Received: 30 June 2021 | Accepted: 6 August 2021

DOI: 10.1049/ccs2.12032

Cognitive Computation and Systems

ORIGINAL RESEARCH PAPER

Cross-cultural analysis of the correlation between musical elements and emotion

Xin Wang | **Yujia Wei** | **Dasheng Yang**

School of Music and Recording Art, Communication University of China, Beijing, China

Correspondence

Xin Wang, School of Music and Recording Art, Communication University of China, Beijing, China.
Email: metero_wx@cuc.edu.cn

Funding information

Humanity and Social Science Youth foundation of Ministry of Education of China, Grant/Award Number: 15YJCZH178; Science Foundation of Communication University of China, Grant/Award Number: HG1608-1; Major Project of Beijing Social Science Foundation, Grant/Award Number: 20ZD20

Abstract

In a cross-cultural context, exploring musical elements' cultural specificity and universality that affect various types of music is conducive to personalised emotion recognition. In this study, high-level musical elements are introduced to explore their influence on emotional perception. By comparing music emotion recognition (MER) models of varied cultural music, musical elements with cultural universality and cultural specificity are further determined. Participants rated valence, tension arousal, and energy arousal on labelled nine-point analogical–categorical scales for four types of classical music: Chinese ensemble, Chinese solo, Western ensemble, and Western solo. Fifteen musical elements in five categories—timbre, rhythm, articulation, dynamics, and register were annotated through manual evaluation or the automatic algorithm. The relationship between music emotion and musical elements was analysed through partial least squares regression. Results showed that tempo, rhythm complexity, and articulation are culturally universal; musical elements related to timbre, register, and dynamics features are culturally specific. By increasing tempo, rhythm complexity, staccato, perception of valence, tension arousal, and energy arousal can be effectively improved. Based on the Partial least squares regression (PLSR) model's results for the datasets, the combination of manual and automatic annotation for musical elements can improve the MER system's performance.

KEYWORDS

cross-culture, emotion perception, musical elements, music emotion recognition

1 | INTRODUCTION

Recently, music has entered the streaming era, and the media are presenting many musical varieties from all over the world. With such diversification, music emotion recognition (MER) and recommendation systems have emerged. MER models can be used in music information retrieval, music recommendation, and music therapy [1]. According to major previous studies, musical emotion perception relates closely to musical elements [2–4]. For example, a smooth rhythm can bring us happiness, and a complex rhythm can make us uneasy, as shown by Gabrielsson and Lindström, who summarised the relationship between musical elements and emotional expression [5].

In MER research, many studies have focussed on the association between low-level features and emotional perception. Juslin explored the correlation between emotional expressions (happiness, sadness, fear, and anger) and low-level features (tempo, sound level, spectrum, and articulation); results showed that the relative importance of each acoustic feature differed for different emotional expressions [6]. Lu, Liu, and Zhang presented an automatic emotion detection system for Western classical music from three acoustic feature sets—intensity, timbre, and rhythm [7]. Yang, Dong, and Li summarised acoustic features related to MER, the centroid of loudness, low energy rate, mel-frequency cepstral coefficients, spectrum shape, spectral contrast, rhythm strength, rhythm regularity, and so on [8]. Despite the many related studies, improving the MER system's performance is difficult because of the huge difference between audio signals' low-level features and people's understanding of music semantics. In fact, in research published at the International Conference of Acoustics, Speech, and Signal Processing (ICASSP), existing MER models' recognition accuracy is lower than 70% [9, 10].

To improve MER's performance, researchers have begun to introduce into MER high-level features related to music semantics and musical elements. Leman et al. explored whether musical affect attribution can be predicted by a linear combination of high-level features. From selective listening, musicological experts rated as high-level features manual structural cues and acoustic structural cues obtained from feature extraction algorithms. Results showed that manual structural cues work better than acoustical structural cues [11]. Yongli et al. also utilised manual annotation for high-level musical elements in MER such as chords, metre, texture, tempo, mode, instrument, and melody [12]. Compared with low-level features, MER performance improved from 59.6% to 69.8%. Until now, semantic-based musical elements in MER have remained mostly unused. In current research, which musical elements should be labelled, how to label them, and how they relate to emotional perception are still difficult problems.

In the cross-cultural context, both perceptual and cultural cues influence musical emotion. In fact, the more cultural-specific and perceptual cues the music presents, the stronger the music's expression of emotion [13]. However, culturally varied music has both universality and diversity in its features' influence. By classifying the emotion of Chinese and Western classical music, Wu and Xie found that feature sets of rhythm, pitch, and timbre had cross-cultural attributes [14]. By selecting a set of features related to pitch, rhythm, and timbre, Zhao et al. compared MER models of Western and Chinese classical music based on a Bayesian network classifier [15]. Their research indicated that Chinese traditional music's detection rate was lower than that of Western classical music. Hu and Yang focussed on a cross-dataset's generalisability to reveal that the arousal regression model had better cross-cultural adaptability [16, 17]. After asking Chinese and Western participants to measure discrete and dimensional emotions in Chinese and Western classical music, Cowen et al. found that 13 dimensionally discrete emotions had cross-cultural attributes [18]. In cross-cultural MER, introducing high-level features and further analysing high-level features' universality and specificity for culturally varied music can effectively improve the MER model's performance and personalisation.

Few studies have introduced high-level features into the cross-cultural MER model [19], although there are some good studies to discuss the similarities and differences between cross-cultural [20] and high-level music features [21–23]. In this study of Chinese and Western classical music, we introduce high-level musical elements to explore their influence on emotional perception. By comparing MER models of varied cultural music, we further determine musical elements with cultural universality and cultural specificity, respectively. We chose Chinese and Western classical music as the research object because these two types have great differences in tonality, harmony, and timbre, thus easing investigation of musical elements' cultural universality and specificity that affect emotional perception in a cross-cultural context.

The rest of this study is organised as follows: We first describe the ground-truth dataset and subjective emotion perception experiment and introduce musical elements' manual and automatic annotation methods. Partial least squares regression (PLSR) is adopted to obtain MER models for different music types, followed by detailed discussion of the results. Finally, we summarise the study's findings and envision future research.

2 | DATASET AND METHODS

2.1 | Dataset

The experiment adopted three criteria for selecting musical excerpts: 1) Each should be non-semantic, which means there is no text content such as lyrics; 2) each should convey a certain strong emotion; and 3) all of them should be evenly distributed among the four emotional quadrants of the valence-arousal model plane. To avoid emotion perception difference caused by the differing number of musical instruments between solo and ensemble, the cross-cultural dataset contains four types of classical music, namely Chinese solo and ensemble and Western solo and ensemble. From the dataset constructed by [24], Chinese classical music was selected; the dataset consists of 500 musical excerpts from famous Chinese classical music albums, national instrumental music compilation discs, and online collected folk music albums, including instrumental solo and ensemble music of bowed chordophones, plucked chordophones, aerophones, and membranophones. The Western classical music excerpt was gathered from the MediaEval Database for Emotional Analysis in Music (DEAM) dataset [25], the Emotify dataset [26], the AMG1608 dataset [27], and the Free Music Archive (FMA) [28, 29]. Detailed music tracks' information can be seen in Supplementary Material.

Based on the integrity of music structure and considering Bachorik's research that participants could make accurate emotional judgments for music excerpts with an average duration of 8.31 s [30], all Chinese and Western music were manually segmented into excerpts with a duration of 10–15 s. The cross-cultural dataset included 515 music excerpts (Table 1). All stimuli were first calibrated based on a loudness measurement algorithm [31] and then finely adjusted by the ear by three volunteers.

