Using machine learning analysis to interpret the relationship between music emotion and lyric features

Liang Xu [1], **Zaoyi Sun** [2], **Xin Wen** [1], **Zhengxi Huang** [1], **Chi-ju Chao** [3], **Liuchang Xu** Corresp. [4, 5]

[1] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China

[2] College of Education, Zhejiang University of Technology, Hangzhou, China

[3] Department of Information Art and Design, Tsinghua University, Beijing, China

[4] Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Zhejiang A&F University, Hangzhou, China

[5] College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou, China

Corresponding Author: Liuchang Xu
Email address: xuliuchang@zafu.edu.cn

Melody and lyrics, reflecting two unique human cognitive abilities, are usually combined in music to convey emotions. Although psychologists and computer scientists have made considerable progress in revealing the association between musical structure and the perceived emotions of music, the features of lyrics are relatively less discussed. Using linguistic inquiry and word count (LIWC) technology to extract lyric features in 2372 Chinese songs, this study investigated the effects of LIWC-based lyric features on the perceived arousal and valence of music. First, correlation analysis shows that, for example, the perceived arousal of music was positively correlated with the total number of lyric words and the mean number of words per sentence and was negatively correlated with the proportion of words related to the past and insight. The perceived valence of music was negatively correlated with the proportion of negative emotion words. Second, we used audio and lyric features as inputs to construct music emotion recognition (MER) models. The performance of random forest regressions reveals that, for the recognition models of perceived valence, adding lyric features can significantly improve the prediction effect of the model using audio features only; for the recognition models of perceived arousal, lyric features are almost useless. Finally, by calculating the feature importance to interpret the MER models, we observed that the audio features played a decisive role in the recognition models of both perceived arousal and perceived valence. Unlike the uselessness of the lyric features in the arousal recognition model, several lyric features, such as the usage frequency of words related to sadness, positive emotions, and tentativeness, played important roles in the valence recognition model.

Using Machine Learning Analysis to Interpret the Relationship between Music Emotion and Lyric Features

Liang Xu[1], Zaoyi Sun[2], Xin Wen[1], Zhengxi Huang[1], Chi-ju Chao[3], Liuchang Xu[4,5]

[1] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China
[2] College of Education, Zhejiang University of Technology, Hangzhou, China
[3] Department of Information Art and Design, Tsinghua University, Beijing, China
[4] Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Zhejiang A&F University, Hangzhou, China
[5] College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou, China.

Corresponding Author:
Liuchang Xu[1]
[1] No. 666, Wusu Rd., Lin'an District, Hangzhou, Zhejiang, 311300, China.
Email address: xuliuchang@zafu.edu.cn

Abstract

Melody and lyrics, reflecting two unique human cognitive abilities, are usually combined in music to convey emotions. Although psychologists and computer scientists have made considerable progress in revealing the association between musical structure and the perceived emotions of music, the features of lyrics are relatively less discussed. Using linguistic inquiry and word count (LIWC) technology to extract lyric features in 2372 Chinese songs, this study investigated the effects of LIWC-based lyric features on the perceived arousal and valence of music. First, correlation analysis shows that, for example, the perceived arousal of music was positively correlated with the total number of lyric words and the mean number of words per sentence and was negatively correlated with the proportion of words related to the past and insight. The perceived valence of music was negatively correlated with the proportion of negative emotion words. Second, we used audio and lyric features as inputs to construct music emotion recognition (MER) models. The performance of random forest regressions reveals that, for the recognition models of perceived valence, adding lyric features can significantly improve the prediction effect of the model using audio features only; for the recognition models of perceived arousal, lyric features are almost useless. Finally, by calculating the feature importance to interpret the MER models, we observed that the audio features played a decisive role in the recognition models of both perceived arousal and perceived valence. Unlike the uselessness of the lyric features in the arousal recognition model, several lyric features, such as the usage

39 frequency of words related to sadness, positive emotions, and tentativeness, played important
40 roles in the valence recognition model.
41

Introduction

43 The pursuit of emotional experience is a vital motivation for listening to music (Juslin &
44 Sloboda, 2001; Juslin & Laukka, 2004), and the ability to convey emotions ensures the important
45 role of music in human life (Yang, Dong, & Li, 2018). Therefore, the relationship between music
46 and perceived emotional expression has attracted increasing academic attention in recent decades
47 (Swaminathan & Schellenberg, 2015). Most of the related studies have focused on investigating
48 the association between musical structure and perceived emotions. For example, psychologists
49 have made considerable progress in revealing structural factors (e.g., tempo, pitch, and timbre),
50 indicating different emotional expressions (Gabrielsson, 2016), and computer scientists have
51 focused on extracting features from audio (audio most commonly refers to sound, as it is
52 transmitted in signal form; e.g., Mel-frequency cepstrum coefficients and Daubechies wavelet
53 coefficient histograms) to automatically identify music emotion (Yang, Dong, & Li, 2018).
54 Previous works have shown that sound features were highly correlated with music emotions
55 (Gabrielsson, 2016; Yang, Dong, & Li, 2018), but the lyric features have been relatively less
56 discussed. Besson et al. (1998) proved that melodic and lyrical components in music are
57 processed independently. Although melodic information may be more dominant than lyrics in
58 conveying emotions (Ali & Peyrincioğlu, 2006), investigating the relationship between lyrical
59 structure and the perceived emotion of music in detail is still necessary.
60

