www.cadanda.com
Computer-Aided Design AND APPLICATIONS

Timbre Classification Method based on Computer-Aided Technology for Internet of Things

Dingding Liu[1,*] and Su Bu[2]

[1]Arts Education Centre, Zhengzhou University of Light Industry, Zhengzhou 450002, China, 2002045@zzuli.edu.cn
[2]Academy of Fine Arts, Shangqiu Normal University, Shangqiu 476000, China, busu1978@163.com

Corresponding author: Dingding Liu, 2002045@zzuli.edu.cn

Abstract. The timbre of different melodies may have different personalities, so as to express different feelings and artistic styles. The effective extraction of timbre information is the key to successfully identify musical instruments. In order to solve the problems of low recognition accuracy and high time cost in the current timbre recognition system, an intelligent musical instrument timbre classification system is proposed and designed in combination with the computer-aided technology of the Internet of things. Firstly, a 5-Dimensional emotion space is determined by MDS method. According to the 5-Dimensional emotion space, the emotion evaluation experiment is carried out, and the reliability and validity of the experimental data are tested and the noise is eliminated. Then, the effects of performance content, time domain characteristics and instrument type on the relationship between timbre perception characteristics and emotion are studied. It is found that the time domain characteristics and performance content have little impact on the relationship between timbre perception characteristics and emotion, and the instrument type will have a certain impact on the relationship between timbre perception characteristics and emotion. Finally, five emotion prediction models are established by using multiple linear regression algorithm, and the models have good prediction ability for the five emotions. Simulation and experimental results show that the proposed system can quickly extract the characteristics of harmonic structure of musical signal, and the timbre recognition system based on this can well reflect the timbre characteristics of musical instruments, which provides a new idea for the feature extraction of musical signal.

Keywords: Internet of things; computer aided design; timbre; feature extraction; musical instrument classification.
DOI: https://doi.org/10.14733/cadaps.2023.S2.167-179

1 INTRODUCTION

Timbre is a subjective attribute of sound perception, not a purely physical attribute. From the perspective of acoustics, the timbre of an instrument is determined by the vibration state of its pronunciation part, and the harmonic ratio of each order in overtone determines its timbre. Research based on psychological music timbre judgment, sound physical feature analysis and machine learning methods all show that timbre is a multi-attribute sound feature including spectrum and time. People usually have the ability to listen to sounds and distinguish people, because everyone pronounces different timbres. For example, the male voice is generally deep and thick, while the female voice is clear and loud. The same is true of music. Different musical instruments produce different timbres; the same piece of music played on the piano is different from that played on the violin or other musical instruments. Based on this, the research and analysis of the essence of sound timbre and the identification of the characteristics that can represent timbre have an extraordinary role and significance for the classification and recognition of audio signals.

Timbre estimation is widely used in content-based music transcription, structured audio coding, music recommendation and query engine, music commentary and so on. In terms of musical instrument recognition itself, timbre is the fundamental basis for realizing musical instrument recognition. Existing instrument recognition systems focus on the use of features with timbre meaning, or multi feature fusion, combined with classifiers. Benetos et al. [1] mentioned that although researchers have explored the extraction of timbre information from multiple perspectives, the mathematical model related to timbre is still not perfect because of the complex relationship between timbre and the sound mechanism of musical instruments and the auditory process of human ears. In addition, the subjective color of timbre is also one of the obstacles. Researchers usually explore the possibility of improving the performance of musical instrument timbre recognition from the perspective of timbre features and classifiers. With the passage of time, the performance of musical instrument timbre recognition is also on the rise, and the time for its wide application is becoming more and more mature. Musical instrument timbre recognition can be applied to music search engine, music teaching, music creation, music evaluation and so on. Music is full of all aspects of people's life, and everyone has different needs for music. Musical instrument timbre recognition is also the key to the realization and optimization of personal music recommendation. In addition, the development of musical instrument timbre recognition can also promote the progress of related fields, such as multi fundamental frequency estimation of music, music genre recognition, speaker recognition and so on. Many phonetic timbre features are used to study the timbre of musical instruments, such as spectral centroid, spectral flux, spectral roll off, cepstrum, etc. However, because the sounding mechanism and pitch range of musical instruments are different from those of voice, it is necessary to verify whether the voice timbre features are effective for the direct use of musical instruments.

Human society is advancing, and the requirements for music are also improving. With the accumulation of time, the music content will be more and more abundant, and the music range will be more and more extensive. The amount of music data is exploding. Jiang et al. [2] mentioned that in the era of big data, how to efficiently and accurately classify music, quickly retrieve music content of interest to individuals and make accurate personalized recommendations for music lovers is a difficult problem. This paper studies and analyzes music from the basic elements of music. The ultimate goal is to try to find an efficient and reliable timbre classification method, so that individuals can accurately and efficiently find out the music they want to listen to from a large number of music in a short time. In this paper, through the analysis of music timbre, some parameters representing music timbre characteristics are verified, so as to realize the classification of music based on musical instrument timbre, and design a timbre classification system based on computer-aided technology for the Internet of things, which plays an extremely important role and significance in the follow-up projects in music retrieval and recommendation related fields.