2.2 | Listening test for emotion ratings

To facilitate description and calculation of emotion, researchers have constructed an emotion perception model that includes categorical and dimensional models [32–34]. Russell proposed a

TABLE 1 Number of stimuli in each of the four datasets

Datasets	Chinese classical ensemble	Chinese classical solo	Western classical ensemble	Western classical solo	All
Number of stimuli	143	101	138	133	515

two-dimensional emotion model composed of valence and arousal [35]. Based on this model, Schimmack and Grob divided arousal into tension arousal and energy arousal and constructed a new three-dimensional emotional model used to improve the arousal's perceptual ambiguity and reduce loss of emotional information accuracy in the two-dimensional model [36]. Therefore, this study adopted the three-dimensional model of affect.

Our previous research has shown that a participant's musical background has little effect on emotional perception [37], so 30 participants (aged 21–23 years, 15 female) not required to have a professional music education background (music study years: $M = 2.4$, $SD = 3.9$) were recruited for this listening test. All were born and live in China, and their ethnicity is Han Chinese.

Due to the COVID-19 pandemic's impact, the subjective evaluation experiment was completed online. Stimuli were stored on Baidu cloud, and participants could open and listen to the stimuli independently. We randomly and evenly divided all stimuli into five groups; participants listened to one group each time and completed all experiments within a week. During the listening tests, participants could adjust the volume by themselves in a quiet room and wear a superior-quality headphone. They reported ratings of valence, tension arousal, and energy arousal on labelled nine-point analogical–categorical scales. They were instructed to focus on emotions conveyed by music, not on what they were feeling because it is easier to agree on the emotion conveyed by music than on the emotion evoked in listeners [38]. The end points of each dimension are described in Table 2. After the first listening test, participants completed the biographical questionnaire.

2.3 | Musical elements' annotation

To investigate the relationships among musical structures' various features and emotions that listeners perceived, Wedin used a five-point scale to rate music's structural properties (e.g. intensity, pitch, rhythm, tempo, rhythmic articulation, harmony, tonality, modality, and melody) [39]. Juslin and Lindstrm also focussed on interactions among eight musical features (pitch, mode, melodic progression, rhythm, tempo, sound level, articulation, and timbre) [40]. Using six primary musical cues (mode, tempo, dynamics, articulation, timbre, and register), Eerola, Friberg and Bresin asked listeners to rate 200 musical examples according to four perceived emotional characteristics (happy, sad, peaceful, and scary) [41]. Among the factors affecting emotional expression in music, tempo is considered important and is also the most widely studied [5, 42, 43]. By summarising a large amount of relevant literature and interviews with music experts, we divided musical elements into five categories, each containing varied sub-elements (Table 3).

In our research, we used two methods to annotate musical elements: automatic annotating by algorithms and manual annotating based on subjective evaluation. In music information retrieval (MIR), musical elements well-fitted by algorithms with subjective perception were automatically labelled by algorithms. Musical elements still difficult to fit with algorithms were manually annotated by music experts.

2.3.1 | Automatic algorithm annotation

In this section of the study, ppmBatch and MIRToolbox were adopted for musical element extraction and processing [44]. Table 4 displays detailed annotation methods for each element. According to various loudness standards, ppmBatch[*] specialises in normalising audio files (mono- and multi-channel). In addition, ppmBatch enables application of batch processing to multiple audio files for efficiency [45]. Power spectrum estimation was applied to calculate the spectrum in MIRToolbox, which used the short-time Fourier transform length of 8192 sample points, with a Hann-windowed analysis of 50 ms and an overlap of 50% between successive frames [46]. The final calculated spectrum was the linear magnitude spectrum.

2.3.2 | Manual annotation

As Table 5 shows, musical elements were subjectively evaluated and manually annotated by 54 musical experts (aged 19–25 years, 38 female). All of them had more than 10 years of professional musical training (music study years: $M = 13.7$, $SD = 4.0$). Figure 1 displays the web-based subjective evaluation interface for musical elements' annotation, including five pages of musical element categories. All stimuli were randomly divided into 18 groups, and the musical experts annotated only one group at a time. They rated each musical element on a continuous Likert scale from 1 to 9. After completing all the annotations, they completed the biographical questionnaire. At least three experts annotated each excerpt.

3 | RESULTS

3.1 | Validity analysis of experimental results

To check the validity of emotion perception, we conducted Cronbach coefficient alpha reliability tests, which indicated that all the scales had good internal consistency (for 30 participants: 0.93 for valence, 0.88 for tension arousal, and 0.92 for energy arousal).

TABLE 2 Evaluation terms for the three-dimensional emotion model

Emotional perception dimension	Left end point (1 point) description	Right end point (9 point) description
Valence	Displeasure	Pleasure
Tension arousal	Relaxation	Tension
Energy arousal	Tired	Awake

[*]The link to ppmBatch: https://products.zplane.de/products/ppmbatch/.

TABLE 3 Musical elements and sub-elements

Musical elements	Musical sub-elements	Description
Timbre	Clarity	The sound is clear, with details and clear layers
	Pleasantness	The sound is euphonious
	Spaciousness	Spatial sense
	Brightness	The sound is bright
	Thickness	The sound is thick
	Richness	The timbre is full
Rhythm	Tempo	Perceived tempo
	Pulse clarity	Rhythmic or metrical pulsation
	Rhythm complexity	Degree of alternation of different rhythm patterns
Articulation	Articulation	Played with staccato or legato
Dynamics	Loudness	Loudness
	Loudness range	Loudness variation
	Dynamics	Dynamics range
Register	Pitch range	Range of register
	Average pitch	Average pitch

TABLE 4 Method of automatic algorithm annotation

Musical elements	Definition	Tools
Brightness	Spectral brightness is used to calculate the ratio of spectrum energy above cut-off frequency f to total energy (cut-off frequency $f = 1.5$ KHz in this experiment) to represent the audio's brightness. The higher the energy ratio of the high-frequency spectrum, the greater the music's brightness.	MIRToolbox
Thickness	Spectral brightness is used to calculate the ratio of spectrum energy below cut-off frequency f to total energy (cut-off frequency $f = 300$ Hz in this experiment) to represent the audio's thickness. The greater the proportion of the low-frequency spectrum energy, the greater the music's darkness.	MIRToolbox
Pitch range	To obtain all segments' single tones, mirpitch was used. The difference between the highest and lowest pitches is the pitch range, expressed by the cent scale.	MIRToolbox
Average pitch	To obtain all segments' single tones and to take all single tones' average value, we calculated the average pitch by using mirpitch. This was expressed by the cent scale.	MIRToolbox
Loudness	We used the ITU-R bs.1770-3 loudness model to calculate loudness.	ppmBatch
Loudness range	We used the ITU-R bs.1770-3 loudness model to calculate loudness. The difference between the greatest and least loudness is the loudness range.	ppmBatch

TABLE 5 Manual annotated musical elements

Musical elements	Musical sub-elements	Left end point (1 point) description	Right end point (9 point) description
Timbre	Clarity	Indistinct	Clear
	Richness	Thin	Full
	Pleasantness	Unpleasant	Pleasant
	Spaciousness	Not spacious	Spacious
Rhythm	Tempo	Slow	Fast
	Pulse clarity	Weak pulse	Strong pulse
	Rhythm complexity	Simple	Complex
Articulation	Articulation	Staccato	Legato
Dynamics	Dynamics	Small	Large

To test the consistency of the manual annotation's experimental results, we adopted the method of Zacharopoulou and Kyriakidou [47]. When the mean standard deviation between experts' ratings of each musical element was fewer than two scale points, experts' results were considered to have achieved good consensus. Based on this consistent calculation result, stimuli with poor consistency were eliminated; finally, 146 stimuli were retained (Table 6).

3.2 | Correlation analysis of musical elements and emotion perception

PLSR combines principal component analysis and multiple linear regression and also allows collinearity among variables [48], so PLSR was used to examine the relationship between musical elements and perceived emotion ratings. A six-fold cross-validation model was applied to the PLSR model, and R^2 and Q^2 were used to evaluate the PLSR model's performance. R^2 evaluates the model's explanatory power and Q^2 its predictive power. The variable importance in projection (VIP) score was applied to assess each independent variable's importance, and a VIP score greater than one was generally considered a significant contribution [49]. The SIMPLS algorithm was applied to the PLSR model and implemented in MATLAB [50]. To investigate musical elements' cultural universality and specificity, we first conducted a PLSR analysis on the entire cross-cultural dataset and then on each of the four datasets independently.

