



ARTICLE



<https://doi.org/10.1057/s41599-025-06211-x>

OPEN

# How does Chinese traditional instrumental music convey emotions? The power of affective timbres

Liang Xu<sup>1✉</sup>, Yishan Liu<sup>2</sup>, Zehua Jiang<sup>2</sup>, Yulong Tang<sup>1</sup> & Xiangming Wu<sup>1✉</sup>

This study explores how timbral features shape perceived emotions in Chinese traditional instrumental music, offering insights into the interplay between music and culture. Using a dataset of 273 music excerpts annotated by 168 Chinese participants, we analyzed the relationships between 54 timbral features and perceived emotions, considering both the dimensional model of emotion (valence and arousal) and discrete emotions (anger, sadness, happiness, peacefulness, transcendence, gentleness, and solemnness). Linear regression analyses revealed that timbral features commonly used in Western music studies can also effectively predict the perceived emotions in Chinese traditional music, with adjusted  $R^2$  values ranging from 0.302 (solemnness) to 0.751 (arousal). Combining linear regression and random forest models, we identified distinct emotional expression patterns unique to Chinese traditional music. Most notably, happiness was idiosyncratically associated with increased noise energy and inharmonicity, a pattern that contrasts with findings in many Western music studies. By contrast, sadness corresponded to reduced noise energy and lower inharmonicity, solemnness was characterized by a narrower range of spectral variability and shorter sound duration, and transcendence was expressed through reduced spectral variability, a wider range of spectral kurtosis, and a narrower range of spectral flatness. These findings may be closely connected to Chinese traditional culture, warranting further in-depth investigation.

<sup>1</sup>Zhejiang University of Technology, Hangzhou, China. <sup>2</sup>Zhejiang University, Hangzhou, China. ✉email: [xuliang\\_psy@zju.edu.cn](mailto:xuliang_psy@zju.edu.cn); [wxm@zjut.edu.cn](mailto:wxm@zjut.edu.cn)

## Introduction

Understanding how musical features shape the perception of emotion is a core area of research in music psychology (Gabrielsson & Lindström, 2010). Perceived emotion, defined as the emotions that listeners recognize as being expressed by music, is central to how music communicates meaning (Feng et al., 2003; Gabrielsson, 2001; Schubert, 2013). In contrast, felt emotion refers to the subjective emotional experience elicited in listeners (Gabrielsson, 2001), which can vary significantly depending on individual factors such as mood, personality, and context (Juslin & Laukka, 2004; Xu, Wen et al., 2021). Although these two dimensions are interconnected, perceived emotion focuses on the communicative intent of the music itself, making it particularly relevant for understanding how musical elements, such as melody, harmony, rhythm, and especially timbre, contribute to emotional expression (Korsmit et al., 2024; Xu, Sun et al., 2021). By shaping listeners' interpretations of emotion, these elements enable music to convey nuanced emotional content independent of listeners' personal reactions.

While most research has predominantly focused on Western music, there is a growing recognition of the need to explore this relationship across different musical traditions (Hu & Yang, 2017; Jacoby et al., 2020; Trehub et al., 2015). Chinese traditional instrumental music, characterized by its distinctive pentatonic system and unique timbral qualities (Nan & Guan, 2023), offers a rich context for such exploration. This study seeks to deepen our understanding of how timbral features contribute to perceived emotion in Chinese traditional music, broadening the global perspective on music's emotional expressiveness.

Chinese traditional music has a long-standing history that sets it apart from Western music, both in terms of instruments and aesthetics (Rao, 2002; Wu et al., 2024). Instruments such as the Erhu, Pipa, and Sheng are crafted from natural materials, giving them distinctive tonal qualities that significantly impact the emotional experience of the music (Hao, 2023). The pentatonic system, which forms the basis of much Chinese traditional music, creates a spacious and open sound that contrasts with the heptatonic system of Western music (Zhang et al., 2022). These differences extend beyond the structural level and are deeply intertwined with cultural philosophies, such as Confucianism and Taoism, which emphasize harmony and balance (Hao, 2023). The unique cultural and aesthetic context of Chinese traditional music thus shapes the emotional expression and perception, offering an opportunity to explore how timbral features may evoke emotions differently than in Western music.

In the broader study of musical emotion, timbre has been recognized as a critical factor influencing listeners' affective responses (Korsmit et al., 2024), although its importance is sometimes overlooked (Filipic et al., 2010). Research highlights that timbre plays a fundamental role in conveying musical emotion (McAdams, 2019; Schutz et al., 2008). However, findings about the emotional associations of different timbres are often inconsistent. For example, timbres have been linked to both anger and fear as well as positive emotions (Grimaud & Eerola, 2022; Xu, Wen et al., 2021). Hence, how do these relationships manifest in Chinese traditional instrumental music? This study seeks to investigate this question in depth. Moreover, recent advancements in computational analysis have enabled more precise explorations of how specific audio features relate to perceived emotion (Panda et al., 2020). Timbre, often characterized by parameters such as brightness, harmonicity, and spectral features (Peeters et al., 2011; Korsmit et al., 2024), plays a crucial role in shaping emotional perception. For instance, a higher spectral centroid is associated with brighter, more energetic emotions, whereas lower centroids evoke darker, more subdued feelings (Peeters et al., 2011). By employing advanced computational

techniques, this study will investigate how timbral features in Chinese traditional music influence perceived emotion.

