Received May 30, 2021, accepted June 21, 2021, date of publication July 6, 2021, date of current version July 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3095197

Research on the Correlation Between the Timbre Attributes of Musical Sound and Visual Color

JINGYU LIU, ANNI ZHAO, SHUANG WANG, YIYANG LI, AND HUI REN
School of Information and Communication Engineering, Communication University of China, Beijing 100024, China

Corresponding author: Hui Ren (renhui@cuc.edu.cn)

This work was supported by the National Key Research and Development under Grant 2018YFB1403703.

ABSTRACT This article investigated the timbre of musical sound and visual color; comprehensively used experimental psychology, audio information processing and other experimental methods and technical means; and conducted research on the relationship between timbre attributes and visual color. First, we designed and implemented a subjective perception experiment based on audiovisual cross-modal timbre-color correlation. Through the statistical analysis of the experimental data, we obtained a complete timbre-color cross-modal correlation dataset. Second, through the visual processing and correlation analysis of the timbre-color correlation data, it is proven that there is a certain correlation between the timbre dimension and the color dimension. On this basis, we used three algorithms to construct a timbre-color correlation model, namely, multiple linear regression, BP neural network and SVR, and we verified the accuracy of the three models. The timbre-color correlation dataset constructed in this paper can provide basic data support for audiovisual cross-modal research. The timbre-color correlation model constructed in this paper can provide a theoretical basis for cross-modal audiovisual applications. In addition, the timbre-color cross-modality research method in this paper can provide new research ideas for audio-visual cross-modality research.

INDEX TERMS Timbre, color, audiovisual cross-modality, timbre feature extraction, timbre-color correlation model.

I. INTRODUCTION

A. BACKGROUND AND SIGNIFICANCE OF THE RESEARCH

The connection between vision and hearing refers to a phenomenon of mutual association between vision and hearing in people's mind, which is a synaesthesia phenomenon. Synaesthesia is widespread in nature and has a long history. It has been widely used in painting, architecture, environmental layout, pattern design and so on. In psychology, synesthesia is defined as mental activity that induces another feeling from one sense; that is, stimulation of one sense organ causes other sense organs to be stimulated [1]. Synaesthesia involves a variety of senses, such as vision, hearing, touch, taste. Audiovisual synaesthesia belongs to one of them. The connection between music and color belongs to one kind of audiovisual synaesthesia. People associate music with color psychologically. The famous master of color composition and modern French composer Messiaen once described this feeling "when I think of chords and sound complexes, they always have a combination of color".

At present, much research has been conducted in the physiology and psychology fields on the phenomenon, process, mechanism and law of multisensory information integration; these research endeavors and have led to the proposal of a variety of theoretical hypotheses, strategies and models for multisensory information integration. For example, Calvert *et al.* used fMRI to investigate the neural mechanism that integrates visual and auditory speech signals. As a result, a significant integration effect was detected in the posterior part of the left superior temporal sulcus [2]. Later, by presenting synchronous and asynchronous visual (black and white alternate checkerboard stimuli) and auditory white noise stimuli, as well as audiovisual single sensory stimuli, they found that synchronous and asynchronous audiovisual stimuli produced super-increased and inhibited responses in the superior temporal sulcus [3]. In the PET study by Bushara *et al.*, participants were asked to determine whether pure auditory tones and colored circles presented in visual form were present at the same time. It was found that the

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

prefrontal lobe and the posterior parietal lobe are part of the multisensory brain area network that detects the synchronization of audiovisual stimuli [4]. Therefore, many research results show that the interactive integration effect of vision and hearing is universal for people of different ages, cultures, and genders, and has considerable stability.

Timbre and color have similar characteristics in some respects, which are manifested in the blending and connection of the two fields of music and painting. The famous musician Marion once said that ''Sound is audible color, and color is visible music'' [5]. This kind of cross-art communication makes audiovisual synaesthesia more fully integrated with artistic expression. In the art world, the color of painting is often described in the language to describe timbre, such as harmony and disharmony. A group of music painters have also emerged. These individuals try to express music with painting and use musical terms to improve themselves, and form the special expression language of modern painting. In the field of music, the form, style and content of music can also be expressed in language describing colors [6]. Some famous composers have put forward specific views on the relationship between color and tonality, such as the Western composer Rimsky-Korsakov believed that the key of C is white, the key of G major is yellow, and the key of D major is golden [7]. The composer Alexander-Skryabin thought the key of C major is red, the key of G major is orange or rose, the key of D major is warm gray and black [8]. Although they have their own views on tonal color, the color sense for certain tones is basically the same. In fact, as early as the seventeenth century, the emergence of ''color music'' in the west can show that the relationship between vision and music can be established in some inevitable or symbolic way [9]. Alexander-Laszlo invented the color piano with color music, Frederick-Benton invented the color console with color music, people invented the music color therapy with the direct influence of color concert on human physiology and psychology, the appearance of music fountain, etc. This series of rational applications of color music not only provides an important way to enhance aesthetic education, but also provides a deep understanding of the spiritual side of individuals on another level, which has a very important enlightening effect in art therapy.

Music and color, as two forms of art, are closely related to our daily life. Related research on timbre and color also has important practical significance. In the field of audio-visual synaesthesia, the correlation study of timbre and color is the basis of audio-visual association, which provides technical means and data support for audio-visual synaesthesia, thus providing ideas for multi-modal emotion recognition, classification and prediction. From the psychological point of view, the correlation between timbre and color can provide basic data for art (music) therapy theory, improve the effect of art therapy, and provide new inspiration for psychology. In terms of aesthetic education, the correlation study of timbre and color can be used as the basis of aesthetic education, enhance people's ability to understand beauty, love beauty and create beauty, and enhance people's ability to understand and feel art. In the field of performing arts technology, the relationship between music and color provides a theoretical basis. For example, the music fountains use the principle of the correlation between timbre and color to reasonably integrate the lighting and sound effects.

B. RESEARCH STATUS

For the study of the correlation between timbre and color, there are two main correlation mechanisms, namely, the direct correlation of timbre-color and the indirect correlation mediated by emotion. Direct correlation refers to the direct correspondence between sound elements and color elements. Early studies on the direct correlation between timbre and color were explained by the physical isomorphism of timbre and color [6]. The physical isomorphism here means timbre and color similarities in physical structure. Physicist Newton compared the vibration of light with the vibration of air, the former stimulates different color sensations based on the wavelength of light, while the latter stimulates different sound sensations based on the length of the air. He speculated that the harmony or disharmony of colors depended on the proportion of vibrations transmitted through the visual nerve, just as harmony or disharmony of sound came from the proportion of air vibrations [10]. And Newton put forward the synaesthesia theory of color and music. He believed that the seven tones of ''do, re, mi, fa, sol, la, si'' in music are the same as the seven colors of ''red, orange, yellow, green, cyan, blue, and purple'', and he confirmed the connection between music and color through data calculations. The vibration frequencies of the seven sounds have a certain range, and the wavelengths of the seven colors are also distributed in a certain range. The ratio of the minimum wavelength to the maximum wavelength of the seven basic colors is similar to the ratio of the minimum and maximum frequencies of the seven basic musical levels [7]. In addition, for the production of color and music, the reflection of light produces color, vibration produces sound, and to some extent, color and sound are derived from the physical effects of external forces [5]. For the representation of color and sound, both can be digitized in a spectral manner. In addition, sound and color have the characteristics of synthesis and decomposition; sound has pitch and overtone, and color has monochromatic light and composite light. Usually, we hear a compound sound and see a compound color [7].

With the continuous deepening of the research, the direct correlation between timbre and color has become a major research method. In reference [1], Zhou Haihong used the method of experimental psychology to prove through qualitative research that there is a certain correlation between visual and auditory attributes. The auditory attributes used in the experiment were pitch, sound intensity, time change rate, auditory tension and new heterosexual experience. The author found that when people have different perceptions of music, the corresponding visual perceptions are also different, which proves that there is a certain connection between vision and hearing. The experimental results show that there is a certain relationship between vision and hearing, but it does not specify what influencing factors are related to vision and hearing, and which characteristics of vision are related to which characteristics of hearing.

The indirect association mediated by emotion means that music and color are respectively associated with emotion. Color is visual art, and music is auditory art. Although the two perceptual experiences differ in their relevant sensory organs, they have the same inner emotional connection. Different sound creates will present different auditory effects, such as concord, incongruous, thick, soft, and light. Different color combinations will also produce a visual experience similar to the abovementioned auditory effects.

Researchers at the University of California, Berkeley conducted a series of studies on the association between music and color. In the early stage of the research, the relationship between music and color was studied with emotion as a medium. In the reference [11], Palmer and others used Mozart, Bach and Brahms' melodies; changed the length of the melody and made the rhythm faster to obtain 18 audio segments; and specified four emotional dimensions. Participants were required to choose the emotion that best matched the audio segment and then determine the matching color based on the emotion. They calculated the correlation based on the experimental data and believed that both music and color are related to a specific emotion. Based on the previous researches, the researchers proved this association mechanism, that is, the emotional mediation hypothesis. The reference [12] proved this point. In the reference [12], Palmer *et al.* used four kinds of Mozart melody, and changed the melody in major and minor, fast and slow tempo, note density, high and low level. Then, they got 64 audio materials. The experimental method is like the experimental method in the reference [11]. Finally, they proved the emotional mediation hypothesis.

With the deepening of the research, the experimental materials, experimental methods and experimental subjects were adjusted accordingly. Reference [13] adopted the same research idea. Different from previous research, this study included subjects who do not have synesthesia ability. Subjects are asked to choose a more matching one from the two color pairs according to the sound they hear, and it is necessary to score each sound and color on five emotional dimensions. Finally, according to the correlation analysis, the authors concluded that non-synthetic participants showed color associations consistent with musical sounds. They proposed studying the connection between more different auditory characteristics and visual characteristics. Reference [14] continued to use the previous research ideas. On the basis of the previous research, William S. Griscom and Stephen E. Palmer used 16 different musical instrument audio, and increased two characteristics, that is, edge contrast and temporal dynamics. According to the analysis of experimental results, it can be concluded that subjects without synaesthesia ability still show consistent visual associations in terms of color, edge contrast and temporal dynamics. Emotional intermediary can better explain color associations, but the correlation with edge contrast and temporal dynamics is affected by low-level features, such as attack time. In addition, this research also proposes to investigate and more connection of different visual features, such as texture and shape.

Based on the previous research results, the reference [15] explored the cross-modal correspondence in the music context, and researched the following questions: Whether people have consistent connections between different colors, musical intervals, and chords. Whether there is a consistent visual association between timbre and music. What is the relationship between normal cross-modal correspondence and a neurological condition called "sympathetic". Using experimental methods similar to previous studies, the researchers found that the subjects were highly consistent in demonstrating the connection between visual features and musical features; at the same time, they proved the role of semantic features and emotional features in guiding these cross-modal correspondences.

Researchers from Cornell University also used emotion as an intermediary in their experiment. In reference [16], the experiment explored the extent to which music color associations could be explained by music selections and color emotional associations. The experimental subjects included non-musicians, musicians, absolute pitch owners, and musical color synesthetes. Through three experiments, the author supported the view that music color association can be explained by the association between music and emotion, color and emotion.

In addition, domestic researchers have also used similar methods to study this concept. For example, reference [17] conducted a more in-depth study on the basis of a certain relationship between vision and hearing. It focused on color in vision and music in hearing to study the relationship between music and color. The author also used the method of experimental psychology to conduct qualitative research. Through three experiments, the author proved that there is a connection between music and color, and emotions play a bridge role in it, that is, emotion is the influencing factor of the connection between music and color.

Researchers at Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University conducted a serious of audio-visual cross-modal analysis studies. For example, visual signals often come along with audio, so it is necessary to investigate the influence of audio on visual attention. In reference [18], Min X *et al.* focused on the problem of when will audio influence visual attention during video watching. They performed eye-tracking experiments on a set of 60 videos in audio-visual (AV) and visual (V) conditions. Based on the eye movement data, they found that the influence of audio on visual attention depends on the consistency between the visual and audio signals. If the salient objects from the visual perspective are not consistent with the salient objects from the audio perspective, audio will influence visual attention. Otherwise, audio has little influence on visual attention.

Audio could have influence on visual attention and such influence has been widely investigated and proofed by many psychological studies. But traditional visual attention models generally make the utmost of stimuli's visual features, while discarding all audio information. In reference [19], they fused both audio and visual information to predict fixations. At last, the audio, spatial and temporal attention maps are fused, and they generated their final audio-visual saliency map. Experiment results show that we can achieve better performance when considering both audio and visual cues. Besides, they conducted a subjective study of audio and video (A/V) quality, which is used to compare and develop A/V quality measurement models and algorithms [20].

Based on the previous researches, they proposed a novel multi-modal saliency (MMS) model for videos containing scenes with high audio-visual correspondence. And they detected the audio saliency map from both audio and visual modalities by localizing the moving-sounding objects using cross-modal kernel canonical correlation analysis, which is first of its kind in the literature. Experimental results on audio-visual attention databases show that the introduced models incorporating audio cues have significant superiority over state-of-the-art image and video saliency models which utilize a single visual modality. The results of this research institute on audio-visual cross-modal analysis provide ideas for our experiment [21].

In summary, at present, experimental psychology methods are mainly used to study the relationship between visual features and auditory features. Visual features mainly include color, texture, shape, and so on. Auditory features mainly include music, major and minor tones, rhythm, and so on. The current research mainly uses emotion as the medium; that is, it is mainly qualitative research. The research results can prove that music is related to color or music and color are related through emotion and perception as the media. However, there are few quantitative studies on the relationship between the objective characteristics of music and color characteristics, and there are relatively few studies on the establishment of the relationship model between music and color from the characteristic level. Since quantitative research will provide new theoretical ideas for future research, our research will focus on quantitative analysis, extract the objective characteristic parameters of timbre, and study the relationship between timbre and color by constructing models. In addition, most of the timbre materials used in the current related research on timbre and color come from Western musical instruments. According to previous research results, namely reference [22], it can be seen that the timbre of Chinese musical instruments and Western musical instruments are quite different, therefore, in this study, in order to make the sound material library more abundant and complete, the sound materials of Chinese national musical instruments and Chinese minority musical instruments have been added.

In response to the above problems, in this study, a material library containing Western and Chinese musical instruments was selected as the experimental material, and a database related to timbre and color was constructed to support the experimental data. The relationship between timbre perception attributes and color attributes was studied, and the relationship between timbre objective characteristic parameters and color attributes was studied.

This paper focuses on the relationship between timbre and color using experimental psychology and using both signal and information processing methods, respectively. The specific chapters are arranged as follows: the first part is the introduction, which introduces the research background and significance of the related research on the timbre of music sound and visual color and then analyzes and summarizes the current research. The second part introduces the process of establishment of the timbre-color association database, including two parts, namely, the construction of the timbre perception feature dataset and the perceptual experiment of timbre-color association. The third part is a qualitative analysis of the relationship between timbre and color, which is based on the dataset constructed in the second part. The fourth part describes the construction of the timbre-color correlation models and quantitatively analyzes the correlation between timbre and color. The fifth part summarizes the research content and conclusions of this article and discusses the next research plan. The chapter arrangement of this paper is shown in figure 1.