In this study, the PLSR independent variables were 15 musical elements for each of 146 stimuli, and the three dependent variables were the mean ratings of participants for valence, tension arousal, and energy arousal. Two principal components (PCs) were considered in the PLSR models of all three emotion dimensions based on the Q^2 criterion computed by cross-validation [51], meaning that PC was significant and selected when predicted variance Q^2 was greater than 0.05. PLSR results for the cross-cultural dataset are listed in Table 7.

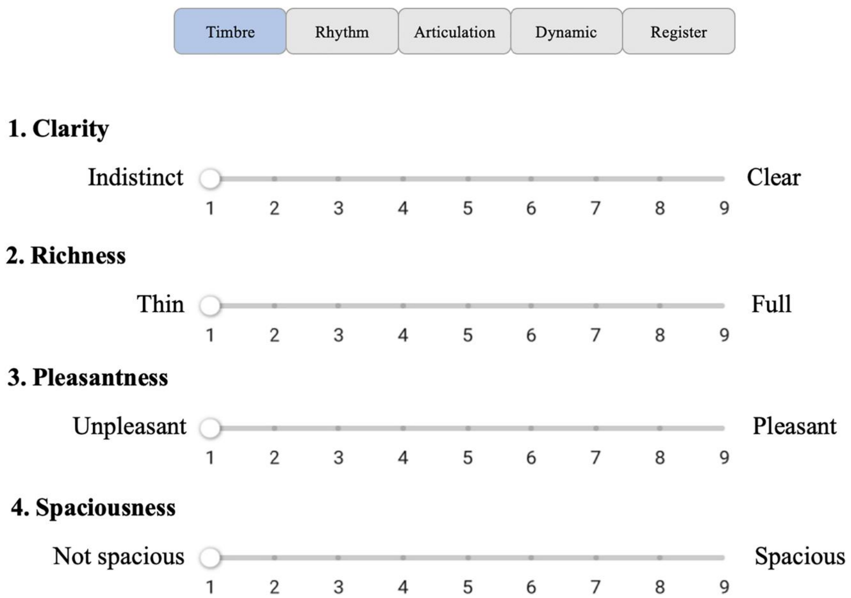


FIGURE 1 Listening test interface

TABLE 6 Number of stimuli after reliability test

Datasets	Chinese classical ensemble	Chinese classical solo	Western classical ensemble	Western classical solo	All
Number of stimuli	28	40	43	35	146

TABLE 7 PLSR performances of the cross-culture dataset on valence, tension arousal, and energy arousal

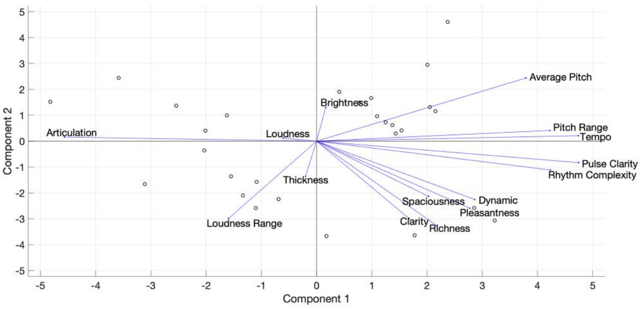
	R^2				Q^2				PC1				PC2			
Dependent variable	CE	CS	WE	WS	CE	CS	WE	WS	CE	CS	WE	WS	CE	CS	WE	WS
Valence	0.85	0.64	0.62	0.80	0.73	0.30	0.40	0.64	0.36	0.22	0.23	0.25	0.14	0.16	0.17	0.14
Tension	0.77	0.54	0.60	0.77	0.62	0.17	0.44	0.64	0.35	0.21	0.28	0.27	0.21	0.17	0.14	0.14
Energy	0.91	0.76	0.81	0.90	0.87	0.50	0.70	0.87	0.37	0.21	0.29	0.27	0.14	0.12	0.14	0.16

Abbreviations: CE, Chinese classical ensemble music dataset; CS, Chinese classical solo music dataset; WE, Western classical ensemble music dataset; WS, Western classical solo music dataset.

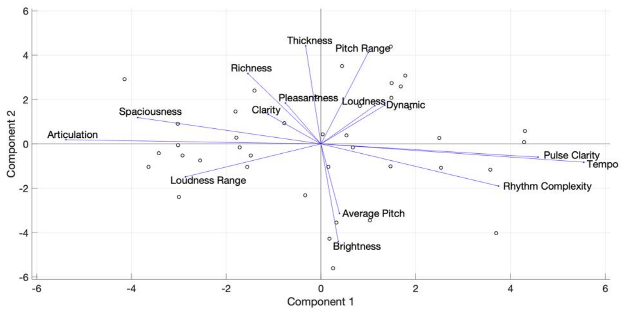
3.2.1 | Valence

Figure 2 shows PLSR loadings (vectors) and scores (circles) for valence across two PCs for different datasets. Longer vectors indicate that musical element loadings contributed more strongly, and orientations indicate which PCs they primarily influenced. PC1 for different datasets was similar and mainly related to two factors: rhythmic features, described by tempo and pulse clarity and the articulation feature. The music excerpt with positive coordinates in PC1 is performed with staccato and also has a fast tempo and definite rhythm. For different datasets, musical elements related to PC2 greatly differed. For Chinese classical ensemble music, PC2 appeared to be influenced by timbre features (richness and clarity) and the loudness range falling slightly oblique to the PC axis. For Chinese classical solo music, PC2 was highly related to timbre features, such as thickness and brightness. PC2 was influenced by dynamics and timbre features for Western classical ensemble music, such as dynamics, loudness, thickness, and brightness. For Western classical solo music, PC2 was related to spaciousness and dynamics.

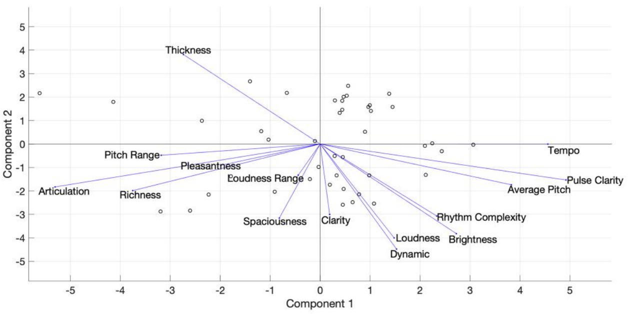
Table 8 displays musical elements with VIP scores greater than one, with each column ranked from high to low according to the VIP scores. The underlined musical elements negatively correlated with perceived valence. Results show articulation as the most important element for all datasets; pulse clarity and tempo were also influenced by valence perception for all datasets. The timbre feature, such as spaciousness, was an important element for classical solo music, whereas the pitch feature was important for classical ensemble music.



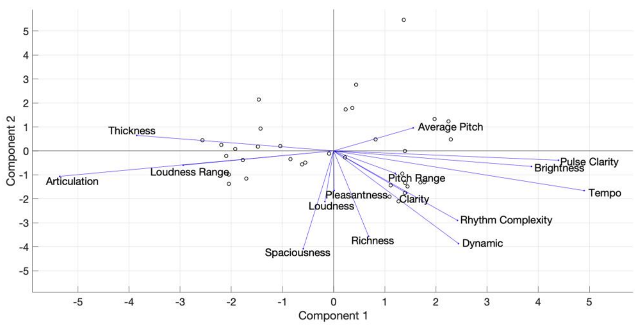
(a) Chinese classical ensemble music dataset.



(b) Chinese classical solo music dataset.



(c) Western classical ensemble music dataset.



