A Study on the Relationship Between Timbre Perception Features and Emotion in Musical Sounds

Jiying Guo[1-4], Jingyu Liu[1-4], Zijin Li[5], Jiaxing Zhu[5], Wei Jiang[1-4*]
[1] State Key Laboratory of Media Convergence of Communication
[2] Key Laboratory of Acoustic Visual Technology and Intelligent Control System
(Communication University of China), Ministry of Culture and Tourism
[3] Beijing Key Laboratory of Modern Entertainment Technology (Communication University of China)
[4] School of Information and Communication Engineering, Communication University of China
[5] Musicology Department China Conservatory of Music
Beijing, China
guojiying@163.com, drumking@126.com, lzj@ccmusic.edu.cn, jiaxingzhu@tju.edu.cn, jw@cuc.edu.cn

***Abstract*—This paper mainly studies the relationship between timbre perception features and emotion in musical sounds. Firstly, a five - dimensional emotional space was determined by multidimensional scaling analysis. Then emotional labeling subjective evaluation experiments were carried out for 72 scale materials and 48 melody materials of timbre to obtain emotional data of timbre materials. The correlation between timbre perception features and emotion is studied according to the playing content, temporal characteristics, and musical instrument types. Finally, a set of mathematical models of the relationship between timbre perception features and emotion was established by the method of multiple linear regression. The experimental results show that the features of timbre perception are an important effect on musical emotion, and these models have good predictive ability for the five emotion dimensions.**

Keywords-timbre emotional association; timbre perception features; emotional space; multiple linear regression

I. INTRODUCTION

The music sound is one of the important factors that can arouse emotions. Composers create different music with different emotions. Listeners can feel strong emotions in the music. Timbre and emotion are the two most important aspects of musical sound, both of which are complex, multidimensional and closely related.

The features of timbre can be divided into the objective features of the lower level, the perceptual features of the middle level, and the semantic features of the higher level. Studies of timbre and emotion usually establish timbre space based on timbre features and then relate it to emotion. Most scholars agree that emotions can be better sensed through human voices, but there is little research on the timbre of non-phonetic sounds (instrumental music). In 2009, Hailstone et al. proved that on the premise of not changing other musical features, different timbre (Musical Instruments) would affect the emotional perception results [1]. In 1977, Scherer and Oshinsky found that timbre is an important factor in synthetic timbre rating [2]. Moreover, researchers have found that compared with other features in music, such as harmony and melody, timbre can help to accelerate the classification of emotions [3]. In 2012, Eerola et al. conducted a hearing test to investigate the correlation between emotions and spectral characteristics [4], and confirmed the strong correlation between some features and emotional dimension, valence and arousal. In order to study the relationship between the objective spectrum characteristics of timbre and emotion[5, 6], Wu Bin et al. judged whether the timbre is affected by changing the slope of the spectrum. The study proved that changing the slope of the spectrum will affect the timbre. Recognition has a significant impact, which means that the spectral centroid will affect the tone. In the same year, Wu demonstrated that the spectral center of mass in the timbre of an instrument is a major factor affecting emotion. Most studies on the relationship between timbre space and emotion were based on the objective features of the underlying timbre. At present, studies have established a timbre space by extracting the objective features of timbre, which can describe the timbre of Musical Instruments in a more comprehensive way [7, 9]. At present, few studies have analyzed emotions from the timbre perception features, and most studies have analyzed emotions through the underlying objective features of timbre.

This study first established an emotional space suitable for timbre evaluation, and carried out emotional labeling experiments based on the emotional space. Then the different influences of playing content, temporal characteristics and musical instrument type on timbre perception characteristics and emotional associations were studied. In the fourth part, emotion modeling was carried out for the timbre perception features with high correlation to realized the quantitative description of emotion through timbre perception features. Figure 1 is the main actor research idea of this paper.

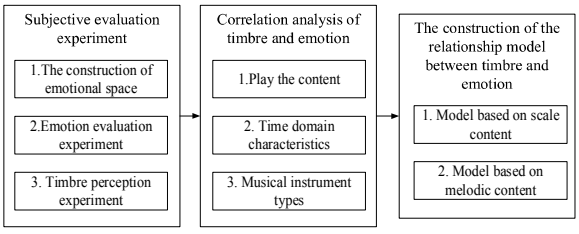

Figure 1. The article structure.

