelings of these 30 people were collected. The music emotion recognition model will be tested by this musical music sample, and finally the test results will be compared with the feelings of 30 voluntary participants.

*4.2. Experimental Settings.* The experimental test samples are Chinese musicals, and Chinese emotional data sets are used. From the matching selection in the database to the data suitable for this musical, it is divided into pretest, development set, and test set. *Note.* The main purpose of pretest is to detect the interference caused by sudden factors such as garbled code and disordered data, and all the extracted data should be tested by pretest. As shown in Table 3, it is the specific data situation.

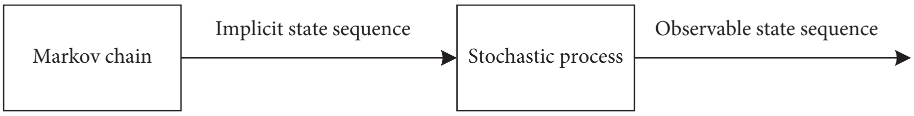

Figure 6: Schematic diagram of hidden Markov process.

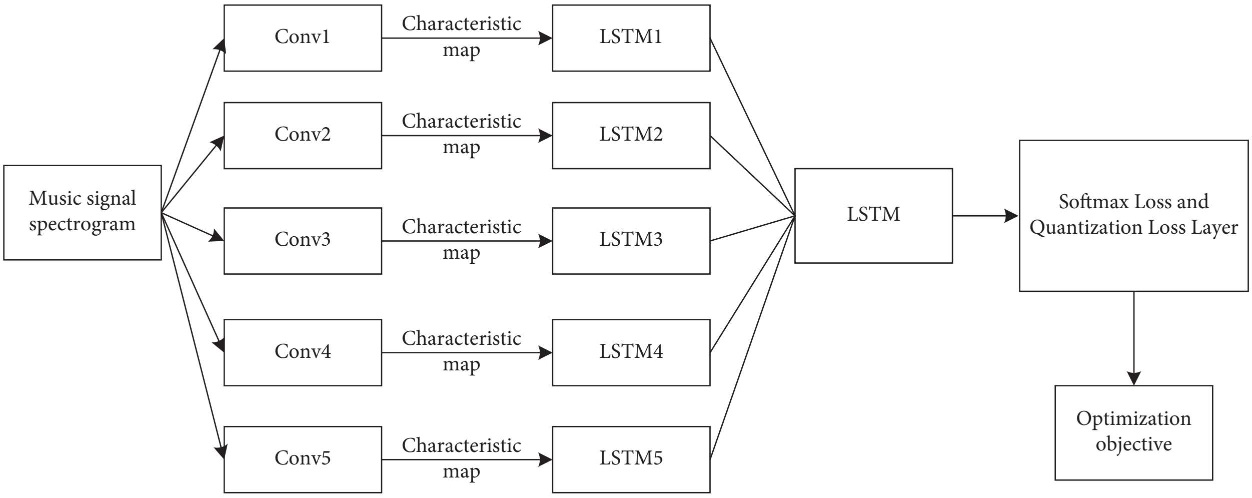

Figure 7: Flow chart of CRNN music recognition framework.

Table 3: Pretest, development set, and test set dataset settings.

| | Pretest | Development set | Test set |
|---|---|---|---|
| Is there any situation | No abnormality | No abnormality | No abnormality |
| Total number of sentences (sentences) | 60 | 120 | 360 |

*4.3. Parameter Optimization.* In the experiment, parameters are the part that needs special attention. Without adjustment and optimization, the final result may not achieve the best experimental effect. This will cause trouble to the experimental results. The relevant parameters are shown in Table 4.

We carry out 20,000 iterations on music samples. When the learning rate is very small (such as 0.001), the learning process is extremely slow and the recognition is unstable; high learning rate (such as 0.1) leads to unstable conditions and even reduces performance as shown in Figure 8.

Figure 8 is a discussion of recognition rate for different iteration times of learning rate in experimental parameters, and the most suitable learning rate interval range is selected. The reason of performance deviation is that the learning rate affects the recognition stability of recognition rate.

Momentum coefficient can speed up the learning process. The deviation of coefficient will cause oscillation in the initial stage, and the fluctuation will cause performance degradation, as shown in Figure 9.