An additional important aspect of studying musical emotion is the selection of an appropriate emotion model. While many studies have applied general human emotion models, such as Ekman's discrete emotion model (Ekman, 1992) or Russell's dimensional model (Russell, 1980), these frameworks may not fully capture the specific nuances of music-related emotions (Korsmit et al., 2023). To address these limitations, music-specific models, like Zentner et al.'s (2008) nine-factor model and the three-dimensional model proposed by Greenberg et al. (2016), have been developed. In the context of Chinese traditional music, Shi (2015) proposed a seven-factor model of musical emotions, encompassing anger, sadness, happiness, peacefulness, transcendence, gentleness, and solemnness. The development of this discrete emotion model followed a methodological framework similar to that of Zentner et al. (2008), involving three key steps: compiling music-related emotion terms, conducting exploratory factor analysis to identify the underlying emotional dimensions, and employing confirmatory factor analysis to validate the structure (Shi, 2015). Many of the factors identified in Shi's model align with well-established dimensions of musical emotions, underscoring cross-cultural commonalities in emotional experiences. For instance, anger, happiness, and sadness are basic emotions extensively studied in the field of affective science (Laukka et al., 2013), while gentleness and peacefulness are often used to describe neutral emotional states (Zentner et al., 2008). Notably, solemnness and transcendence stand out as prominent aesthetic emotions, reflecting the deeper, often spiritual dimensions of musical experience (Akkermans et al., 2018; Zentner et al., 2008). This study will adopt Shi's (2015) seven-factor model to explore the relationship between timbral features and the perceived emotions in Chinese traditional music.

In sum, we conducted an exploratory study to investigate the associations between affective timbres and the perceived emotions of Chinese traditional music. By integrating audio feature extraction techniques and machine learning (ML) methods, this study aimed to address the following three questions: (a) Can timbral features that have been shown to predict different emotions in Western music also predict perceived emotion in Chinese traditional music? (b) Based on the results of computational modeling, which timbral features are most effective in predicting perceived emotion in Chinese traditional music? (c) How do these findings compare to those from studies on Western music, highlighting similarities and differences? Answering these questions will deepen our understanding of the emotional expressiveness of Chinese traditional music and contribute to a broader cross-cultural perspective in music research.

## Methods

**Dataset.** This study employed music excerpts and corresponding emotion annotations from the Chinese Traditional Instrumental Music (CTIM) dataset (Wu et al., 2024), which was specifically designed to comprehensively represent the diversity and emotional depth of traditional Chinese instrumental music. While earlier datasets (e.g., Li et al., 2012; Xu, Yun et al., 2022) included some Chinese instrumental pieces, they often lacked genre-specific focus and sufficient coverage of traditional repertoire. In contrast, the CTIM dataset was curated through a rigorous process led by an expert panel comprising seasoned musicians and psychology graduate students. This panel selected 145 ensemble performances featuring traditional *bayin* instruments, spanning a historical timeline from the Qin dynasty (221 BCE–206 BCE) to

the 20th century. The selection emphasized emotional richness and stylistic diversity.

To ensure consistency and scientific utility, each musical piece was edited into one to four 10 s excerpts (Wu et al., 2024). These excerpts were carefully segmented at phrase boundaries containing core melodies, with particular attention to preserving emotional continuity and minimizing variations in musical elements such as rhythm, timbre, and dynamics. The final dataset was processed uniformly: all excerpts were sampled at 44 kHz, encoded at a bit-rate of 192 kbps, and standardized in sound intensity. For the current study, all the 273 excerpts from the CTIM dataset were utilized as stimuli, each lasting 10 s.

Given that this research retrospectively used publicly available data, the Research Ethics Committee confirmed that ethical approval was not required. All data collection procedures and analytical methods adhered strictly to relevant ethical guidelines and standards.

**Music emotion annotations.** Emotion annotations were provided by 168 Chinese participants (Wu et al., 2024), with each excerpt being rated by 56 individuals. Wu et al. (2024) employed an adjective-based rating system on a 7-point Likert scale to evaluate emotions. Participants were instructed to assess the intensity of each discrete emotion (anger, gentleness, happiness, peacefulness, sadness, solemnness, and transcendence; Shi, 2015), from 1 (“nonexistent”) to 7 (“extremely intense”). For the dimensional model, valence ranged from 1 (“extremely negative”) to 7 (“extremely positive”), and arousal from 1 (“not at all aroused”) to 7 (“extremely aroused”). Further details on the annotation process are available in Wu et al. (2024).

**Timbre feature extraction.** Timbre features were computed using the Timbre Toolbox (Kazazis et al., 2021), developed from the work of Peeters et al. (2011). In line with Korsmit et al. (2024), timbre features were derived from the short-term fast-Fourier transform in the spectral domain (including spectral centroid, spectral spread, spectral skewness, spectral kurtosis, spectral flatness, spectral crest, spectral slope, spectral decrease, spectral roll off, spectral variation, and spectral flux), the harmonic partials (including fundamental frequency, harmonic spectral deviation, Tristimulus 1, Tristimulus 2, Tristimulus 3, harmonic odd to even ratio, inharmonicity, harmonic energy, noise energy, noisiness, harmonic to noise energy, and partials to noise energy), and the temporal energy envelope (including attack time, log attack time, attack slope, decrease slope, temporal centroid, effective duration, frequency of energy modulation, and amplitude of energy modulation). Time-varying descriptors were summarized using the median and interquartile range (IQR) across each 10 s excerpt. A total of 54 descriptors, as outlined by Korsmit et al. (2024), were used for predicting emotional perception. The specific details of these features are provided in the Supplementary Table S1.

**Analytical approach.** To predict musical emotions, both linear and nonlinear regression techniques were applied (e.g., Korsmit et al., 2024; Wen et al., 2022), following previous findings that suggest nonlinear relationships might better capture the interaction between timbre and emotion (McAdams & Goodchild, 2017; Xu, Wen et al., 2021). Following the method of Korsmit et al. (2024), Lasso regression was first used for variable selection, after which standard linear regression was applied to predict different emotion ratings.