II. SUBJECTIVE EVALUATION EXPERIMENT

This part firstly determines the five-dimensional emotion space, then carries out the emotion evaluation experiment

978-1-7281-8138-7/20/$31.00 ©2020 IEEE
DOI 10.1109/ICCST50977.2020.00010

based on the series category method to the timbre material, and carries on the reliability test, the validity test, the noise elimination and the normal distribution verification to the experimental data. The experimental results will be used in the correlation analysis of timbre perception features and emotions as well as the parametric modeling of emotions.

A. The Construction of Emotional Space

Most of the famous theories on the semantic description of music emotion are western studies, and there are few studies on music emotion in Chinese semantic background in China. If western emotion descriptors are used directly, there will be semantic differences, so this paper selects the emotional evaluation words in the Chinese emotional evaluation thesaurus [8]. Firstly, based on the relevant research at home and abroad, 70 words were determined from the Chinese context for the semantic survey of music emotion. The survey results selected 25 emotional evaluation words with word frequency over 29% and added an emotional word -- helpless. Due to the evaluation of 26 words between synonyms and antonyms, through the subjective experiment analysis of 26 emotional words, experiment material chooses the 39 western famous music clips and the emotion tendency obvious music clips, with level 5 evaluation scale for emotion scores, psychological scale calculation experimental data, by 39 music clips on the emotional words psychological scale can get 26 the correlation matrix of emotional words.

1) Emotion evaluation words based on cluster analysis

To reduce information redundancy, cluster analysis is used to analyze the correlation matrix of emotion evaluation words, because some emotion words have a high correlation. The system cluster genealogy chart showed the degree of correlation between 26 affective evaluation words. Through factor analysis, the cumulative variance contribution rate of the first five factors is 99.87%. Therefore, it is appropriate to determine the cluster number of 26 evaluation words as five categories.

2) Emotional evaluation words based on multidimensional scale analysis

To obtain the spatial distance relations of 26 emotional evaluation words, the correlation matrix of them can be analyzed using MDS(Multidimensional Scaling Analysis) method. Figure 2 shows the results of MDS analysis, indicating that the 26 evaluation words have obvious clustering classification and some evaluation words are close to each other. Then, five emotional clusters are formed according to the results of factor analysis.

Emotion words representing the five emotional tendencies refer to the Hevner emotional circle, and the emotion words that are more suitable to describe the timbre are selected in combination with the opinions of the music majors. Then the emotion words that are more easily understood are selected according to the opinions of the non-music majors. Finally, 5 emotional evaluation words are obtained. Table □ shows the emotional space composed of five evaluation words, which will be used for the experiment of emotional evaluation based on the serial category method and the parametric modeling of emotional space.

TABLE I. FIVE EMOTIONAL EVALUATION WORDS AND THEIR ORIGINAL TEXTS (CHINESE)

Emotion Categories
Quiet (宁静)
Holy (神圣)
Happy (愉快)
Sad (悲伤)
Passion (激情)

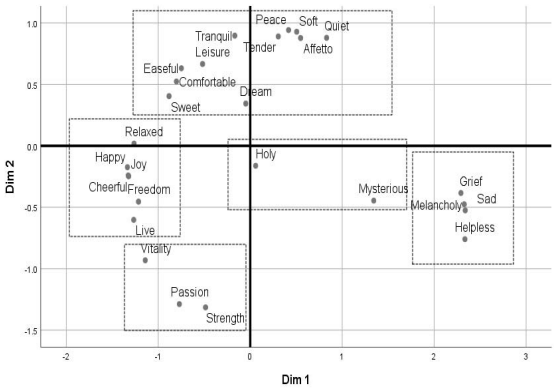

Figure 2. MDS results of 26 emotional evaluation words.

B. Emotional Evaluation Experiment

Firstly, the serial category method was used to conduct an emotional evaluation experiment on the materials in the material library, and reliability and validity test was conducted on the experimental results to ensure the reliability and validity of the experimental data.