61 Music emotion studies related to lyrics have often focused on investigating the differences
62 between the presence and absence of lyrics (Ali & Peyrincioğlu, 2006; Brattico et al., 2011; Yu et
63 al., 2019), the effects of consistency or differences in melodic and lyrical information (Morton &
64 Trehub, 2001; Vidas et al., 2019), or the effects of lyrics with different meanings (Batcho et al.,
65 2008; Stratton & Zalanowski, 1994). While lyric structures and features have been rarely studied,
66 melodic information has been processed in previous psychology studies (Gabrielsson, 2016;
67 Swaminathan & Schellenberg, 2015; Xu, Wen et al., 2020). On the other hand, with the
68 development of natural language processing (NLP) technology, different lyric features have been
69 widely extracted and analyzed in music emotion recognition (MER) studies (e.g., Malheiro et al.,
70 2016; Delbouys et al., 2018), a field investigating computational models for detecting music
71 emotion (Aljanaki, Yang, & Soleymani, 2017; Chen, Lee et al., 2015). These MER studies have
72 typically focused on improving the prediction effect of constructed models but have not
73 interpreted the model and variables. Thus, can the structural factors of lyrics be analyzed in more
74 detail by combining NLP technology? If so, this may facilitate the understanding of the
75 relationship between lyrics and perceived emotions. Therefore, the present study investigated the
76 effects of various lyric features on the perceived emotions in music.
77

Related work

79	**Emotion Perception in Music with Lyrics**
80	Melody and lyrics, reflecting two unique human cognitive abilities, are usually combined in
81	music to convey various information (Gordon et al., 2010). A melody is a linear succession of
82	musical tones that the listener perceives as a single entity (van Waesberghe, 1955); and lyrics is
83	the composition in verse which is sung to a melody to constitute a song. Besson et al. (1998) and
84	Bonnel et al. (2001) have shown that the melodic and lyrical components are processed
85	independently. However, they are often integrated in such a way that melodic information is
86	enhanced by the simultaneous presence of lyrics (Serafine, Crowder, & Repp, 1984; Serafine et
87	al., 1986). Studies have been continuously updated to investigate the interaction or independence
88	of melody and lyrics by using melody only (Bonnel et al., 2001; Brattico te al., 2011; Kolinsky et
89	al., 2009), lyrics only (Fedorenko et al., 2009; Poulin-Charronnat et al., 2005; Yu, Wu et al.,
90	2019), or both (Gordon et al., 2010; van Besouw, Howard, & Ternström, 2005). The role of
91	melody and lyrics in music emotion perception is a vital research focus.
92	
93	As the soul of music, perceived emotion has been widely discussed in recent decades
94	(Swaminathan & Schellenberg, 2015). Ali and Peynircioğlu (2006) investigated differences in
95	melodies and lyrics conveying the same and mismatched emotions and confirmed the dominance
96	of melody in music emotional information processing. Additionally, they observed that lyrics can
97	strengthen the perception of negative emotions but weaken the perceived positive emotions. A
98	computational study (Hu, Downie, & Ehmann, 2009) found that negative emotion classification
99	accuracy was improved by adding lyric information, while the opposite effect was obtained for
100	the classification of positive emotions. In contrast, the results of Laurier, Grivolla, and Herrera
101	(2008) showed that lyrics can facilitate the recognition of happy and sad musical emotions but
102	not angry and violent emotions. Explanatory studies were also conducted instantly following the
103	observed phenomena. Pieschl and Fegers (2016) advocated using short-term effects on cognition
104	and affect to explain the power of lyrics. By comparing music with and without lyrics, evidence
105	from functional magnetic resonance imaging also indicated the importance of lyrics for negative
106	musical emotions (Brattico et al., 2011). Following the work of Brattico et al. (2011), neural
107	mechanisms have been continually studied in recent years (e.g., Greer et al., 2019; Proverbio, De
108	Benedetto, & Guazzone, 2020). In sum, although subtle conflicts exist in different studies, the
109	substantial role of lyrics in music emotion perception is consistent.
110	
111	The aforementioned theoretical findings have also been supplemented or utilized in other fields.
112	For instance, developmental psychology studies have proven that lyrical information dominates
113	children's judgment of music emotion (Morton & Trehub, 2001; Morton & Trehub, 2007), but
114	adults rely on melody (Ali & Peynircioğlu, 2006; Vidas et al., 2019); music therapy studies have
115	widely conducted lyric analyses to extend the understanding of clients' emotional and health
116	states (Baker et al., 2009; O'Callaghan & Grocke, 2009; Silverman, 2020; Viega & Baker,
117	2016); and in computational studies, lyrical information used as additional inputs can
118	significantly improve the predictive effects of MER models (Laurier, Grivolla, & Herrera, 2008;