For the nonlinear approach, random forest regression (RFR) was used to assess the contribution of timbre descriptors. Each timbre feature served as input, while emotion ratings were treated

as the output (ground truth) for building separate RFR models for each emotion (Xu et al., 2024). A grid search was performed to fine-tune the model’s parameters, and tenfold cross-validation was implemented to validate the model’s generalizability. Model performance was assessed using the  $R^2$  statistic, while Gini importance (Strobl, Malley, & Tutz, 2009) was utilized to rank the importance of variables in predicting emotions. Gini importance efficiently identifies key features by measuring their contribution to impurity reduction at each decision tree split (Archer & Kimes, 2008), making it a suitable choice for exploratory analysis. However, it is not without limitations, such as potential biases toward features with higher variability or more categories, and it provides limited interpretive value regarding feature relationships. To address this, correlation analysis was incorporated to complement the rankings and enhance the interpretability of the model results.

In addition, to visualize the similarity among timbral correlates of different emotion dimensions, we applied Principal Component Analysis (PCA). Specifically, we first computed the correlation coefficients between each of the 54 timbral features and each emotion dimension, yielding a 54-dimensional timbral vector for every emotion. These vectors were then submitted to PCA, and the first two principal components were plotted to provide a two-dimensional visualization of how timbral profiles of different emotions cluster or diverge. This analysis was used solely for visualization and interpretation purposes, without being part of the main inferential analyses.

## Results

**Linear regressions.** To explore the linear relationships between timbral features and various emotions, we first applied Lasso regression for feature selection, followed by standard linear regression. The full results of the feature selection and standard linear regression for each emotion are presented in Supplementary Tables S3–S11. Table 1 summarizes the predictive performance of the linear regression models for each emotion category, along with the top five features with the highest absolute standardized regression coefficients.

We observed that for the dimensional model (valence and arousal), the selected timbral features predicted arousal more effectively, with an adjusted  $R^2$  of 0.751. Partial Noise Energy and the IQR of Inharmonicity were significant negative predictors of perceived arousal, whereas Fundamental Frequency and Spectral Variation positively predicted arousal. In contrast, the model for valence had a lower predictive performance (adjusted  $R^2 = 0.492$ ). Interestingly, both the median and IQR of Inharmonicity played key roles in predicting valence and arousal, indirectly supporting previous research that found a strong positive correlation between these two dimensions (Chen et al., 2015).

For the discrete emotion models, transcendence had the highest predictive accuracy (adjusted  $R^2 = 0.642$ ), followed by happiness (adjusted  $R^2 = 0.530$ ), peacefulness (adjusted  $R^2 = 0.505$ ), anger (adjusted  $R^2 = 0.406$ ), sadness (adjusted  $R^2 = 0.404$ ), gentleness (adjusted  $R^2 = 0.307$ ), and solemnness (adjusted  $R^2 = 0.302$ ). We also observed that certain timbral features were consistently important predictors across different emotions. For example, the median of Spectral Spread positively predicted gentleness and happiness, while negatively predicting anger, solemnness, and transcendence. Additionally, the median of Tristimulus 3 positively predicted anger and solemnness but negatively predicted gentleness. Notably, noisiness was strongly negatively correlated with peacefulness ( $\beta = -0.975$ ) and positively correlated with anger ( $\beta = -0.561$ ). These findings provide valuable insights into the linear relationships between various timbral features and perceived emotions in Chinese traditional instrumental music.

**Table 1 Results of linear regressions.**

**Top five predictors ( $\beta$ )**

**Emotions (N)**  
[R<sup>2</sup>, adjusted R<sup>2</sup>]

Valence (35) [0.557, 0.492]	Spectral Spread MED (0.532)	Inharmonicity MED (0.502)	Spectral Roll Off MED (-0.402)	Spectral Crest MED (-0.364)	Inharmonicity IQR (-0.276)
Arousal (28) [0.777, 0.751]	Partials Noise Energy MED (-0.371)	Inharmonicity IQR (-0.247)	Fundamental Frequency IQR (0.221)	Spectral Variation IQR (0.214)	Inharmonicity MED (0.204)
Anger (32) [0.475, 0.406]	Tristimulus 3 MED (0.643)	Noisiness MED (0.561)	Spectral Spread MED (-0.543)	Spectral Flux MED (-0.369)	Spectral Slope IQR (0.366)
Gentleness (34) [0.394, 0.307]	Spectral Spread MED (1.245)	Spectral Roll Off MED (-0.751)	Inharmonicity MED (0.614)	Tristimulus 3 MED (-0.504)	Spectral Flux MED (0.399)
Happiness (32) [0.586, 0.530]	Spectral Spread MED (0.581)	Inharmonicity MED (0.523)	Spectral Roll Off MED (-0.346)	Inharmonicity IQR (-0.314)	Spectral Variation IQR (0.282)
Peacefulness (31) [0.562, 0.505]	Noisiness MED (-0.975)	Partials Noise Energy MED (-0.706)	Spectral Skewness IQR (0.447)	Spectral Skewness MED (-0.377)	Spectral Variation IQR (-0.364)
Sadness (38) [0.487, 0.404]	Spectral Roll Off MED (0.627)	Spectral Crest MED (0.604)	Inharmonicity MED (-0.425)	Partials Noise Energy MED (0.371)	Spectral Skewness IQR (-0.364)
Solemnness (33) [0.387, 0.302]	Spectral Spread MED (-0.865)	Inharmonicity MED (-0.708)	Tristimulus 3 MED (0.707)	Spectral Variation IQR (-0.511)	Inharmonicity IQR (-0.312)
Transcendence (24) [0.674, 0.642]	Spectral Variation IQR (-0.455)	Spectral Slope IQR (-0.345)	Spectral Spread MED (-0.292)	Spectral Skewness IQR (0.289)	Effective Duration (-0.214)

Note: N = number of selected predictors from Lasso regression;  $\beta$  = Standardized coefficients of linear regression; MED = median values; IQR = interquartile range.