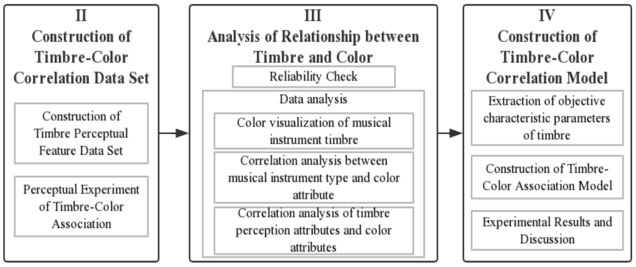

II
Construction of Timbre-Color Correlation Data Set
Construction of Timbre Perceptual Feature Data Set
Perceptual Experiment of Timbre-Color Association
III
Analysis of Relationship between Timbre and Color
Reliability Check
Data analysis
Color visualization of musical instrument timbre
Correlation analysis between musical instrument type and color attribute
Correlation analysis of timbre perception attributes and color attributes
IV
Construction of Timbre-Color Correlation Model
Extraction of objective characteristic parameters of timbre
Construction of Timbre-Color Association Model
Experimental Results and Discussion

FIGURE 1. Chapter of the paper.

II. CONSTRUCTION OF TIMBRE-COLOR CORRELATION DATA SET

The construction of timbre-color correlation data set in this part is mainly divided into two parts. The first is the construction of timbre perceptual feature data set. The five perceptual dimensions of timbre are evaluated by series category method, and the evaluation score of each timbre material in five dimensions of timbre is got. It can provide data for the third part of timbre and color correlation analysis. The second is the perceptual experiment of timbre-color association, which requires the subjects to match the timbre material with the color material, to get the matching result of timbre and color, and then to process the data, so that the perceptual data associated with timbre and color can be got. It can provide data support for the construction of the fourth part of the model.

A. CONSTRUCTION OF TIMBRE PERCEPTUAL FEATURE DATA SET

1) EXPERIMENTAL MATERIAL

In the experiment, we should select the representative timbre material, and consider the following three factors when selecting the timbre material [23]: First, the experimental timbre material should have a suitable sample number, too little material can not guarantee the accuracy of multi-dimensional scale analysis, too much material will make the subjects tired, and the judgment criteria are easy to change, which leads to the reduction of the reliability of the evaluation results. Second, the timbre material should have a wide range of variations. Third, after determining the range of timbre changes, we must pay attention to the uniformity of the distribution of timbre materials. Considering the above three factors, 72 kinds of musical instruments were selected in this experiment, including 24 kinds of western musical instruments and 48 kinds of Chinese musical instruments, among which 36 kinds of Chinese national musical instruments and 12 kinds of Chinese minority musical instruments were selected. These 72 kinds of musical instrument materials can also be divided into two categories according to the time domain characteristics of musical instruments: sustainable musical sound signal and non-sustainable musical sound signal. The proportion of musical instruments in the timbre library is shown in figure 2, and the list of specific instruments is shown in Appendix 2.

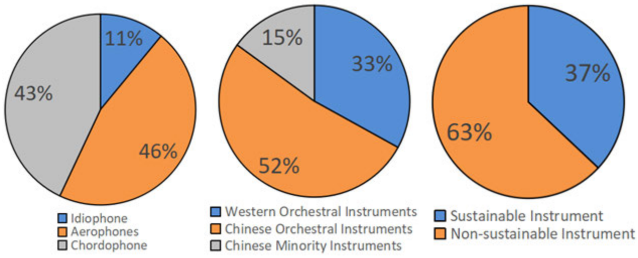


FIGURE 2. Proportion of musical instrument composition of timbre material library.

In addition, this subjective evaluation experiment adopts five timbre evaluation dimensions, which are based on the previous results of the laboratory, and the specific calculation process is in reference [22]. The five evaluation dimensions of timbre are bright-dark, raspy-mellow, sharp-vigorous, coarse-pure, hoarse-consonant, each pair of evaluation dimensions is opposite, and can describe the five main dimensions of timbre perception. The correlation coefficient of each timbre evaluation dimension is shown in TABLE 1.

All the timbre materials used in the experiment were recorded in the recording studio by players with professional performance level, and the contents were single tone, scale and music fragment of medium strength (mf) content. The Protools software is used to clip each material segment for 6-10 seconds, the sampling rate is 44100 Hz, the quantization accuracy is 16 Bit, and saved into wav format.

TABLE 1. Correlation coefficient of five evaluation dimensions of timbre in their original language(chinese) and the associated english translations.

The five evaluation dimensions of timbre	Abbreviation	Relative coefficient
明亮-暗淡(Bright-Dark)	B-D	-0.99
干瘪-柔和(Raspy-Mellow)	R-M	-0.82
尖锐-浑厚(Sharp-Vigorous)	S-V	-0.93
粗糙-纯净(Coarse-Pure)	C-P	-0.92
嘶哑-协和(Hoarse-Consonant)	H-C	-0.86

In order to avoid the influence of playing content on timbre perception, the timbre material used in this experiment is scaled. In addition, loudness also has an important effect on timbre. For example, in reference [24], Zhu Siyu studied the pitch loudness and timbre brightness of typical Chinese and Western musical instruments, research shows that for typical Chinese and Western musical instruments, timbre brightness increases with the increase of loudness, but the degree of influence of loudness on the brightness of the timbre is affected by the range of loudness. The results have been verified in the previous pre-experiment. And we designed the loudness balance experiment for Chinese and Western musical instruments. Through the analysis of the data, we summarized the loudness balance mode of different kinds of musical instruments. The experimental materials are normalized by this model. The specific operation method can be referred to reference [25].

2) EXPERIMENTAL METHODS

This experiment adopts the series category method, which assumes that the psychological perception is a random variable from normal distribution, and converts the cumulative probability into the psychological scale (category boundary). The boundary of each category in the series category method is not a given definite value in advance, but a random variable based on experimental data. Therefore, in the experiment, the boundary of each category is determined according to Thurstone algorithm, and then the relationship between the psychological scale and the boundary of each evaluation object is calculated, so as to determine the distribution of each object in each category.

3) EXPERIMENTAL CONDITIONS

The experiment was carried out in the standard listening room. The reverberation time of the listening room was 0.3 seconds, the sound field distribution was uniform, and there was no bad acoustic phenomenon and body noise. Genelec 1038B three-way active midfield monitoring speaker was used to replay the experimental signals. Its parameters are shown in TABLE 2. It conform to international standard [26].

Because the experimental results will be affected by the listening sound pressure level, it is necessary to ensure that the subjects listen at the standard listening sound pressure level, and the sound pressure level of the listening sound remains unchanged throughout the experiment. The equipment used

TABLE 2. **The parameters of Genelec three-way active midfield 1038B monitoring speaker.**

Parameter	
Frequency response	35 Hz - 20 kHz (± 2.5 dB)
Maximum sound pressure	124（dB）
Long-term sound pressure	120（dB）

in the calibration test system is: Lenovo T460 notebook computer, BK4231 sound calibrator, BK2250 sound level meter, YAMAHA 01V96i digital mixer. The actual listening pressure level is 75 dBA, which conforms to the international listening standard [26]. Adobe Audition software for playing experimental material. The seats in the listening room are arranged in triangles. In the process of listening, the ear height of the subjects should be at the same level as the midpoint of the vertical line in the high and low sounds of the speakers.

4) EXPERIMENTAL SUBJECTS

The number and category of experimental subjects are the factors to be considered in the selection process. In theory, the number of subjects can be determined according to the distribution of test results and the accuracy required for measurement, but in practical experiments, the distribution of test results is difficult to determine, and there are many factors affecting the distribution of measurement results. So the general number of subjects is selected according to the empirical value [23], Tuliis and Wood indicated that the average correlation coefficient between the results of 20 to 30 evaluations and 168 evaluations is 0.95 [27]. So it is more appropriate to select 20-30 subjects when selecting the number of subjects. This experiment selected 31 subjects, including 11 men, aged 20-30 years old. The subjects had no hearing impairment, they had certain listening experience, and they had a good understanding of each attribute of timbre.

5) EXPERIMENTAL PROCESS

Experiment 1 consists of three stages: familiarity stage, training stage and formal experiment stage. Familiarity stage: Play all the audio samples used in the experiment in sequence, so that the participants are familiar with and perceive their range of changes. Training stage: Select 3 out of the total timbre material randomly and let the subjects evaluate the tone color according to the subjective feeling of the timbre material. The purpose of this step is to make the subjects more familiar with the experimental process and avoid the impact caused by unfamiliarity with the experimental process in the formal experimental stage. This part of the data is not used for the analysis of the final results. Formal experimental stage: Play each tone pigment material in turn, the subjects are asked to evaluate and score the five evaluation dimensions of timbre, namely, bright-dark, raspy-mellow, sharp-vigorous, coarse-pure, hoarse-consonant, and there are nine grades. After scoring, the subjects need to fill in the results in the form. In order to avoid the fatigue of listening for a long time, the experimental materials were divided into three groups, each group was not more than 30 minutes, and the rest between each group was 15 minutes. According to the above steps in turn experiment, and complete the data collection work.

6) CHECKING THE COMPLETENESS OF THE DATA SET

After processing the collected data, it is found that 72 kinds of musical instrument sample materials are distributed in five timbre perception dimensions, and the sample distribution of timbre material in the specific timbre evaluation scale is shown in Appendix 1. On the whole, the timbre of different musical instruments is distributed evenly in the dimension of timbre evaluation, but there are also some differences, such as the western musical instruments are brighter, more vigorous, mellower and pure, compared with the Chinese national musical instruments, the Chinese minority musical instruments are raspier, coarser and hoarser. Therefore, it can be seen that the timbre material library of Chinese musical instruments and Western musical instruments is richer and more complete than that of Western musical instruments alone.

B. PERCEPTUAL EXPERIMENT OF TIMBRE-COLOR ASSOCIATION

1) EXPERIMENTAL MATERIAL

The timbre material of this experiment is the same as that of the first experiment. For color material construction, we choose the HSV color space, because HSV color space can reflect visual perception, and it is closely related to visual perception. Among them, H means hue, it is the most intuitive feature to distinguish different colors. The color block diagram used in this subjective evaluation experiment selects eight colors, corresponding to 0°, 30°, 60°, 90°, 120°, 180°, 240° and 300° of the hue ring. Red, green, yellow and blue are four primary colors. Red and green, yellow and blue are opposite. Orange, yellow green, blue green, purple are four middle colors. The color circle system used in this experiment is two-dimensional, two features are selected to represent the hue, namely, red-green and yellow-blue. S means saturation, it refers to the brightness of color. Considering the experimental data workload, saturation only selects 50% and 100% levels. V means value, it mainly depends on the intensity of the light. Also considering the experimental data workload, value only choose 50% and 100% two levels. Figure 3 shows the HSV used in this experiment, the upper layer represents 100% value, the following layer represents 50% value, the outside circle represents 100% saturation, the middle circle represents 50% saturation. The four dimensions of color are shown in TABLE 3.

Taking red as an example, figure 4 shows the principle of the design color block diagram. In figure 4, the upper left corner is red with saturation and value of 100%, the upper right corner is red with saturation of 50% and value of 100%, the lower left corner is red with saturation of 100% and value of 50%, and the lower right corner is red with saturation

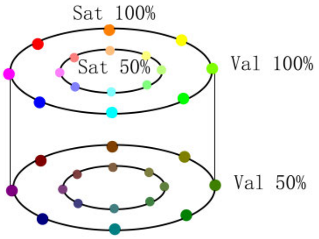


FIGURE 3. Distribution of color material.

TABLE 3. Four dimensions list of colors.

The four dimensions of color	Abbreviation
Red-Green	R-G
Yellow-Blue	Y-B
Saturation	Sat
Value	Val

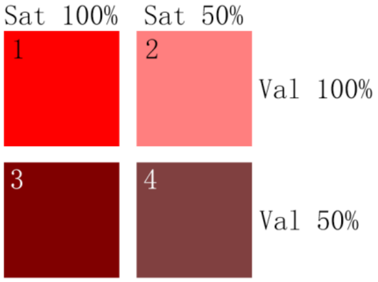


FIGURE 4. The design principle of the color block diagram.

and value of 50%. According to this design method, each hue corresponds to four colors. According to the arrangement and combination, there are 32 color blocks corresponding to the eight hues. And the hue, saturation, value of each color block change separately, so as to ensure that the experimental results are universal. The specific color block diagram is shown in figure 5.

2) EXPERIMENTAL METHODS

According to the timbre material, the subjects selected the first matching color, the second matching color, and the third matching color, and then chose the first mismatch color, the second mismatch color, and the third mismatch color. In experimental data processing, the difference between the matched weighted average and the mismatched weighted average is used as the criterion to judge the timbre. Specific calculation formula see formula (1), (2).

$$C_{d,m} = \left(3C_{1,d,m} + 2C_{2,d,m} + C_{3,d,m}\right)/6 \qquad (1)$$

$$I_{d,m} = \left(3I_{1,d,m} + 2I_{2,d,m} + I_{3,d,m}\right)/6 \qquad (2)$$

Among them, $\mathrm{C}_{\mathrm{j,d,m}}$ represents the value of the j-th matching color dimension for the m-th timbre material selected by the subjects. For instance, the $\mathrm{C}_{1,\mathrm{R-G,clarinet}}$ represents the value of the first matching color's Red-Green degree selected by the subjects for the timbre material of the clarinet; $\mathrm{I}_{\mathrm{j,d,m}}$ represents the value of the j-th mismatching color dimension for the m-th timbre material selected by the subjects, $\mathrm{I}_{2,\mathrm{Y-B\ ,clarinet}}$ represents the value of the second mismatched color's Yellow-Blue degree selected by the subjects for the timbre material of the clarinet. The value range of j is 1-3, and the value range of m is 1-72.

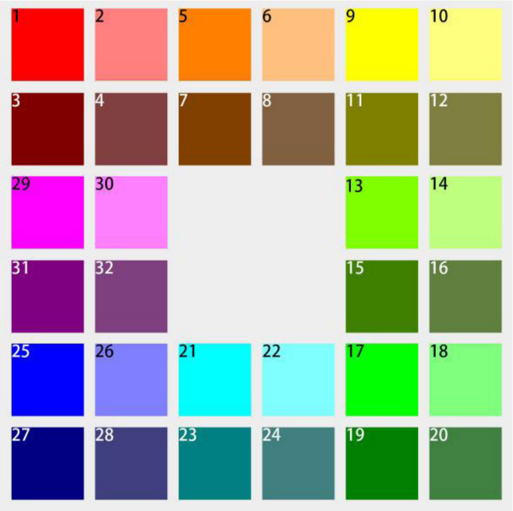


FIGURE 5. Color block diagram for subjective evaluation of timbre-color association.