119   Malheiro et al., 2018; Yu, Tang et al., 2019). These studies provide a development direction and
120   practical value of basic lyrics research and encourage the optimization of basic research.
121
122   One of the limitations in previous behavioral research is that the lyric was always treated as a
123   complete object. Studies usually investigated the differences between the presence and absence
124   of lyrics (Ali & Peyrincioğlu, 2006; Yu, Wu et al, 2019) or the effects of lyrics with a certain
125   meaning (e.g., lyrics expressing homesickness in Batcho et al., 2008, or happy and sad lyrics in
126   Brattico et al., 2011) but rarely analyzed the elements extracted from lyrics. In melody-related
127   research, various musical structural factors (e.g., mode, timbre, tempo, and pitch) have been
128   studied (Juslin & Laukka, 2004), and the association between perceived emotion and structural
129   factors has been repeatedly verified in recent decades (Gabrielsson, 2016; Hunter &
130   Schellenberg, 2010). Therefore, to better understand the relationship between lyrics and music
131   perception emotions, can similar methods be used to analyze the structural factors of lyrics? In
132   addition, unlike musical structural factors, which have been summarized in musicology,
133   analyzing lyrical factor types remains a challenge. We noticed that recent linguistic and NLP-
134   based computational studies can provide inspiration.
135
136   **NLP-based Lyric Features**
137   NLP technology has been widely used to analyze the emotions expressed or perceived in texts,
138   such as book reviews (Zhang, Tong, & Bu, 2019), opinions on social media platforms (Xu, Li et
139   al., 2020), movie reviews (Kaur & Verma, 2017; Lu & Wu, 2019), party statements (Haselmayer
140   & Jenny, 2017), and song lyrics (Rachman, Samo, & Fatichah, 2019). Knowledge-based
141   approaches and machine learning-based approaches are two common approaches used for
142   emotional analysis (Liu & Chen, 2015). The knowledge-based approach is an unsupervised
143   approach that uses an emotional dictionary or lexicon (a list of words or expressions used to
144   express human emotions) to label emotional words in text (Liu, 2012). Thus, a high-quality
145   emotional dictionary is the basis of this approach. In contrast, the machine learning approach is
146   usually a supervised approach that requires a labeled dataset to construct emotion recognition
147   models (Peng, Cambria, & Hussain 2017). It usually constitutes a process of (a) extracting text
148   features (including lexical features, syntactic features and semantic features), (b) using machine
149   learning algorithms to construct the relationship between extracted features and labeled
150   emotions, and (c) predicting the emotions of untagged texts.
151
152   When conducting emotion analyses of song lyrics, machine learning-based approaches have been
153   more prevalent in the past two decades. Laurier, Grivolla, and Herrera (2008) used lyric feature
154   vectors based on latent semantic analysis (LSA) dimensional reduction and audio features to
155   conduct music mood classification. They found that standard distance-based methods and LSA
156   were effective for lyric feature extraction, although the performance of lyric features was inferior
157   to that of audio features. Petrie, Pennebaker, and Sivertsen (2008) conducted linguistic inquiry
158   and word count (LIWC) analyses to explore the emotional changes in Beatles' lyrics over time.

159 Panda et al. (2013) used support vector machines, K-nearest neighbors, and naïve Bayes
160 algorithms to map the relationship between music emotion and extracted audio and lyric features.
161 In a recent study, Zhou, Chen, and Yang (2019) applied unsupervised deep neural networks to
162 perform feature learning and found that this method performed well for audio and lyric data and
163 could model the relationships between features and music emotions effectively. Notably,
164 traditional MER research focuses on improving the prediction effect of the MER models, while
165 our study attempted to use an interpretable way to investigate the relationship between lyrics
166 features and music emotions.
167
168 Previous studies have shown a variety of methods for extracting lyric features. Although the lyric
169 feature vectors and the deep learning-based features performed well in MER studies, the
170 meaning of these features is often difficult to understand. Thus, considering the interpretability
171 of lyric features, this study selected the LIWC-based method to extract lyric features. The LIWC
172 software package was first developed to analyze text for more than 70 language dimensions by
173 Pennebaker, Francis, and Booth (2001). It has been applied for text analysis in psychological
174 health (Slatcher & Pennebaker, 2006), physical health (Pennebaker, 2004), and lyric studies
175 (Petrie, Pennebaker, & Sivertsen, 2008; Pettijohn & Sacco Jr, 2009). The simplified Chinese
176 version of LIWC (SC-LIWC; Gao et al., 2013; Zhao et al., 2016), which expanded text features
177 for more than 100 dimensions, was also developed in recent years. This technology was
178 considered for lyric feature extraction in this study.
179
180 **The Present Research**
181 In sum, the present study investigates the association between LIWC-based lyric features and the
182 perceived emotions of Chinese songs. First, the direct relationships between the independent
183 variables (lyric features) and the dependent variables (perceived emotions of music) are
184 investigated through correlation analysis. Then, a computational modeling method is considered
185 to examine the effects of lyric features on music emotion perception. Since melody and lyrics are
186 inseparable in music, we use the audio and lyric features extracted in music to predict the
187 perceived emotions. By comparing the prediction effects of the models that use lyric features as
188 input and that lack lyric features, we can intuitively witness the effect of lyrics. Moreover, using
189 interpretable and nonlinear machine learning methods to construct prediction models, different
190 forms of association between lyrics and perceived emotions can be observed (Vempala & Russo,
191 2018; Xu, Wen et al., 2020). Finally, the constructed MER models are also of practical value
192 because the recognized music emotion information can be used in various fields, such as music
193 recommendation (Deng et al., 2015), music information retrieval (Downie, 2008), and music
194 therapy (Dingle et al., 2015).
195