**Machine learning analysis.** We then used RFR to explore the nonlinear relationship between timbre features and various emotions. Figures 1a–i present the 12 most important timbre features for each emotion recognition model, with the complete feature importance results available in Supplementary Table S12. The RFR model successfully captured the nonlinear associations between timbre characteristics and perceived emotions. For example, in the RFR model for valence, the median of Noisiness emerged as the most crucial predictor of valence. This feature, however, was excluded in the linear models using Lasso regression. In the case of arousal, the median of Partials Noise Energy was identified as the most significant feature, contributing 29.11% to the model’s total predictive power as measured by Gini importance, a finding consistent with linear regression results.

For discrete emotion models, the median of Partials Noise Energy was also the most important feature for recognizing anger, contributing 9.49% to the model’s total predictive power. This was followed by the median of Noisiness (7.94%) and the IQR of Noisiness (7.81%), indicating that noise-related timbre features play a key role in predicting anger. Similarly, the median of Noisiness was the most important predictor for sadness (explaining 7.12%) and happiness (12.81%). The negative relationship between Noisiness and sadness reveals that Chinese traditional music tends to use less noise when conveying sadness. Conversely, the positive relationship between Noisiness and happiness suggests that more noise elements are incorporated into music to express happiness.

For gentleness, the most significant predictor in the RFR model was Effective Duration, contributing 6.74% to the model’s total predictive power. The positive association between Effective Duration and gentleness reflects a tendency in Chinese traditional music to use longer perceived sounds when expressing gentle emotions. Regarding peacefulness, both the median and IQR of Spectral Variation played a crucial role, contributing 16.63 and 13.03% to the model’s total predictive power, respectively. This aligns with findings from Western music, where peaceful compositions often exhibit less spectral variation. A similar pattern was observed in the RFR model for transcendence, where the IQR of Spectral Variation accounted for 32.80% of the variance, suggesting that transcendence is often associated with reduced spectral variation.

Finally, for solemnness, the IQR of Spectral Variation and Noisiness contributed 8.54 and 6.82% to the model’s total predictive power, respectively. These results indicate that lower variability in both spectral variation and noisiness is linked to a heightened perception of solemnness in Chinese traditional music. In other words, more consistent spectral properties and reduced noisiness variability may contribute to a more solemn emotional tone in the music.

**Comparison of top timbre features in linear and nonlinear models.** To better understand the differences in results between the linear regression (LR) and RFR models and clarify the influence of timbre features on different perceived emotions, Table 2 presents a comparison of the most important features identified by the two models. As shown in Table 2, for a few emotions (such as arousal and peacefulness), the key features identified by both models were similar. For instance, in the prediction model for peacefulness, both LR and RFR highlighted Spectral Variation IQR, Partials to Noise Energy MED, and Noisiness MED as important features.

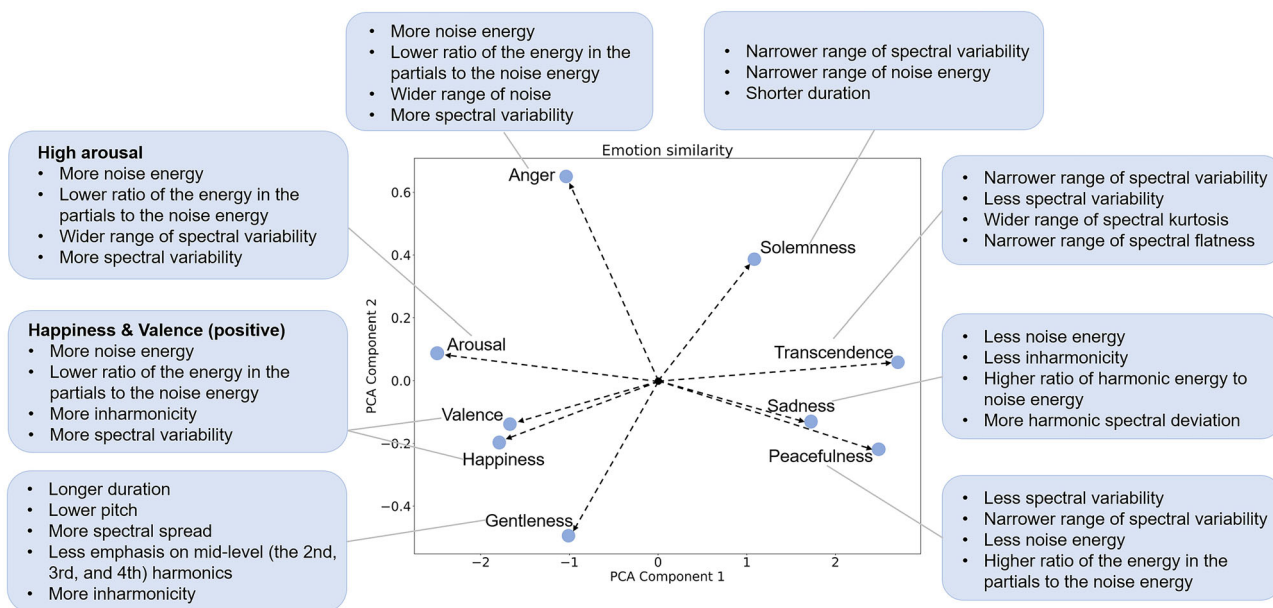
However, for most emotions, RFR captured key timbre features that differed significantly from those identified by LR. For example, in the prediction of solemnness, only one feature—Spectral Variation IQR—was shared among the top five features



