Research Article

Lute Acoustic Quality Evaluation and Note Recognition Based on the Softmax Regression BP Neural Network

Lili Liu
Taiyuan University of Technology Taiyuan City, Taiyuan 030024, China
Correspondence should be addressed to Lili Liu; liulili01@tyut.edu.cn
Received 22 February 2022; Revised 14 March 2022; Accepted 21 March 2022; Published 12 April 2022

Abstract
Note recognition technology has very important applications in instrument tuning, automatic computer music recognition, music database retrieval, and electronic music synthesis. This paper addresses the above issues by conducting a study on acoustic quality evaluation and its note recognition based on artificial neural networks, taking the lute as an example. For the acoustic quality evaluation of musical instruments, this paper uses the subjective evaluation criteria of musical instruments as the basis for obtaining the results of the subjective evaluation of the acoustic quality of the lute, similar to the acoustic quality evaluation, extracts the CQT and MFCC note signal features, and uses the single and combined features as the input to the Softmax regression BP neural network multiclassification recogniser; the classification coding of standard tones is used as the target for supervised network learning. The algorithm can identify 25 notes from bass to treble with high accuracy, with an average recognition rate of 95.6%; compared to other recognition algorithms, the algorithm has the advantage of fewer constraints, a wider range of notes, and a higher recognition rate.

1. Introduction

Traditional Chinese music is an important component of music that makes up the world, containing a rich resource of historical, cultural, and folk traditions, representing the accumulation of national history and ideology, a living tradition. In the world of musical instruments, western instruments dominate the field, and our national instruments are not yet comparable to them. Our national musical instruments need to move out of the country to promote our excellent traditional culture, as well as vigorously promote national musical instruments [1]. At present, the musical instrument manufacturing industry of China lags behind that of the west by 20 years and there is a need to change the current situation. The study of the acoustic quality of musical instruments will help the inheritance, development, and promotion of folk musical instruments, play a vital role in improving the quality of musical instruments, promote the development of musical instrument manufacturing and related cultural industries, and provide guidance to the buyers of musical instruments [2]. In today’s context, the development of science and technology and the prosperity of culture and art have brought about the integration of technology and culture, resulting in a prosperous scene of cultural development. The new discipline of music technology is a product of the combination of music and technology. The recent emergence of music information retrieval (MIR) technology is an important part of the music technology field. Note recognition is an important branch of music information retrieval (MIR) [3]. Note recognition is an important area of research in the field of music signal analysis and processing, and note recognition technology has important applications in the tuning of musical instruments, automatic computerised score recognition, music database retrieval, and electronic music synthesis. Note recognition is of great importance in promoting the development of music technology and new electronic industries [2].

With the development of science and technology, new tools and technologies are changing day by day and artificial intelligence (AI) has become synonymous with the new era [4]. In this paper, we focus on the subjective evaluation of the
acoustic quality of musical instruments, which does not objectively and comprehensively reflect the acoustic quality of musical instruments due to the difference in individual’s musicianship and preferences, as well as the following problems in the study of note recognition: the estimated pitch is difficult to correspond to the standard pitch, the range of recognisable pitches is narrow, the recognition process is not robust, and the recognition rate is low. Using machine learning in the field of artificial intelligence (AI), a new idea is proposed: an artificial neural network-based assessment of the acoustic quality of the pipa and its note recognition, which eliminates the human subjective factor and the uncertainty that arises in the subjective evaluation process. The pipa is the most ethnically distinctive of China’s traditional musical instruments, and it is known as the “king of musical instruments” due to its complexity, versatility, and representativeness [4]. Therefore, the pipa was chosen as the object of this study to facilitate the subsequent study of other instruments.