(d) Western classical solo music dataset.

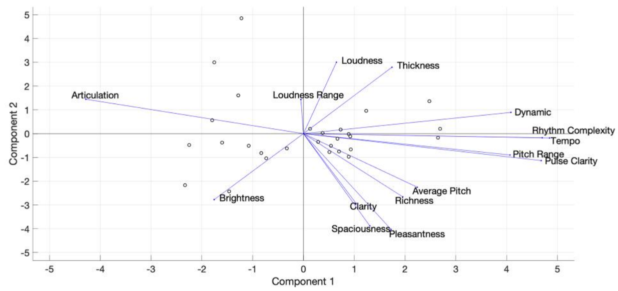
FIGURE 2 Loadings and scores across two PCs of PLSR for valence

3.2.2 | Tension arousal

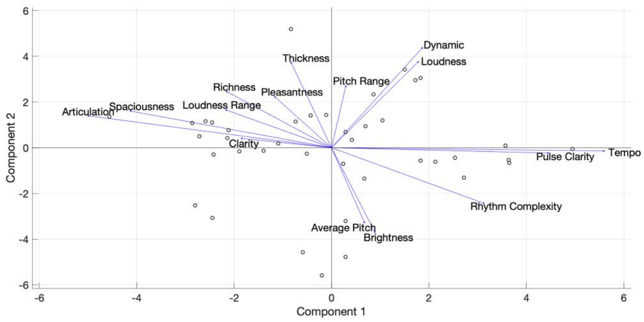
Figure 3 illustrates PLSR loadings and scores for tension arousal across two PCs. For all datasets, PC1 was highly related to rhythmic and articulation features, such as tempo, pulse clarity, and articulation. With the exception of Chinese classical solo music, PC1 was also related to dynamics. Although for all datasets, PC2 was related to timbre features, for each dataset, musical elements related to PC2 differ. For Chinese classical ensemble and Western classical solo music, pleasantness was highly related to PC2, whereas clarity was important for Western classical ensemble and solo music.

TABLE 8 Important musical elements on valence for the four datasets

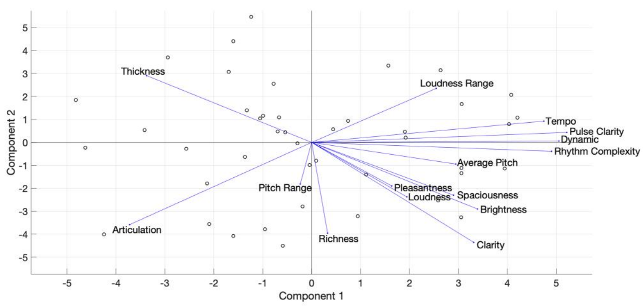
Chinese classical ensemble music	Chinese classical solo music	Western classical ensemble music	Western classical solo music
Articulation	Articulation	Articulation	Articulation
Average pitch	Tempo	Tempo	Pulse clarity
Tempo	Loudness range	Pulse clarity	Tempo
Pitch range	Pulse clarity	Average pitch	Rhythm complexity
Pulse clarity	Spaciousness		Brightness
Rhythm complexity	Rhythm complexity		Thickness Spaciousness



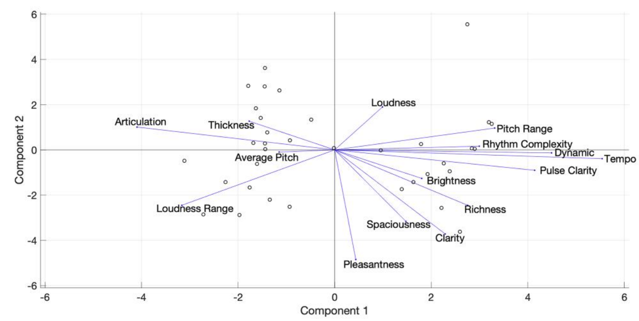
(a) Chinese classical ensemble music dataset.



(b) Chinese classical solo music dataset.



(c) Western classical ensemble music dataset.



(d) Western classical solo music dataset.

FIGURE 3 Loadings and scores across two PCs of PLSR for tension arousal

For Chinese classical solo music, PC2 was also influenced by dynamics.

Table 9's lists of important musical elements for tension arousal, tempo, articulation, pulse clarity, and dynamics are important for all datasets. Musical stimuli with positive tension arousal have a fast tempo, definite rhythm, large dynamics, and performance with staccato. Rhythm complexity is important for classical ensemble music, indicating that rhythmic variations can arouse people's tension

TABLE 9 Important musical elements in tension arousal for the four datasets

Chinese classical ensemble music	Chinese classical solo music	Western classical ensemble music	Western classical solo music
Tempo	Tempo	Dynamics	Tempo
Rhythm complexity	Articulation	Pulse clarity	Dynamics
Dynamics	Loudness	Articulation	Pleasantness
Pulse clarity	Pulse clarity	Rhythm complexity	Pulse clarity
Articulation	Spaciousness	Tempo	Articulation
Spaciousness	Dynamics	Loudness range	Loudness range
Pitch range			Dynamics

3.2.3 | Energy arousal

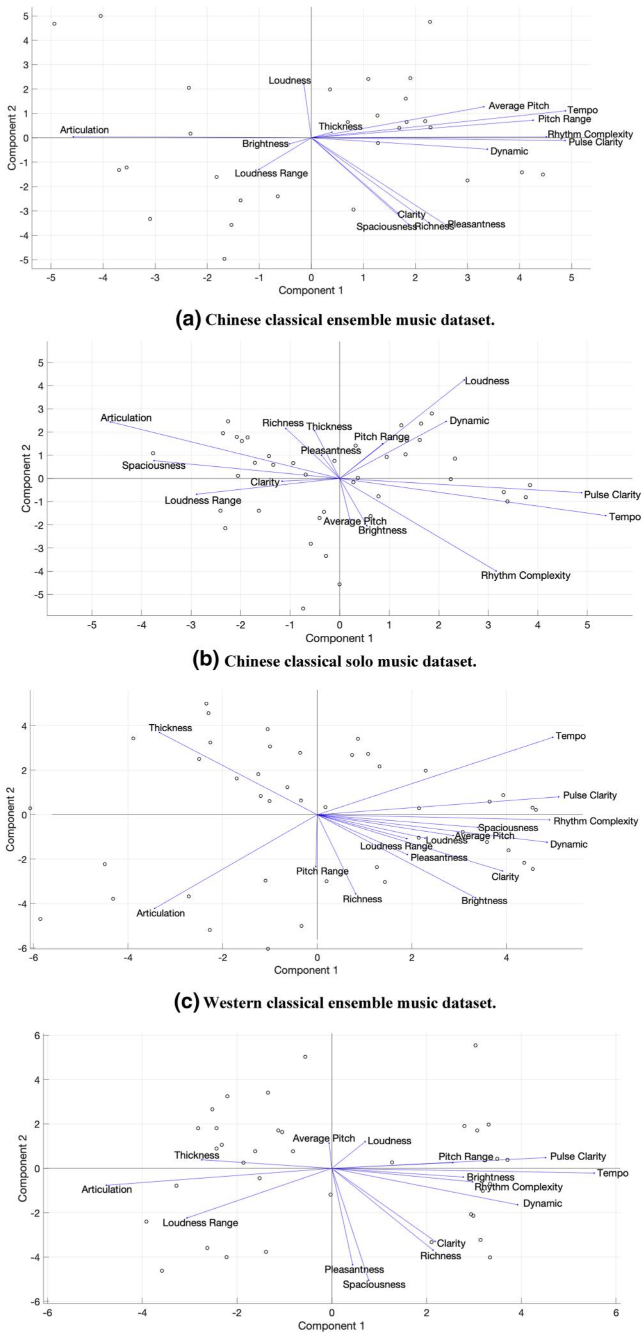
Figure 4 displays PLSR loadings and scores for energy arousal across two PCs for different datasets. For all datasets, PC1 was highly related to rhythmic features such as tempo and pulse clarity. For ensemble music datasets, PC1 was also influenced by rhythm complexity. Except in Western classical ensemble music, PC1 was highly related to articulation. For the four datasets, musical elements that related to PC2 were quite different. For Chinese classical ensemble and Western classical solo music, PC2 was related to timbre features, described by spaciousness, richness, clarity, and pleasantness. For Chinese classical solo music, PC2 was influenced by loudness and rhythm complexity falling slightly oblique to the PC axis. For Western classical ensemble music, PC2 was related to articulation and timbre features such as thickness, richness, and brightness.