2 RELEVANT THEORIES AND TECHNOLOGIES

2.1 Rudimentary Knowledge of Music

Music is a kind of sound signal and one of the ideological carriers of people's life. People often use music to express their thoughts and emotions. From the perspective of sound waves, music is the synthesis of a group of regular sound signals, with rich frequency components. The frequency range of sounds emitted by different musical instruments is different, and the music emitted by various musical instruments is different, mainly because the harmonic energy distribution of different frequency components is inconsistent, and the combination of different harmonic components forms a variety of different music. From the perspective of subjective feeling, music consists of several basic elements, such as pitch, value, volume, timbre, etc. Then these basic elements constitute the common form elements of music, such as rhythm, harmony and so on. The elements of music are shown in Figure 1.

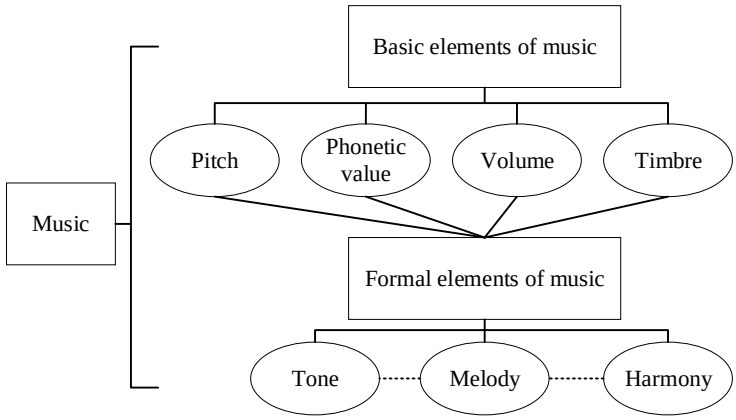

Figure 1: Elements of music.

It can be seen from the figure that the pitch is the height of the sound, which is determined by the vibration frequency of the pronunciation object. The higher the frequency, the higher the pitch. Pitch is one of the decisive phonemes of the tone of a sound. In music theory, the height of the tune tone is called tone. For pure tone, the tone increases with the increase of the vibration frequency of the pronunciation body, and the volume also has a certain impact on the tone.

Volume refers to the intensity of the sound, which is used to describe the vibration amplitude of the sound body. The greater the vibration amplitude of the speaker, the greater the volume. Erickson [3] mentioned that the influence of volume on tone is as follows: the tone of low-frequency tone decreases with the increase of volume, while the tone of high-frequency pure tone increases with the increase of volume.

Tone value refers to the length of sound, which is used to represent the length of continuous vibration time of the pronunciation body. The longer the object vibrates, the greater the sound value. Tone value is the main element of music melody, and the combination of different length tones constitutes the beat and rhythm of music, thus forming melody.

Timbre refers to the color of sound. Generally, the pronunciation of musical instruments will produce one pitch and several overtones, and the energy distribution is different. Timbre is used to describe people's subjective feelings about different sounds. People can easily distinguish whether a song is played by piano or violin because the timbre of the two instruments is different. The timbre is determined by the number and strength of overtones produced by the pronunciation body. The distribution of overtones and overtones produced by different instruments is different. These overtones and pitch together form the unique timbre of various instruments.

Timbre is the basis of identifying musical instruments, a kind of characteristic information different from other musical instruments. At present, there is no exact physical quantity that directly reflects the timbre, but the spectrum can be used to characterize the timbre. Musical instruments usually produce sound waves of multiple frequencies, which are called harmonics or overtones. In addition to the fundamental frequency, the proportion of these harmonic components determines the timbre, also known as tonal timbre. The difference in hearing of different musical instruments is mainly based on the different timbres. The human auditory system can distinguish which musical instrument the sound comes from. The reason is that the high-frequency overtone components of musical instruments are different, and different musical instruments have different overtone proportions.

2.2 Formation Mechanism of Musical Instrument Timbre

Timbre is a simple and vague word. It is a subjective auditory attribute, involving many complex physiological, psychological and musical problems. Each musical instrument, each person's voice band and all vibrating objects in the world will produce different sounds because they can emit different vibration frequencies. They all have their own distinctive characteristics, which can be expressed by instruments. Timbre refers to a certain attribute of sound produced in hearing. The listener can judge the difference between two sounds that are presented in the same way and have the same pitch and loudness according to this attribute. In other words, timbre is used to describe the auditory quality and sound properties of a sound with a specific pitch and loudness. Closely related to timbre is the fundamental frequency, the lowest frequency of sound wave; Secondary and higher frequencies are called harmonics. The difference in human auditory perception of different musical instruments is mainly due to the different timbre of each other. The human ear can recognize the sounds of these musical instruments, but this recognition process is too subjective, which is not conducive to the automatic classification and recognition of musical instruments. Although listeners can make quantitative analysis on the pitch and loudness of music, the timbre is subjective and difficult to be described by an exact mathematical model.