2. Related Work

From the relevant domestic and international literature reviewed so far on the research methods for evaluating the acoustic quality of musical instruments, they can be broadly divided into three categories: first, subjective evaluation; second, objective evaluation; third, a combination of subjective and objective evaluation, with subjective evaluation being the main focus and objective evaluation being supplementary. However, most of the literature focuses on the human subjective sense of listening and finally gives the instrument a corresponding evaluation through subjective feelings. In the absence of a unified objective evaluation standard, it is not yet possible to use an advanced and independent objective evaluation for the assessment of the acoustic quality of musical instruments and scholars at home and abroad are currently exploring a scientific method that can replace subjective evaluation. Alternatively, it may be possible to combine the evaluation of some measurable physical parameters (frequency, amplitude, time, etc.) and the physical characteristics of the instrument (material, mechanics, size and resonance characteristics, etc.) with subjective perception. The study in [5, 6] systematically discusses terminology, technical preparation, and methods related to evaluation, giving specific evaluation methods; for the first time, it proposes the use of seven subjective parameters for scoring and evaluation. The study in [7] presents a comprehensive overview of the issues to be taken into account in the behind-the-scenes appraisal (subjective evaluation) of the acoustic quality of musical instruments; the study in [8] proposes a system for evaluating the acoustic quality of musical instruments according to the different purposes, positions, and perspectives of people [9]. A two-channel FFT analyser, a 0.62 cm condenser microphone, and a preamplifier were used to analyse the sound waves generated by the instrument and thus to evaluate its performance. The study in [10] objectively evaluated the acoustic quality of musical instruments from a mechanical perspective by means of simulation analysis. The study in [11] used a new objective acoustic quality evaluation metric, the difference in fractal dimensionality in the time-frequency domain (DFDTF), which is no longer an objective evaluation metric in the traditional simple sense of frequency, amplitude, and time. The study in [12] established a link between subjective evaluation and objective quantitative analysis and described the importance of establishing subjective and objective evaluation methods. Different evaluations of the violin by the study in [13] ultimately make an assessment of the instrument and the musician from the subjective perception, while the physicist from the vibrational properties of the violin. A system of subjective and objective sound quality evaluation methods for the improvement of bass string instruments was investigated by the study in [14], where the objective evaluation was done by analysing the frequency spectrum of the sound signal, the acoustic sound pressure level, and the dynamic range of the sound intensity [15].

In summary, subjective evaluation is still the most important method for evaluating the acoustic quality of musical instruments, and there is no scientific evaluation method that can replace it; the study of objective evaluation is also challenging, and it is not easy to achieve the effect of subjective evaluation [16]. The ultimate goal of research into methods for evaluating the acoustic quality of musical instruments is to replace human subjective perceptions, to replace subjective evaluation as far as possible, and to achieve artificial intelligence. It is hoped that the results of this paper will make a small contribution to this goal.

3. Pipa Music Signal Library

The music signal is acquired using the audio signal acquisition system set up for the subjective evaluation of the music used in the evaluation, and both the subjective evaluation process and the pipa music signal acquisition process are carried out simultaneously. Each music file is acquired in 30 s, and each lute is subjectively evaluated and acquired three times. The principle process of building a pipa music signal library is shown in Figure 1.

3.1. Note Signal Library. The audio signal acquisition system is used for the acquisition of the pipa note signals. The acquisition process is carried out after the pipa music signal has been acquired and a library of the various types of pipa note audio to be identified is created. Prior to note acquisition, the timing of each single note audio needs to be determined to facilitate subsequent experimental studies. In terms of string vibration, the longer the vibrating string length, the longer the duration of the audio, because more overtones are produced; conversely, the shorter the vibrating string length, the shorter the duration of the audio and the fewer the overtones produced [17]. To ensure that the four phases of each single note (silent section, transition section, musical section, and ending section) are all within the determined time, the captured notes should be complete. The duration of each single note was determined to be 3 s. To facilitate the control of each note within 3 s and reduce
repetitive recording operations and improve efficiency, each type of note was played three times at intervals of about 2 s. The acquisition time was set to 20 s to ensure the integrity of each type of note signal [18]. The specific principle process of building a pipa note signal library is shown in Figure 2.

4. Evaluation of the Acoustic Quality of the Lute and Its Note Recognition Model

The determination and establishment of a model for the evaluation of the acoustic quality of the lute and its note recognition is the core of this research and a key part of the solution to the research problem. The model will directly affect the results of the research; therefore, the selection and construction of a suitable model is the focus of this chapter.

4.1. Modelling Ideas for the Evaluation of the Acoustic Quality of the Lute. The ultimate goal of the research on the lute acoustic quality evaluation method is to replace the subjective feelings of human beings, to replace subjective evaluation as far as possible, and to achieve artificial intelligence. Using artificial neural networks with the function of mimicking the behavioural characteristics of the human brain, a modelling analysis is carried out using BP neural networks and a model based on BP neural networks is constructed for the evaluation of the acoustic quality of the pipa, the basic idea of which is shown in Figure 3. A library of subjectively evaluated pipa music signals is established, which contains samples to be trained, test samples, and validation samples. The parameters (CC, CQT, and MFCC) that are more representative and closer to human ear perception are extracted from the time, frequency, and cepstrum domains of the lute signal and fed into the evaluation model for learning and training, resulting in the best predicted evaluation results.