Table 10 lists musical elements for energy arousal. Tempo, articulation, and pulse clarity were important for all datasets. For Western classical music, dynamics was also important. Except for Western classical ensemble music, rhythm complexity greatly affected on energy arousal perception in the datasets. For classical solo music, the loudness range negatively influenced energy arousal perception

4 | DISCUSSION

4.1 | Analysis of musical elements with cultural universality

Based on VIP scores greater than one, musical elements important in the three emotional dimensions for all datasets are shown in Figure 5, in which each element's VIP score



(a) Chinese classical ensemble music dataset.

(b) Chinese classical solo music dataset.

(c) Western classical ensemble music dataset.

(d) Western classical solo music dataset.

FIGURE 4 Loadings and scores across two PCs of PLSR for energy arousal

is the average of all datasets. Results reveal that tempo, pulse clarity, articulation, and dynamics greatly influenced emotional perception for all datasets; thus indicating these four elements' cultural universality. Among them, tempo, pulse clarity, and articulation significantly affected the three emotional dimensions, whereas the dynamics significantly impacted perception of tension arousal. Music stimuli with fast tempo, definite rhythm, and performance with staccato can usually convey pleasure and tension and awaken

TABLE 10 Important musical elements for energy arousal in the four datasets

Chinese classical ensemble music	Chinese classical solo music	Western classical ensemble music	Western classical solo music
Tempo	Loudness	Tempo	Tempo
Pulse clarity	Tempo	Pulse clarity	Pulse clarity
Articulation	Pulse clarity	Rhythm complexity	Articulation
Pitch range	Articulation	Dynamics	Dynamics
Rhythm complexity	Spaciousness	Articulation	Loudness range
Average pitch	Rhythm complexity	Spaciousness	
	Loudness range		

feelings. Music with a great dynamics range can also increase people's tension. Of the five musical elements' categories we have examined, the influence of rhythm, dynamics, and articulation on emotional perception for both Chinese and Western classical music had good consistency, corresponding with previous studies on the perception of emotion in music [13, 16, 36, 52, 53].

Further analysis, however, found that different musical elements had slightly different effects on the perception of the three emotional dimensions. For valence, the most important element was articulation, followed by tempo and pulse clarity. VIP scores among tempo, pulse clarity, and articulation for energy arousal revealed no great differences, although tempo scored the highest. Based on VIP scores, important musical elements for tension arousal, ranked from high to low, were tempo, articulation, pulse clarity, and dynamics.

4.2 | Analysis of musical elements with cultural specificity

For cultural specificity, important elements that influence emotional perception for the four datasets are displayed according to VIP scores greater than one (Figure 6). For Chinese classical ensemble music, register features played a positive role in the perception of the three emotional dimensions. Its higher pitch range significantly impacted all three emotional dimensions, indicating that increased pitch range can cause people's emotional perception to be more positive. In addition, higher pitch components can make Chinese classical ensemble music more active and pleasant. Compared with Western classical ensemble music that emphasises the fusion of vertical and horizontal harmony, Chinese classical ensemble music focusses on a horizontal melody structure, achieving a greater pitch variety thanks to the multiple and different instruments involved and paying more attention to the music personality's

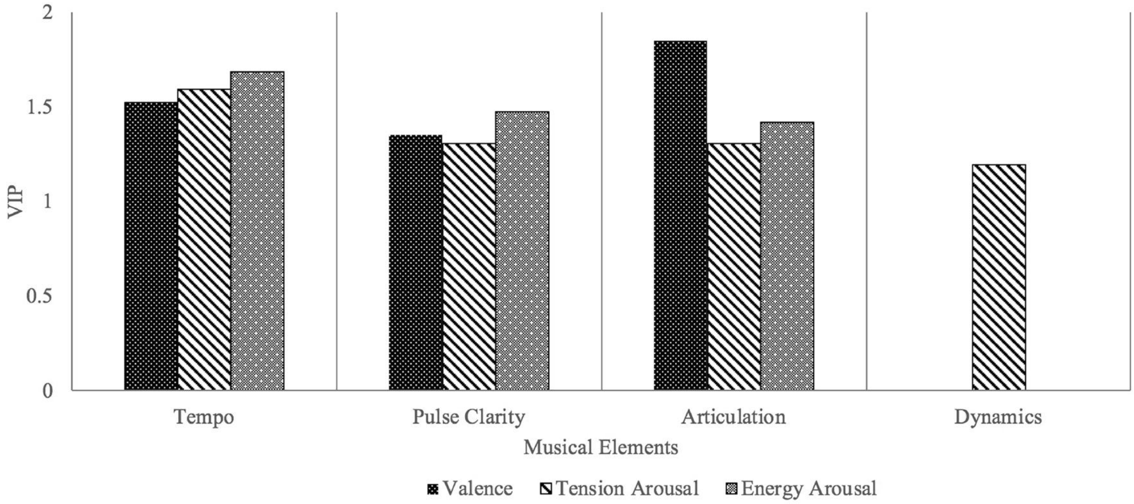


FIGURE 5 Important musical elements with cultural universality for three emotional dimensions

expression. Such aesthetic characteristics bring more pitch changes to the performance of Chinese music. Rhythm complexity also greatly influenced the perception of the three emotional dimensions.

For Chinese classical solo music, spaciousness was more prominent for emotional perception. There was negative correlation between spaciousness and the three dimensions of emotional perception, indicating that playing Chinese classical solo music in a small space can obtain higher emotional perception. Because Chinese musical instruments' pronunciation is based on imitating the human voice [54], Chinese classical solo music in a small space can reveal the instrument's details and timbre characteristics, thus adding more flavour to the overall performance. In addition, higher loudness significantly impacts the perception of tension and energy arousal, whereas the loudness range affects valence.

Western classical solo music showed great differences in musical elements that affect the three emotional dimensions. For valence, important elements included timbre and rhythm features, such as spaciousness, thickness, brightness, and rhythm complexity. The main elements that affected energy arousal were dynamics features such as loudness range and dynamics. Loudness range, pitch range, and pleasantness mainly influenced the perception of tension arousal. Like Western classical solo music, dynamics significantly impacted the perception of energy arousal in Western classical ensemble music. Therefore, whether it was Western classical solo or ensemble, dynamics played an important role in energy arousal perception. Due to the emphasis on timbre's consonance in Western classical music, symphony is the main ensemble form to create magnificence and shock value. This characteristic gives Western classical music high dynamics that arouses full, active emotions. Rhythm complexity was also an important element for energy and tension arousal perception.

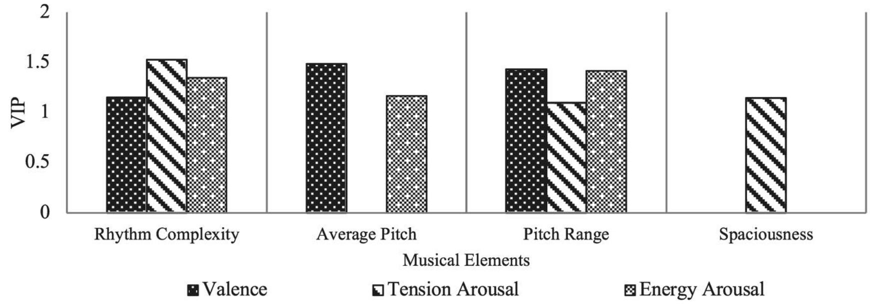
These results revealed differences in culturally specific musical elements for the four cultural datasets. Register features played an important role for Chinese classical ensemble music. Spaciousness was important to all emotional perceptions of Chinese classical solo music and for Western classical solo music, timbre features were important for valence perception.