196 # Materials & Methods
197 **Data Collection**

198   The original music files and emotion annotation results were obtained from the PSIC3839
199   dataset, a public-free dataset for MER studies (Xu, Yun et al., 2020, unpublished data). In this
200   dataset, arousal and valence scores of 3839 songs popular in China were manually annotated by
201   87 university students using 5-point Likert scales. Based on the multi-dimensional emotion space
202   model, Lang (1995) suggested that emotions can be categorized in a two-dimensional space by
203   valence and arousal; valence ranges from negative to positive, and arousal ranges from passive
204   (low) to active (high). In the PSIC3839 dataset, valence was evaluated from -2 (*negative*) to 2
205   (*positive*), and arousal was evaluated from -2 (*not at all*) to 2 (*very much*). We then downloaded
206   the lyrics of songs from NetEase Cloud Music (https://music.163.com/), a popular music site in
207   China. Considering that the annotators of the PSIC3839 dataset are all native Chinese, only 2372
208   songs with Chinese lyrics were retained for subsequent analysis.
209
210   **Lyric Feature Extraction**
211   To extract the lyric features, the raw data need to be preprocessed. First, since the raw data of
212   lyrics were downloaded online, they contained a large amount of unwanted information, such as
213   singer, composer, and lyricist names and the words "man" and "woman" in duet songs. Thus, we
214   manually filtered these unwanted information elements. Second, unlike English texts that are
215   directly composed of separated words, Chinese texts require special tools to divide them into
216   separate words for analysis. For example, the sentence "he feels happy" in English text is
217   "hefeelshappy" in Chinese text. Therefore, this study used the Chinese word segmentation tool in
218   the Language Technology Platform (Che, Li, & Liu, 2010) for text segmentation. Through the
219   above steps, the raw lyrics of each song were processed into Chinese words arranged in order.
220
221   After the above data preprocessing, we used SC-LIWC (Gao et al., 2013; Zhao et al., 2016) to
222   extract the lyric features. A total of 98 types of lyric features were calculated for each song, such
223   as the total number of words (*WordCount*), the proportion of positive emotion words (*PosEmo*),
224   and the proportion of swear words (*Swear*). For example, the lyric feature *PosEmo*, reflecting the
225   usage frequency of positive emotion words in each song, is calculated by dividing the number of
226   positive emotion words in SC-LIWC by the total number of words. All the extracted features are
227   listed and introduced in Supplemental Materials Table S1
228
229   **Audio Feature Extraction**
230   For audio features, this study considered both rhythmic features (by beat and tempo detection)
231   and spectral features related to timbre, pitch, harmony, and so forth (e.g., Mel-frequency
232   cepstrum coefficients, MFCCs). Audio signal preprocessing was first conducted by (a) using a
233   22050 Hz sampling rate to sample each song and (b) using a short-term Fourier transform to
234   obtain the power spectrogram. Then, using the librosa toolkit (McFee et al., 2015), a total of nine
235   low or middle features were extracted, including MFCCs, spectral centroid, spectral bandwidth,
236   spectral roll-off, spectral flatness, spectral contrast, tonal centroid features (tonnetz),
237   chromagram, and tempo. Different spectral features were calculated in different ways. For

238   instance, MFCCs were calculated by selecting the lower cepstral coefficients of a Mel-scaled
239   spectrogram, which was generated by using a Mel filter bank to filter the spectrogram (Meyer &
240   Kollmeier, 2009).
241
242   Since the extracted features of each song were represented in a subspace of high dimensionality,
243   we conducted feature reduction of each type of feature to reduce the storage and computational
244   space. Principal component analysis (PCA), widely used in MER studies (Xu, Wen et al., 2020;
245   Yang, Dong, & Li, 2018), was applied to reduce the dimensionality of features. After conducting
246   the PCA, we selected and combined the top 50 dimensions of each type of audio feature as
247   model inputs. Finally, we scaled each continuous audio feature to a value of 0 to 1 via min-max
248   scaling (Kahng, Mantik, & Markov, 2002). The above processed features were used as the final
249   audio inputs of the computational models.
250

251   **Construction of Computational Models**
252   The proposed modeling method is shown in Figure 1. Since the annotation results of the
253   perceived valence and arousal values are continuous variables, this study formulated the
254   construction of computational models as a regression problem, which predicts a real value from
255   observed features (Sen & Srivastava, 2012). We used the audio and lyric features as model
256   inputs and the perceived emotion values as ground truth. To explore the effects of lyric features,
257   three types of input sets were considered: (a) audio features only; (b) lyric features only; and (c)
258   combining audio and lyric features. In addition, the ground truth values were also scaled to a
259   value of 0 to 1 via min-max scaling before modeling.
260

261   Two machine learning algorithms were then considered to map the inputs and perceived emotion
262   values (ground truth data). Multiple linear regression (MLR) was used as the baseline algorithm.
263   Random forest regression (RFR), showing good performance in MER tasks (e.g., Xu, Wen et al.,
264   2020; Xu et al., 2021), was used as the main algorithm. For each RFR model, we used a grid
265   parameter search to obtain the best modeling parameters. The performances of our models were
266   evaluated by the tenfold cross-validation technique. The prediction accuracy of each regressor
267   was measured by R^2 statistics and the root mean-squared error (RMSE) as follows:
268

269
$$R^2 = 1 - \frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{\sum_{i=1}^{N}(X_i - \overline{X})^2}$$

270

271
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2}$$

272   where X_i is the perceived emotion value (ground truth) of each song, $\overline{X}$ is the mean value of the
273   perceived emotion values, Y_i is the predicted result of each song, and N is the number of testing
274   samples in the tenfold cross-validation technique.