1) The experimental material

To include as many timbre types as possible, the experimental materials include 72 scale materials for 72 timbres and 220 melodic materials for 48 timbres, with a length of about 10S-15s. Musical Instruments include Chinese ethnic instruments, Chinese minority Musical Instruments and Western Musical Instruments. The main objective of this study is timbre, but some studies have proved that timbre and tone are inseparable, so the timbre material in this study also includes the tonal factor. To exclude the influence of the loudness factor on the experiment, the loudness of the experimental materials was normalized.

2) Subjects

In this experiment, 66 non-hearing impaired subjects aged between 22 and 35 took part in the questionnaire. There were 33 participants in both scales of emotional experiment

and melody of emotional experiment, including 14 males. Before the formal experiment, all the subjects had received professional audio training, had a certain listening experience, and had a good understanding of musical emotion.

3) Listening Test Design

The subjective evaluation experiment was conducted in the standard listening room. The reverberation time of the room was 0.3 seconds, which was in line with the international listening standard. Before the experiment, the experimental process and matters needing attention were explained to the subjects, and the emotion and timbre in the experiment were annotated. To enable the subjects to grasp the evaluation criteria and scoring scale, before the formal experiment, the scales/melody of the same form outside the three material Banks were played for pre-experiment. The formal experimental procedure was to play the material, and the subjects judged the emotional psychological scale of the material once and scored it on a nine-level scale. In the experiment, each footage was played 2-3 times at an interval of 5s, with a break of 10 minutes every half an hour.

4) The data processing

Firstly, the reliability test was carried out on the sample data in the experiment. In the study, the reliability coefficient method was used for the reliability test. The test results showed that the reliability coefficient of all the data was greater than 0.8, indicating that the experimental data were ideal and highly reliable.

The validity test was conducted on the experimental results through factor analysis, and the data of the validity test had to meet the following two conditions: KMO value (Kaiser - Meyer - Olkin) is greater than 0.6 and the Bartlett spherical P value is less than 0.05. The factor analysis results showed that KMO values were all greater than 0.8, P values were all less than 0.05, and the cumulative percentage was all greater than 50%. Therefore, it was considered that the experimental data were valid and the next data analysis could be carried out.

In this study, the noise data were processed using Laida criterion method, the original data were standardized by z-score, and noise data were eliminated according to the multiple of Z-score and standard deviation of 1.5 times the standard deviation. For the following correlation analysis, the data were verified by the normal distribution, and the verification results showed that the emotional data of melody and scales were in line with the normal distribution. Therefore, the Pearson correlation coefficient was used in the next correlation analysis.

C. Timbre Perception Experiment

The features of timbre perception in this paper are determined by timbre perception experiments. Firstly, a glossary of timbre evaluation terms is constructed and a subjective timbre evaluation experiment based on a forced selection method is carried out. The subjective evaluation experiment was conducted in the standard listening room. The reverberation time of the room was 0.3 seconds, which met the listening standard. A total of 41 subjects with a professional musical background, aged between 18 and 35, had a male to female ratio of nearly 1: 1 and normal hearing. In the subjective evaluation experiment, the forced selection method was adopted: the audio in the material library was played successively, and the subjects judged whether each evaluation word was suitable to describe the audio segment. Then the distant relation of 32 timbre evaluation terms in two-dimensional space is obtained by analyzing the experimental data with the MDS method, and 16 representative timbre evaluation terms are selected from the results of cluster analysis. Finally, correlation analysis was used to calculate the correlation of 16 evaluation terms. According to the analysis results, 6 evaluation terms with high correlation (correlation coefficient > 0.85) were eliminated, and the remaining 10 evaluation terms were paired with 5 pairs with opposite polarity (absolute value of correlation coefficient > 0.81). Table □ shows five timbre perception features.