4.3 | Performance analysis for combined annotations

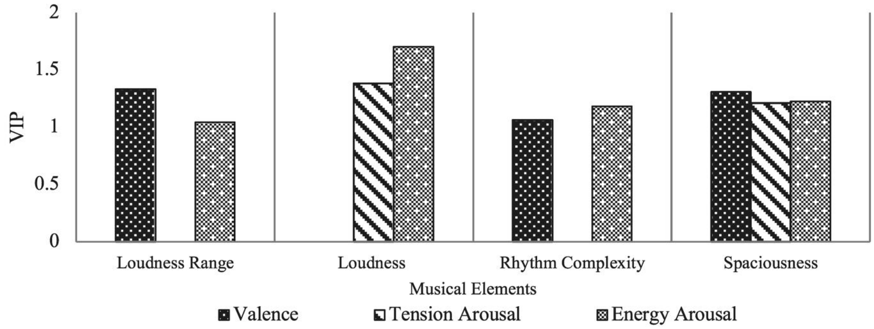
Based on the PLSR model's results for the datasets, performances of the PLSR model (R^2) were greater than 0.6 except for tension arousal in Chinese classical solo music, thus indicating that the regression models based on 15 musical elements performed well. Musical elements have higher explanatory and predictive abilities in energy arousal, followed by valence and tension arousal. Compared with automatic algorithm annotation, manual annotation can truly reflect participants' perceptual results and cover more effective information with actual audition perception. Therefore, the combination of manual and automatic annotation can, to a certain extent, improve the music emotion model's recognition accuracy.

4.4 | Interaction of musical elements

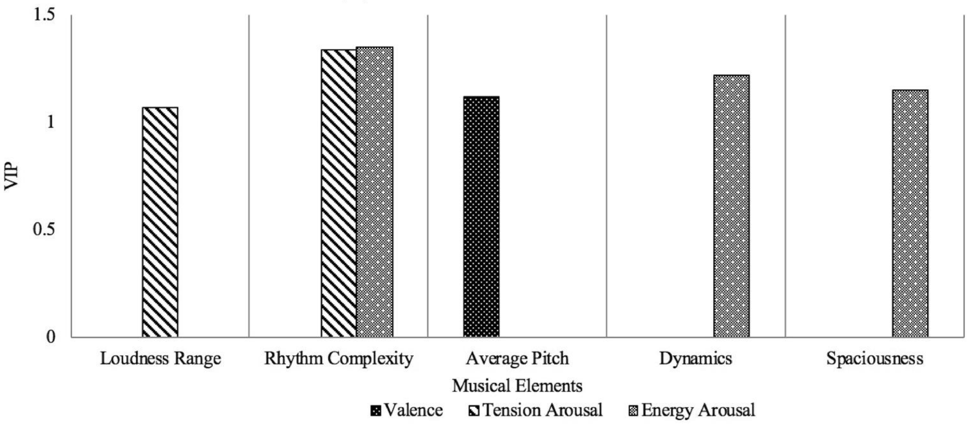
To explore musical elements' interaction, we conducted the Pearson correlation analysis on the cross-cultural dataset and only the musical elements that were significantly correlated are shown in Table 11, in which correlation coefficient above 0.6 are underlined. In rhythm features, rhythm complexity correlated strongly and positively with pulse clarity. In addition, tempo correlated strongly and positively with pulse clarity and rhythm complexity but strongly and negatively with articulation. By analysing the correlation between articulation and



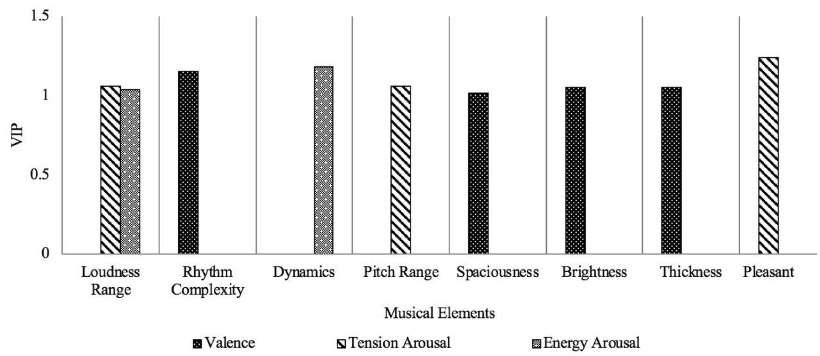
(a) Chinese classical ensemble music dataset.



(b) Chinese classical solo music dataset.



(c) Western classical ensemble music dataset.



(d) Western classical solo music dataset.

FIGURE 6 Musical elements with cultural specificity for the four datasets

rhythm features, we infer that staccato articulation can strengthen the sense of rhythm. In timbre features, richness correlated strongly and positively with clarity, pleasantness, and spaciousness. For register features, average pitch and brightness correlated strongly and positively; thus, pitch is affected by the frequency component.

TABLE 11 Pearson correlation results of significantly correlated musical elements

	Clarity	Pleasantness	Spaciousness	Pulse clarity	Rhythm complexity	Tempo	Articulation	Brightness	Richness
Pleasantness	0.45**								
Spaciousness	0.42**	0.52**							
Pulse clarity	0.23**	0.26**	0.11						
Rhythm complexity	0.26**	0.21*	0.13	0.61**					
Tempo	0.20*	0.09	0.05	0.76**	0.65**				
Articulation	−0.08	0.06	0.15	−0.57**	−0.47**	−0.74**			
Brightness	0.23**	−0.21*	−0.09	0.15	0.18*	0.14	−0.18*		
Richness	0.67**	0.64**	0.61**	0.17*	0.20*	0.09	0.16	−0.02	
Average pitch	0.10	−0.19*	−0.10	0.21**	0.15	0.20*	−0.26**	0.64**	−0.21*

Note: $df = 144$. Correlation coefficient above 0.6 are underlined.
*$p < 0.05$, **$p < 0.01$.

5 | CONCLUSION AND FUTURE WORK

In this study, we explored the correlations between emotion perception and musical elements using a cross-cultural approach, and results showed that the combination of manual and automatic annotation of musical elements can improve the MER system's accuracy. Based on our cross-cultural dataset, tempo, rhythm complexity, and articulation are culturally universal musical elements. With increasing tempo, rhythm complexity, and staccato, perception of valence, tension arousal, and energy arousal effectively improves. Musical elements related to timbre, register, and dynamics features are culturally specific. According to correlation results, staccato articulation can strengthen the sense of rhythm.

The annotation of datasets still needs improvement. Due to the experiment's restriction to online experimentation due to the pandemic, our annotation results were not ideal, causing much music data to be removed. In cross-cultural MER, we can appropriately enrich Chinese and Western music styles as experimental signals, such as jazz and electronic music, and also increase the dynamics annotation of musical elements and emotional perceptions to lay a solid foundation for the development of an improved emotion recognition model. In addition, for future experiments, it is worth discussing to ask Western musicians to do the same task with Western and Chinese music. The comparison of the results will be helpful to explore more cross-cultural findings.

ACKNOWLEDGMENT

The authors would like to thank the participants of the listening test. Funding was provided by the Humanity and Social Science Youth Foundation of the Ministry of Education of China (15YJCZH178) and the Science Foundation of the Communication University of China (HG1608-1) to XW and Major Project of Beijing Social Science Foundation (20ZD20) to Xiuqing Jia.