Define $\mathrm{CTA}_{\mathrm{d,m}}$ as the standard to measure the association between timbre and color, where C stands for color, T stands for timbre, A stands for analyze, and $\mathrm{CTA}_{\mathrm{d,m}}$ stands the timbre-color associated value of the m-th timbre material in the color dimension d, for example, $\mathrm{CTA}_{\mathrm{R-G,clarinet}}$ represents the Red-Green values of the corresponding color of the clarinet. The specific calculation formula is shown in formula (3).

$$CTA_{d,m} = C_{d,m} - I_{d,m} \qquad (3)$$

3) EXPERIMENTAL CONDITIONS

In this experiment, the experimental conditions of sound field are the same as that of experiment 1. We used SONY KD-75 × 9400D to render the color blocks. Its screen ratio is 16:9, the screen resolution is 3840 × 2160, the high-definition format is 2160P, and the backlight performance is LED backlight. It conform to international standard. Digital TV display equipment and experimental process conform to international standard [28]. The connection of the experimental system is shown in figure 6. One end of the notebook computer is connected to the display to realize the presentation of the color material, and the other end is connected to the left speaker and the right speaker respectively to realize the play of the music material.

The subjects were the same as the subjects in experiment 1, and the subjects had no visual impairment.

4) EXPERIMENTAL PROCESS

The experiment consists of three stages: familiarity stage, training stage, and formal experiment stage. In the familiar stage, all the audio samples used in the experiment can be

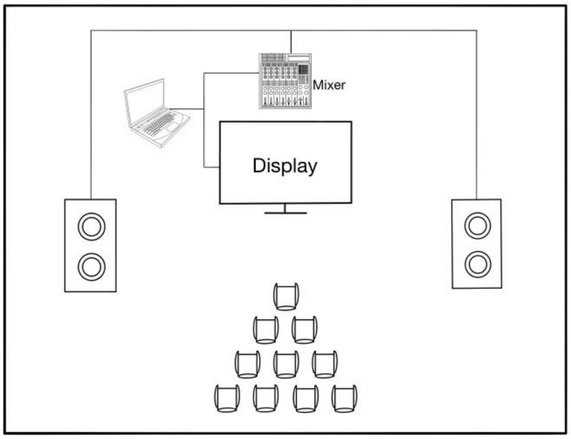


FIGURE 6. The connection diagram of listening system in experiment 2.

played sequentially, the subjects can feel the range of tone color change, in order to avoid too much deviation in the color selected in the formal experiment. We use the display screen to present the color block diagram so that the subjects are familiar with the color block diagram in advance. In the training stage, three pieces are randomly selected from the musical instrument material to make the subjects try to choose the color chart according to the subjective feeling of the timbre material. The purpose of this step is to make the subjects more familiar with the experimental process and avoid the influence caused by the unfamiliar experimental process in the formal experiment stage. This part of the data is not used for the analysis of the final results. In the formal experiment stage, each tone pigment material was played in turn. According to the timbre material, the subjects selected the first matching color, the second matching color and the third matching color, and then selected the first mismatch color, the second mismatch color and the third mismatch color, they also need to fill in the selection in the form used by the subjects. In order to ensure the fatigue of listening for a long time, the experimental time of each group was not more than 30 minutes, and the rest between each group was 15 minutes. The experiment was divided into three groups. According to the above method, the experiment was completed, and the data were collected.

Through the processing and statistics of the experimental data, a perceptual data set containing 31 subjects on 72 musical instrument timbre materials was constructed. The data set can be used for qualitative analysis of timbre and color association in the third part (Analysis of Relationship between Timbre and Color). It can also be used for the construction of timbre - color correlation model in the fourth part (Construction of Timbre-Color Correlation Model), and then the correlation between timbre and color can be quantitatively analyzed.

III. ANALYSIS OF RELATIONSHIP BETWEEN TIMBRE AND COLOR

The main content of this part is to process the data and analyze the relationship between timbre and color qualitatively. First, the reliability test of the data obtained from the subjective evaluation experiment is carried out, and then the data is analyzed from three aspects: color visualization of musical instrument timbre, correlation analysis of musical instrument type and color attribute, correlation analysis of timbre perception attribute and color attribute.

A. RELIABILITY CHECK

The subjective judgement of the correlation between the timbre attributes of musical sound and visual color is subjective. This can also be seen from the experimental data, namely, the judgement of different subjects may be quite different. Therefore, in order to ensure the reliability of experimental data, firstly, we conducted a reliability calibration test on the selection results of 31 subjects. The reliability statistical results are shown in TABLE 4.

TABLE 4. Reliability calibration test results.

Variable	R-G	Y-B	Sat	Val
Cronbach's alpha	0.76	0.75	0.67	0.93
Participants	31	31	31	31
Samples	216	216	216	216

Cronbach's alpha are all greater than 0.67. It can be concluded that the experimental data are reliable, the subjective judgement of different subjects are highly reliable.

In order to facilitate the visualization of the data, the data were arranged neatly in a 6×5 matrix, so we need to eliminate one data from the first match, second match, third match, first mismatch, second mismatch, and third mismatch corresponding to each instrument. Firstly, the histogram corresponding to the first matching, the second matching, the third matching, the first mismatch, the second mismatch and the third mismatch are drawn in the data culling link. There are 72 × 6 histograms. The color sequence number farthest from the highest frequency in the histogram was selected for culling.

Taking the first mismatched of Gehu as an example, figure 7 is the histogram of the first mismatched color. According to the culling principle, we need to eliminate the

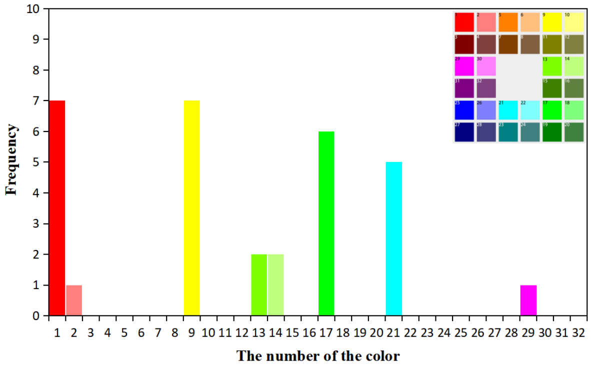


FIGURE 7. Histogram of the first mismatched color of Gehu.

lowest frequency color serial number, but there are two lowest frequency color serial numbers 2 and 29, the color farthest away from the color with the highest frequency on the tonal diagram is selected for cull processing. The frequency of color serial numbers 1 and 9 are the same and highest, and colors serial number 29 are further away from colors with numbers 1 and 9 on the tonal diagram, so let's get rid of the color with the number 29.

According to the above method, all data are processed.

B. DATA ANALYSIS

The data analysis of this part is divided into three parts. First, visualize the color of the musical instrument timbre, and the correlation between timbre and color can be proved by drawing 72 kinds of musical instruments. Then, based on the correlation between timbre and color, the correlation between musical instrument type and color attribute is analyzed. Finally, on the basis of the former two, the correlation between the perceptual attribute of timbre and the attribute of color is analyzed.

1) COLOR VISUALIZATION OF MUSICAL INSTRUMENT TIMBRE

In order to make the data more intuitive, we need to draw the matching color block diagram and the non-match color block diagram corresponding to each instrument. According to the calculation method described in the experimental method in the second chapter, the average matching color and the average mismatch color of each instrument can be calculated. The calculation formula is as follows:

$$R_C = \left(3\overline{R_{C1}} + 2\overline{R_{C2}} + \overline{R_{C3}}\right)/6 \tag{4}$$

$$R_I = \left(3\overline{R_{I1}} + 2\overline{R_{I2}} + \overline{R_{I3}}\right)/6 \tag{5}$$

$$G_C = \left(3\overline{G_{C1}} + 2\overline{G_{C2}} + \overline{G_{C3}}\right)/6 \tag{6}$$

$$G_I = \left(3\overline{G_{I1}} + 2\overline{G_{I2}} + \overline{G_{I3}}\right)/6 \tag{7}$$

$$B_C = \left(3\overline{B_{C1}} + 2\overline{B_{C2}} + \overline{B_{C3}}\right)/6 \tag{8}$$

$$B_I = \left(3\overline{B_{I1}} + 2\overline{B_{I2}} + \overline{B_{I3}}\right)/6 \tag{9}$$

Among them, R_{Cj}, G_{Cj} and B_{Cj} represent the RGB value of the j-th matching color, R_{Ij}, G_{Ij} and B_{Ij} represent the RGB value of the j-th mismatching color, and $\bar{R}\ \bar{G}\ \bar{B}$ represent the R, G, B average value of the color selected by all subjects, R_C, G_C and B_C represent the weighted average color of the j-th matching color, R_I, G_I and B_I represent the RGB value of the weighted average color of the j-th mismatching color.

According to the above calculation method, the color visualization results of 72 musical instrument timbre can be drawn, see Appendix 3. Taking chimes as an example, C1, C2, C3, I1, I2, I3 represent the first matching color, the second matching color, the third matching color, the first mismatch color, the second mismatch color and the third mismatch color, the weighted average color of the matching and mismatch is drawn at the bottom of the color block diagram. As can be seen from figure 8, the matching color of the chimes is green, while the mismatching color is purple.

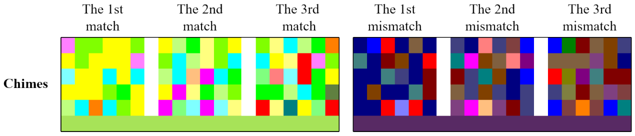


FIGURE 8. Color visualization of the timbre(chimes).

Combined with the clear and pleasant sound characteristics of the chimes, it can be proved that the matching color selected by the subjects is related to the timbre of the chimes. According to this method, the color visualization results of all musical instruments can be analyzed, which can prove that timbre is related to color.

2) CORRELATION ANALYSIS BETWEEN MUSICAL INSTRUMENT TYPE AND COLOR ATTRIBUTE

Through the experiment of the second part, we can get the red-green, yellow-blue, value and saturation of each instrument's matching color. We can analyze the relationship between the timbre of the instrument and the four color attributes by drawing the scattered plot, and we can also compare the similarities and differences between the timbre and the color attribute of different kinds of instruments.

Because the color space of the HSV includes three parts: hue, saturation and value, and the red-green, yellow-blue degree can represent hue, so in this part, we will draw three pictures, which are the average hue scatter plot of the instrument, the value scatter plot of the instrument and the saturation scatter plot of the instrument. The abscissa of the scatter plot of the average hue of musical instruments is the $\text{CTA}_{\text{Yellow}-\text{Blue}}$ degree of the timber color of each instrument, and the ordinate is the $\text{CTA}_{\text{Red}-\text{Green}}$ degree of the timber color of each instrument. Fill the background color of the scatter diagram with the corresponding yellow-blue and red-green degrees, then the hue at the position of each point is the hue of the instrument. See Figure 9 for the average scatter

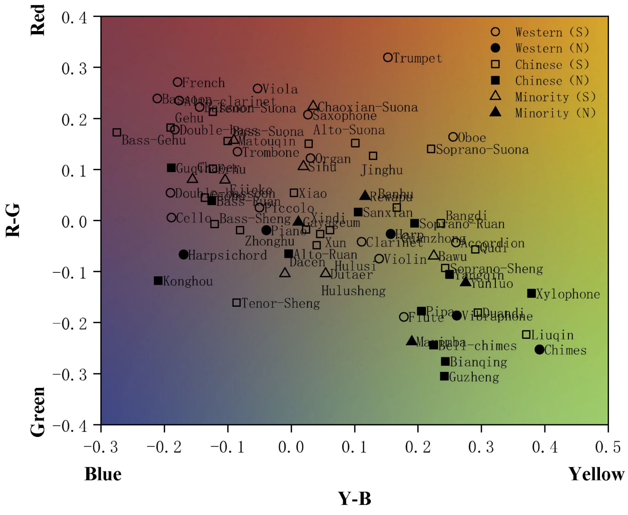


FIGURE 9. Scatter plot of average hue.

diagram of hue. The abscissa of the scatter diagram of the value of instruments is the CTA$_{\text{Value}}$ of the timber color of each instrument, and the ordinate is the instrument number. The background color of the scatter diagram is filled with color according to its corresponding value, so the value at the position of each point is the value of the instrument. See Figure 10 for the scatter diagram of the value. The abscissa is the CTA$_{\text{Saturation}}$ value of the tone color of each instrument, and the ordinate is the instrument number. The background color of the scatter diagram is filled with the corresponding color's saturation, so the saturation at the position of each point is the saturation of the instrument. See Figure 11 for the saturation scatter diagram. The dotted lines in the scatter plot correspond to average value, with orange for Western instruments, blue for Chinese national musical instruments, and green for Chinese minority instruments. Among them, CTA is the standard for measuring timbre, and the calculation formulas of CTA are shown in Formula (1), Formula (2) and Formula (3).

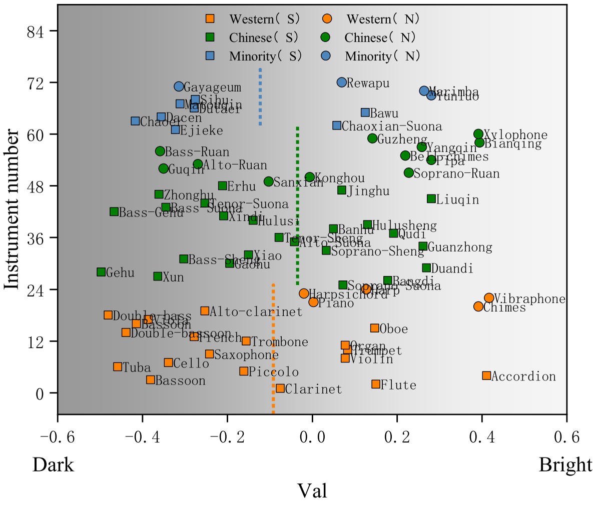


FIGURE 10. Scatter plot of average value.

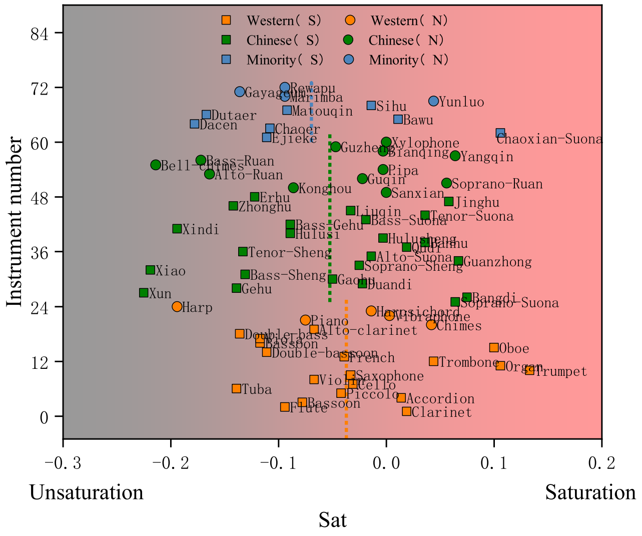


FIGURE 11. Scatter plot of average saturation.