4.2. Pipa Note Recognition Modelling Ideas. The idea of multiclassification recognition is used to recognise the notes of the lute. As it is considered that BP neural network has a good recognition effect for binary classification problems, for multiclassification problems, the recognition effect is not good and it cannot meet the requirements of this paper for note recognition. The Softmax regression model has a great advantage in the multiclassification problem, and combined with the good nonlinear mapping ability, self-learning ability, and fault tolerance ability of BP neural networks, therefore this paper proposes a method combining Softmax regression with BP neural network that has the ability of multiclassification recognition and constructs a multiclassification note recognition model of BP neural network based on Softmax regression. The basic idea of the modelling is shown in Figure 4. A pipa note signal library is established, which contains samples to be trained and test samples. The frequency domain features (CQT) and inverse frequency domain features (MFCC) of the lute note signal are extracted and input as feature parameters into the multiclassification recognition model for learning and training, and the optimal recognition and classification results are obtained.
4.3. BP Neural Network Based on Softmax Regression. Based on the above analysis of the modelling idea of lute note recognition, the advantages of Softmax regression model in multiclassification problems are utilised and combined with the good nonlinear mapping ability, self-learning ability, and fault tolerance of BP neural networks. With reference to the structure of multilayer BP neural networks, a Softmax regression-based BP neural network multiclassification recogniser is constructed and the structure is shown in Figure 5. Based on the experimental sample size, the single implicit layer is unable to meet the experimental requirements and there are problems of large computational effort, long training time, and low recognition rate. Therefore, the number of neurons in the input layer is determined by the dimensionality of the input features, two layers are used in the hidden layer (which has been verified to be the optimal number of layers), and the Softmax regression model is used in the output layer.

The multiclassification problem solved in this paper is a one-vs-all classification problem, where the output $y$ is expected to be a vector:

$$y \in \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{25 \times 25}.$$

The immediate reason for using one-hot coding is that the output layer of Softmax classification outputs a probability distribution and therefore requires that the input target labels also appear as probability distributions, which makes it easier to calculate the cross entropy for loss-supervised network learning. One-hot coded labels give the sample the true probability distribution, where only one occurs with probability 1 and the others occur with probability 0.

5. Experimental Results and Performance Analysis

5.1. Model Evaluation Indicators. The purpose of the model evaluation indicators is to analyse and evaluate the performance of the lute acoustic quality evaluation model and the lute note recognition model constructed in this paper and whether they are suitable for the object of study in this paper.

Average accuracy is used as the evaluation index of the pipa acoustic quality evaluation model, which is defined as follows:

$$\text{average accuracy} = \left(1 - \frac{\sum_{i=1}^{N} |Y_i - Y'_i|/|Y'_i|}{N}\right) \times 100\%.$$ (2)

Five evaluation metrics were used to evaluate the classification performance of this classifier: the Confusion Matrix $P$ 84l, Accuracy, i.e., recognition rate, Recall, Precision, and the composite metric $F$-Score.

The recognition rate (accuracy) is defined by the following equation:

$$A = \frac{O_{cd}}{O_{cd} + O_{fn} + O_{fp}}.$$ (3)

Precision is defined by

$$P = \frac{O_{cd}}{O_{cd} + O_{fp}}.$$ (4)

Recall is defined by the following equation:

$$R = \frac{O_{cd}}{O_{cd} + O_{fn}}.$$ (5)

The composite indicator $F$-Score is defined by the following formula:

$$F = \frac{2PR}{P + R}.$$ (6)

In equations (3)-(6), $O_{cd}$ is the number of correct classifications, $O_{fn}$ is the number of unrecognised classifications, and $O_{fp}$ is the number of incorrect classifications. Precision ($P$) and recall ($R$) are two mutually constraining metrics, and the $F$-value is a combination of $R$ and $P$ [19].
5.2. Experiment to Evaluate the Acoustic Quality of the Lute.
The fused features (MFCC + CQT + CC) were used as feature parameters input into the lute acoustic quality evaluation network model for learning and training in the MATLAB R2016a environment. In the experiments, the subjective evaluation results were used as expectation values to supervise the learning and training of this network model; of 144 sets of samples, 110 sets were used as training samples.