Timbre is a special member of music attributes. The different timbres make vocal music works show various features and meet the aesthetic needs of different listeners. Vocal music art contains a variety of different singing categories. As an important component, bel canto's development and timbre cannot be underestimated. Timbre is a necessary means to convey the content and connotation of vocal music works. The beauty of sound is the intermediary of emotional transmission. In the performance, the use of different timbres is enough to affect the transmission of music content. Different timbres cross each other in the expression to achieve the performance of music works. At the same time, the use of timbre is also the expression of the singer's personal aesthetic habits and thinking. Jathal [4] mentioned that only by injecting emotion into timbre can the audience share. It is through this element that the soul and connotation are shaped during bel canto singing. In front of singers who have accumulated experience and time, through various forms of vocal training, they can acquire the ability to master different timbres, and combine timbre and music image through personal input in singing, use timbre to express the content and verve of vocal music works, and show the beauty in sound.

2.3 Connotation of Computer-Aided Technology for Internet of Things

Generally speaking, there are three key technologies in the operation of the computer Internet of things: wireless sensing technology, wireless radio frequency identification technology and intelligent processing technology. These three key technologies basically cover the three important links of sensing, identification and processing in the operation of the computer Internet of things. First, wireless sensor technology. Wireless sensor technology is actually a kind of detection device. It can generate a signal source through the measurement of an object, so as to react to the information source, so as to feedback and output the information, and finally achieve the purpose of information transmission, storage and recording processing. With the progress of science and technology, especially the development of nanotechnology, wireless sensor technology is gradually

developing towards miniaturization and miniaturization. Secondly, RFID technology. Radio frequency identification (RFID) technology can help people identify objects in different states, especially in harsh environments. The application of radio frequency identification technology is very popular, for example, in the automatic toll collection system of Expressway and the logistics supply chain management, this technology has a considerable popularity and application. The composition of RFID technology is relatively simple. It is mainly composed of reader, tag and antenna. The three components bear the functions of identification, writing and transmission respectively. Finally, intelligent processing technology. Intelligent processing technology is the core of Internet of things technology and the key to realize information combination between objects. All kinds of objects and devices realize information interaction and instruction execution on the basis of intelligent processing technology. Through intelligent processing technology, the flexibility of objects can be realized, so as to achieve the purpose of intelligent operation.

3 TIMBRE FEATURE PARAMETER EXTRACTION AND ANALYSIS SYSTEM DESIGN

3.1 Analysis of Time-frequency Domain Characteristics of Musical Instrument Timbre

Sound is produced by the vibration of an object in a medium. The combination of overtone and pitch brings a special feeling to the listener, and this special subjective feeling is called the timbre of sound. Paquette et al. [5] mentioned that timbre, the color of sound, is one of the four basic elements of music signal. Through the perception of timbre, people can easily distinguish what instrument a piece of music is played by and what kind of emotion the music wants to express. From the perspective of musical instrument timbre, not only different musical instruments have different timbres, but even the same musical instrument will show different timbres due to different manufacturing processes and materials. From the perspective of people's subjective feelings, timbre represents the ideological content of music. Through hearing, listeners can feel the content and thoughts expressed by a piece of music, such as positive, low, slow, lightness, depression, etc. When the color of sound changes in sequence with any time scale, the texture of sound changes accordingly. Different musical instruments emit different sounds. In the time domain, the duration of single tone emitted by different musical instruments is different, and the signal energy distribution in different time periods is also different. Generally, the single tone signal of an instrument is divided into four stages in time domain according to its envelope, including playing, attenuation, persistence and disappearance.

For a single tone played by different instruments, the duration of each stage of the time domain envelope is different. For example, for string percussion instruments, sound is generated through the vibration of the strings. Lega et al. [6] mentioned that the attenuation and duration of each single tone is long, and the amplitude changes slowly with time until the single tone disappears, which is a relatively slow attenuation process. The auditory feeling of this kind of instruments is rich and surrounding timbre. For percussion instruments, external force is used to strike the instrument itself to make the instrument vibrate and pronounce. This kind of instrument has no tone change and gives people a simple and straightforward sound. In the time domain, the monophonic change of this kind of musical instrument is relatively rapid. The duration of attenuation phase is very short, and there is basically no continuous phase. It is a relatively rapid attenuation process. As for noise instruments, there is no rule in their pronunciation. Just like noise, the energy distribution of time-domain waveform is irregular. Therefore, no change can be seen from the time domain envelope of this kind of musical instrument, and even the four stages of music signal cannot be felt. Timbre has very obvious time domain characteristics. There are some characteristic parameters in the time domain that can be used to describe timbre, so as to distinguish different types of musical instrument timbre

According to the definition of timbre, timbre also has some frequency domain characteristics in addition to the time domain characteristics. In the frequency domain, timbre is a combination of pitch and overtone. Generally speaking, the higher the frequency of the signal, the clearer the timbre; the lower the frequency, the lower the tone. In terms of frequency relationship, the

frequency of overtones that enhance the pitch is close to the integer multiple of the pitch frequency, and the combination of multiple overtones affects people's subjective feelings. Different frequency bands have different effects on the timbre. Different musical instruments have their own unique timbre because their sound effects are in different frequency bands and their proportion of overtone components is also different. This difference in timbre makes it easy for people to distinguish the music played by different musical instruments. Different timbres have different frequency characteristics. Analyzing the frequency components of musical instrument timbre and finding out the frequency domain characteristic parameters of their timbre are helpful to accurately classify music by timbre [7].