As can be seen from the above figure, the color corresponding to sustainable instruments has a larger red-green degree, while the color corresponding to non-sustainable instruments has a smaller red-green degree. In other words, the color corresponding to sustainable instruments tends to be red, while the color corresponding to non-sustainable instruments tends to be green. The value distribution of the color corresponding to the musical instrument is relatively dispersed, and the average value of the Chinese national musical instruments is greater than that of the Western musical instrument and that of the Chinese minority musical instrument. The saturation distribution of the color corresponding to the musical instrument is relatively concentrated, which indicates that the saturation has no significant influence on the sound color. The average saturation of the Western musical instrument is higher than that of the Chinese national musical instrument.

3) CORRELATION ANALYSIS OF TIMBRE PERCEPTION ATTRIBUTES AND COLOR ATTRIBUTES

In order to get the relationship between the timbre perception attribute and the color attribute, we use the method of correlation analysis to analyze the two. The five perceptual attributes of timbre are Bright-Dark, Raspy-Mellow, Sharp-Vigorous, Coarse-Pure and Hoarse-Consonant. The four attributes of color are red-green, yellow-blue, saturation and value. The correlation coefficient is shown in TABLE 5. The absolute value of the correlation coefficient in TABLE 5 is greater than or equal to 0.6.

TABLE 5. Correlation coefficient tables.

	Red-Green	Yellow-Blue	Saturation	Value
Bright-Dark	0.41	**-0.77**	**-0.67**	**-0.88**
Raspy-Mellow	-0.52	0.44	0.10	0.57
Sharp-Vigorous	0.25	**-0.65**	**-0.61**	**-0.67**
Coarse-Pure	**-0.60**	**0.60**	0.25	**0.71**
Hoarse-Consonant	-0.55	0.49	0.17	**0.62**

In order to further present the results, we draw scatter plots, whose ordinate is CTA and abscissa is timbre perception attribute. Western musical instruments, Chinese national musical instruments and Chinese minority musical instruments are represented by graphs of different colors. Different shapes are used to represent the sustainable and non-sustainable instruments. According to the correlation coefficient matrix, 0.7 is selected as the threshold value. If the correlation coefficient is greater than 0.7, linear fitting will be carried out for each point; otherwise, nonlinear fitting will be carried out. In the scatter plots, the orange curve represents the trend line of Western instruments, the green curve represents the trend line of Chinese national instruments, the blue curve represents the trend line of Chinese minority instruments, the solid black line represents the trend line of sustainable instruments, and the black line represents the trend line of non-sustainable instruments. Considering that the trend line of each type of musical instrument is not convenient for practical application, the trend line of all musical instruments is fitted linearly and the expression is given.

According to the above method, the scatter plot and its fitting curve are drawn for five timbre perception attributes and four color attributes (only the correlation coefficient is not less than 0.6). The specific analysis is as follows:

Figure 12 is Red-Green and Coarse-Pure scatter plot. It can be seen from the diagram that coarse-pure of timbre is negatively correlated with red-green.

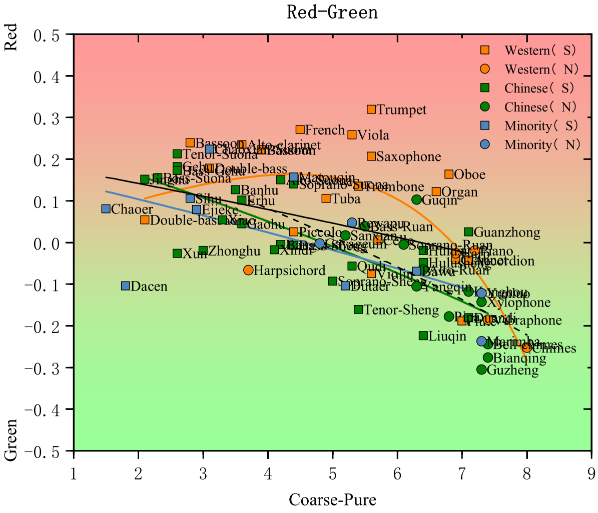
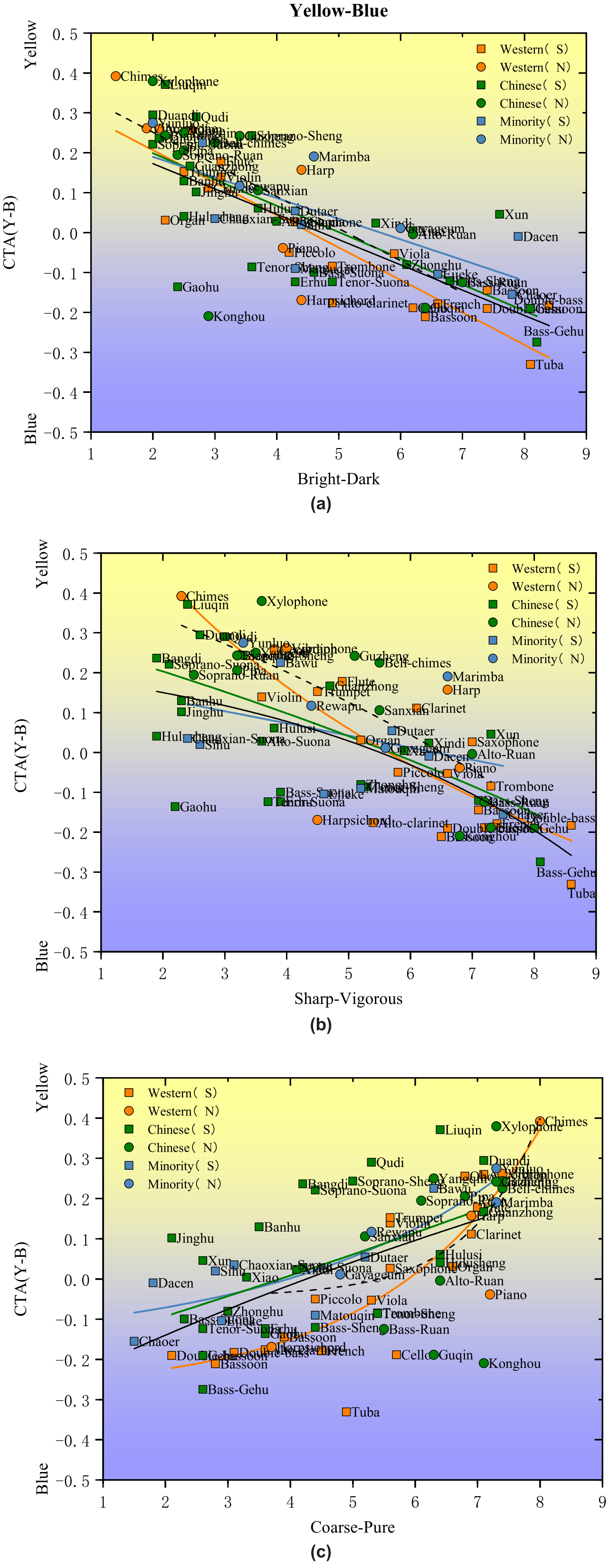


FIGURE 12. Red-green and coarse-pure scatter plot.

Overall, most Western instruments are located above Chinese instruments, that is, Western instruments tend to favor red colors. For more pure sounds, they tend to associate brighter colors such as green. In addition, sustainable instruments are rough relative to non-sustainable instruments.

Figure 13 shows the scatter plot of the yellow-blue degree of color and the three perceptual attributes of timbre, respectively. As can be seen from the figure, the coarse-pure of timbre is positively correlated with the yellow-blue. The bright-dark of timbre is negatively correlated with the yellow-blue, and the correlation coefficient is 0.77, showing a strong linear correlation. The sharp-vigorous of timbre is negatively correlated with the yellow-blue.

In general, most Western instruments are located below the Chinese instruments, namely, Western instruments make people think of blue, Chinese instruments make people think of yellow. For instruments with purer and bright timbre, it is easier to think of bright colors such as yellow; for instruments with a thicker timbre, it is easier to think of blue. Most of the non-sustainable instruments are distributed in the upper half of the picture, that is, non-sustainable instruments can be associated with yellow, sustainable instruments can be associated with blue, and, non-sustainable instruments are



(a)

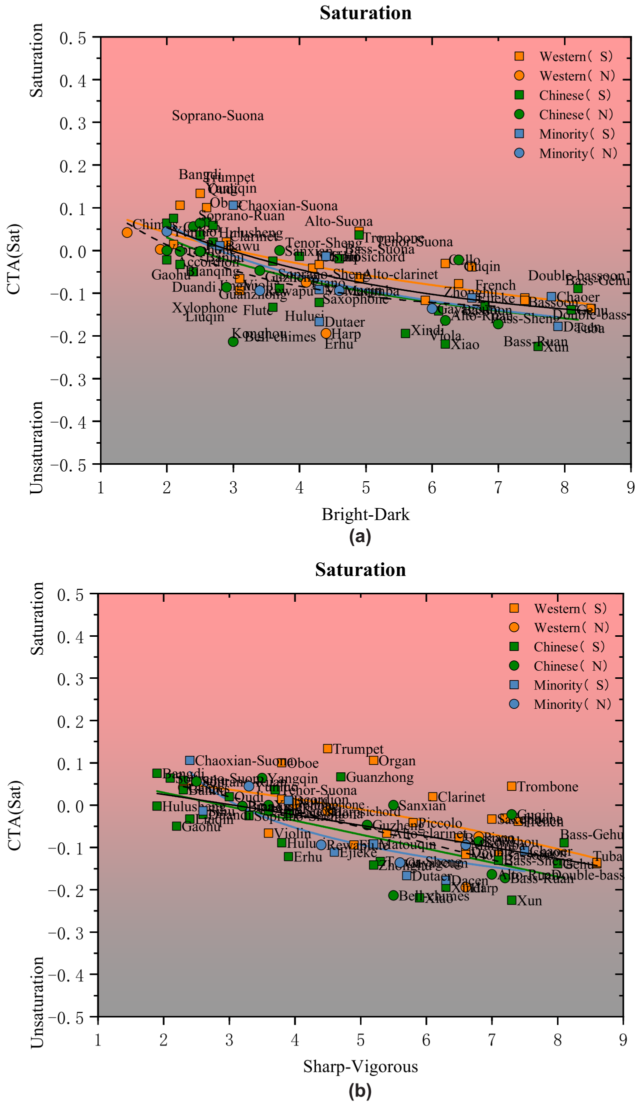
(b)

(c)

FIGURE 13. Yellow-blue and timbre dimensions scatter plot. (a) Yellow-blue and bright-dark scatter plot. (b) Yellow-blue and sharp-vigorous scatter plot. (c) Yellow-blue and coarse-pure scatter plot.

concentrated in the right half of the Yellow-Blue and Coarse-Pure Scatter Plot, indicating that non-sustainable instruments are more pure than sustainable instruments.

Figure 14 shows the scatter plot of color saturation and timbre dimensions. It can be seen from the diagram that bright-dark, sharp-vigorous of timbre are negatively correlated with saturation.



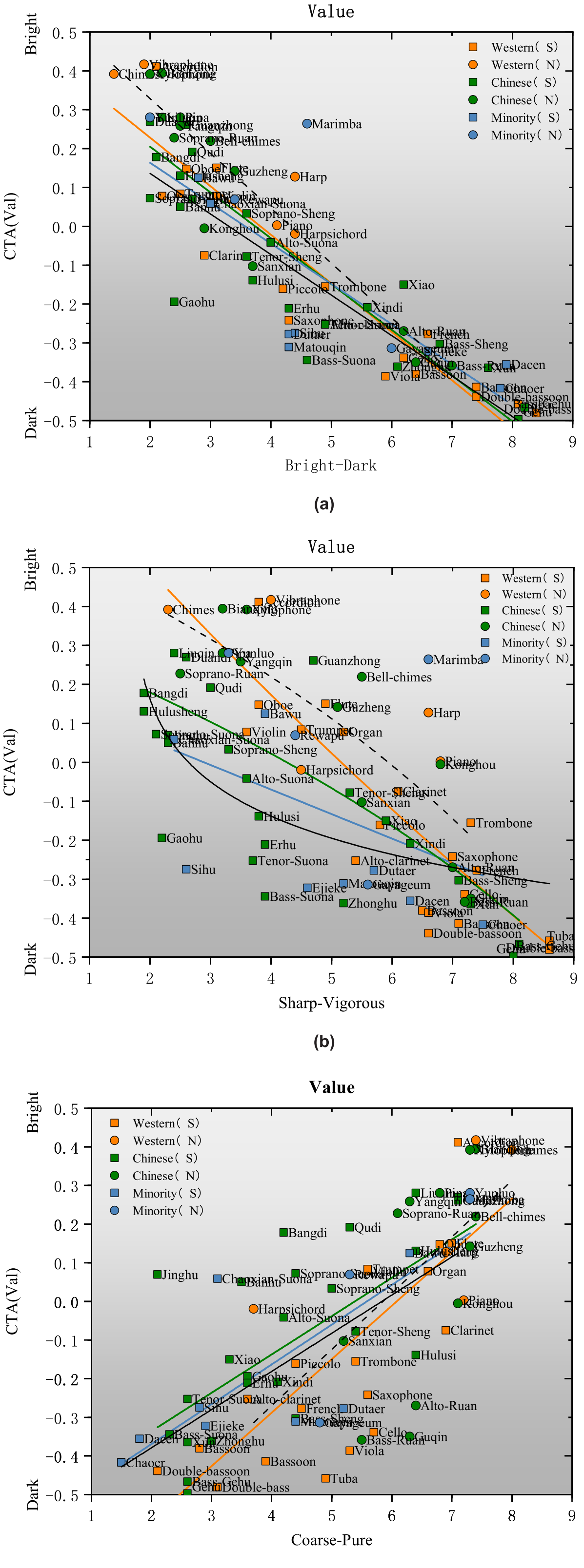
(a)

(b)

FIGURE 14. **Saturation and timbre dimensions scatter plot. (a) Saturation and bright-dark scatter plot. (b) Saturation and sharp-vigorous scatter plot.**

In general, the distribution of Western instruments and Chinese instruments is more uniform and concentrated in the center of the scattered plot, indicating that saturation and timbre dimension are not significantly related. And, relative to non-sustainable instruments, sustainable instruments are dimmer.