**Table 2 Top five timbral features in different music emotion recognition models.**

Emotion models		Top five timbral features in different models		
Valence	LR RFR	Spectral Spread MED (+) <i>Noisiness MED</i>	<b>Inharmonicity MED (+)</b> <i>Partials: Noise Energy MED</i>	Spectral Roll Off MED (-) <b>Inharmonicity MED</b>
Arousal	LR RFR	<b>Partials Noise Energy MED (-)</b> <b>Partials: Noise Energy MED</b>	Inharmonicity IQR (-) <i>Noisiness MED</i>	Fundamental Frequency IQR (+) <b>Spectral Variation IQR (+)</b> <i>Spectral Variation MED</i>
Anger	LR RFR	Tristimulus 3 MED (+) <i>Partials: Noise Energy MED</i>	<b>Noisiness MED (+)</b> <b>Noisiness MED</b>	Spectral Flux MED (-) <i>Spectral Variation MED</i>
Gentleness	LR RFR	<b>Spectral Spread MED (+)</b> <i>Effective Duration</i>	Spectral Roll Off MED (-) <b>Spectral Spread MED</b>	Tristimulus 3 MED (-) <b>Inharmonicity MED (+)</b> <i>Tristimulus 2 MED</i>
Happiness	LR RFR	Spectral Spread MED (+) <i>Noisiness MED</i>	<b>Inharmonicity MED (+)</b> <i>Partials: Noise Energy MED</i>	Inharmonicity IQR (-) <i>Spectral Crest MED</i>
Peacefulness	LR RFR	<b>Noisiness MED (-)</b> <i>Spectral Variation MED</i>	<b>Partials Noise Energy MED (-)</b> <b>Spectral Variation IQR</b>	Spectral Skewness MED (-) <b>Noisiness MED</b>
Sadness	LR RFR	Spectral Roll Off MED (+) <i>Noisiness MED</i>	Spectral Crest MED (+) <b>Inharmonicity MED</b>	Partials Noise Energy MED (+) <i>Harmonic Spectral Deviation IQR</i>
Solemnness	LR RFR	Spectral Spread MED (-) <b>Spectral Variation IQR</b>	Inharmonicity MED (-) <i>Noisiness IQR</i>	<b>Spectral Variation IQR (-)</b> <i>Effective Duration</i>
Transcendence	LR RFR	<b>Spectral Variation IQR (-)</b> <b>Spectral Variation IQR</b>	Spectral Slope IQR (-) <i>Spectral Variation MED</i>	Spectral Skewness IQR (+) <b>Spectral Spread MED</b>

Note. LR = linear regression; RFR = random forest regression. (+) and (-) indicate whether the feature is positively or negatively correlated with the corresponding emotion in the LR model. Bolded features are among the top five features in both models, while italicized features are in the top five only in the RFR model.



**Fig. 2 The associations between timbre features and perceived emotions in Chinese traditional music.** The more similar the emotional timbre features are, the closer the vectors appear in the figure.

and vibrant social dynamics typical of Chinese festivals. Future studies could directly test these cultural interpretations, for instance by asking listeners to rate celebratory feelings beyond general happiness, or by experimentally manipulating timbral features to examine whether they elicit *Re Nao*-related responses.

By contrast, studies of Western music have often highlighted the role of harmonic consonance and melodic structure in conveying positive emotions (Webster & Weir, 2005). These are not timbral features per se, but there is also evidence that timbre plays a role in Western contexts. For instance, brightness, spectral centroid, and attack time have been associated with joy or positive affect in Western classical and popular music (Eerola, Ferrer, & Alluri, 2012; Eerola, Friberg, & Bresin, 2013). This suggests that timbre contributes to emotional expression across cultures, although the specific features emphasized may differ. In addition, recent cross-cultural studies suggest that Western music may rely more on harmonic consonance and pitch-based cues to convey joy, whereas Chinese music emphasizes timbral cues such as loudness and spectral variability (Wang, Wang, & Xie, 2022). This contrast could stem from differences in instrumentation, performance practices, or aesthetic preferences, such as the prominence of percussive timbres in Chinese ensembles versus the harmonic resources emphasized in Western traditions. At the same time, cultural concepts like *Re Nao*, which value communal energy, may also provide a useful lens for interpreting these findings, though such connections remain speculative and require further empirical validation.

In contrast to happiness, Chinese traditional music expresses sadness through reduced noise energy and lower inharmonicity, portraying a more subdued and restrained form of sorrow. The restrained expression of sadness in Chinese music might reflect a more introspective and inward-focused emotional style, aligning with cultural values that emphasize emotional balance and social harmony (Reilly, 2017). These musical features—reduced inharmonicity and smoother timbre—may represent a form of acceptance or reflection rather than overt grief, consistent with collectivist values that prioritize emotional restraint (Ip et al., 2021) and maintain harmony within the group (Chiu & Kosinski, 1994). Interestingly, this subdued and introspective mode of expressing sadness is also found in certain Western musical

traditions (Juslin & Laukka, 2004; Juslin & Sloboda, 2011), where slower tempos, softer dynamics, and smoother timbres are often employed to convey sadness in a more understated manner.

For solemnness, we found that solemn music exhibited a narrower range of spectral variability, reduced noise energy, and shorter durations, which points to a more focused and controlled sonic texture. These acoustic features likely contribute to a perception of solemnity by limiting excessive variation and complexity, aligning with the expectation of emotional restraint typically associated with solemn contexts. Recognizing that cultural factors significantly influence the perception of complex emotions (Matsumoto & Hwang, 2012) like solemnness, we further examined solemn excerpts from the CTIM database. These selections shared similarities with music used in Chinese Buddhist ceremonies (Zhang et al., 2016), suggesting a potential cultural connection. It is plausible that solemn music in this context draws upon traditional religious soundscapes, where simplicity and clarity in sound are integral. However, this resemblance does not necessarily confirm a direct relationship and should be investigated in the future.