CONFLICT OF INTERESTS

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Xin Wang https://orcid.org/0000-0002-8108-2911
Yujia Wei https://orcid.org/0000-0002-6445-2431

REFERENCES

1. Yang, Y., Chen, H.: Machine recognition of music emotion. ACM Trans. Intell. Syst. Technol. 3(3), 1–30 (2012)
2. Zentner, M., Grandjean, D., Scherer, K.R.: Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion. 8(4), 494–521 (2008)
3. Trochidis, K., Bigand, E.: Emotional responses during music listening. In: Guide to Brain-Computer Music Interfacing, pp. 105–132. (2014)
4. Bresin, R., Friberg, A.: Emotion rendering in music: range and characteristic values of seven musical variables. Cortex. 47(9), 1068–1081 (2011)
5. Gabrielsson, A., Lindström, E.: The role of structure in the musical expression of emotions. In: Juslin, P.N., Sloboda, J.A. (eds.) Handbook of Music and Emotion: Theory, Research, Applications, pp. 367–400. Oxford University Press, Oxford (2010)
6. Juslin, P.N.: Cue utilization in communication of emotion in music performance: relating performance to perception. J. Exp. Psychol. Hum. Percept. Perform. 26(6), 1797–1812 (2000)
7. Lu, L., Liu, D., Zhang, H.: Automatic mood detection and tracking of music audio signals. IEEE Trans. Audio Speech Lang. Process. 14(1), 5–18 (2006)
8. Yang, X., Dong, Y., Li, J.: Review of data features-based music emotion recognition methods. Multimed. Syst. 24(4), 365–389 (2017)
9. Chen, Y., et al.: Linear regression-based adaptation of music emotion recognition models for personalization [Internet]. In: ICASSP, pp. 2149–2153 (2014). [cited 2020 Jul 14]. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6853979&isnumber=6853544 https://doi.org/10.1109/ICASSP.2014.6853979

10. Wang, J., Wang, H., Lanckriet, G.: A histogram density modeling approach to music emotion recognition [Internet]. In: ICASSP, pp. 698–702. (2015). [cited 2020 Jul 14]. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7178059&isnumber=7177909 https://doi.org/10.1109/ICASSP.2015.7178059
11. Leman, M., et al.: Prediction of musical affect using a combination of acoustic structural cues. J. New Music Res. 34(1), 39–67 (2007)
12. Deng, Y., et al.: Music emotion recognition based on middle and high level features [Internet]. Comput. Eng. Des. 38(4), 1029–1034 (2017). [cited 2020 Jul 1]. https://xueshu.baidu.com/usercenter/paper/show?paperid=1b9a5cbb8bca8acff62990fdf4047cb1&site=xueshu_se&hitarticle=1. https://doi.org/10.16208/j.issn1000-7024.2017.04.034
13. Balkwill, L., Thompson, W.: A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. Music Percep. 17(1), 43–64 (1999)
14. Wu, W., Xie, L.: Discriminating mood taxonomy of Chinese traditional music and Western classical music with content feature sets [Internet]. In: CISP, pp. 148–152. (2008). [cited 2020 Jul 14]. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4566804&isnumber=4566766. https://doi.org/10.1109/CISP.2008.272
15. Zhao, Z., et al.: The analysis of mood taxonomy comparison between Chinese and Western music [Internet]. In: ICSPS, pp. 606–610. (2010) https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5555502&isnumber=5555236 https://doi.org/10.1109/ICSPS.2010.5555502
16. Hu, X., Yang, Y.: Cross-dataset and cross-cultural music mood prediction: A case on Western and Chinese pop songs. IEEE Trans. Affect. Comput. 8(2), 228–240 (2017)
17. Hu, X., Yang, Y.: The mood of Chinese pop music: Representation and recognition. J. Assoc. Inform. Sci. Technol. 68(8), 1899–1910 (2017)
18. Cowen, A., et al.: What music makes us feel: At least 13 dimensions organize subjective experiences associated with music across different cultures. Proc. Natl. Acad. Sci. Unit. States. Am. 117(4), 1924–1934 (2020)
19. Davies, S.: Emotions expressed and aroused by music: philosophical perspectives. In: Juslin, P.N., Sloboda, J.A. (eds.) Handbook of Music and Emotion: Theory, Research, Applications, pp. 52. Oxford University Press, Oxford (2010)
20. Thompson, W.F., Balkwill, L.: Cross-cultural similarities and differences. In: Juslin, P.N., Sloboda, J.A. (eds.) Handbook of Music and Emotion: Theory, Research, Applications, pp. 755–788. Oxford University Press, Oxford (2010)
21. Mearns, L., Tidhar, D., Dixon, S.: Characterisation of composer style using high-level musical features [Internet]. In: MML (2010). https://doi.org/10.1145/1878003.1878016
22. Kim, Y.E., et al.: Music Emotion Recognition: A State of the Art Review. Music Emotion Recognition A State of the Art Review Researchgate (2010)
23. MacDorman, K.F., Stuart, O., Karl, F.: Automatic emotion prediction of song excerpts: index construction, algorithm design, and empirical comparison. J. New Music Res. 36(4), 281–299 (2007)
24. Ma, X., Gao, Y., Meng, Z.: The aesthetic recognition of Chinese folk songs [Internet]. Tech. Acoust. 37(6), 18–19 (2018). [cited 2020 Jul 3]. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CPFD&dbname=CPFDLAST2019&filename=OGSM201811014009&v=co4Hl0TwMMzyQd6ZeDVBsJF31H1dYD%25mmd2BmC9PdccGKMTVYagZeyHj07veX4OSWH%25mmd2BiesXNFIj6FzXo%3d
25. Soleymani, M., Aljanaki, A., Yang, Y.: DEAM: MediaEval Database for Emotional Analysis in Music. [Internet]. Swiss Center for Affective Sciences, University of Geneva, Swnd. [updated 2018 Apr; cited 2021 July 10]. https://cvml.unige.ch/databases/DEAM/manual.pdf.itzerla
26. Aljanaki, A., Wiering, F., Veltkamp, R.: Studying emotion induced by music through a crowdsourcing game. Inf. Process. Manag. 52(1), 115–128 (2016)
27. Chen, Y., et al.: The AMG1608 dataset for music emotion recognition [Internet]. In: ICASSP, pp. 693–697. (2015). https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7178058&isnumber=7177909. https://doi.org/10.1109/ICASSP.2015.7178058
28. Soleymani, M., et al.: 1000 songs for emotional analysis of music [Internet]. In: ACM, pp. 1–6. (2013). https://doi.org/10.1145/2506364.2506365
29. Aljanaki, A., Yang, Y., Soleymani, M.: Developing a benchmark for emotional analysis of music. PLoS One. 12(3), e0173392 (2017)
30. Bachorik, J., et al.: Emotion in motion: Investigating the time-course of emotional judgments of musical stimuli. Music Percep. 26(4), 355–364 (2009)
31. Itu-RBS: Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level. International Telecommunication Union (2015). [updated 2015 Oct; cited 2021 July 10]. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-4-201510-I!!PDF-E.pdf
32. Hevner, K.: Expression in music: A discussion of experimental studies and theories. Psychol Rev. 42(2), 186–204 (1935)
33. Watson, D., Clark, L., Tellegen, A.: Development and validation of brief measures of positive and negative affect: The PANAS scales. J. Personal. Soc. Psychol. 54(6), 1063–1070 (1988)
34. Thayer, R.: Activation-deactivation adjective check list: Current overview and structural analysis. Psychol. Rep. 58(2), 607–614 (1986)
35. Russell, J.: A circumplex model of affect. J. Personal. Soc. Psychol. 39(6), 1161–1178 (1980)
36. Schimmack, U., Grob, A.: Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. Eur. J. Personal. 14(4), 325–345 (2000)
37. Dai, M., et al.: Study on the relationship between timbre and emotional perception of Western musical instruments sounds. J. Commun. Univ. China (Sci. Technol.). 27(4), 43–50 (2020)
38. Schubert, E.: Modeling perceived emotion with continuous musical features. Music Percep. 21(4), 561–585 (2004)
39. Wedin, L.: A multidimensional study of perceptual-emotional qualities in music. Scand. J. Psychol. 13(1), 241–257 (1972)
40. Juslin, P., Lindstrom, E.: Musical expression of emotions: Modelling listeners judgements of composed and performed features. Music Anal. 29(1–3), 334–364 (2011)
41. Eerola, T., Friberg, A., Bresin, R.: Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. Front. Psychol. 4, 487 (2013)
42. Swaminathan, S., Schellenberg, E.: Current emotion research in music psychology. Emot. Rev. 7(2), 189–197 (2015)
43. Gagnon, L., Peretz, I.: Mode and tempo relative contributions to happy-sad judgements in equitone melodies. Cognit. Emot. 17(1), 25–40 (2003)
44. Lartillot, O., Toiviainen, P.: MIR in Matlab (II): A toolbox for musical feature extraction from audio [Internet]. In: ISMIR, pp. 127–130. (2007). https://www.researchgate.net/profile/Olivier-Lartillot/publication/220723440_MIR_in_Matlab_II_A_Toolbox_for_Musical_Feature_Extraction_from_Audio/links/09e4150c636c524d07000000/MIR-in-Matlab-II-A-Toolbox-for-Musical-Feature-Extraction-from-Audio.pdf
45. Zplane: ppmBatch Quickstart Guide v1.0. [Internet]. Zplane. development GmbH & Co KG, zplane (2016). [updated 2016 Jan; cited 2021 July 10]. https://products.zplane.de/wp-content/downloads/ppmBatch/ppmBatch_manual
46. Lartillot, O.: MIRtoolbox 1.7.2 User's Manual. [Internet]. University of Oslo, Norway (2019). [updated 2019 Jan 8; cited 2021 July 10]. https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7-2.pdf
47. Zacharopoulou, K., Kyriakidou, A.: A cross-cultural comparative study of the role of musical structural features in the perception of emotion in Greek traditional music. J. Interdiscip. Music Stud. 3(1–2), 1–15 (2009)
48. Geladi, P., Kowalski, B.: Partial least-squares regression: A tutorial. Anal. Chim. Acta. 185(1), 1–17 (1986)
49. Janes, K., Reinhardt, H., Yaffe, M.: Cytokine-induced signaling networks prioritize dynamic range over signal strength. Cell. 135(2), 343–354 (2008)
50. De Jong, S.: SIMPLS: An alternative approach to partial least squares regression. Chemometr. Intell. Lab. Syst. 18(3), 251–263 (1993)
51. Nengsih, T., et al.: Determining the number of components in PLS regression on incomplete data set. Stat. Appl. Genet. Mol. Biol. 18(6), 20180059 (2019)