TABLE II. FIVE FEATURES OF TIMBRE PERCEPTION INCLUDE THE ORIGINAL TEXT (CHINESE) AND THE ABBREVIATIONS

Timbre Perception Features	Abbreviations
Bright-Dark (明亮 - 暗淡)	B-D
Raspy-Mellow (干瘪 - 柔和)	R-M
Sharp-Vigorous (尖锐 - 浑厚)	S-V
Coarse-Pure (粗糙 - 纯净)	C-P
Hoarse – Consonant (嘶哑 - 协和)	H-C

III. A STUDY ON THE CORRELATION BETWEEN TIMBRE PERCEPTION FEATURES AND EMOTION

This part studies the relationship between timbre and emotion according to the experimental data of subjective evaluation of emotion. Firstly, the correlation coefficient of each timbre perception feature and each emotion is calculated to determine whether there has a correlation between timbre perception feature and emotion. Then the material is divided into scales and melody according to the content of the instrument. According to the different features of time-domain, the instruments are divided into continuous instruments and non-continuous instruments. According to the different types of Musical Instruments, Musical Instruments are divided into Chinese and Western Musical Instruments to further study the impact of different timbre perceptual features on emotion. Here the principle of continuous and non-continuous instrument classification by with the Musical Instruments sound way, non-durable instruments sound attenuation and unsustainable, play immediately after general instruments can be divided into playing Musical Instruments, string instruments, plucked instruments, and percussion, wind instruments and string instrument for continuous instruments, percussion instruments and plucked instrument are unsustainable.

To present the correlation between timbre perception features and emotion, firstly, the correlation coefficients between timbre perception features and emotion are calculated according to the classification of three factors, namely playing content, time-domain characteristics and musical instrument type. Then a line chart is drawn according to the correlation coefficients. To analyze the influence of three factors on timbre perception features and emotional association, correlation coefficients of different factors and the same timbre perception features were analyzed. Table □ shows the correlation coefficients of various timbre perception features of different factors. If the correlation coefficient is greater than 0.9, it indicates that this factor has no significant influence on the timbre perception features and emotional relationship. If the correlation coefficient is less than 0.9, it indicates that this factor has a certain influence on the timbre perception features and emotional relationship.

TABLE III. CORRELATION BETWEEN TIMBRE PERCEPTION FEATURES AND THREE ELEMENTS

Timbre Perception Features	**Three Kinds of Elements**			
	Playing content	Time domain properties of scales	Time domain properties of melody	Musical instrument types
Bright-Dark	0.85	0.98	0.97	0.98
Raspy-Mellow	0.43	0.92	0.97	0.78
Sharp-Vigorous	0.93	0.91	0.99	0.93
Coarse-Pure	0.77	0.99	0.96	0.87
Hoarse-Consonant	0.70	0.95	0.98	0.87
The average	0.73	0.95	0.97	0.89

A. The Influence of Playing Content on Timbre Perception Features and Emotional Correlation

The material is classified according to the playing content, and the line diagram of the correlation coefficient between timbre perception features and emotion is drawn. Figure 3 is the line diagram.

The following conclusions can be drawn from the figure above:

It can be seen from Table □ that the correlation between scales and melody of sharp-vigorous is as high as 0.93, which is greater than 0.9, indicating that the playing content has very little influence on the correlation of sharp-vigorous and emotional.

Whether melody material or scale material, bright-dark and coarse-pure are very high to the correlation of sad and happy, and have the opposite effect, that is a kind of tone color more bright, purer sounded happier, more gloomy, rougher, sounds more sadder. This is consistent with the emotional perception of music.

In addition to the emotional dimension of holy, the correlation coefficients between the bright-dark and coarse-purity of the scale and his four emotions are distributed at the top and the bottom respectively, indicating that the bright-dark and coarse-purity of the scale are the two timbre perception features that have the greatest impact on emotion and opposite effects.

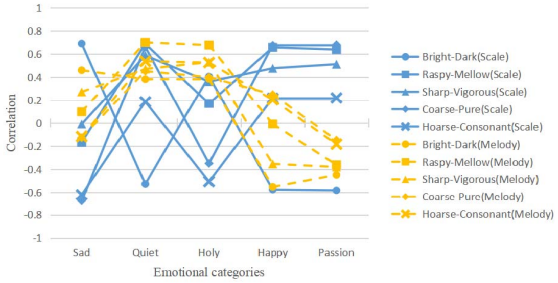

Figure 3. The influence of playing content on timbre perception features and emotional correlation.