Similar cultural phenomena are also in the expression of the emotion of transcendence. Our findings reveal that the musical expression of transcendence in Chinese traditional music is closely tied to natural sounds, characterized by a narrower range of spectral variability and overall less spectral variability. The recurring theme of stable timbral patterns associated with transcendence might reflect a philosophical resonance with Daoist ideas of harmony between humans and nature (Lun, 2012; Verellen, 1995). In Daoist thought, transcendence is not an escape from reality but a state of attunement with nature’s rhythms and cycles. The use of stable and consistent spectral properties in transcendent music may symbolize a sense of unity and balance, reflecting the Daoist ideal of “*Wu Wei*” (non-action) and an effortless existence in accordance with the natural order (Loy, 1985; Slingerland, 2000).

In summary, this study deeply explores the relationship between timbre features and perceived emotions in Chinese traditional music, and discusses these findings from a Chinese cultural perspective. However, the study has several limitations. First, the CTLM dataset (Wu et al., 2024) used in this study is

unbalanced, with some emotions (such as happiness) represented by more excerpts than others (such as anger and gentleness). This imbalance may influence the machine learning results and limit their generalizability (Kaur et al., 2019). Furthermore, the relatively small number of music pieces for certain emotions (i.e., gentleness, solemnness, and transcendence) raises questions about the reliability of the findings for those specific categories. Future studies should aim to construct more balanced datasets and incorporate larger sample sizes (Krawczyk, 2016) to ensure robust and reliable modeling of emotional associations in music. Second, as all participants in the CTLM database (Wu et al., 2024) were Chinese, this study cannot comprehensively address cross-cultural differences in the perception of emotions in Chinese traditional music. While our findings suggest that certain timbral features are associated with specific emotions, these associations may be shaped by cultural factors, such as collectivist values (Hu, 2024). For instance, timbral features linked to happiness in this study may not evoke the same emotional responses in listeners from cultures with more individualist values (Wang et al., 2022). Future research should include participants from diverse cultural backgrounds to explore whether such associations are consistent across cultures or culturally specific. Comparative studies examining how listeners from different cultures interpret timbre and emotion in Chinese traditional music would provide stronger evidence to validate or challenge these claims.

Third, the connections drawn between timbral features and Chinese cultural concepts, such as Daoist harmony or Buddhist ideas, are speculative and not empirically tested. These associations were inferred based on theoretical considerations rather than direct evidence from the data or participant feedback. Future research should employ empirical methodologies, such as balanced experimental designs and participant ratings of cultural concepts (Cowen et al., 2020), to rigorously validate these claims. Such approaches would provide a more robust foundation for understanding the interplay between timbre and cultural interpretations in traditional Chinese music. Fourth, although machine learning methods like RFR provide powerful predictive capabilities, their interpretability remains a key limitation (Krishnan, 2020; Murdoch et al., 2019). Feature importance metrics, including Gini importance, can be influenced by feature correlations and may not fully capture causal relationships. This highlights the importance of combining such models with interpretable methods, as done in this study, to provide a more comprehensive understanding of the relationships between timbre features and perceived emotions. Finally, the present study does not fully account for the historical evolution of Chinese traditional instrumental music or the different contexts in which it is performed and experienced. The emotional expression in music can be influenced by its historical background (Xu et al., 2023), regional styles (Argstatter, 2016), and performance settings (Rocke et al., 2022), all of which have evolved over time. A deeper exploration of these contextual factors is needed in future.

## Conclusion

In conclusion, this study provides important insights into how timbral features influence emotional perception in Chinese traditional instrumental music, revealing both shared patterns and unique cultural expressions. The findings show that happiness is conveyed through increased noise energy, inharmonicity, and spectral variability, which may reflect collectivist values of communal celebration and shared joy. Sadness, on the other hand, is expressed with reduced noise energy and smoother timbre, possibly aligning with cultural ideals of emotional restraint and social harmony. Solemnness is characterized by a narrower range of spectral variability, reduced noise energy, and shorter durations,

potentially suggesting a controlled and focused sonic texture that resembles features of traditional religious soundscapes. The portrayal of transcendence, through minimal spectral variability, might resonate with Daoist philosophy, emphasizing harmony with nature and balance between humans and the natural world. These findings highlight the potential ways in which Chinese traditional music expresses emotion, shaped by cultural and philosophical influences that warrant further exploration.

## Data availability

The data used in this study were sourced from the publicly available database by Wu et al. (2024), accessible at [https://osf.io/tzlx7/?view\\_only=c31867bc2ddf413583286bc0a582635a](https://osf.io/tzlx7/?view_only=c31867bc2ddf413583286bc0a582635a).

## Code availability

The codes used in this study can be found in the OSF repository ([https://osf.io/w7g35/?view\\_only=83a1f9aec7354774b3f79218bf361688](https://osf.io/w7g35/?view_only=83a1f9aec7354774b3f79218bf361688)).

Received: 9 February 2025; Accepted: 31 October 2025;