However, if the weight attenuation coefficient is too large (such as 0.005), the stability of the learning process will be destroyed. On the contrary, a relatively small weight attenuation coefficient will be more stable and safe, as shown in Figure 10.

*4.4. Key Elements of Musical Emotion.* There are some fragments in the music. It is useful for experiments, including emotional parts, and some fragments are not important, which may cause suspension, transition, and other effects. Let us take time as an example and look at the emotional information carried by each sentence as a characteristic, one is unimportant and the other is important.

As shown in Figure 11, the music sentence part is divided into five parts, and the emotional information is mainly distributed in the first, second, and third parts, which takes more time to test.

Table 4: Parameters related to experiments.

| Parameter | Learning rate | Batch size | Momentum coefficient | Weight attenuation coefficient | Dropout coefficient |
|---|---|---|---|---|---|
| Value | 0.01 | 16 | 0.9 | 0.0005 | 0.5 |

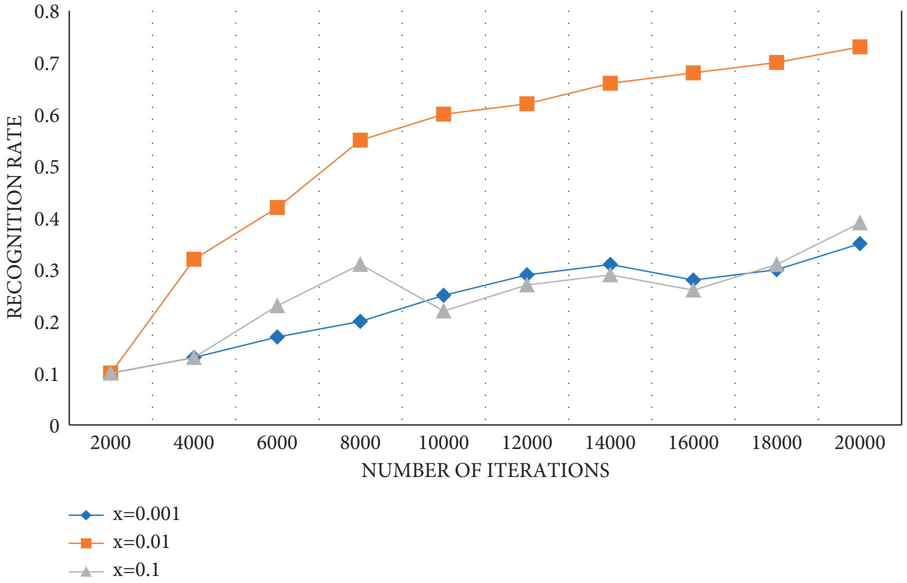

Figure 8: Learning rate-related effects.

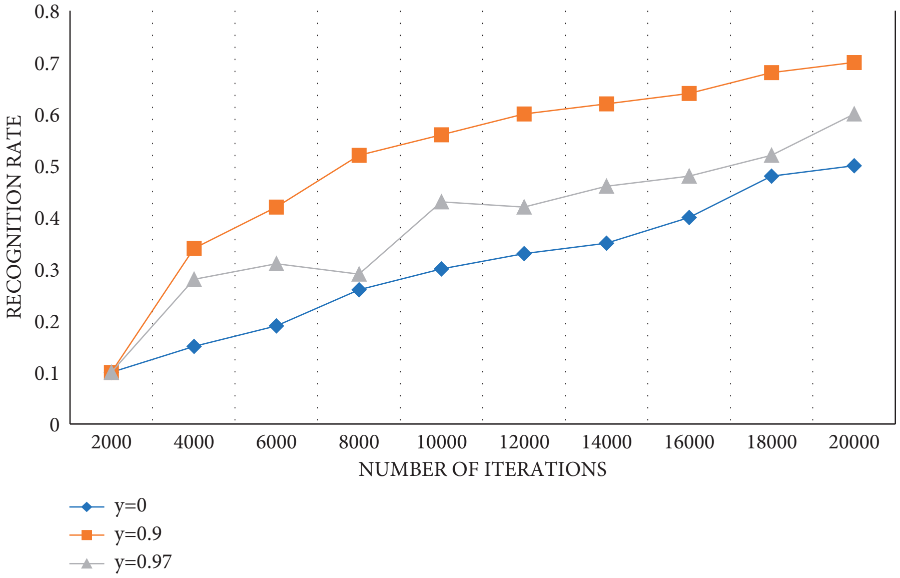

Figure 9: Correlation effect of momentum coefficient.