B. The Influence of Time Domain Characteristics on Timbre Perception Features and Emotional Correlation

Depending on how the instrument sounds, musical scale material and melodic material can be divided into two categories: continuous instruments and non-continuous instruments. Using the same method in the previous section, timbre perception features and emotional correlation coefficients of continuous and non-continuous instruments were obtained respectively. Figure 4 and Figure 5 are scale and melody line charts respectively.

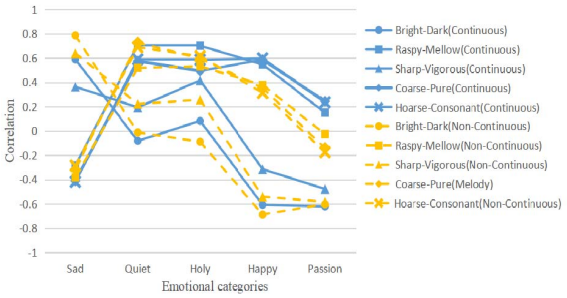

Figure 4. The influence of temporal characteristics on timbre perception features and emotional correlation of scales.

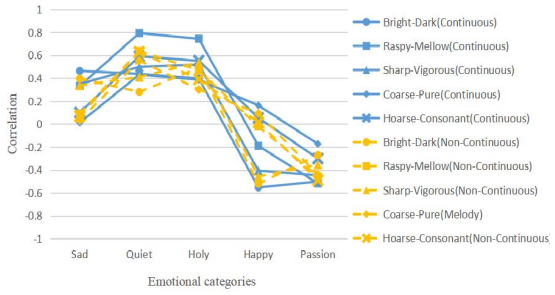

Figure 5. The influence of temporal characteristics on timbre perception features and emotional correlation of melody.

The following conclusions can be drawn from the two figures:

The trends of lines in the two figures are relatively consistent, and the distribution of timbre perception features of continuous and non-continuous instruments for the same emotion is also consistent. For example, in Figure 4, both persistent and non-persistent Musical Instruments are positively correlated with sad in the sense of bright-dark and sharp-vigorous. The remaining three-dimensional timbre perception features were negatively correlated with sad. The largest positive correlation was bright-dark, while the largest negative correlation was coarse-pure. In Table □, the correlation coefficient of timbre perception features corresponding to temporal scales and melodic features is higher than 0.9. Therefore, it can be considered that temporal characteristics have very little influence on timbre perception features and emotional correlation.

The broken lines of coarse-pure and hoarse-consonant in the two figures tend to coincide, so these two timbre perception features have the same impact on emotion.

C. The Influence of Chinese and Western Musical Instruments on Timbre Perception Features and Emotional Correlation

This part mainly analyzes the different influence of timbre of Chinese and Western Musical Instruments on emotion. Due to cultural differences, in addition to different timbre, the different composition structure of melodies in Chinese and Western music also has a greater impact on emotions. Therefore, this part excludes the influence of this factor on emotions, and only studies scale materials without melodies. In the same way, timbre perception features and emotional correlation coefficients of Chinese and Western Musical Instruments were calculated. Figure 6 is a line diagram of the emotional influence of Chinese and Western Musical Instruments.

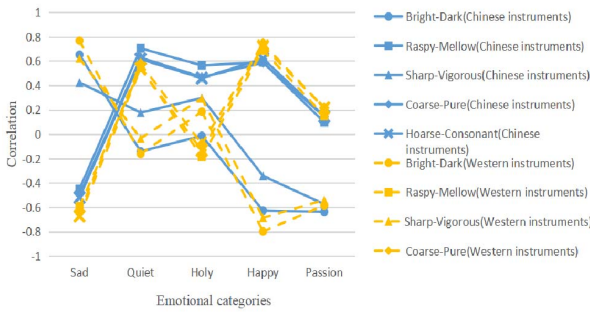

Figure 6. The influence of Chinese and Western Musical Instruments on timbre perception features and Emotional correlation of scales.

The following conclusions can be drawn from figure 6:

It can be seen from Table □ that the correlation of bright-dark and sharp-vigorous features of different musical instrument types are higher than 0.9, indicating that musical instrument types have very little influence on the emotional correlation between the two features.