Published online: 17 December 2025

## References

- Akkermans J, Schapiro R, Müllensiefen D, Jakubowski K, Shanahan D, Baker D, Busch V, Lothwesen K, Elvers P, Fischinger T, Schlemmer K, Frieler K (2018) Decoding emotions in expressive music performances: a multi-lab replication and extension study. *Cogn Emot* 33(6):1099–1118
- Archer KJ, Kimes RV (2008) Empirical characterization of random forest variable importance measures. *Comput Stat data Anal* 52(4):2249–2260
- Argstatter H (2016) Perception of basic emotions in music: culture-specific or multicultural? *Psychol Music* 44(4):674–690. <https://doi.org/10.1177/0305735615589214>
- Chen, YA, Yang, YH, Wang, JC, & Chen, H (2015). The AMG1608 dataset for music emotion recognition. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 693–697). IEEE. <https://doi.org/10.1109/ICASSP.2015.7178058>
- Chiu RK, Kosinski FA (1994) Is Chinese conflict-handling behavior influenced by Chinese values? *Soc Behav Pers Int J* 22(1):81–90
- Cowen AS, Fang X, Sauter D, Keltner D (2020) What music makes us feel: at least 13 dimensions organize subjective experiences associated with music across different cultures. *Proc Natl Acad Sci USA* 117(4):1924–1934
- Eerola T, Ferrer R, Alluri V (2012) Timbre and affect dimensions: evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Percept* 30(1):49–70
- Eerola T, Friberg A, Bresin R (2013) Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Front Psychol* 4:487
- Ekman P (1992) An argument for basic emotions. *Cogn Emot* 6(3/4):169–200
- Feng, Y, Zhuang, Y, & Pan, Y (2003). Popular music retrieval by detecting mood. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 375–376). New York, NY: ACM
- Filipic S, Tillmann B, Bigand E (2010) Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bull Rev* 17(3):335–341
- Gabrielsson A (2001) Emotion perceived and emotion felt: same or different? *Musica Sci* 5(1\_suppl):123–147
- Gabrielsson, A, & Lindström, E (2010). The role of structure in the musical expression of emotions. In Juslin PN and Sloboda JA (Eds.), *Handbook of music and emotion: Theory, Research, Applications* (pp. 367–400). London, UK: Oxford University Press
- Greenberg DM, Kosinski M, Stillwell DJ, Monteiro BL, Levitin DJ, Rentfrow PJ (2016) The song is you: preferences for musical attribute dimensions reflect personality. *Soc Psychol Pers Sci* 7(6):597–605
- Grimaud AM, Eerola T (2022) Emotional expression through musical cues: a comparison of production and perception approaches. *PLoS ONE* 17(12):e0279605
- Hao W (2023) A comparative study of chinese and western music. *Highlights Art Des* 3(1):80–82
- Hu Y (2024) Cross-cultural perspectives in music: analyzing the impact of cultural differences on music preferences and practices. *Int J Educ Humanities* 15(1):170–173
- Hu X, Yang YH (2017) Cross-dataset and cross-cultural music mood prediction: a case on western and chinese pop songs. *IEEE Trans Affect Comput* 8(2):228–240