Table 1: Average accuracy of varying the number of training samples for different feature inputs.

| Input              | 10 sets of samples | 30 sets of samples | 50 sets of samples | 70 sets of samples | 90 sets of samples | 110 sets of samples |
|--------------------|-------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| MFCC               | 91.63             | 95.09             | 95.48             | 95.64             | 94.92             | 95.70             |
| CQT                | 92.57             | 97.33             | 96.81             | 97.81             | 98.41             | 98.50             |
| MFCC + CQT         | 96.16             | 96.1              | 97.24             | 97.21             | 98.68             | 99.14             |
| MFCC + CC          | 94.79             | 95.7              | 95.09             | 96.11             | 96.15             | 96.72             |
| CQT + CC           | 92.16             | 97.25             | 97.77             | 98.06             | 98.55             | 98.82             |
| MFCC + CQT + CC    | 94.61             | 94.78             | 97.89             | 98.7              | 99.36             | 99.68             |

Figure 6: Predicted output and error analysis curves of the lute acoustic quality evaluation network.
24 sets as testing samples, and 10 sets as validation samples. According to the characteristics of the experimental samples and the structure of the network model, logsig (Sigmoid function) was used for the activation function of the implicit layer and linear purelin was used for the activation function of the output layer; the training function of the model network was trainlm, and the training algorithm was Levenberg–Marquardt algorithm. The optimal network parameter configuration was obtained by adjusting the network parameters through several experimental comparisons. With the optimal prediction evaluation results obtained, the number of layers of the hidden layers and the number of neurons in each layer were \{10, 20, 150, 50, 10, and 10\} [20, 21].

For further validation, the use of fused features (MFCC+CQT+CC) as feature parameters input to the BP neural network-based lute acoustic quality evaluation model was the best way to fuse the features. Additional comparison experiments were conducted, using single features and different combinations of features as well as exploratory experiments varying only in the number of training samples.

The results obtained in the preliminary experiments are shown in Figure 6. As can be seen in Figure 6, the predicted output values are very similar to the desired output values, with some individual samples not being predicted very well, but overall, the predictions are very good. The average accuracy of the test samples was 99.68%, and the average accuracy of the validation samples was also 99.49%.

The results of further exploratory experiments are shown in Table 1 and Figure 7, and the results obtained are all under the optimal network parameter configuration. From Table 1 and Figure 7, it can be seen that the average accuracy shows an overall increasing trend with the increase in the number of training samples and the prediction effect of the fused features (MFCC+CQT+CC) is stronger than several other feature fusions after the number of samples reaches 50 groups; on the whole, the single feature MFCC and its combined features (MFCC+CC) are not as effective in prediction.

The experimental results show that the best predictive evaluation is obtained by fusing the features (MFCC+CQT+CC), which maximizes the characterization of the pipa sound quality; the BP neural network-based pipa acoustic quality evaluation model constructed in this paper is reliable and has good predictive evaluation performance; the pipa acoustic quality evaluation method proposed in this paper is very novel and feasible.

6. Conclusions

Based on the subjective evaluation criteria of musical instruments, this paper proposes Softmax regression BP neural network multiclassification recogniser input based on obtaining the results of subjective evaluation of lute acoustic quality, similar to acoustic quality evaluation; the classification code of standard tones is used as the target label for supervised network learning, resulting in an optimised model for lute note recognition. The experimental results show that this scheme can achieve high precision recognition of 25 notes from bass to treble, with an average recognition accuracy of 95.6%.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.
References

[1] V. Eyharabide, I. E. I. Bekkouch, and N. D. Constantin, “Knowledge graph embedding-based domain adaptation for musical instrument recognition,” Computers, vol. 10, no. 8, p. 94, 2021.

[2] P. Peso Parada, D. Sharma, J. Lainez, D. Barreda, T. V. Waterschoot, and P. A. Naylor, “A single-channel non-intrusive C50 estimator correlated with speech recognition performance,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 4, pp. 719–732, 2016.

[3] M. Walczyński and J. Fica, “Comparison of selected acoustic signal parameterization methods in the problem of machine recognition of classical music styles,” in Progress in Image Processing, Pattern Recognition and Communication Systems, pp. 181–190, Springer, Cham, 2021.