Whether it is a Chinese instrument or a Western instrument, sad and happy are highly correlated with the five perceptual features. There was a strong positive correlation between quiet and three perceptual features: raspy-mellow, coarse-pure and hoarse-consonant. Passion has a high negative correlation with bright-dark and sharp-vigorous.

IV. THE CONSTRUCTION OF TIMBRE PERCEPTION FEATURES AND EMOTIONAL ASSOCIATION MODEL

In this part, multiple linear regression was used to construct the emotion prediction model. The stepwise regression method was adopted in multiple linear regression, and timbre perception features with collinear problems were eliminated. The input of the model was timbre perception features and emotion, and the output was the expression of emotion. The correlation coefficient between each timbre perception feature and emotion is calculated. When the correlation coefficient is less than 0.2, it can be considered that the timbre perception feature and emotion are not correlated, and the timbre perception feature will not be used as the input of the emotion prediction model. Equation (1-10) are the output of the model:

$$Y_{S,SAD} = 4.15 + 0.35X_{B-D} - 0.24X_{H-C} \tag{1}$$

$$Y_{S,QUI} = 2.65 + 0.58X_{R-M} \tag{2}$$

$$Y_{S,HOL} = 1.57 + 0.46X_{R-M} + 0.25X_{S-V} \tag{3}$$

$$Y_{S,HAP} = 5.71 \quad 0.41X_{B\ D} \quad 0.72X_{C\ P} + 0.95X_{H\ C} \tag{4}$$

$$Y_{S,PAS} = 6.05 - 0.35X_{B-D} \tag{5}$$

$$Y_{M,SAD} = 3.46 + 0.78X_{B-D} - 0.44X_{S-V} \tag{6}$$

$$Y_{M,QUI} = -2.25 + 0.60X_{B-D} + 0.72X_{R-M} + 0.94X_{C-P} - 0.67X_{H-C} \tag{7}$$

$$Y_{M,HOL} = 0.37X_{B,D} + 0.41X_{R-M} + 0.24X_{C-P} \tag{8}$$

$$Y_{M,HAP} = 7.28 - 0.94X_{B-D} + 0.51X_{S-V} - 1.45X_{S-P} \tag{9}$$

$$Y_{M,PAS} = 8.04 - 0.88X_{B-D} - 0.53X_{R-M} + 0.56X_{S-V} \tag{10}$$

$Y_{S,SAD}, Y_{S,QUI}, Y_{S,HOL}, Y_{S,HAP}, Y_{S,PAS}$ represents the five dimensions of sad, quiet, holy, happy and passion of the scale; $Y_{M,SAD}, Y_{M,QUI}, Y_{M,HOL}, Y_{M,HAP}, Y_{M,PAS}$ represents the five dimensions of sad, quiet, holy, happy and passion of the melody; $X_{B-D}, X_{R-M}, X_{S-V}, X_{C-P}, X_{H-C}$ represents the five timbre perception features of bright-dark, dry-soft, sharp-rich, rough-pure and hoarseness-harmony respectively.

Correlation coefficient R and Mean Absolute Error (MAE) were used to evaluate the accuracy of the above

models. The larger R was, the better the fitting of the model was and the smaller the MAE value was, the better the prediction accuracy of the model was.

Figure 7 shows the correlation coefficients of the five emotional perception models of scales and melodies. It can be seen that the correlation coefficients of the five emotional models of scales and melody are all above 0.6, among which the correlation coefficient of happy of scales is as high as 0.82, and the quiet and holy of melody are both higher than 0.8, indicating that this model can well predict the five emotions. The average absolute error of the five emotion models is below 1.1, which further indicates that the model has a good predictive ability for the five emotions.

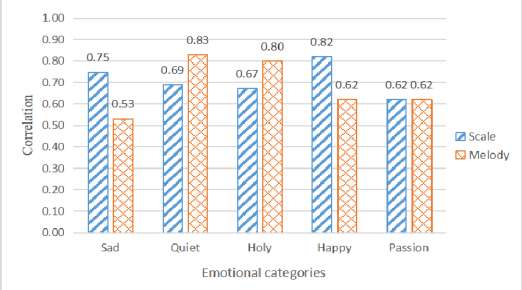

Figure 7. Correlation coefficient of model evaluation.