The timbre characteristics in frequency domain reflect the physical characteristics of sound from different aspects. The spectral centroid is a measure of the brightness of the sound. The spectral roll off usually indicates the frequency asymmetry in a frame. Spectral roll off and spectral centroid reflect the distribution of signal energy in frequency. Spectral flux is a measure of the change of spectral energy between successive tone tones, which reflects the dynamic characteristics of tone signals. The music has obvious harmonic structure, and the energy is mainly concentrated in the lower harmonics. Different musical instruments contain different harmonic times. The music with rich high-order harmonics and large amplitude sounds brighter. The harmonic number of sound reflects the timbre. Generally, the lower order harmonics are the most important, while the higher order harmonics do not contribute significantly to the timbre.

3.2 Timbre Feature Extraction Method

Because the original music signal may contain noise components, it is not suitable to directly extract the timbre feature parameters, so it is necessary to preprocess the target music signa. Music signal pre-processing includes sampling quantization, pre-emphasis processing, sub frame windowing and so on. Each sample must be quantized at the sampling point. The more quantized digits, the higher the precision. Generally, 8 or 16 digits are taken [8].

Because of the principle of sound pronunciation, the amplitude of high-frequency formant is lower than that of low-frequency formant. The higher the frequency, the smaller the spectral value. In order to improve the high-frequency resolution of music signal, a pre-emphasis processing method is proposed by analyzing the overall spectrum of the whole frequency band. Before extracting the characteristic parameters, the pre-emphasis is usually realized by a first-order digital filter.

Like voice signals, music signals can be considered stable in a relatively short period of time. Here, the feature extraction of music signal is based on steady-state signal. Therefore, before extracting the features of music signals, it is usually necessary to perform frame segmentation, that is, the signals are divided into a small segment of signals with stable statistical characteristics, and each small segment of signals requires a frame [9].

The workflow of the voice classification system is shown in Figure 2. It can be seen from the figure that the work is mainly divided into two parts: first, spectrum analysis. This part of the function flow first calculates the energy spectrum of the music signal according to the spectrum, then calculates the energy component ratio of each frequency band according to the energy spectrum, and finally processes and outputs the results. Second, timbre classification. Firstly, the input signal is extracted, and then the extracted feature vector is classified according to the learning samples in the BP neural network classifier, and the classification results are obtained and output. At the same time, judge whether the classification results are accurate manually according to the classification results. If the results are not accurate, the information will be edited and classified as a part of the learning sample set. When the sample set changes greatly compared with the original sample set, the second learning training will be carried out, so that the system can adjust itself according to the situation.

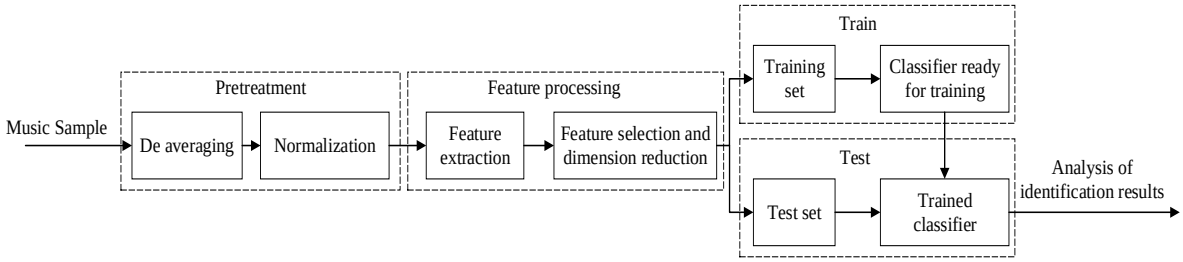

Figure 2: General framework of musical instrument timbre recognition.

3.3 Design of Timbre Classification System

The extraction framework of timbre descriptor features includes time-domain features, frequency-domain features and harmonic features, but not all features are conducive to music modeling. The timbre classification system mainly includes two functional modules. One is the timbre component analysis module of music, which includes the extraction, preprocessing and FFT transformation of the music signal, calculates the energy component ratio of each frequency band in the music signal according to the spectrum diagram, and extracts the timbre feature parameter vector in the music signal; the second is the musical instrument classification module. After establishing the music database, BP neural network algorithm is used to design an appropriate training model to train and learn the timbre feature vector in the database, so as to realize the correct musical instrument classification of the input music signal.