Results

Data Distribution

277   As the first step in exploring the data, we created a preliminary description of the emotional
278   annotation results of all songs and the content of the lyrics. Figure 2a shows the distribution of
279   annotated emotions of songs in the valence-arousal emotion space. We observed that a large
280   proportion of the songs fell under the third quadrant (37.94%; low arousal and negative),
281   followed by the first (34.49%; high arousal and positive) and fourth quadrants (17.07%; high
282   arousal and negative). Pearson correlation analysis shows that the perceived arousal values are
283   positively correlated with valence ($r(2371) = 0.537, p < .001$).

285   We then calculated the usage frequency of different words in song sets with different emotions.
286   Excluding commonly used personal pronouns (e.g., "I" and "you") and verbs (e.g., "is" and
287   "are"), the top used words are presented in the word clouds (see Figure 2b). We observed that the
288   words "love" and "world" frequently appear in every quadrant of the valence-arousal emotion
289   space, meaning that these are two popular song themes. The words "happy" and "beautiful"
290   frequently appear in positive songs, whereas the words "lonely" and "recall" frequently appear in
291   negative songs. The above results allow us to intuitively see the difference in word usage in
292   different emotional songs.

Correlation Analysis between Perceived Emotions and Lyric Features

295   In this part, we analyzed how well the independent variables (lyric features) accounted for the
296   dependent variables (perceived arousal and valence values). The lyric features most relevant to
297   arousal and valence are shown in Table 1 and Table 2, respectively. For instance, using Pearson
298   correlation analysis, we found that the perceived arousal values were positively correlated with
299   the total number of words in songs (*WordCount*, $r(2371) = 0.206, p < .01$), the mean number of
300   words per sentence (*WordPerSentence*, $r(2371) = 0.179, p < .01$), the ratio of Latin words
301   (*RateLatinWord*, $r(2371) = 0.183, p < .01$), and the proportion of words related to achievement
302   (*Achieve*, $r(2371) = 0.111, p < .01$) and were negatively correlated with the proportion of words
303   related to the past (*tPast*, $r(2371) = -0.124, p < .01$) and the proportion of words related to insight
304   (*Insight*, $r(2371) = -0.122, p < .01$). For valence, we observed that the perceived valence values
305   were negatively correlated with the proportion of negative emotion words (*NegEmo*, $r(2371) = -0.364, p < .01$) and proportion of words related to sadness (*Sad*, $r(2371) = -0.299, p < .01$). The
307   entire correlation results are presented in Supplemental Materials TabelS2. The correlation
308   results only reveal the linear relationships between perceived emotions and lyric features. Thus,
309   we then used machine learning methods to investigate other types of relationships.

Model Prediction Results

312   In this section, we used MLR and RFR algorithms, different input sets (audio features, lyric
313   features, and combined features), and different ground truth data (arousal and valence) to
314   construct MER models. To obtain relatively good models, a grid parameter search was first
315   conducted to obtain the best performing parameters for each RFR model (results are shown in
316   Table 3).
317
318   After parameter searching, the performances of the constructed models, evaluated by the tenfold
319   cross-validation technique, are presented in Figure 3. Since the tenfold cross-validation
320   technique, using 10% data as the testing data and using the remaining 90% instances as training
321   data to train regressor, used the same folds in the evaluation of different models, a tenfold paired
322   sample t-test can be applied to compare the model results. For the algorithms, the RFR algorithm
323   performed better than the MLR algorithm in all the constructed models. For example, ten-fold
324   paired sample t-test showed that, using combined features to predict perceived arousal values,
325   the RFR-based model reached a mean R^2 value of 0.631 and a mean RMSE value of 0.147,
326   significantly better than the MLR-based model ($R^2 = 0.532$, t(9) = 4.206, $p < .01$, $d = 1.883$;
327   RMSE = 0.165, t(9) = -4.012, $p < .01$, $d = -1.960$). Therefore, the subsequent analysis was only
328   conducted on the models based on the RFR algorithm.
329
330   For the recognition models of perceived arousal values, paired sample t-test showed that the
331   model using audio features as inputs performed significantly better than the model using lyric
332   features (R^2: t(9) = 36.335, $p < .001$, $d = 15.219$; RMSE: t(9) = -34.693, $p < .001$, $d = -12.167$).
333   Although the model using combined features ($R^2 = 0.631$, RMSE = 0.147) performed slightly
334   better than the model using audio features ($R^2 = 0.629$, RMSE = 0.147), there was no significant
335   effect (R^2: t(9) = 0.134, $p = .896$, $d = 0.063$; RMSE: t(9) = -0.231, $p = .822$, $d = -0.114$). These
336   results revealed that the perceived arousal of music mainly depends on audio information, while
337   lyrics are not important.
338
339   For valence, paired sample t-test showed that, although the best performing recognition model of
340   perceived valence values performed worse than the arousal model (R^2: t(9) = -6.700, $p < .001$, d
341   = -3.271; RMSE: t(9) = 11.539, $p < .001$, $d = 6.094$), both lyric and audio features played
342   important roles. The RFR-based model using audio features only reached a mean R^2 value of
343   0.371 and a mean RMSE value of 0.214, and the model using lyric features as inputs reached a
344   mean R^2 value of 0.370 and a mean RMSE value of 0.214. When the audio and lyric features
345   were combined as inputs, the new model achieved a mean R^2 value of 0.481 and a mean RMSE
346   value of 0.194, significantly better than the previous two models. These results indicated that the
347   perceived valence of music was influenced by both audio and lyric information.
348
349   **Model Interpretability**
350   In the last step, we attempted to explain the best performing RFR-based models by examining
351   the information gain of features. Quiroz, Geangu, & Yong (2018) noted that models constructed