52. Fernández-Sotos, A., Fernández-Caballero, A., Latorre, J.: Influence of tempo and rhythmic unit in musical emotion regulation. Front. Comput. Neurosci. 10, 80 (2016)
53. Bresin, R., Friberg, A.: Emotion rendering in music: Range and characteristic values of seven musical variables. Cortex. 47(9), 1068–1081 (2011)
54. Yi, L.: Research on timbre style in Chinese and Western music culture. Art Edu. 61 (2010)

How to cite this article: Wang, X., Wei, Y., Yang, D.: Cross-cultural analysis of the correlation between musical elements and emotion. Cogn. Comput. Syst. 4(2), 116–129 (2022). https://doi.org/10.1049/ccs2.12032

APPENDIX

The supplementary material lists parts of music tracks' names in the datasets.

	Chinese classical ensemble	Chinese classical solo	Western classical ensemble	Western classical solo
1	Bamboo Rhizome	Moon on Guan Mountain	Piano Concerto No. 2 In C Minor, Op. 18.1. Moderato	Fantasias (7 Piano Pieces), Op. 116.3. Capriccio In G Minor
2	On surging wave	Remembering playing Xiao on Tower Phoenix	Trio Sonata in D minor-3. Allegro (Telemann)	Hungarian Rhapsody No. 2
3	Yue Fei	Flicking of Candles	Sonata II in D minor – Allegro	Vladimir Ashkenazy_12 Etudes, Op. 10_ No. 12. in C Minor Revolutionary
4	Reluctant to Bid Farewell	Jin Zhu Er Ma	Matteis: Ground After the Scotch Humour	Piano Sonata No. 17 In D Minor, Op. 31 No. 2 – "Tempest" – 3. Allegretto
5	Su Wu Shepherd	Vain Longing	Bughici – Suite for Violin, 9 Hora, vivace	Études, Op. 25-No. 11 in A Minor
6	Wu You Dai	Legend of Little White Horse	Der Nuβknacker, Op. 71a – II. Danses charac. Marche	Waltz KK IVa no. 12 in E major
7	Nezha's Mandate	Mountain Reverie	Leonard Bernstein – Carmen Suite No. 2 – Chanson du Toredor. Allegro moderato (Act II)	BWV 1001/IV. Presto
8	Rush to Harvest Crops	Chuanjiang Charm	Sonata No. 2 in G major Op. 50 – Allemanda (allegro)	Romance in B Major Opus 28 No. 3 (Robert Schumann)
9	Mottled Bamboo	The Melody of the Great Wall	Oboe Quartet in F (K370) – Allegro	Sonatina No. 1 in C Major – I. Allegro
10	Soul of the Yellow River	General's Mandate	Sinfonia In Re Minore (Luigi Boccherini)	Winterreise, D. 911.1. Gute Nacht
11	Breaking the Blockade Line	A Vulture Pounced on The Swan	Phenix – Allegro	Guido Nielsen – Original Rags
12	The mountain village has changed	Festival in The Tian Mountains	Concerto No. 1 in G major Op. 26 for Organ – Allegro	Concertino for Trombone and Piano in Eb Major, Op. 4 – I. Allegro maestoso
13	Mountain area	Dance of the Golden Snake	Handel – Entrance to the Queen of Sheba for Two Oboes, Strings, and Continuo allegro	Intermezzo in B flat Minor Opus 117 No. 2 (Johannes Brahms)
14	A New Song of Herdsmen	Morning Glow	Ballo Del Gran Ducca (Giovanni Battista Buonamente)	Intermezzo in E Major Opus 116 No. 6 (Johannes Brahms)
15	Journey to Qinchuan	Ten Thousand Steeds Gallop	Chansons sans paroles (1989) op. 2 Pastorale	Chopin Prelude In A Flat Major Opus 28
16	Reed Pipe	Two Horses of Genghis Khan	Triosonate fur Flute Violine und Basso – Dolce (Telemann)	Romance in F sharp Major Opus 28 No. 2 (Robert Schumann)
17	Whip the Horse to Carry Grain	The Green Lotus	Wheel of fortune (with Les Gauchers Orchestra)	J. S. Bach: Prelude in C – BWV 846
18	Spring in Mulberry Field	A Cinematic Journey	Ouverture – Lee Maddeford	Intermezzo in E Major Opus 116 No. 4 (Johannes Brahms)

(Continued)

	Chinese classical ensemble	Chinese classical solo	Western classical ensemble	Western classical solo
19	The Fishing Songs in East China Sea	A Night of Flowers and Moonlight by the Spring River	Dmitri Shostakovich – VI. Waltz 2 from Jazz Suite No. 2 (Eyes Wide Shut). 3	Chopin Prelude In B Minor Opus 28
20	Lament at Changmen Palace	Lotus Emerging out of Water		Les Saisons, Op37b. Octobre
21	Bid Farewell to a Departing Friend	Sinking Moon		Suite No. 3 In D, BWV 1068
22	Running Water	The Two Phoenixes Fly Side by Side		The Seasons, for piano, Op. 37 Barcarole (June)
23	Li Sao			Waltz No. 10 in B Minor, Op. 69 No. 2
24	The Night Rain Outside the Window			
25	Litchi Turns Red			
26	Hunting Tigers up on the Mountain			
27	River Water			