Figure 15 shows the scatter plot of the value of the color and the four perceptual attributes of the timbre. It can be seen from the diagram that the coarse-pure of timbre is positively



(a)

(b)

(c)

FIGURE 15. **Value and timbre dimensions scatter plot. (a) Value and bright-dark scatter plot. (b) Value and sharp-vigorous scatter plot. (c) Value and coarse-pure scatter plot. (d) Value and hoarse-consonant scatter plot.**

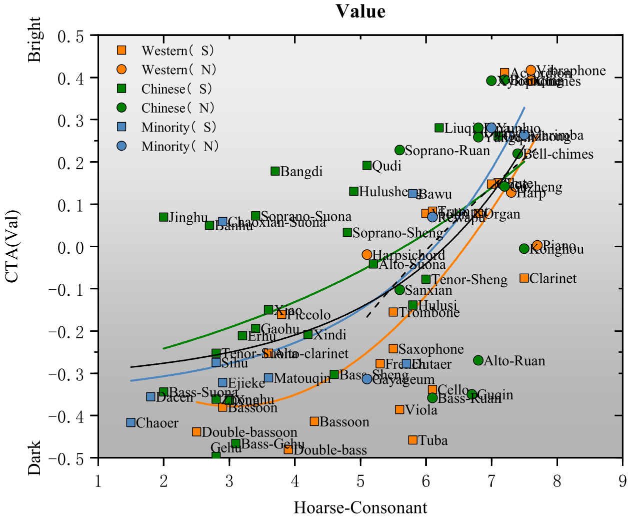


FIGURE 15. ***(Continued.)*** **Value and timbre dimensions scatter plot. (a) Value and bright-dark scatter plot. (b) Value and sharp-vigorous scatter plot. (c) Value and coarse-pure scatter plot. (d) Value and hoarse-consonant scatter plot.**

correlated with value, and the correlation coefficient is 0.71, showing a strong linear correlation. The hoarse-consonant of timbre is positively correlated with value. The sharp-vigorous of timbre is negatively correlated with value. And the bright-dark of timbre is negatively correlated with value, and the correlation coefficient is 0.88, showing a strong linear correlation.

On the whole, for the purer, bright, hoarser, sharper instruments, people will think of a preference for bright colors. Moreover, most non-sustainable instruments are distributed in the upper half of the picture, indicating that non-sustainable instruments make people think of brighter colors.

To sum up, there is no obvious correlation between saturation and the five dimensions of timbre; Western instruments tend to favor red and blue colors; non-sustainable instruments tend to think of brighter colors such as yellow; for purer, soft, hoarse instruments, people think of brighter colors such as green and yellow; and compared with non-sustainable instruments, sustainable instruments are purer, brighter and hoarser.

From a psychological point of view, colors can make people feel warm and cold psychologically. Such as red, yellow, orange and other colors give people a warm, exciting feeling. People call this series of colors warm; blue, green, green and other colors give people a cold, quiet feeling. People call this series of colors cold; while gray, purple and other colors give people feel not cold and warm, people call it ''neutral color'' [29]. The color material used in this experiment can be divided into cold color, warm color and neutral color according to this classification method. In the experiment, the subjects will have corresponding psychological feelings about these colors, which will echo the psychological feelings brought by timbre to the subjects.

In addition, the more crisp tone will give people a lighter feeling, the thicker tone will give people a heavier feeling, similar to the color also has a sense of weight. This weight is not the actual weight, but the weight that the color produces in the senses [30]. The same tone, the color with low value is heavier than the color with high value, and the color with low brightness is heavier than the color with high brightness. The weight of the color depends mainly on the value and saturation of the color. The color with high value and saturation generally has a lighter feeling. From this point of view, the light and heavy feeling of musical instrument timbre will correspond to the weight feeling of color, and the choice of subjects in the experiment is also affected by this factor.

TABLE 6. Trend line list of red-green and coarse-pure.

	Trend line expression	R-Square(COD)	Pearson's r
Red-Green and Coarse-Pure	$Y_{R\text{-}G}$=0.27-0.05$X_{C\text{-}P}$	0.35	-0.59

TABLE 7. Trend line list of yellow-blue and timbre dimensions.

	Trend line expression	R-Square(COD)	Pearson's r
Yellow-Blue and Bright-Dark	$Y_{Y\text{-}B}$=0.34-0.07$X_{B\text{-}D}$	0.60	-0.77
Yellow-Blue and Sharp-Vigorous	$Y_{Y\text{-}B}$=0.35-0.06$X_{S\text{-}V}$	0.43	-0.65
Yellow-Blue and Coarse-Pure	$Y_{Y\text{-}B}$=-0.27+0.06$X_{C\text{-}P}$	0.37	0.61

IV. CONSTRUCTION OF TIMBRE-COLOR CORRELATION MODEL

In order to predict the relationship between timbre and color in practical application, this part mainly constructs the model of timbre and color association. Firstly, we extracted the objective characteristic parameters of timber, and then we built the correlation model by using these objective characteristic parameters and the CTA value of color perception data. After model testing, the conclusion can be drawn by comparing and analyzing the model at last.

A. EXTRACTION OF OBJBCTIVE CHARACTERISTIC PARAMETERS OF TIMBRE

In order to obtain the parameters needed in the model, we extracted the objective characteristic parameters of the timbre that can describe the timbre. These objective characteristic parameters of timbre can be divided into four categories: time domain objective characteristic parameters, frequency domain objective characteristic parameters, harmonic objective characteristic parameters and time-frequency objective characteristic parameters. The name and classification of the specific feature parameters are shown in TABLE 10.

The calculation method of the important objective characteristic parameters of timbre is as follows.

TABLE 8. Trend line list of saturation and timbre dimensions.

	Trend line expression	R-Square(COD)	Pearson's r
Saturation and Bright-Dark	Y_{Sat}=0.08-0.03$X_{\mathrm{B\text{-}D}}$	0.44	-0.67
Saturation and Sharp-Vigorous	Y_{Sat}=0.09-0.03$X_{\mathrm{S\text{-}V}}$	0.37	-0.61

TABLE 9. Trend line list of value and timbre dimensions.

	Trend line expression	R-Squares (COD)	Pearson's r
Value and Bright-Dark	Y_{Val}=0.44-0.12$X_{\mathrm{B\text{-}D}}$	0.78	-0.88
Value and Sharp-Vigorous	Y_{Val}=0.40-0.09$X_{\mathrm{S\text{-}V}}$	0.45	-0.67
Value and Coarse-Pure	Y_{Val}=-0.61+0.11$X_{\mathrm{C\text{-}P}}$	0.51	0.71
Value and Hoarse-Consonant	Y_{Val}=-0.54+0.09$X_{\mathrm{H\text{-}C}}$	0.38	0.62

TABLE 10. Summary of objective features of timbre.

Symbol	Parameters	dimension
Time-domain objective characteristic parameters	Temporal centriod	1
	Zero-crossing rate	1
	Attack time Decrease	1
	Decrease time	1
	Release time	1
	Log-attack-time	1
	Amplitude modulation	1
	Attack slope	1
	Decrease slope	1
	Frequency modulation	1
	Effective duration	1
Frequency-domain objective characteristic parameters	Spectral centroid	1
	Spectral Spread	1
	Spectral decrease	1
	Spectral skewness	1
	Spectral kurtosis	1
	Spectral roll-off	1
	Spectral-flatness measure	1
	Spectral crest measure	1
	Root mean square energy	1
Harmonic objective characteristic parameters	Harmonic Energy	1
	Noisiness Part Energy	1
	Tristimulus	3
	Harmonic spectral deviation	1
	odd-to-even harmonic energy ratio	1
	Noisiness	1
	f0	1
	Inharmonicity	1
Time-frequency objective characteristic parameters	Spectral flux	1

1) TIME-DOMAIN OBJECTIVE CHARACTERISTIC PARAMETERS

Temporal centriod(TC) is the center of gravity of the signal's energy on the time axis, in seconds(s), which reflects the area where the signal's main energy is concentrated [31], and is an important parameter for describing high-transient shock signals. The formula is as follows:

$$TC = \frac{\sum_{n=n_1}^{n=n_2} t_n \cdot e(t_n)}{\sum_n \mathrm{e}(t_n)} \tag{10}$$

Among them, n1 and n2 are the first and last values of n that can make $\mathrm{e}(\mathrm{t_n})$ greater than 15% of its maximum value, respectively, in order to avoid including silent segments in the TC calculation.

Zero-crossing rate(ZCR) is the number of times the signal passes through the zero value in each frame, reflecting the speed of the signal change [32]. When calculating the zero-crossing rate, first subtract the amount of DC slices of the signal in each frame, and then normalize the zero-crossing rate of each frame with the window length $\mathrm{L_t}$ (in seconds) as the normalization factor. The formula is as follows:

$$ZCR = \frac{1}{2}\sum_{\mathrm{m=0}}^{N} |\mathrm{sgn}\,[\mathrm{x}_n(\mathrm{m}+1)] - \mathrm{sgn}\,[\mathrm{x}_n(\mathrm{m})]| \tag{11}$$

Among them, N is the length of the window function; n represents the number of sampling points in each frame of signal; x(n) is the positive or negative of the amplitude of a certain frame of signal; x(m) and $\mathrm{x}(\mathrm{m}-1)$ are the positive and negative signal amplitudes of the adjacent two sampling points, and sgn[] is the sign function.

$$\mathrm{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \tag{12}$$

Attack slope(AS) is the ratio of the energy difference between the end point and the starting point of the sound signal to the time difference [33]. This parameter can better distinguish between steady-state sound and impact sound (percussion sound). The formula is as follows:

$$AS = \frac{E_{\mathrm{max}} - E_{\mathrm{min}}}{t_{\mathrm{end}} - t_{st}} \tag{13}$$

Among them, $\mathrm{E_{max}}$ is the energy value at the time of $\mathrm{t_{end}}$, $\mathrm{E_{min}}$ is the energy value at the time of $\mathrm{t_{st}}$.

2) FREQUENCY-DOMAIN OBJECTIVE CHARACTERISTIC PARAMETERS

Root mean square energy(RMS) describes the total energy value of the sound signal in the time domain, which is related to the loudness of the sound. The formula is as follows [34]:

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x(n)^2} \tag{14}$$

Spectral centroid is the first moment of the spectrum energy distribution, representing the geometric center of the spectrum, the unit is Hz. This parameter can describe the brightness of the sound. The higher the center of mass, the brighter the sound [35]. The formula is as follows:

$$Centroid\ (t_m) = \frac{\sum\limits_{f=f_{\min}}^{f_{\max}} f \cdot a(f)}{\sum\limits_{f=f_{\min}}^{f_{\max}} a(f)} \tag{15}$$

Spectral Spread is the second-order moment of the spectrum energy distribution, indicating the degree of spectrum diffusion near the center of mass of the spectrum, the unit is Hz. This parameter can reflect the frequency range of the energy concentration of the sound near the center of mass of the spectrum [36]. The formula is as follows:

$$Spread = \sqrt{\frac{\sum\limits_{f=f_{\min}}^{f_{\max}} (f - C_t)^2 \cdot a(f)}{\sum\limits_{f=f_{\min}}^{f_{\max}} a(f)}} \tag{16}$$

Spectral skewness is the third-order moment of the energy distribution of the spectrum. It describes the asymmetry of the spectrum near the centroid of the spectrum. It is a measure of the direction and degree of the skewness of the statistical data distribution [37]. When the skewness coefficient is 0, the spectrum is symmetrical. When the skewness coefficient is greater than 0, when the heavy tail is on the right, the distribution is skewed to the right, that is, the spectrum energy is more biased toward the low frequency end. When the skewness coefficient is less than 0, when the heavy tail is on the left, the distribution is left skewed, that is, the spectrum energy is more biased towards the high frequency end, the calculation formula is as follows:

$$Skewness\ (t_m) = \frac{\left(\sum\limits_{k=1}^{K} (f_k - Centroid\ (t_m))^3 \cdot p_k\ (t_m)\right)}{Spread^3} \tag{17}$$

Spectral kurtosis is the fourth-order moment of the spectrum energy distribution, describing the flatness of the spectrum near the center of mass. When the kurtosis coefficient is 3, the spectrum is normal (Gaussian distribution); when the kurtosis coefficient is greater than 0, the spectrum is more energy concentrated near the mean; when the kurtosis coefficient is less than 0, the spectrum is flat, that is, the energy distribution of the spectrum energy concentrated near the mean is more average [38]. The formula is as follows:

$$Kurtosis\ (t_m) = \frac{\left(\sum\limits_{k=1}^{K} (f_k - Centroid\ (t_m))^4 \cdot p_k\ (t_m)\right)}{Spread^4} \tag{18}$$

Spectral decrease indicates the degree of spectrum amplitude decline [32]. The formula is as follows:

$$\text{decrease}\ (t_m) = \frac{1}{\sum\limits_{k=2}^{K} a_k\ (t_m)} \sum_{k=2}^{K} \frac{a_k\ (t_m) - a_1\ (t_m)}{k-1} \tag{19}$$

Spectral-flatness measure(SFM) describes the flatness of the spectrum distribution, which is the ratio of the geometric mean of the spectrum to its arithmetic mean, and is used to extract the speech signals of turbid and non-turbid sounds [23]. The unit is decibels (dB). For modulated signals, the SFM value is close to 0(peak spectrum), and for noisy signals, the SFM value is close to 1(flat spectrum). The formula is as follows:

$$SFM\ (t_m) = \frac{\left(\prod\limits_{k=1}^{K} \mathrm{a}_k\ (t_m)\right)^{\frac{1}{K}}}{\frac{1}{K} \sum\limits_{k=1}^{K} \mathrm{a}_k\ (t_m)} \tag{20}$$

Spectral crest measure(SCM) refers to the ratio of the peak value of the spectrum energy to the arithmetic mean [23]. The formula is as follows:

$$SCM\ (t_m) = \frac{\underset{k}{\overset{a}{\mathrm{maxk}}}\ (t_m)}{\frac{1}{K} \sum\limits_{k=1}^{K} a_k\ (t_m)} \tag{21}$$

3) HARMONIC OBJECTIVE CHARACTERISTIC PARAMETERS

Harmonic Energy($\mathrm{E_H}$) is the sum of energy for all detected harmonic components [33]. The formula is as follows:

$$E_H\ (\mathrm{t}_m) = \sum_{h=1}^{H} a_h^2\ (\mathrm{t}_m) \tag{22}$$

Among them, $\mathrm{a_h(t_m)}$ represents the amplitude corresponding to $\mathrm{t_m}$ time, H represents the total number of components.

Noisiness Part Energy($\mathrm{E_N}$) is the total energy of the non-harmonic part of the sound signal, which can be approximated as the total energy minus the harmonic energy [39]. The formula is as follows:

$$E_N\ (t_m) = E_T\ (t_m) - E_H\ (t_m) \tag{23}$$

Among them, $\mathrm{E_T}$ represents total energy, $\mathrm{E_H}$ represents harmonic energy.