352   using the RFR algorithm can be interpreted by calculating feature importance. Thus, the feature
353   importance of the best performing recognition models of the perceived arousal and valence was
354   calculated and is presented in Figure 4. Since tenfold cross-validation was used to evaluate the
355   models, the coefficients of feature importance might differ when predicting different test sets
356   (Xu, Wen et al., 2020). Thus, the distribution of feature importance was arranged in descending
357   order of the mean value, and only the top 30 features were included for visibility.
358
359   In the arousal recognition model, the audio features played a decisive role, which accounted for
360   95.01% of the model. The first PCA components of spectral flatness, spectral contrast,
361   chromagram, MFCCs, and spectral bandwidth are the five most contributing features, accounting
362   for 30.71%, 12.38%, 8.77%, 8.62%, and 4.26%, respectively. For the lyric features, the feature
363   importance results are similar to the results of the correlation analysis. The total number of words
364   in songs (*WordCount*) and the mean number of words per sentence (*WordPerSentence*) were the
365   top two contributing lyric features, accounting for 0.19% and 0.15% of the model.
366
367   For valence, the audio features explained 73.44% of the model. The first PCA components of
368   spectral contrast and spectral flatness also showed good predictive effects on the perceived
369   valence, accounting for 24.82% and 10.42% of the model, respectively. The proportion of
370   negative emotion words (*NegEmo*) was the most important lyric feature (accounting for 5.32%),
371   followed by the proportion of words related to sadness (*Sad*, 0.56%), positive emotions
372   (*PosEmo*, 0.39%), tentativeness (*Tentat*, 0.23%), and so on. These findings also support the
373   opinion that lyric features can provide more information for recognition models of valence than
374   for recognition models of arousal.
375

Discussion

377   This study investigated the effects of LIWC-based lyric features on the perceived arousal and
378   valence of music. We first explored the data distribution, and several interesting results were
379   found. First, the emotional distribution (in the valence-arousal emotion space) of music in this
380   study is similar to previous works. Various studies have shown that the perceived valence and
381   arousal of music were positively correlated (e.g., Chen, Yang et al., 2015; Greenberg et al., 2016;
382   Speck et al., 2011). This reveals that the relationship between valence and arousal in music is
383   relatively constant. Second, analyzing the songs' word usage frequency in different quadrants of
384   the valence-arousal emotion space, we then found that the word "love" frequently appears in
385   each quadrant. The cross-cultural study of Freeman (2012) has shown that "romantic love" is the
386   top topic category of Chinese-language pop songs (82.5%), while Western pop songs with the
387   topic of "romantic love" accounted for only 40%. The high usage frequency of the word "love"
388   in this study further confirms that love-themed songs are the mainstream in Chinese pop music.
389   In addition, we observed that the words "happy" and "beautiful" frequently appear in positive
390   songs, whereas the words "lonely" and "recall" frequently appear in negative songs. This

391   intuitive phenomenon shows that, in general, perceived music emotions are related to lyrics,
392   which encourages us to further explore lyric features.
393
394   Although LIWC-based lyric features have been considered in computational studies to improve
395   the model effect (e.g., Hu, Chen, & Yang, 2009; Malheiro et al., 2016), the role of LIWC-based
396   lyric features has never been analyzed and discussed individually. Thus, this study then
397   investigated the linear relationship between each lyric feature and the perceived emotion of
398   music. In general, valence is more correlated with the features reflecting the meaning of the lyric
399   text, while arousal is more correlated with the features reflecting the structure of the lyrics. For
400   example, we observed that the perceived valence values were negatively correlated with the
401   usage frequency of words related to negative emotions (e.g., "sad", "horrible", "angry", and
402   "bitter"), tentativeness (e.g., "seem", "dim", "guess", and "dubitation"), insight (e.g.,
403   "understanding", "notice", "analyze", and "know"), and exclusiveness (e.g., "exclude", "forget",
404   "ignore", and "cancel"). It is obvious that the usage of emotional words is related to perceived
405   emotions because emotional words shape emotional percepts (Gendron et al., 2012). Words
406   related to tentativeness are often used in sad love songs to express doubts about love (e.g., "In
407   the next few days, I guess you won't show up either"), and insight words are often used with
408   negative words to portray the sad atmosphere (e.g., "no one notices me, only a small raindrop
409   accompanies me to wait for dawn"). We believe that some words in lyrics are often used to
410   describe certain behaviors, feelings or scenes, which are related to negative emotions. For
411   instance, nostalgia, characterized by sadness, insomnia, loss of appetite, pessimism, and anxiety
412   (Batcho et al., 2008; Havlena & Holak, 1991), is one of the themes of songs. Thus, words that
413   describe nostalgia in lyrics are often related to negative emotions. This phenomenon may appear
414   in songs with various themes, such as farewell, war, and tragic love stories.
415
416   For arousal, we observed that the perceived arousal values were positively correlated with the
417   total number of lyric words (*WordCount*) and the mean number of words per sentence
418   (*WordPerSentence*). We wonder that the size of music may play an important role, because the
419   duration of music is positively correlated with the total number of lyric words ($r(2371) = 0.219$, p
420   $< .01$); that is, the longer the music, the more words are in the lyrics. Unfortunately, as shown in
421   Supplemental Materials Figure S1, the duration of music was negatively correlated with the
422   arousal values ($r(2371) = -0.111$, $p < .01$). In addition, Holbrook and Anand (1990) found that
423   the tempo of music is positively correlated with listeners' perceived arousal. Thus, another
424   assumption is that fast-paced songs tend to match more lyric words. Unfortunately, the above
425   hypothesis was not confirmed in the current data set ($r(2371) = -0.039$, $p = .058$). This result
426   reminds us to analyze the relationship between audio features and the number of words. We
427   found that the total number of lyric words was positively correlated with chromagram and
428   negatively correlated with MFCCs (see Supplemental Materials Table S3). Chromagrams reflect
429   the pitch components of music over a short time interval (Schmidt, Turnbull, & Kim, 2010).
430   Previous work on screams has found a significant tendency to perceive higher-pitched screams