*4.5. Recognition Performance.* The performance calculation formula of music recognition is as follows:

$$\text{rejection rate} = \frac{\text{number of successfully recognized music sample fragments}}{\text{total number of sample fragments of test music}}. \quad (20)$$

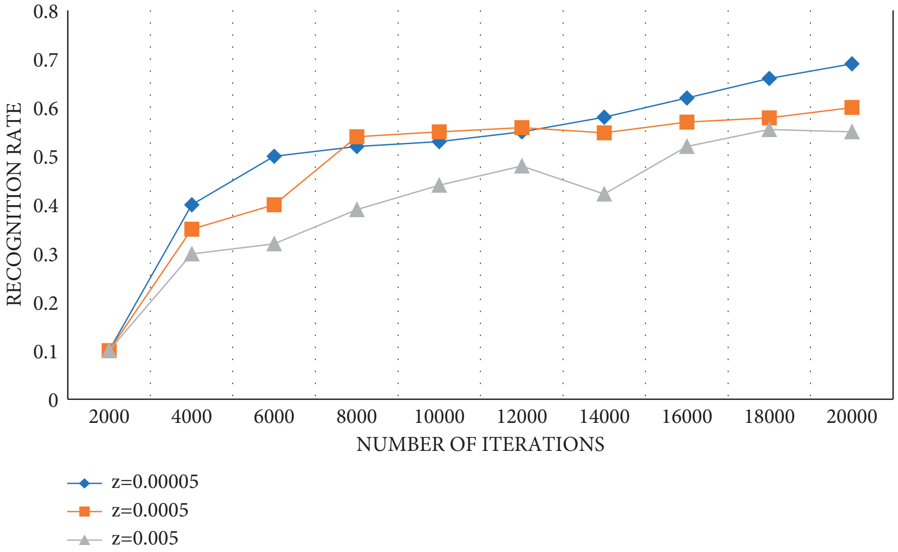

Figure 10: Correlation effect of weight attenuation coefficient.

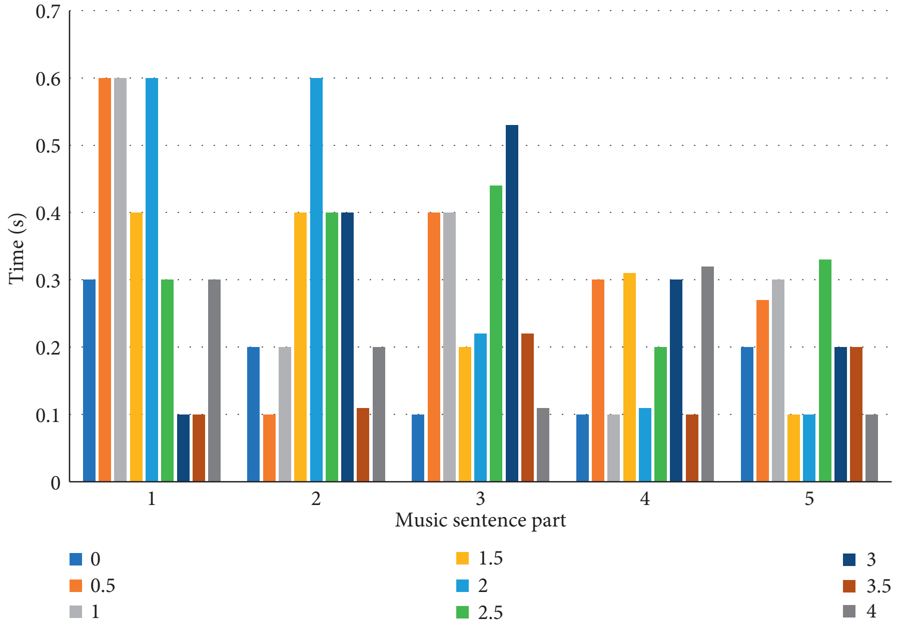

Figure 11: Visual emotional information distribution map.

The specific identification is shown in Table 5.