Tristimulus is the timbre equivalent of color attribute in vision introduced by Pollard and Jansson [39]. T1 represents the ratio of the first harmonic energy to the total harmonic energy, T2 represents the ratio of the energy of the first three harmonics to the total harmonic energy, T3 represents the ratio of the energy of the first five harmonics to the energy of the total harmonics. The calculation formula is as follows:

$$T1\ (t_m) = \frac{a_1\ (t_m)}{\sum\limits_{h=1}^{H} a_h\ (t_m)} \tag{24}$$

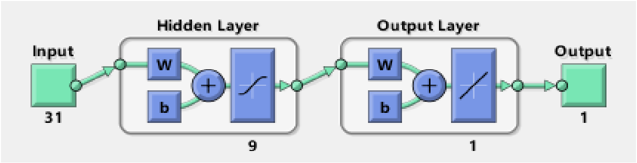


FIGURE 16. Neural network diagram.

$$T2\left(t_m\right) = \frac{a_2\left(t_m\right) + a_3\left(t_m\right) + a_4\left(t_m\right)}{\sum_{h=1}^{H} a_h\left(t_m\right)} \tag{25}$$

$$T3\left(t_m\right) = \frac{\sum_{h=5}^{H} a_h\left(t_m\right)}{\sum_{h=1}^{H} a_h\left(t_m\right)} \tag{26}$$

Among them, H is the total number of harmonics, usually taking H = 20.

Odd-to-even harmonic energy ratio(OER) distinguishes between odd harmonics and smooth spectral envelopes [22]. This parameter can distinguish the timbre with outstanding odd harmonics from the timbre with smooth spectrum envelope. The formula is as follows:

$$OER\left(t_m\right) = \frac{\sum_{h=1}^{H/2} a_{2h-1}^2\left(t_m\right)}{\sum_{h=1}^{H/2} a_{2h}^2\left(t_m\right)} \tag{27}$$

Noisiness is a numerical value proportional to the "noisy" degree of people's subjective judgment of noise. The unit is noy. It is defined as the ratio of noise energy to total energy and reflects the proportion of signal noise components to total energy. The formula is as follows:

$$noisiness\left(\mathrm{t}_m\right) = \frac{E_N\left(\mathrm{t}_m\right)}{E_T\left(\mathrm{t}_m\right)} \tag{28}$$

4) TIME-FREQUENCY OBJECTIVE CHARACTERISTIC PARAMETERS

Spectral flux(SF) is a time-varying descriptor calculated by STFT values, indicating the extent to which the spectrum varies with time [40], defined as one minus the normalized mutual coefficient of the amplitude spectrum of t_m and $\mathrm{t}_{\mathrm{m}-1}$ at two consecutive moments. If the spectrum changes little at the adjacent time, the spectrum flux is close to 0, and if the similarity of the spectrum at the adjacent time is very high, the spectrum flux is close to 1. The formula is as follows:

$$SF = 1 - \frac{\sum_{k=1}^{K} a_k\left(t_{m-1}\right) a_k\left(t_m\right)}{\sqrt{\sum_{k=1}^{K} a_k\left(t_{m-1}\right)^2}\sqrt{\sum_{k=1}^{K} a_k\left(t_m\right)^2}} \tag{29}$$

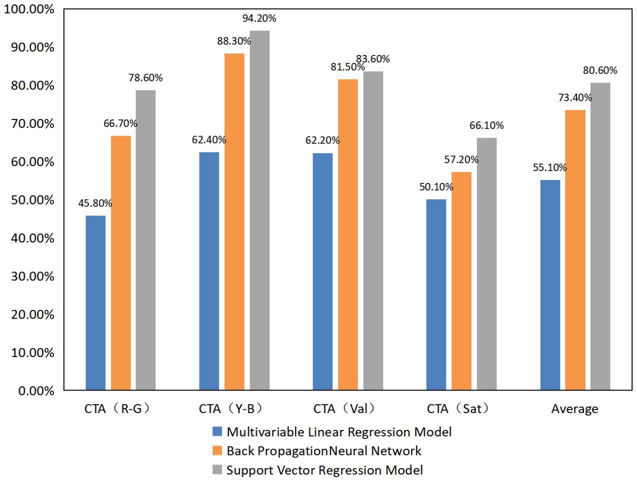


FIGURE 17. Comparison of the accuracy of three algorithms to build models.

TABLE 11. Experimental results.

	Multivariate Linear Regression Model	Back Propagation Neural Network	Support Vector Regression Model
CTA(R-G)	45.8%	66.7%	78.6%
CTA(Y-B)	62.4%	88.3%	94.2%
CTA(Val)	62.2%	81.5%	83.6%
CTA(Sat)	50.1%	57.2%	66.1%
Average	55.1%	73.4%	80.6%

B. CONSTRUCTION OF TIMBRE-COLOR ASSOCIATION MODEL

When constructing the correlation model of timbre and color, in order to predict the correlation between timbre and color more accurately, three algorithms are selected: multivariate linear regression model, Back Propagation neural network and Support vector regression model.

When people have a fuller understanding of the internal characteristics of the research object and the relationship between the various factors, they generally use the mechanism analysis method to establish a mathematical model. The statistical regression model is a very commonly used mathematical model, and when there is more than one independent variable associated with the dependent variable, then the least squares criterion should be considered to establish a multiple linear regression model [41]. Suppose the linear regression model of random y and general variable x_k is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{30}$$

Among them, β_0, $\beta_1, \ldots, \beta_\mathrm{k}$ are unknown parameters, β_0 is called regression constant, $\beta_1, \ldots, \beta_\mathrm{k}$ are called regression coefficients; y is called the explained variable; x_1, $\mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{k}$ are general variables that can be precisely controlled, called explanatory variables. Use the least square method to estimate the estimated regression coefficient β_0, $\beta_1, \ldots, \beta_\mathrm{n}$ in the above formula, obtain the β value, and then use the multiple linear regression model to predict.

Back Propagation neural network(BP neural network) is a multi-layer feedforward network trained according to the

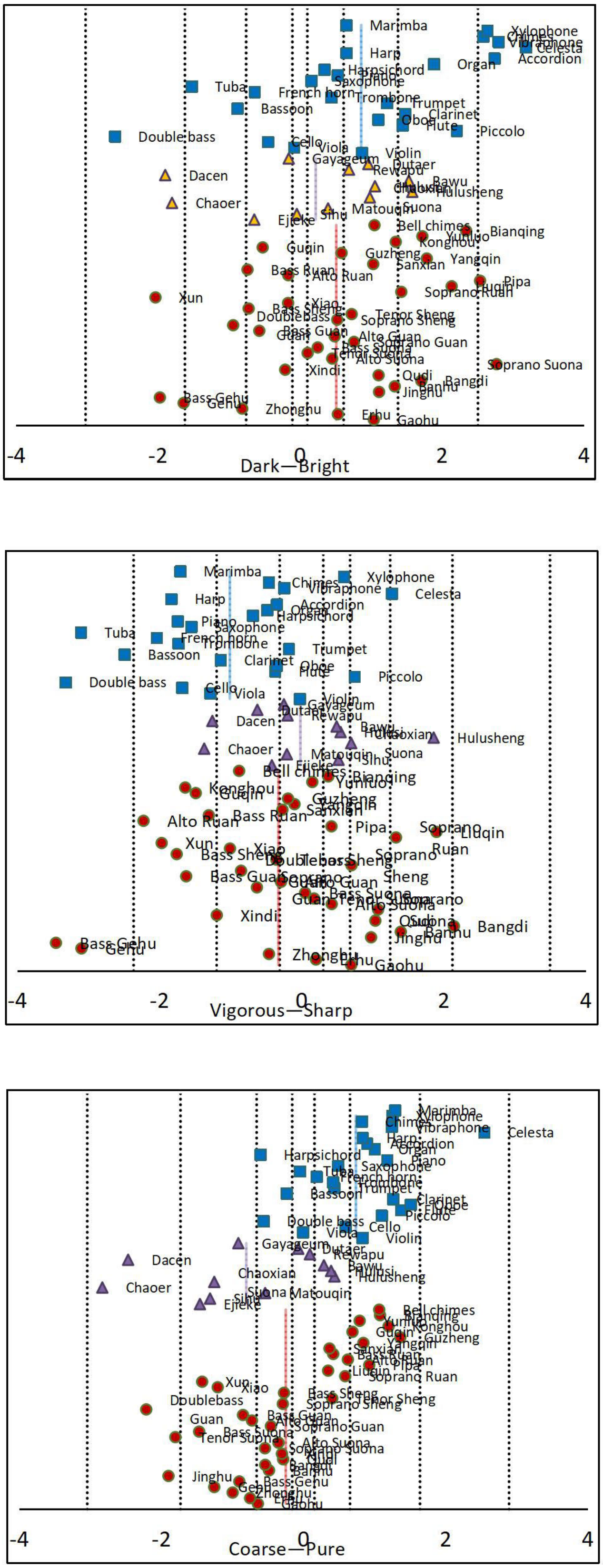


FIGURE 18. The sample distribution of timbre material in the specific timbre evaluation scale.

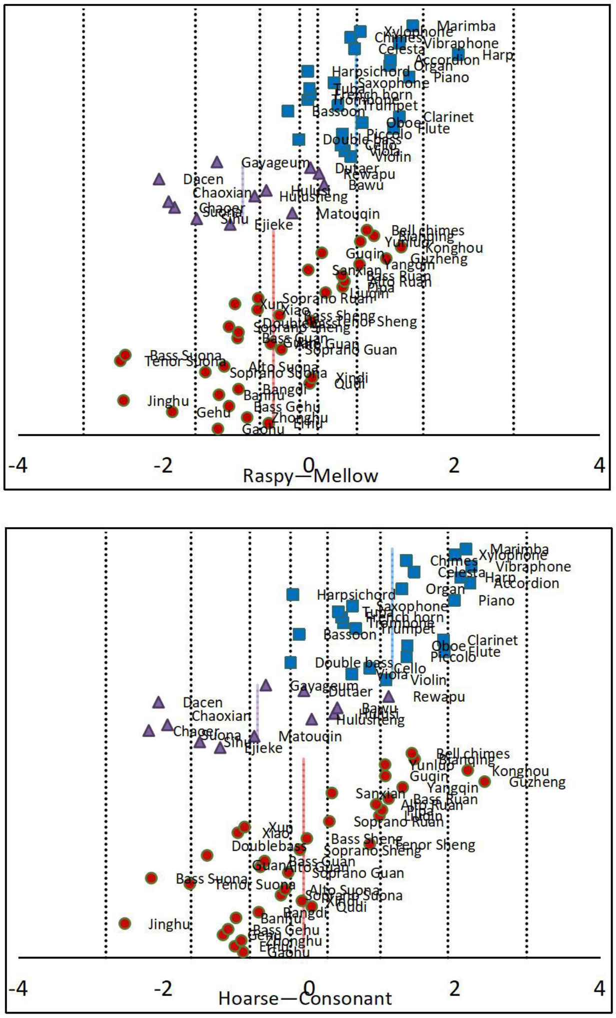


FIGURE 18. *(Continued.)* The sample distribution of timbre material in the specific timbre evaluation scale.

error back propagation algorithm. Its principle is to input learning samples and train the deviation and weight of the network repeatedly through back propagation algorithm. Make the final output value as close as possible to the expected value [42]. The BP neural network model in this article sets a total of three input layers, one hidden layer and one output layer structure. The number of neurons in the three layers is 31, 9, and 1, respectively. The learning rate and the number of iterations are set as 0.01, 1000. Figure 16 shows the neural network diagram of the algorithm.

Support vector regression model(SVR) is based on the VC dimension theory of statistical learning theory and the structural risk minimization principle, the support vector machine approach seeks the best compromise between the complexity of the model (the learning accuracy for a specific training sample) and the learning ability (the ability to identify arbitrary samples without error) based on the limited sample information in order to obtain the best generalization ability. In solving the problem of small

1) COLOR VISUALIZATION RESULTS OF WESTERN MUSICAL INSTRUMENT

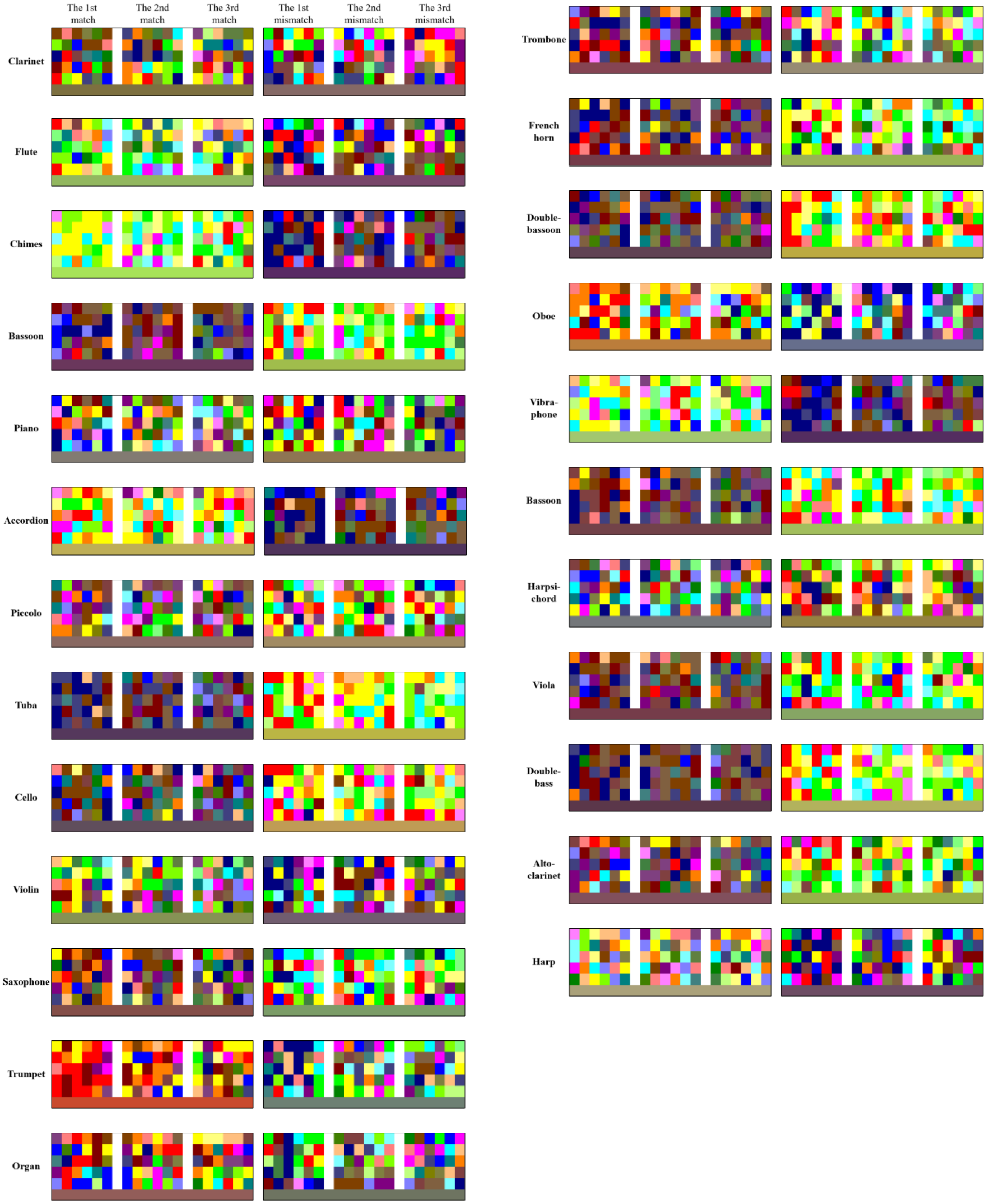


FIGURE 19. Color visualization results of western musical instrument.