431   as more emotionally arousing than lower-pitched screams (Schwartz & Gouzoules, 2019).
432   However, whether the above phenomenon holds in music is still unknown, and it is worthy of
433   further research. While MFCCs reflect the nonlinear frequency sensitivity of the human auditory
434   system (Wang et al., 2012), it is difficult to map well-known music features in conventional
435   musical writing. The low-level audio features in this study may not directly explain the
436   relationship between melody features and lyrics. In fact, McVicar, Freeman, and De Bie (2011)
437   found it hard to interpret the correlations between arousal and lyric features. Therefore, how to
438   map the relationships among arousal, melody, and lyrics still needs further investigation.
439

440   The above correlation analysis reflects the direct connection between lyric features and perceived
441   emotions. We then used audio features and lyric features to construct MER models. By
442   comparing the results of models and calculating feature importance to interpret the constructed
443   models, we investigated the role of lyric features and obtained two major discoveries. First, we
444   found that, compared with lyric features, audio features played a decisive role in the MER
445   models for both perceived arousal and perceived valence. From the perspective of computational
446   modeling, this finding confirms previous conclusions that melodic information may be more
447   dominant than lyrics in conveying emotions (Ali & Peynircioğlu, 2006). However, previous
448   works used individual moods affected by music to evaluate the ability of music to convey
449   emotions (Ali & Peynircioğlu, 2006; Sousou, 1997), which is not equivalent to the perceived
450   emotions of music. Thus, our study provided more direct evidence that melody information plays
451   a decisive role in the perception of music emotions, and we believe that this result can be
452   generalized to all countries. The second major finding was that, unlike the uselessness of the
453   lyric features in the arousal recognition model, lyric features can significantly improve the
454   prediction effect of the valence recognition model. Feature importance analysis also shows that
455   lyric features, such as the proportion of words related to sadness (*Sad*), positive emotions
456   (*PosEmo*), and tentativeness (*Tentat*), played important roles in the valence recognition model.
457   This finding was consistent with that of Hu, Downie, & Ehmann (2009), which showed that
458   lyrics can express the valence dimension of emotion but usually do not express much about the
459   arousal dimension of emotion, rather than the opposite finding shown by Malheiro et al. (2016).
460   We hypothesize that the main reason for the difference in results is that our study and the study
461   of Hu, Downie, & Ehmann (2009) both focused on Chinese music and participants, but the study
462   of Malheiro et al. (2016) was conducted in Portugal. Cross-cultural studies have shown that
463   although listeners are similarly sensitive to musically expressed emotion (which is facilitated by
464   psychophysical cues; Argstatter, 2016; Balkwill & Thompson, 1999), differences still exist
465   (Zacharopoulou & Kyriakidou, 2009). Therefore, we believe that in the Chinese environment,
466   perceived music valence is affected by lyrics, although its influence is not as strong as that of
467   melody information.
468

469   As mentioned before, this study is also of practical value. The computational modeling method
470   was first proposed in the field of MER, which aims to automatically recognize the perceptual

471   emotion of music (Yang, Dong, & Li, 2018). There are many existing songs, but it is difficult for
472   people to manually annotate all emotional information. Thus, MER technology is urgently
473   needed and has made great progress in the past two decades. Recognized emotion information
474   can be used in various application scenarios (Deng et al., 2015; Dingle et al., 2015; Downie,
475   2008). The collected data and proposed methods in this study can also provide references for
476   future MER research. Notably, the computational modeling methods in MER studies pursue
477   model effects and prediction accuracy, but when they are applied in music psychology research,
478   the interpretability of the model should be taken into account (Vempala & Russo, 2018; Xu, Wen
479   et al., 2020). Therefore, we chose MLR and RFR to construct MER models. How to integrate
480   machine learning methods into music psychology research more effectively still needs more
481   exploration.
482

483   **Conclusions**
484   The present work investigated the effects of LIWC-based lyric features on the perceived arousal
485   and valence of music by analyzing 2372 Chinese songs. Correlation analysis shows that, for
486   example, the perceived arousal of music was positively correlated with the total number of lyric
487   words (*WordCount*, $r(2371) = 0.206$, $p < .01$) and the mean number of words per sentence
488   (*WordPerSentence*, $r(2371) = 0.179$, $p < .01$) and was negatively correlated with the proportion of
489   words related to the past (*tPast*, $r(2371) = -0.124$, $p < .01$) and insight (*Insight*, $r(2371) = -0.122$,
490    $p < .01$). The perceived valence of music was negatively correlated with the proportion of
491   negative emotion words (*NegEmo*, $r(2371) = -0.364$, $p < .01$) and the proportion of words related
492   to sadness (*Sad*, $r(2371) = -0.299$, $p < .01$). We then used audio and lyric features as inputs to
493   construct MER models. The performance of RFR-based models shows that, for the recognition
494   models of perceived valence, adding lyric features can significantly improve the prediction effect
495   of the model using audio features only; for the recognition models of perceived arousal, lyric
496   features are almost useless. Calculating the importance of features to interpret the MER models,
497   we observed that the audio features played a decisive role in the recognition models of both
498   perceived arousal and perceived valence. Unlike the uselessness of the lyric features in the
499   arousal recognition model, several lyric features, such as the proportion of words related to
500   sadness (*Sad*), positive emotions (*PosEmo*), and tentativeness (*Tentat*), played important roles in
501   the valence recognition model.
502