V. CONCLUSIONS

In this paper, the correlation and modeling of timbre perception features and emotion were studied. First, a 5-dimensional emotional space was determined by the MDS method, and an emotional evaluation experiment was conducted according to the 5-dimensional emotional space. The reliability, validity test and noise elimination were conducted on the experimental data. The influence of playing content, time-domain characteristics and musical instrument type on timbre perception features and emotional relationship is studied. The results show that temporal characteristics have little influence on the timbre and emotional relationship, and playing content and musical instrument type have a certain influence on the timbre and emotional relationship. Then the multivariate linear regression algorithm is used to distinguish 5 emotion models of scale and melody. The results show that the model has a good predictive ability for all the 5 emotions.

In future studies, the following points can be highlighted: Firstly, to explore the influence of musical instrument types on emotion, melodic materials of western Musical Instruments can be added based on the existing timbre database to improve the universality of the model. Second, the model can be compared to other machine learning methods, and then refined to improve its predictive power. Finally, the model of perceptual features and emotion can be compared with the model of underlying physical features and emotion, so as to provide a theoretical and experimental basis for solving the semantic gap between underlying objective features of timbre and the existence of subjective emotional space.

ACKNOWLEDGMENT

We are grateful to China Conservatory of Music for providing part of the scale material of Chinese instruments for this experiment.

This work was supported by project of Key Laboratory of Acoustic Visual Technology and Intelligent Control System, Ministry of Culture and Tourism (WLB - ZDSYS - 2019 - 06), the Fundamental Research Funds for the Central Universities (CUC19ZD005 - Intelligent Matching Theory and Typical Application of Music and Scene Content Based on Acoustic and Visual Perception) and National Endowment for the Arts (01020120180529564031 - Digital Museum of Chinese Musical Instrument Construction).

REFERENCES

[1] J. C. Hailstone, R. Omar, S. M. Henley, C. Frost, M. G. Kenward, and J. D. Warren, "It's Not What You Play, It's How You Play It: Timbre Affects Perception of Emotion in Music," The Quart. J.Experimental Psychology, vol. 62, no. 11, pp. 2141–2155 (2009)

[2] K. R. Scherer and J. S. Oshinsky, "Cue Utilization in Emotion Attribution from Auditory Stimuli," Motivation and Emotion, vol. 1, no. 4, pp. 331–346, 1977.

[3] I. Peretz, L. Gagnon, and B. Bouchard, "Music and Emotion: Perceptual Determinants, Immediacy, and Isolation after Brain Damage," Cognition, vol. 68, no. 2, pp. 111–141, 1998.

[4] E Bigand, S Vieillard, F Madurell, J Marozeau, and A Dacquet, "Multidimensional Scaling of Emotional Responses to Music: The Effect of Musical Expertise and of the Duration of the Excerpts," Cognition and Emotion, vol. 19, no. 8, pp. 1113–1139, 2005

[5] Wun S , Horner A , Bin W U . Effect of Spectral Centroid Manipulation on Discrimination and Identification of Instrument Timbres[J]. Journal of the Audio Engineering Society, 2014, 62(9):575-583.

[6] The Emotional Characteristics and Timbre of Nonsustaining Instrument Sounds Horner, Andrew, Wu,et al. The Emotional Characteristics and Timbre of Nonsustaining Instrument Sounds[J]. Journal of the Audio Engineering Society Audio Acoustics Applications, 2015.

[7] W.Jiang,J.Liu,X.Zhang,et al.Analysis and Modeling of Timbre Perception Features in Musical Sounds[J]. Applied Science, 2020,10(3):789.

[8] Z.meng, L,dai. "The Perception and Design of Reverberation," BeiJing: China Building Industry Press, 2018.

[9] W.jiang, J. Liu, Z. Li, J. Zhu, X. Zhang and S. Wang, "Analysis and Modeling of Timbre Perception Features of Chinese Musical Instruments," 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), Beijing, China, 2019, pp. 191-195.