The functional structure of the timbre classification system is shown in Figure 3. It can be seen from the figure that the whole timbre classification system is mainly composed of three parts: background offline sample learning, intermediate data processing layer and front-end display layer. The background offline sample learning layer is mainly used for offline sample learning in the background. The main work of the background offline part is sample collection, feature extraction and BP neural network modeling; at the same time, the background offline part is also responsible for the persistence of the classification model in memory. CRC is used to verify whether the training set has changed. If it has changed, a new round of training and learning will be carried out at this layer. The intermediate data processing layer is actually a background web system, which is mainly responsible for analyzing the music files submitted by users. The middle layer interacts with the background and front end [10]. In the interaction with the bottom layer, the main work is to load the classification model of the background offline layer into the memory for instrument recognition; In the interaction with the front end, there are two main tasks: one is to process and analyze the music files uploaded by users, and save the wrong music timbre characteristic parameters to the database when necessary; The second is to return the data of analysis results to the user. The main work of the front-end display layer is to display the timbre analysis results. After the middle layer processes the input music, it will send the result data to the web front-end. After receiving these data, the front-end will visualize the data using HTML and other data processing tools to display it in a friendly way.

4 SIMULATION ANALYSIS OF DIFFERENT TIMBRE CLASSIFICATION MODELS

This paper selects eight kinds of musical instruments as sound sources, four of which are Western instruments: Piano, violin, saxophone and guitar; There are four kinds of oriental musical instruments: pipa, zither, flute and erhu. In addition, among the eight musical instruments, violin, piano, guitar, pipa, zither and erhu are string instruments, while flute and erhu are wind instruments. In this way, the timbre of musical instruments with similar pronunciation principles can be compared and analyzed. In the selection of music repertoire, the influence of performance style is also taken into account.

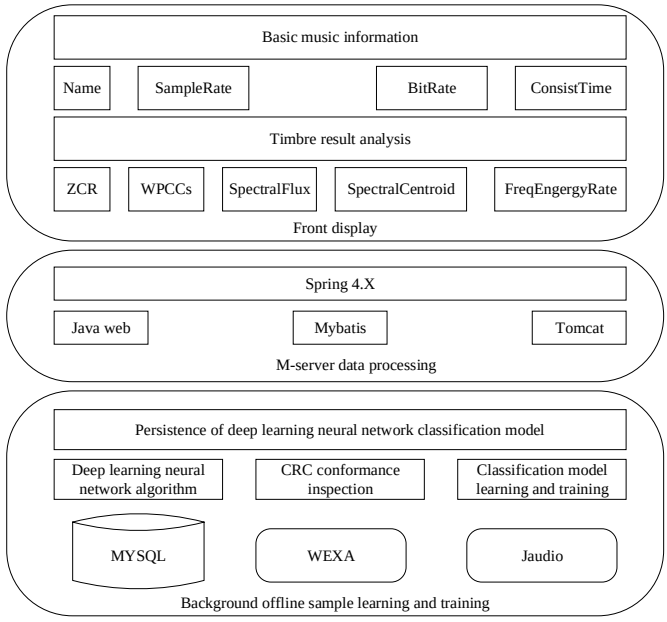

Figure 3: Functional structure of the timbre classification system.

For the repertoire of each kind of musical instrument, a considerable number of repertoires with fast, medium and slow playing rhythm are selected. Figure 4 shows the timbre estimation results obtained by using different classification methods. It can be seen from the figure that this paper uses a total of 950 music clips, each of which is a solo of an instrument. The experiment collected the track data of 8 kinds of musical instruments, including 4 kinds of Western musical instruments: Piano, violin, saxophone and guitar; Four kinds of oriental musical instruments: pipa, zither, erhu and flute. Among the 950 music clips, there are 98 flutes, 158 erhu, 122 piano pieces, 122 zither pieces, 100 guitars, 146 pipas, 132 saxophones and 72 violins.

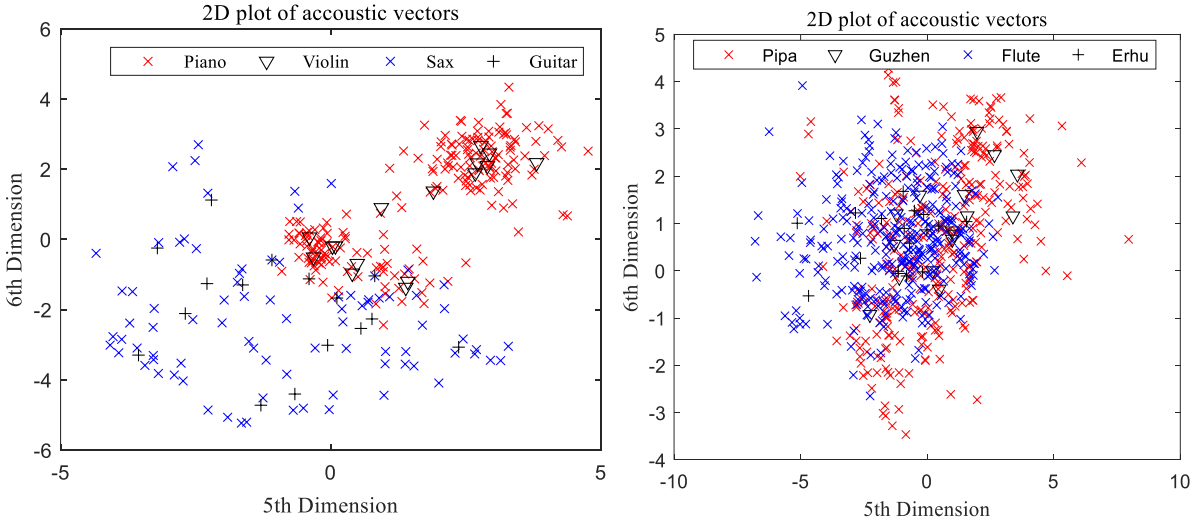

Figure 4: Timbre estimation results obtained by using different classification methods.