sample, nonlinear and high dimensional pattern recognition, it has shown a lot of performance better than the existing methods, and greatly improves the generalization ability of learning methods. Kernel function, as the core of SVR regression algorithm, plays an important role in the accuracy of SVR model [43]. The kernel function used in this paper is the Gaussian radial basis kernel function, see Formula (31).

$$K\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{\sigma^2}\right) \tag{31}$$

When using SVR for regression prediction, we need to adjust the relevant parameters, mainly the penalty parameters c and the kernel function parameter g. The K-CV interactive test method is used to optimize the parameters. First, the original sample data is divided into K groups, each subset data is made a verification set, and the other K − 1 sets of subset data are used as training sets. The range discretization search is needed in the process of determining the model parameter c and parameter g of the K-CV method. In order to improve the efficiency of the operation, it is first determined that both the c and the g parameters are in [2^{-10}, 2^{10}], with a search step of 1 in this exponential coordinate system, locking a rough coordinate of the best parameter combination according to the position of the minimum average square error (Mean Squared Error, MSE) after the interaction test in the whole network, and then reducing the mesh to [2^5, 2^{10}], taking 0.5 as the search step, we find the combination of parameter c and g that makes the interaction verification set MSE minimum. The optimal penalty parameter c and kernel function parameter g of SVR model are optimized by interactive test method. After optimizing the parameters of the support vector machine model through the interactive verification method, use the program to determine the optimal penalty parameter c and the kernel function parameter g. Regressor training is performed on CTA(Red-Green), CTA(Yellow-Blue), CTA(Value), and CTA(Saturation) to obtain the best performance regression prediction model.

In order to evaluate the accuracy of the model, the coefficient of determination R^2 can be used for judgment. R^2 can be understood as how much the predicted value explains the variance of the independent variable, and it measures how well the predicted value fits the true value, which is defined as:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \tag{32}$$

Among them, SSR is the sum of regression squares, SSE is the sum of residual squares, and SST is the sum of total deviations. And $0 < R^2 < 1$, the closer it is to 1, the better the prediction results are.

C. EXPERIMENTAL RESULTS AND DISCUSSION

According to the principle of the above algorithm, the prediction results of the three models are shown in TABLE 11 and figure 17.

TABLE 12. List of musical instruments.

Category	Type	Name of the Instrument
Western Orchestral Instrument (24)	Bowed Instrument(4)	Violin Viola Cello Double Bass
	Woodwind Instrument(6)	Piccolo Flute Oboe Clarinet Bassoon Saxophone
	Brass Instrument(4)	Trombone French horn Tuba Trumpet
	Keyboard Instrument(4)	Piano Organ Harpsichord Accordion
	Plucked Instrument(1)	Harp
	Percussion Instrument(5)	Celesta Vibraphone Chimes Xylopho Marimba

Chinese Orchestral Instrument (37)	Bowed Instrument(7)	Gaohu(高胡) Erhu(二胡) Gehu(革胡)
		Zhonghu(中胡) Bass Gehu(低音革胡)
		Jinghu(京胡) Banhu(板胡)

	Wind Instrument(17)	Bass Sheng(低音笙) Xindi(新笛) Soprano Suona (高音唢呐) Double bass Guan (倍低音管) Bass Guan(低音管) Alto Guan(中音管) Bass Suona(低音唢呐) Bangdi(梆笛) Alto Suona(中音唢呐) Xiao(萧) Soprano Guan(高音管) Xun(埙) Soprano Sheng(高音笙) Qudi(曲笛) Tenor Suona(次中音唢呐) Tenor Sheng(中音笙) Bawu(巴乌)

	Plucked Instrument(10)	Soprano Ruan(小阮) Alto Ruan(中阮)
		Bass Ruan(大阮) Liuqin(柳琴)
		Pipa(琵琶) Yangqin(扬琴)
		Konghou(箜篌) Guzheng(古筝)
		Guqin(古琴) Sanxian(三弦)

	Percussion Instrument(3)	Bell chimes(编钟) Bianqing(编磬) Yunluo(云锣)

Chinese Minority Instrument (11)	Bowed Instrument(4)	Ejieke(艾捷克) Sihu(四胡)
		Matouqin(马头琴) Chaoer(潮尔)

	Wind Instrument(4)	Chaoxian Suona(朝鲜唢呐) Dacen(大岑) Hulusheng(葫芦笙) Hulusi(葫芦丝)

	Plucked Instrument(3)	Rewapu(热瓦普) Dutaer(都塔尔)
		Gayageum(伽倻琴)

When the multiple linear regression model is used to predict each attribute of color, the expression can be obtained. The expression of the prediction of color four attributes (Red-Green, Yellow-Blue, Value, Saturation) are as follows:

$$\begin{aligned} CTA_{(RG)} = {} & 0.44 + 0.39X_{RMS} + 0.62X_{Centroid} \\ & + 0.45X_{Spread} + 0.67X_{Skewness} \\ & + 0.02X_{Kurtosis} + 0.11X_{SCM} \end{aligned} \tag{33}$$

$$\begin{aligned} CTA_{(YB)} = {} & 0.67 + 0.47X_{RMS} + 0.54X_{f0} + 0.25X_{HSD} \\ & + 0.38X_{OER} + 0.30X_{SF} + 0.80X_{SCM} \end{aligned} \tag{34}$$

$$\begin{aligned} CTA_{Val} = {} & 2.23 + 0.53X_{Noisiness} - 0.29X_{T3} + 0.40X_{OER} \\ & - 0.66X_{decrease} - 0.54X_{SFM} \end{aligned} \tag{35}$$

$$\begin{aligned} CTA_{Sat} = {} & 0.22 + 0.45X_{EH} + 0.54X_{EN} + 0.75X_{Centroid} \\ & + 0.15X_{Spread} + 0.91X_{Skewness} + 0.81X_{Kurtosis} \end{aligned} \tag{36}$$

According to the evaluation criteria of the model prediction results and formula (38), we evaluate the above three algorithms. From Table 9 and Figure 17, we can see that the accuracy (R2) of the three models in descending order are: Support Vector Regression Prediction Model(80.6%), BP Neural Network(73.4%), Multiple Linear Regression Model(55.1%), which shows that the use of Support Vector Regression

2) COLOR VISUALIZATION RESULTS OF CHINESE MUSICAL INSTRUMENT

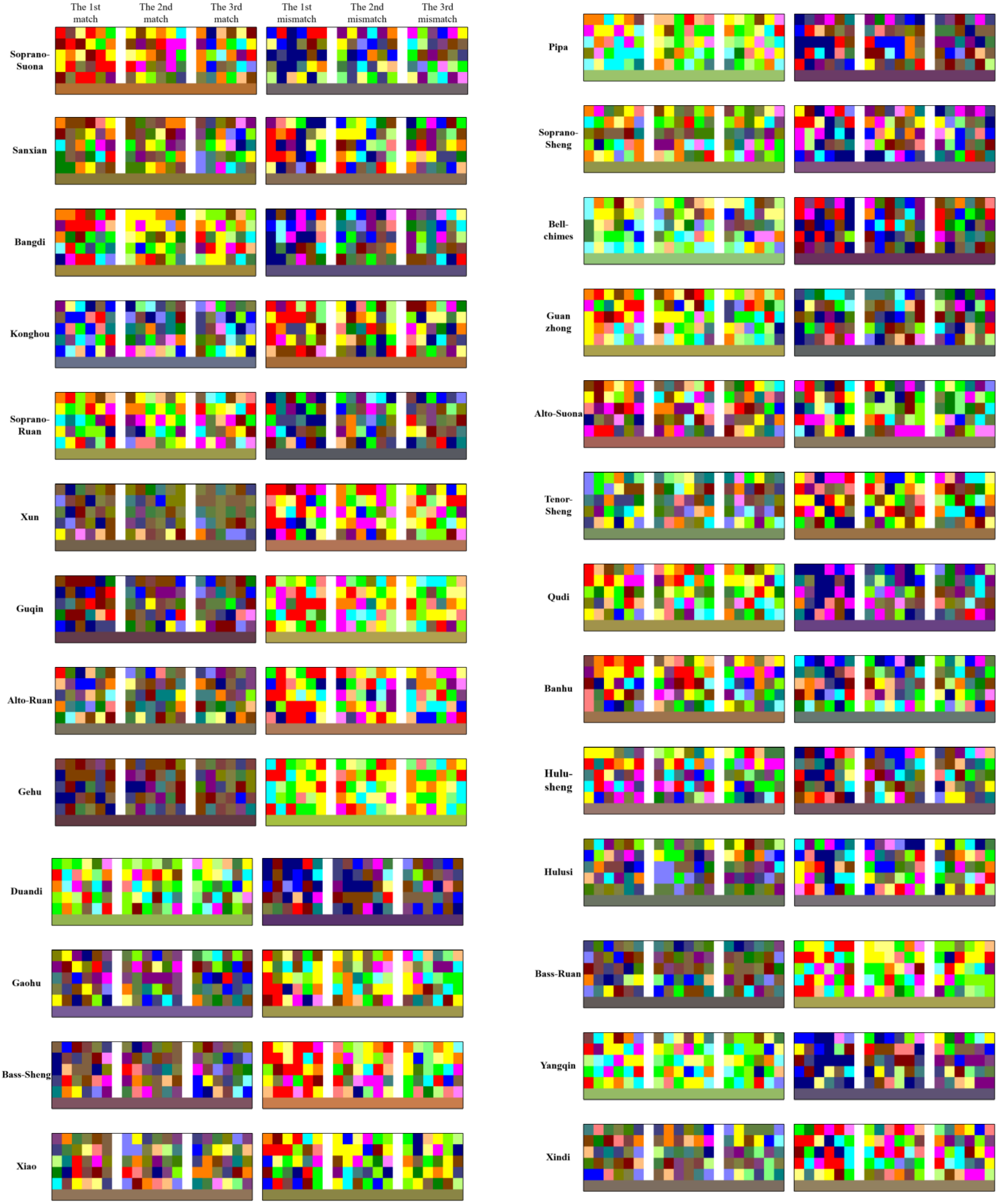


FIGURE 20. **Color visualization results of Chinese musical instrument.**

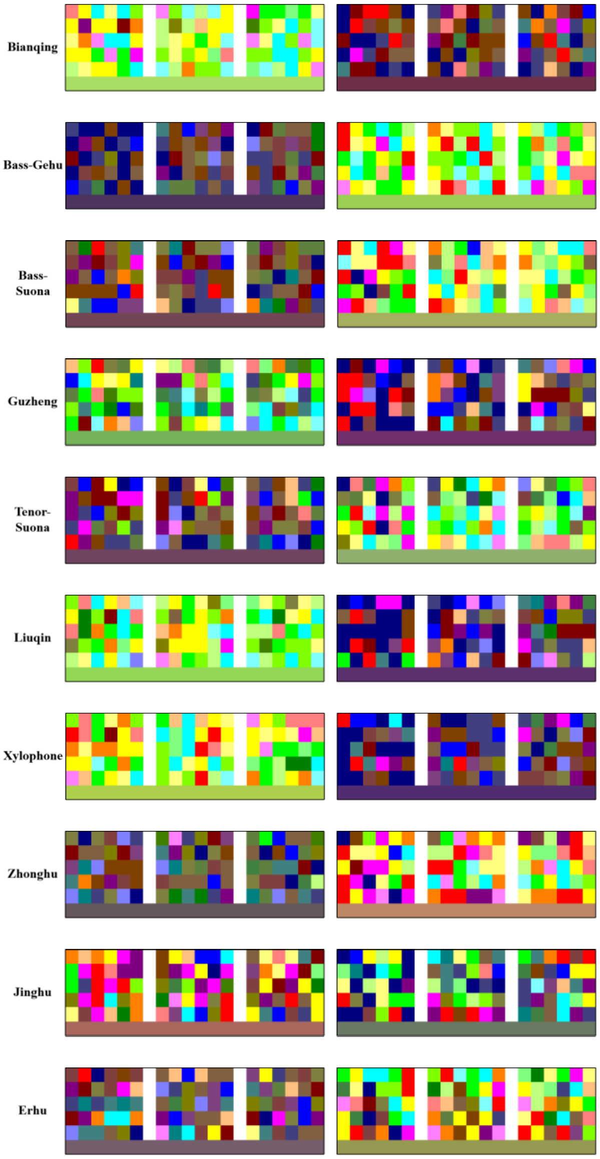


FIGURE 20. ***(Continued.)*** **Color visualization results of Chinese musical instrument.**

Prediction Model to predict color attributes is the best. Using Multiple Linear Regression Prediction Model can reflect the importance of each feature of timbre for color prediction, and can also illustrate the correspondence between feature parameters and colors. However, the accuracy of these three models is not very high, and the amount of data needs to be supplemented to build a more accurate model.