From Table 5, we can see that the recognition performance of this experiment is very good, the final recognition rate is 98.33%, and the final average accuracy rate is as high as 93.22%.

*4.6. Experimental Test Results.* Because the content of this musical is sad as a whole, the emotional reaction of positive energy is excluded. The specific feelings of 30 voluntary participants are collected as shown in Table 6.

The results tested by the music emotion model are shown in Table 7.

After calculating and comparing the experimental results, it can be found that the coincidence degree of emotional response with 30 voluntary participants is as high as 95.68%.

*4.7. Comparison with Other Methods.* Compare WAR and UAR on Chinese emotion database (which can be changed in other tests). The results show that only the model created by us has the highest recognition performance and the most accurate accuracy. Among the three methods in the table, even CRNN-AttGRU with the highest recognition performance has only 71.77% and 71.60% recognition rates, respectively, as shown in Table 8.

Table 5: Table of recognition performance.

| Numbering | 1 (%) | 2 | 3 | 4 (%) | 5 (%) | 6 (%) | 7 (%) | ··· |
|---|---|---|---|---|---|---|---|---|
| False acceptance | 0 | 0% | 0% | 0 | 1.19 | 0 | 0 | ··· |
| False rejection rate | 0 | 20% | 2.14% | 0 | 1.9 | 0 | 0 | ··· |

Table 6: Emotional feelings of volunteers.

| Voluntary participant number | Emotion |
|---|---|
| 1 | Pain, fear |
| 2 | Expectation, fear, anger, panic |
| 3 | Fear, sadness, anger |
| 4 | Acceptance, anger, expectation, disgust |
| 5 | Fear, sorrow |
| 6 | Anger, fear, anxiety |
| 7 | Interest, disgust, pain, fear |
| 8 | Anxiety, surprise |
| 9 | Anger, disgust, fear, sadness, surprise |
| ··· | Boredom, fear, sadness |
| 30 | Surprise, sadness, fear |

Table 7: Test results of the music emotion model.

| Name | Emotion |
|---|---|
| CRNN model | Sadness, sadness, pain, disgust, fear, fear, surprise, anger, surprise, surprise, anxiety, and so on |

Table 8: Comparison of WAR and UAR on CASIA by different methods.

| Model | WAR (%) | UAR (%) |
|---|---|---|
| CRNN-CTC | 70.42 | 69.75 |
| CRNN-GRU | 60.48 | 61.72 |
| CRNN-AttGRU | 71.77 | 71.60 |

## 5. Conclusion

In this paper, based on the changes of the times, electronic instruments are of great significance for musicals to be perfectly moved to mobile devices. The stage effect provided by musicals for the audience leads the audience to immerse themselves in the emotional changes in music. This study needs to restore people's emotional feelings brought by music in musicals with technology. Therefore, the CRNN model is designed for testing. The recognition accuracy of the model for emotion is more accurate than that of other methods. The results show that (1) there is little difference between the emotional results identified by the CRNN model test and the actual feelings of people, and the coincidence degree of emotional responses is as high as 95.68%; (2) the final recognition rate of the model is 98.33%, and the final average accuracy rate is as high as 93.22%; and (3) compared with other methods on the CASIA emotion set, the CRNN-AttGRU has only 71.77% and 71.60% of WAR and UAR, and only this model has the highest recognition degree. Although this paper has some results in music emotion recognition, it still needs further exploration by later workers. This model also needs to update iteration and use other learning methods to learn at different levels, so as to make this model widely used and bring more perfect enjoyment to the audience.

## 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 authors declare that they have no conflicts of interest regarding this work.

## Acknowledgments

This research was supported by the phased achievement of the "Hidden Infiltration of Red Culture in Instrumental Music Teaching" (2021KCSZ01) and the key project of "Curriculum Ideological and Political Research" of Chengdu University in 2021.

## References

[1] J.-J. Wong and S.-Y. Cho, "A local experts organization model with application to face emotion recognition," *Expert Systems with Applications*, vol. 36, no. 1, pp. 804–819, 2009.

[2] S. C. Neoh, L. Zhang, K. Mistry et al., "Intelligent facial emotion recognition using a layered encoding cascade optimization model," *Applied Soft Computing*, vol. 34, pp. 72–93, 2015.