Figure 5 shows the classification results of musical instruments based on combined features. It can be seen from the figure that the classification accuracy of the four classifiers is above 92%, of which SVM classifier is the worst, with an accuracy of 92.4%, while BP neural network is the best, with an accuracy of 99.15%. For a single instrument, 100 guitar samples are all correctly classified in the four classifiers, indicating that the guitar timbre features are obvious and easy to distinguish; there are a few errors in the classification of piano and flute timbres in the four classifiers. Most of the error types of piano timbres are classified as zither, and most of the error classifications of zither are classified as piano, which shows that the timbres of piano and zither are relatively similar and easy to be confused. The error of the flute is more complex, indicating that there is still effective information about the tone of the flute that has not been found. Violin and Saxophone are also easy to be confused, which shows that violin timbre and Saxophone have some things in common, and they can bring similar auditory feelings to people. The musical instrument family generally has a high recognition rate because the timbre characteristics of the musical instrument family are in a certain range, and the recognition of a single musical instrument is to find specific corresponding points in this range. Similar musical instruments are similar in sound, and people cannot distinguish them, such as violin and viola, flute and alto flute; but it is easy to distinguish different musical instruments.

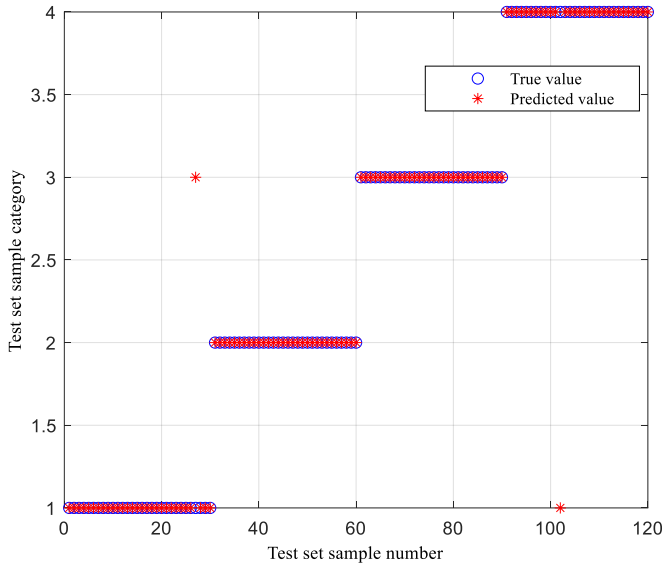

Figure 5: Classification results of musical instruments based on combined features.

Figure 6 shows the timbre analysis results of the test track. It can be seen from the figure that for the music played by the flute with a duration of 10s, the energy proportion of each frequency band is different. The music is also very rich in high-frequency components. The 4kHz-16kHz frequency band accounts for about 0.35 of the whole, which makes the whole music sound very spatial, And the sound is clear and powerful. In addition, according to the characteristic parameter vector of the music, the result of classification using BP neural network classification model is flute, which is consistent with the expectation, which shows that the system has made a correct prediction for the musical instrument of the music. Therefore, time-frequency features are more conducive to the recognition of timbre than cepstrum features. In addition, when the excited state of percussion instruments with pitch plays an absolute role in timbre, if time-domain information is not considered, the recognition of percussion instruments will be poor.

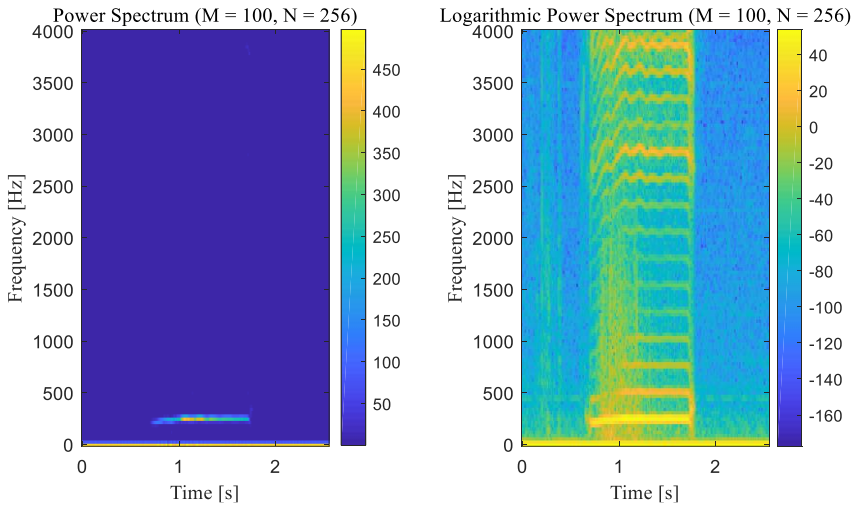

Figure 6: Timbre analysis results of the test track.