- Ip KI, Miller AL, Karasawa M, Hirabayashi H, Kazama M, Wang L, Olson SL, Kessler D, Tardif T (2021) Emotion expression and regulation in three cultures: Chinese, Japanese, and American preschoolers' reactions to disappointment. *J Exp Child Psychol* 201:104972
- Jacoby N, Margulis EH, Clayton M, Hannon E, Honing H, Iversen J, Klein TR, Mehr SA, Pearson L, Peretz I, Perlman M, Polak R, Ravnani A, Savage PE, Steingo G, Stevens CJ, Trainor L, Trehub S, Veal M, Wald-Fuhrmann M (2020) Cross-cultural work in music cognition: challenges, insights, and recommendations. *Music Percept* 37(3):185–195
- Juslin PN, Laukka P (2004) Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. *J N. music Res* 33(3):217–238
- Juslin, PN, & Sloboda, J (2011). *Handbook of music and emotion: Theory, Research, Applications*. Oxford University Press
- Kaur H, Pannu HS, Malhi AK (2019) A systematic review on imbalanced data challenges in machine learning: applications and solutions. *ACM Comput Surv (CSUR)* 52(4):1–36
- Kazakis, S, Depalle, P, & McAdams, S (2021). The Timbre Toolbox User's Manual. <https://github.com/MPCCL-McGill/TimbreToolbox-R2021a>
- Korsmit IR, Montrey M, Wong-Min AYT, McAdams S (2023) A comparison of dimensional and discrete models for the representation of perceived and induced affect in response to short musical sounds. *Front Psychol* 14:1287334
- Korsmit, IR, Montrey, M, Wong-Min, AYT, & McAdams, S (2024). The Acoustic Properties of Affective Timbres: Consistencies and Discrepancies in a Synthesis of Multiple Datasets. *Music & Science*, advance online publication. <https://doi.org/10.1177/20592043241256012>
- Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell* 5(4):221–232
- Krishnan M (2020) Against interpretability: a critical examination of the interpretability problem in machine learning. *Philos Technol* 33(3):487–502
- Laukka P, Eerola T, Thingujam NS, Yamasaki T, Beller G (2013) Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion* 13(3):434–449
- Li DD, Cheng ZB, Dai RN, Wang F, Huang YX (2012) Preliminary establishment and assessment of affective music system. *Chin Ment Health J* 26(7):552–556
- Loy D (1985) Wei-wu-wei: nondual action. *Philos East West* 35(1):73–86
- Lun, VMC (2012). Harmonizing conflicting views about harmony in Chinese culture. In Xu H and Michael BH (Eds.), *Handbook of Chinese Organizational Behavior* (p 560). Edward Elgar Publishing. <https://doi.org/10.4337/9780857933409.00033>
- Matsumoto D, Hwang HS (2012) Culture and emotion: the integration of biological and cultural contributions. *J Cross-Cult Psychol* 43(1):91–118
- McAdams, S (2019). Timbre as a structuring force in music. In Siedenburg, K, Saitis, C, McAdams, S, Popper, AN & Fay RR (Eds.), *Timbre: Acoustics, perception, and cognition*. *Springer Handbook of Auditory Research* (vol 69, pp. 211–243). Springer. [https://doi.org/10.1007/978-3-030-14832-4\\_8](https://doi.org/10.1007/978-3-030-14832-4_8)
- McAdams, S, & Goodchild, M (2017). Musical structure: Sound and timbre. In Ashley R & Timmers R (Eds.), *The Routledge companion to music cognition* (pp. 129–139). Routledge
- Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B (2019) Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci* 116(44):22071–22080
- Nan N, Guan X (2023) Common and distinct quantitative characteristics of Chinese and Western music in terms of modes, scales, degrees and melody variations. *J N. Music Res* 52(2-3):227–244
- Panda R, Malheiro R, & Paiva RP (2020) Audio features for music emotion recognition: a survey. *IEEE Transactions on Affective Computing* 14(1):68–88. <https://doi.org/10.1109/TAFFC.2020.3032373>
- Peeters G, Giordano BL, Susini P, Misdariis N, McAdams S (2011) The timbre toolbox: extracting audio descriptors from musical signals. *J Acoustical Soc Am* 130(5):2902–2916
- Rao, NY (2002). Hearing pentatonicism through serialism: integrating different traditions in Chinese contemporary music. *Perspectives of New Music*, 190–231. <https://www.jstor.org/stable/25164495>
- Reilly R (2017) Review of *Understanding emotion in Chinese culture: thinking through psychology*. *Humanist Psychol* 45(2):199–206
- Rocke, S, Davidson, JW, & Kiernan, F (2022). Emotion and Performance Practices. In McPherson G (Ed.), *The Oxford Handbook of Music Performance* (pp. 456–483). Oxford University Press
- Russell JA (1980) A circumplex model of affect. *J Personal Soc Psychol* 39(6):1161–1178
- Schubert E (2013) Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Front Psychol* 4:837
- Schutz M, Huron D, Keeton K, Loewer G (2008) The happy xylophone: acoustics affordances restrict an emotional palate. *Empir Musicol Rev* 3(3):126–135
- Shi, J (2015). The emotional model of Chinese folk music (Master dissertation). *East China Normal University*
- Slingerland E (2000) Effortless action: the Chinese spiritual ideal of wu-wei. *J Am Acad Relig* 68(2):293–327
- Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 14(4):323–348
- Trehub SE, Becker J, Morley I (2015) Cross-cultural perspectives on music and musicality. *Philos Trans R Soc B: Biol Sci* 370(1664):20140096
- Verellen F (1995) Taoism. *J Asian Stud* 54(2):322–346
- Wang X, Wang L, Xie L (2022) Comparison and analysis of acoustic features of Western and Chinese classical music emotion recognition based on VA model. *Appl Sci* 12(12):5787
- Webster GD, Weir CG (2005) Emotional responses to music: interactive effects of mode, texture, and tempo. *Motiv Emot* 29:19–39
- Wen X, Huang Z, Sun Z, Xu L (2022) What a deep song: the role of music features in perceived depth. *Psych J* 11(5):673–683
- Wu D, Jia X, Rao W, Dou W, Li Y, Li B (2024) Construction of a Chinese traditional instrumental music dataset: a validated set of naturalistic affective music excerpts. *Behav Res Methods* 56:3757–3778
- Xu, L, Xu, B, Sun, Z, & Li, H (2024). Associations between lyric and musical depth in Chinese songs: Evidence from computational modeling. *PsyCh Journal*, advance online publication. <https://doi.org/10.1002/pchj.785>
- Xu L, Sun Z, Wen X, Huang Z, Chao CJ, Xu L (2021) Using machine learning analysis to interpret the relationship between music emotion and lyric features. *PeerJ Comp Sci* 7:e785
- Xu L, Wen X, Shi J, Li S, Xiao Y, Wan Q, Qian X (2021) Effects of individual factors on perceived emotion and felt emotion of music: based on machine learning methods. *Psychol Music* 49(5):1069–1087
- Xu L, Xu M, Jiang Z, Wen X, Liu Y, Sun Z, Li H, Qian X (2023) How have music emotions been described in Google books? Historical trends and corpus differences. *Humanities Soc Sci Commun* 10:346
- Xu, L, Yun, Z, Sun, Z, Wen, X, Qin, X, & Qian, X (2022). PSIC3839: predicting the overall emotion and depth of entire songs. In *Design studies and intelligence engineering* (pp. 1–9). IOS Press. <https://doi.org/10.3233/FAIA220004>
- Zhang D, Zhang M, Liu D, Kang J (2016) Soundscape evaluation in Han Chinese Buddhist temples. *Appl Acoust* 111:188–197
- Zhang Y, Zhou Z, Sun M (2022) Influence of musical elements on the perception of “Chinese style” in music. *Cogn Comput Syst* 4(2):147–164
- Zentner M, Grandjean D, Scherer KR (2008) Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8(4):494–521

## Acknowledgements

This work was supported by Humanities and Social Sciences Youth Foundation, Ministry of Education of the People's Republic of China (grant number 24YJC190037), the National Natural Science Youth Foundation of China (grant number 32400904), and the Fundamental Research Funds for the Provincial Universities of Zhejiang (grant number GB202503006).

## Author contributions

Conceptualization: [Liang Xu], [Xiangming Wu]; Methodology: [Liang Xu], [Yishan Liu]; Formal analysis and investigation: [Liang Xu], [Yishan Liu]; Writing - original draft preparation: [Liang Xu], [Zehua Jiang]; Writing - review and editing: [Yulong Tang], [Xiangming Wu]; Funding acquisition: [Liang Xu], [Yulong Tang]; Supervision: [Xiangming Wu].

## Competing interests

The authors declare no competing interests.

## Ethical Approval

This study is a retrospective analysis based entirely on publicly available data derived from open online sources. It does not involve direct interaction with human participants, nor does it include any identifiable personal information. According to the institutional and national ethical guidelines, studies using fully anonymized and publicly accessible data do not require formal ethical approval. The Research Ethics Committee of College of Education, Zhejiang University of Technology has reviewed the study protocol and provided confirmation that no ethical approval is required. All procedures were conducted in accordance with relevant institutional and national regulations.

## Informed Consent

As the study analyzed publicly available and fully anonymized data, no direct participation, intervention, or collection of identifiable personal information was involved. Therefore, obtaining informed consent from individuals was not applicable. This

approach aligns with institutional and national ethical standards, which specify that research using de-identified, publicly accessible data is exempt from informed consent requirements.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-025-06211-x>.

**Correspondence** and requests for materials should be addressed to Liang Xu or Xiangming Wu.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025