In the prediction of the four attributes of color, the three models have higher prediction accuracy rates for CTA (Val) and CTA (Y-B). This is due to the significant correlation between the characteristics of timbre and the yellow-blue, value of the color. It is consistent with the conclusion obtained in the third part, that is, non-sustainable instruments can remind people of brighter colors such as yellow; for sounds that are purer, softer, and hoarser, they will remind people

3) COLOR VISUALIZATION RESULTS OF CHINESE MINORITY MUSICAL INSTRUMENT

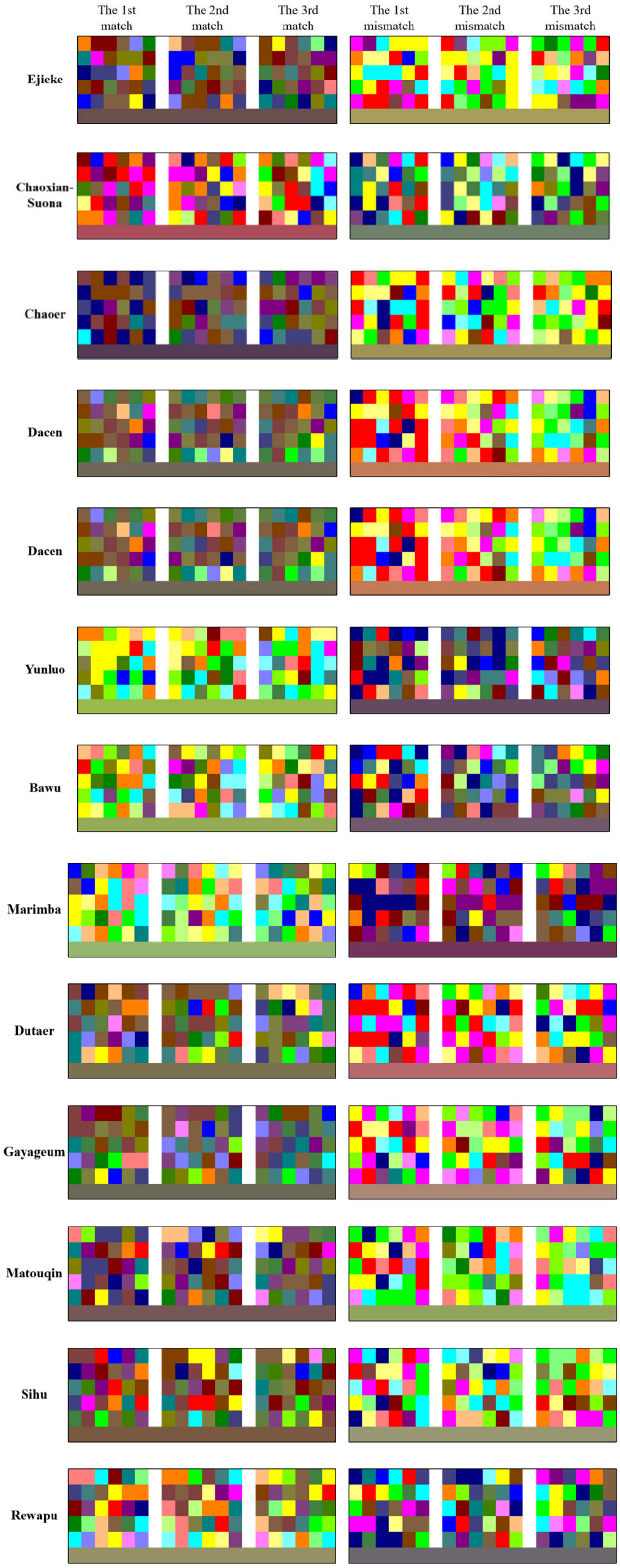


FIGURE 21. **Color visualization results of Chinese minority musical instrument.**

of brighter colors such as yellow. The prediction accuracy of CTA (Sat) of the three models is low, because the characteristics of the timbre and the saturation of the color are not significantly related, which is also consistent with the conclusion drawn in the third part, that is, the relationship between the saturation of the color and the various timbre attributes is not obvious.

V. SUMMARY

This paper focuses on the relationship between the timbre attributes of musical sound and visual color. Through subjective evaluation experiments of cross-modal perception, a dataset related to timbre and color was constructed, and the completeness of the data set was tested. Through qualitative analysis methods, such as correlation analysis and visualization processing, it is further verified that certain attributes of timbre have a strong correlation with certain attributes of color. Based on the constructed timbre-color association database, three algorithms were used to construct a timbre-color association model: a multivariate linear regression model, a back propagation neural network and a support vector regression model. Equations for the multivariate linear regression model and the prediction of the three models were obtained. The results provide guidance for predicting color using objective characteristic parameters of timbre and verify the correlation between timbre and color.

The following research can be carried out to address the following aspects. First, it is necessary to build a richer dataset that involves more situations. Second, because in practical engineering applications, most of the scenarios are real scenes, such as stage performances combining lighting and music, a music fountain and other engineering applications, the color material used in this experiment is artificially generated rather than material with real color images. Therefore, to make the research results more effective in practical engineering applications, real color images should be used as color materials in future research. Third, the sound in practical application is a combination of many timbres. To make the research results more applicable to practical engineering, we will consider including composite timbre material to enrich the experimental timbre material library in future research. Fourth, the models were established on the basis of a small sample data set currently, and we need to further improve the algorithm of the model by increasing the sample data set in the next research endeavor; in this way, we can build a more accurate model and improve the accuracy of the model.

APPENDIX

A. APPENDIX 1: SAMPLE DISTRIBUTION DIAGRAM IN THE PAIR OF TIMBRE EVALUATION SCALES

See Figure 18a and 18b.

B. APPENDIX 2

See Table 12.

C. APPENDIX 3

1) COLOR VISUALIZATION RESULTS OF WESTERN MUSICAL INSTRUMENT

See Figure 19.

2) COLOR VISUALIZATION RESULTS OF CHINESE MUSICAL INSTRUMENT

See Figure 20a and 20b.

3) COLOR VISUALIZATION RESULTS OF CHINESE MINORITY MUSICAL INSTRUMENT

See Figure 21.

REFERENCES

[1] H. Zhou, *The World of Music and Its Expression*. Beijing, China: Central Conservatory of Music Press, 2008.

[2] K. O. Bushara, J. Grafman, and M. Hallett, "Neural correlates of auditory–visual stimulus onset asynchrony detection," *J. Neurosci.*, vol. 21, no. 1, pp. 300–304, Jan. 2001.

[3] L. Busse, K. C. Roberts, R. E. Crist, D. H. Weissman, and M. G. Woldorff, "The spread of attention across modalities and space in a multisensory object," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 51, pp. 18751–18756, Dec. 2005.

[4] M. Eimer and E. Schröger, "ERP effects of intermodal attention and cross-modal links in spatial attention," *Psychophysiology*, vol. 35, no. 3, pp. 313–327, 1998.

[5] H. Jian, "The associative relationship between music and color-the 'synthesis' effect of tone and color and its implementation method (part 1)," *Audio Technol.*, vol. 2009, no. 11, pp. 67–69 and 72, 2009.

[6] G. Jianyong, "Color, music, and composition techniques (part 1)," Qilu Realm of Arts, Jinan, China, Tech. Rep. 1002-2236, 1999, pp. 50–56, no. 1.

[7] P. Zihua, "Analysis and interpretation of music color," *J. Nanjing Univ. Arts, Music Perform. Ed.*, vol. 2011, no. 1, pp. 46–51, 2011.

[8] M. Dongfeng, "Proposal and development of music audiovisual theory," *Symphony-J. Xi'an Conservatory Music*, vol. 28, no. 1, pp. 15–20, 2009.

[9] Z. Wentao, "From 'color music' to 'visual music': The conceptual source of western audiovisual synesthesia and its early art experiments," *J. Nanjing Univ. Arts, Fine Arts Des.*, vol. 2017, no. 5, pp. 96–106, 2017.

[10] J. L. Caivano, "Color and sound: Physical and psychophysical relations," *Color Res. Appl.*, vol. 19, no. 2, pp. 126–133, 2015.

[11] K. B. Schloss, P. Lawler, and S. E. Palmer, "The color of music," *J. Vis.*, vol. 8, no. 6, p. 580, 2010, doi: 10.1167/8.6.580.

[12] S. E. Palmer, T. Langlois, T. Tsang, K. B. Schloss, and D. J. Levitin, "Color, music, and emotion," *J. Vis.*, vol. 11, no. 11, p. 391, 2011.

[13] W. S. Griscom and S. E. Palmer, "The color of musical sounds: Color associates of harmony and timbre in non-synesthetes," *J. Vis.*, vol. 12, no. 9, p. 74, 2012.

[14] W. Griscom and S. Palmer, "Violins are green, pianos are blue: Cross-modal sound-to-sight associations with timbre in synesthetes & non-synesthetes," *J. Vis.*, vol. 13, no. 9, p. 1169, 2013.

[15] W. Griscom, "Visualizing sound: Cross-modal mapping between music and color," M.S. theses, Gradworks, Regina, SK, Canada, 2015.

[16] E. S. Isbilen and C. L. Krumhansl, "The color of music: Emotion-mediated associations to bach's well-tempered clavier," *Psychomusicol., Music, Mind, Brain*, vol. 26, no. 2, p. 149, 2016.

[17] Z. Congcong, "The emotional connection between music and color," East China Normal Univ., Shanghai, China, Tech. Rep., 2014.

[18] X. Min, G. Zhai, Z. Gao, C. Hu, and X. Yang, "Sound influences visual attention discriminately in videos," in *Proc. 6th Int. Workshop Quality Multimedia Exp. (QoMEX)*, Sep. 2014, pp. 153–158.

[19] X. Min, G. Zhai, C. Hu, and K. Gu, "Fixation prediction through multimodal analysis," in *Proc. Vis. Commun. Image Process. (VCIP)*, Dec. 2015, pp. 1–4.

[20] X. Min, G. Zhai, J. Zhou, X.-P. Zhang, X. Yang, and X. Guan, "A multimodal saliency model for videos with high audio-visual correspondence," *IEEE Trans. Image Process.*, vol. 29, pp. 3805–3819, 2020.

[21] X. Min, G. Zhai, J. Zhou, M. C. Q. Farias, and A. C. Bovik, "Study of subjective and objective quality assessment of audio-visual signals," *IEEE Trans. Image Process.*, vol. 29, pp. 6054–6068, 2020.

[22] W. Jiang, J. Liu, X. Zhang, S. Wang, and Y. Jiang, "Analysis and modeling of timbre perception features in musical sounds," *Appl. Sci.*, vol. 10, no. 3, p. 789, Jan. 2020.

[23] C. Keyan, *Auditory Perception and Automatic Recognition of Environmental Sound*. Beijing, China: Science Press, 2014.

[24] Z. Siyu, J. Peifeng, and Y. Jun, "Research on the pitch, loudness and timbre brightness of typical Chinese and western musical instruments," *Appl. Acoust.*, vol. 36, no. 6, pp. 481–489, 2017.

[25] Z. Jiaxing, L. Jingyu, and L. Zijin, "A study on the loudness balance of Chinese national orchestra instruments," in *Proc. Nat. Acoust. Conf. Physiol. Acoust., Psychoacoust., Music Acoust.*, 2018, pp. 34–35.

[26] *The Method of Subjective Assessment of the Sound Quality for Broadcast Programmes and the Technical Parameters Requirements*, Standard GB/T 16463-1996, 1996.

[27] T. T. Tuliis and L. Wood, "How many users are enough for a card-sorting study," in *Proc. UPA Conf.*, Minneapolis, MN, USA, 2004, pp. 1–10.

[28] *Methods for Picture and Audio Subjective Assessment of Digital Television Receiving Equipment*, Standard GB/T 22123-2008, 2008.

[29] H. Qi, "The relationship between warm and cold colors and happiness: A metaphorical perspective," Chin. Psychol. Soc., Shaanxi Normal Univ., Beijing, China, Tech. Rep., 2019.

[30] R. Arnheim, *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA, USA: Univ. of California Press, 1974.

[31] Y. Ando, "Autocorrelation-based features for speech representation," *J. Acoust. Soc. Amer.*, vol. 133, no. 5, p. 3292, 2013.

[32] X. Yang, B. Tan, J. Ding, J. Zhang, and J. Gong, "Comparative study on voice activity detection algorithm," in *Proc. Int. Conf. Electr. Control Eng.*, Jun. 2010, pp. 599–602.

[33] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *J. Acoust. Soc. Amer.*, vol. 130, no. 5, pp. 2902–2916, Nov. 2011.

[34] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A survey of audio-based music classification and annotation," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 303–319, Apr. 2011.

[35] G. Agostini, M. Longari, and E. Pollastri, "Musical instrument timbres classification with spectral features," *EURASIP J. Adv. Signal Process.*, vol. 2003, no. 1, 2003, Art. no. 943279.

[36] W. A. Sethares, R. D. Morris, and J. C. Sethares, "Beat tracking of musical performances using low-level audio features," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 2, pp. 275–285, Mar. 2005.

[37] B. K. Baniya, J. Lee, and Z.-N. Li, "Audio feature reduction and analysis for automatic music genre classification," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2014, pp. 457–462.

[38] S. Shukla, S. Dandapat, and S. R. M. Prasanna, "Spectral slope based analysis and classification of stressed speech," *Int. J. Speech Technol.*, vol. 14, no. 3, pp. 245–258, Sep. 2011.

[39] H. F. Pollard and E. V. Jansson, "A tristimulus method for the specification of musical timbre," *Acta Acustica United Acustica*, vol. 51, no. 3, pp. 162–171, 1982.

[40] J. Saunders, "Real-time discrimination of broadcast speech/music," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Conf.*, May 1996, pp. 993–996.

[41] Y. Wang and S. D. Zhao, "A nonparametric empirical bayes approach to large-scale multivariate regression," *Comput. Statist. Data Anal.*, vol. 156, Apr. 2021, Art. no. 107130.

[42] L. Mingxing, "Music classification model based on BP neural network," *Mod. Electron. Technol.*, vol. 41, no. 5, pp. 136–139, 2018.

[43] J. Liu and L. Xie, "SVM-based automatic classification of musical instruments," in *Proc. Int. Conf. Intell. Comput. Technol. Automat.*, 2010, pp. 669–673.

JINGYU LIU was born in Xinxiang, Henan, China in 1987. He received the B.S. degree in recording engineering from the Communication University of China, in 2010, and the M.S. degree in musical acoustics from the Central Conservatory of Music, in 2014. He is currently pursuing the Ph.D. degree in signal processing with the Communication University of China.

His research interests include musical acoustics, perception and cognition of musical timbre, perceptual foundations of orchestration (acoustics of orchestration), sensory evaluation of sound, and visual auditory fusion information processing.

ANNI ZHAO was born in Fenyang, Shanxi, China, in 2000. She is currently pursuing the bachelor's degree with the Department of Automation, School of Information and Communication Engineering, Communication University of China. Her research interests include automatic control, digital image-processing, and pattern recognition.

SHUANG WANG was born in Taizhou, Jiangsu, China, in 1989. She received the B.S. degree in digital media processing from the Nanjing University of Posts and Telecommunications, in 2012, and the M.S. degree in signal processing from the Communication University of China, in 2015, where she is currently pursuing the Ph.D. degree in signal processing.

Her research interests include visual auditory fusion information processing, visual perceptual feature modeling, and computer vision.

YIYANG LI was born in Zhumadian, Henan, in 2000. She is currently pursuing the bachelor's degree with the Department of Automation, School of Information and Communication Engineering, Communication University of China. She has published two papers in international journals and conferences, of which one have been retrieved EI.

HUI REN was born in Shuozhou, Shanxi, in 1966. He is currently a Doctor, a Professor, and a Ph.D. Supervisor. He is also the Director with the Department of Automation, Communication University of China. He has hosted more than ten national, provincial, and ministerial level scientific research projects. More than 100 academic papers have been published in core journals and international conferences, of which more than 50 articles have been retrieved from SCI and EI. His research interest includes intelligent information processing and control.

He is the Vice Director of the Beijing Electrotechnics Society.