Musical instrument types represent the timbre of different musical instruments and have the greatest impact on timbre, so it is considered that musical instrument types have the greatest impact on timbre and emotional correlation; the time-domain characteristics represent the vibration change of the tone color, which will produce different impact and impact. Figure 7 is the isophonic curve of musical instrument timbre transparency and spectral centroid. It can be seen from the figure that the time domain characteristics will also affect the correlation between timbre and emotion; the performance content is mainly related to the pitch of the music and basically has no impact on the timbre. Therefore, compared with the other two factors, the performance content has little impact on the timbre and emotion. For example, playing scales and melodies with the same instrument will not have a greater impact on the timbre emotion due to different performance content. Bright, crisp, and thin are close, thick and thick are close, indicating a high similarity in perception. Concord, purity and roughness, hoarseness are far away, crisp, bright, thin and thick, dim and thick are far away, indicating a great difference in perception. This also shows that the more features, the higher the accuracy of classification, which also proves the importance of feature selection Based on the other four evaluation indexes, the maximum total accuracy of the algorithm is obtained, which shows the effectiveness of the timbre features proposed in this paper.

Figure 8 shows the timbre differences of musical instruments under different tunes. It can be seen from the figure that for the overall classification results, the classification error rate of CQT frequency domain feature combination is the smallest, and the classification error rate of MCQT frequency domain feature combination is the largest. Within the musical instrument family, the classification error rate of MCQT frequency domain feature combination is the smallest, which is much better than CQT frequency domain feature combination and frequency domain feature combination. Among musical instrument families, the classification error rate of CQT frequency domain feature combination is the smallest. The mean and standard deviation of six different frequency domain feature parameters show the characteristics of the spectrum in different aspects, that is, they reflect different aspects of timbre. The overall classification result obtained by CQT frequency domain feature combination is the best, indicating that if there are appropriate methods to extract the features under the spectrum, the features under CQT are expected to achieve the best overall classification result. To sum up, the timbre of musical instruments is related to the category of musical instruments. The timbre perception of the same category of musical instruments is similar, and the timbre perception of different categories of musical instruments is quite different.

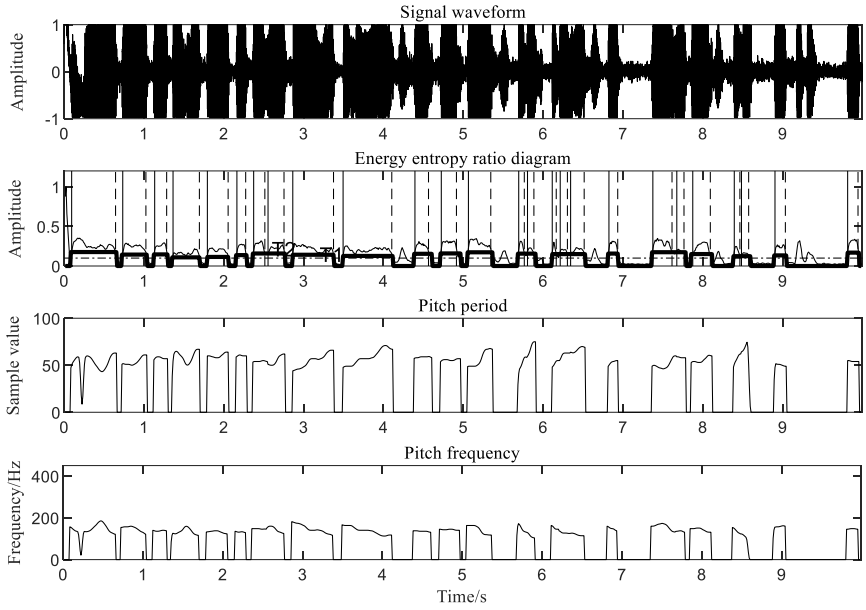

Figure 7: Isophonic curve of musical instrument timbre transparency and spectral centroid.

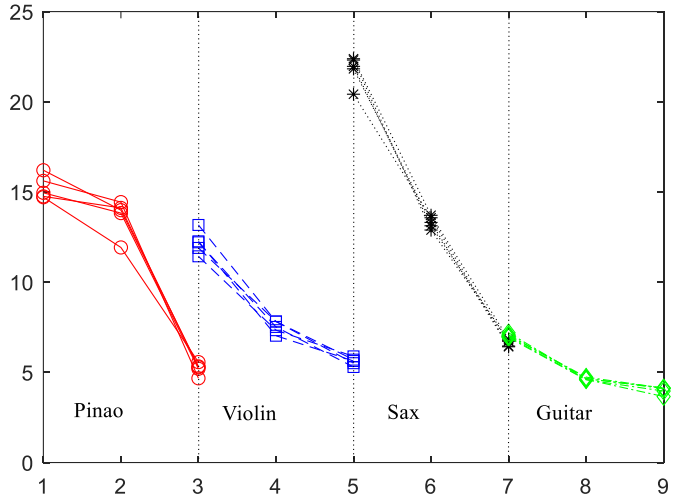

Figure 8: Timbre differences of musical instruments under different tunes.