[3] S.-Y. Park, D.-K. Kim, and M.-C. Whang, "Maximum entropy-based emotion recognition model using individual average difference," *The Journal of the Korean Institute of Information and Communication Engineering*, vol. 14, no. 7, pp. 1557–1564, 2010.

[4] T. Takehara, F. Ochiai, and N. Suzuki, "A small-world network model of facial emotion recognition," *Quarterly Journal of Experimental Psychology*, vol. 69, no. 8, pp. 1508–1529, 2016.

[5] C. Zheng, C. Wang, and N. Jia, "Emotion recognition model based on multimodal decision fusion," *Journal of Physics: Conference Series*, vol. 1873, no. 1, Article ID 012092, 2021.

[6] S. Ulukaya and C. E. Erdem, "Gaussian mixture model based estimation of the neutral face shape for emotion recognition," *Digital Signal Processing*, vol. 32, pp. 11–23, 2014.

[7] C. Li, C. Xu, and Z. Feng, "Analysis of physiological for emotion recognition with the IRS model," *Neurocomputing*, vol. 178, no. 20, pp. 103–111, 2016.

[8] T. Zhang and J. Wu, "Speech emotion recognition with i-vector feature and RNN model," in *Proceedings of the 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pp. 524–528, IEEE, Chengdu China, July 2015.

[9] M. Chen, Y. Li, and R. Li, "Research on neural machine translation model," *Journal of Physics: Conference Series*, vol. 1237, Article ID 052020, 2019.

[10] A. H. Diaz-Perez and D. Rodriguez, "One dimensional cyclic convolution algorithms with minimal multiplicative complexity," in *Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Toulouse, France, May 2006.

[11] A. Rodriguez, "Cyclic convolution algorithm formulations using polynomial transform theory," *Journal of Computers*, vol. 2, no. 7, pp. 40–48, 2007.

[12] R. Sarkar, S. Choudhury, S. Dutta, A. Roy, and S. K. Saha, "Recognition of emotion in music based on deep convolutional neural network," *Multimedia Tools and Applications*, vol. 79, no. 10, pp. 765–783, 2020.

[13] R. Panda, R. M. Malheiro, and R. P. Paiva, "Audio features for music emotion recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 99, p. 1, 2020.

[14] Y. R. Pandeya and J. Lee, "Deep learning-based late fusion of multimodal information for emotion classification of music video[J]," *Multimedia Tools and Applications*, vol. 80, no. 38, pp. 1–19, 2021.

[15] K. Xu, "Recognition and classification model of music genres and Chinese traditional musical instruments based on deep neural networks," *Scientific Programming*, vol. 2021, Article ID 2348494, 8 pages, 2021.

[16] M. Fredrikson and R. Gunnarsson, "Psychobiology of stage fright: the effect of public performance on neuroendocrine, cardiovascular and subjective reactions," *Biological Psychology*, vol. 33, no. 1, pp. 51–61, 1992.

[17] R. Studer, P. Gomez, H. Hildebrandt, M. Arial, and B. Danuser, "Stage fright: its experience as a problem and coping with it," *International Archives of Occupational and Environmental Health*, vol. 84, no. 7, pp. 761–771, 2011.

[18] Y. Lyu, "Research on the influence of music educational psychology on saxophone players' mental state and stage performance," *Journal of Intelligent and Fuzzy Systems*, vol. 2, no. 2, pp. 1–12, 2021.

[19] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.

[20] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Qullien, "Learning hand-eye coordination for robotic Grasping with deep learning and large-scale data collection," pp. 421–436, 2016, https://arxiv.org/abs/1603.02199.

[21] Z. Jian and O. G. Troyanskaya, "Predicting effects of noncoding variants with deep learning-based sequence model," *[J]. Nature Methods*, vol. 12, no. 10, pp. 931–934, 2015.

[22] K. Warburton, "Deep learning and education for sustainability," *International Journal of Sustainability in Higher Education*, vol. 4, no. 1, pp. 44–56, 2003.

[23] J. Kim and E. Andre, "Emotion recognition based on physiological changes in music listening," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.

[24] Y. P. Lin, C. H. Wang, T. P. Jung et al., "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[25] A. D. Patel, "A neurobiological strategy for exploring links between emotion recognition in music and speech," *Behavioral and Brain Sciences*, vol. 31, no. 5, pp. 589-590, 2008.