[4] B. Jia, J. Lv, and D. Liu, “Deep learning-based automatic downbeat tracking: a brief review,” Multimedia Systems, vol. 25, no. 6, pp. 617–638, 2019.

[5] D. Kiela and S. Clark, “Learning neural audio embeddings for grounding semantics in auditory perception,” Journal of Artificial Intelligence Research, vol. 60, pp. 1003–1030, 2017.

[6] D. Herremans, C. H. Chuan, and E. Chew, “A functional taxonomy of music generation systems,” ACM Computing Surveys, vol. 50, no. 5, pp. 1–30, 2017.

[7] L. Verde, G. De Pietro, M. Alrashoud, A. Ghoneim, K. N. Al-Mutib, and G. Sannino, “Dysphonia detection index (DDI): a new multi-parametric marker to evaluate voice quality,” IEEE Access, vol. 7, pp. 55689–55697, 2019.

[8] Z. H. A. N. G. Zhengwan, Z. H. A. N. G. Chunjiong, L. I. Hongbing, and X. I. E. Tao, “Multipath transmission selection algorithm based on immune connectivity model,” Journal of Computer Applications, vol. 40, no. 12, p. 3571, 2020.

[9] L. Su, C.-C. M. Yeh, J.-Y. Liu, J.-C. Wang, and Y.-H. Yang, “A systematic evaluation of the bag-of-frames representation for music information retrieval,” IEEE Transactions on Multimedia, vol. 16, no. 5, pp. 1188–1200, 2014.

[10] X. L. Zhang and J. Wu, “Deep belief networks based voice activity detection,” IEEE Transactions on Audio Speech and Language Processing, vol. 21, no. 4, pp. 697–710, 2012.

[11] D. G. Bhalke, C. B. R. Rao, and D. S. Bormane, “Automatic musical instrument classification using fractional fourier transform based- MFCC features and counter propagation neural network,” Journal of Intelligent Information Systems, vol. 46, no. 3, pp. 425–446, 2016.

[12] V. Lostanlen, J. Andén, and M. Lagrange, “Extended playing techniques: the next milestone in musical instrument recognition,” in Proceedings of the 5th International Conference on Digital Libraries for Musicology, pp. 1–10, Paris, France, 2018, September.

[13] R. Martinek, M. Kelnar, J. Vanus, P. Biliš, and J. Zidek, “A robust approach for acoustic noise suppression in speech using ANFIS,” Journal of Electrical Engineering, vol. 66, no. 6, pp. 301–310, 2015.

[14] C. Halkiopoulos and B. Boutsinas, “Automatic interactive music improvisation based ON data mining,” The International Journal on Artificial Intelligence Tools, vol. 21, no. 04, Article ID 1250016, 2012.

[15] P. Jain, S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, “A review of machine learning applications in wildfire science and management,” Environmental Reviews, vol. 28, no. 4, pp. 478–505, 2020.

[16] I. E. I. Bekkouch, N. D. Constantin, V. Eyharabide, and F. Billiet, “Adversarial domain adaptation for medieval instrument recognition, Lecture Notes in Networks and Systems,” in Proceedings of the SAI Intelligent Systems Conference, pp. 674–687, Amsterdam, 2021 September.

[17] D. G. Bhalke, C. B. R. Rao, and D. S. Bormane, “Fractional fourier transform based features for musical instrument recognition using machine learning techniques,” in Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013, pp. 155–163, Bhubaneswar, Odisha, India, 2014.

[18] C. Dai, H. Che, and M.-F. Leung, “A neurodynamic optimization approach for L1 minimization with application to compressed image reconstruction,” The International Journal on Artificial Intelligence Tools, vol. 30, no. 1, Article ID 2140007, 2021.

[19] Y. Wang, J. Wang, and H. Che, “Two-timescale neurodynamic approaches to supervised feature selection based on alternative problem formulations,” Neural Networks, vol. 142, pp. 180–191, 2021.

[20] J. H. Calvo-Zaragoza and A. Pacha, “Understanding optical music recognition,” ACM Computing Surveys, vol. 53, no. 4, pp. 1–35, 2020.

[21] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghameemaghami, and C. Fookes, “A robust interpretable deep learning classifier for heart anomaly detection without segmentation,” IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 6, pp. 2162–2171, 2020.