5 CONCLUSION

Timbre feature extraction plays a key role in the research of music signal feature extraction and musical instrument recognition. According to the 5-Dimensional emotion space, the emotion evaluation experiment is carried out, and the reliability and validity of the experimental data are tested and the noise is eliminated. The effects of performance content, time-domain characteristics and instrument types on the relationship between timbre perception characteristics and emotion are studied. The results show that the time-domain characteristics and performance content have

little effect on the relationship between timbre perception characteristics and emotion, the type of musical instrument will have a certain impact on the relationship between timbre perception and emotion. Then, five emotion prediction models are established by using multiple linear regression algorithm. The results show that the model has good prediction ability for the five emotions. Finally, from the two aspects of the similarity analysis of some dimensions of each inverted frequency domain feature in different pitches of the same instrument and the same pitch of different instruments, and the importance of each dimension of each inverted frequency domain feature, the most favorable voice color classification part of each inverted frequency domain feature is obtained. The extracted features are used to classify, combine and compare the classification results.

The work of this paper is only a small part of musical instrument timbre recognition. Although it improves the effect of musical instrument recognition to a certain extent, there are still many problems to be further solved. For example:

(1) When extracting the harmonic structure of music signal, we can consider combining Mel auditory mechanism and adopting nonlinear discrete harmonic transformation to make the calculated discrete harmonic coefficient more in line with the auditory characteristics of human ears, in order to better reflect the timbre characteristics of musical instruments.

(2) In terms of the application of musical instrument recognition, it is considered to discuss more machine learning methods, such as big data, deep learning, artificial intelligence, etc., and to introduce the timing characteristics of musical signal into feature extraction and classification, in order to further improve the accuracy of musical instrument recognition.

(3) In this paper, the object of instrument recognition is pure single tone, but the actual music is mostly compound tone. The recognition of musical instruments with compound tones not only needs to consider the problems of sound source mixing and note overlap, but also needs to consider the content of music theory.

Dingding Liu, https://orcid.org/0000-0001-7682-1604
Su Bu, https://orcid.org/0000-0001-9533-2961

REFERENCES

[1] Benetos, E.; Dixon, S.; Ewert, S.: Automatic music transcription an overview, IEEE Signal Processing Magazine, 36(1), 2019, 20-30. https://sci-hub.et-fine.com/10.1109/msp.2018.2869928

[2] Jiang, W.; Liu, J.-Y.; Jiang, Y.-J.: Analysis and modeling of timbre perception features in musical sounds, Applied Sciences-Basel, 10(3), 2020, 2-23. https://sci-hub.et-fine.com/10.3390/app10030789

[3] Erickson, M.-L.: Inexperienced listeners' perception of timbre dissimilarity within and between voice categories, Journal of Voice, 34(2), 2020, 1-13. https://sci-hub.et-fine.com/10.1016/j.jvoice.2018.09.012

[4] Jathal, K.: Real-time timbre classification for tabletop hand drumming, Computer Music Journal, 41(2), 2017, 38-51. https://sci-hub.et-fine.com/10.1162/comj_a_00419

[5] Paquette, S.; Takerkart, S.; Belin, P.: Cross-classification of musical and vocal emotions in the auditory cortex, Annals of the New York Academy of Sciences, 1423(1), 2018, 329-337. https://sci-hub.et-fine.com/10.1111/nyas.13666

[6] Lega, C.; Cattaneo, Z.; Rinaldi, L.: Instrumental expertise and musical timbre modulate the spatial representation of pitch, Quarterly Journal of Experimental Psychology, 73(8), 2020, 1162-1172. https://sci-hub.et-fine.com/10.1177/1747021819897779

[7] Wang, K.-M.; Hui, L.: Effectiveness evaluation of Internet of Things-aided firefighting by simulation, Journal of Supercomputing, 76(3), 2020, 1383-1397. https://sci-hub.et-fine.com/10.1007/s11227-017-2098-3

[8] Bai, L.; Han, R.; Zhang, W.: Random access and detection performance of Internet of Things for smart ocean, IEEE Internet of Things Journal, 7(10), 2020, 9858-9869. https://sci-hub.et-fine.com/10.1109/jiot.2020.2990164

[9] Yang, L.-L.; Zhang, S.: English teaching model and cultivation of students' speculative ability based on internet of things and typical case analysis, Journal of Intelligent & Fuzzy Systems, 37(5), 2019, 5983-5991. https://sci-hub.et-fine.com/10.3233/jifs-179180

[10] Xu, Y.-Z.; Holanda, G.; Reboucas, P.-P.: Deep learning-enhanced Internet of medical things to analyze brain CT scans of hemorrhagic stroke patients: a new approach, IEEE Sensors Journal, 21(22), 2021, 24941-24951. https://sci-hub.et-fine.com/10.1109/jsen.2020.3032897