503   **References**
504   Ali SO, Peynircioğlu ZF. 2006. Songs and emotions: are lyrics and melodies equal partners?
505   *Psychology of Music* 34:511–534 DOI: 10.1177/0305735606067168.
506   Aljanaki A, Yang YH, Soleymani M. 2017. Developing a benchmark for emotional analysis of
507   music. *PLoS ONE* 12:e0173392 DOI: 10.1371/journal.pone.0173392.
508   Argstatter H. 2016. Perception of basic emotions in music: Culture-specific or multicultural?
509   *Psychology of Music* 44:674–690 DOI:10.1177/0305735615589214.

510 Balkwill LL, Thompson WF. 1999. A Cross-Cultural Investigation of the Perception of Emotion
511   in Music: Psychophysical and Cultural Cues. *Music Perception* 17:43–64
512   DOI:10.2307/40285811.
513 Batcho KI, DaRin ML, Nave AM, Yaworsky RR. 2008. Nostalgia and identity in song lyrics.
514   *Psychology of Aesthetics, Creativity, and the Arts*. 2:236–244 DOI:10.1037/1931-
515   3896.2.4.236.
516 Baker F, Wigram T, Stott D, McFerran K. 2009. Therapeutic songwriting in music therapy, Part
517   II: Comparing the literature with practice across diverse clinical populations. *Nordic*
518   *Journal of Music Therapy* 18:32–56 DOI: 10.1080/08098130802496373.
519 Besson M, Faïta F, Peretz I, Bonnel AM, Requin J. 1998. Singing in the Brain: Independence of
520   Lyrics and Tunes. *Psychological Science* 9:494–498 DOI: 10.1111/1467-9280.00091.
521 Bonnel AM, Faita F, Peretz I, Besson M. 2001. Divided attention between lyrics and tunes of
522   operatic songs: Evidence for independent processing. *Perception & psychophysics*.
523   63(7):1201–1213 DOI: 10.3758/BF03194534.
524 Brattico E, Alluri V, Bogert B, Jacobsen T, Vartiainen N, Nieminen S, Tervaniemi M. 2011. A
525   Functional MRI Study of Happy and Sad Emotions in Music with and without Lyrics.
526   *Frontiers in Psychology* 2. DOI: 10.3389/fpsyg.2011.00308.
527 Che W, Li Z, Liu T. 2010. LTP: a Chinese Language Technology Platform. In: *Proceedings of*
528   *the 23rd International Conference on Computational Linguistics: Demonstrations* 13-16.
529 Chen S-H, Lee Y-S, Hsieh W-C, Wang J-C. 2015. Music emotion recognition using deep
530   Gaussian process. In: *2015 Asia-Pacific Signal and Information Processing Association*
531   *Annual Summit and Conference (APSIPA)*. Hong Kong: IEEE, 495–498. DOI:
532   10.1109/APSIPA.2015.7415321.
533 Chen Y-A, Yang Y-H, Wang J-C, Chen H. 2015. The AMG1608 dataset for music emotion
534   recognition. In: *2015 IEEE International Conference on Acoustics, Speech and Signal*
535   *Processing (ICASSP)*. South Brisbane, Queensland, Australia: IEEE, 693–697. DOI:
536   10.1109/ICASSP.2015.7178058.
537 Delbouys R, Hennequin R, Piccoli F, Royo-Letelier J, Moussallam M. 2018. Music mood
538   detection based on audio and lyrics with deep neural net. In: *Proceedings of the 19th*
539   *International Society for Music Information Retrieval Conference (ISMIR)*, 370-375.
540 Deng JJ, Leung CHC, Milani A, Chen L. 2015. Emotional States Associated with Music:
541   Classification, Prediction of Changes, and Consideration in Recommendation. *ACM*
542   *Transactions on Interactive Intelligent Systems* 5:1–36. DOI: 10.1145/2723575.
543 Dingle GA, Kelly PJ, Flynn LM, Baker FA. 2015. The influence of music on emotions and
544   cravings in clients in addiction treatment: A study of two clinical samples. *The Arts in*
545   *Psychotherapy* 45:18–25. DOI: 10.1016/j.aip.2015.05.005.
546 Downie JS. 2008. The music information retrieval evaluation exchange (2005-2007): A window
547   into music information retrieval research. *Acoustical Science and Technology* 29:247–255.
548   DOI: 10.1250/ast.29.247.

549 Fedorenko E, Patel A, Casasanto D, Winawer J, Gibson E. 2009. Structural integration in
550   language and music: Evidence for a shared system. *Memory & Cognition* 37:1–9. DOI:
551   10.3758/MC.37.1.1.
552 Freeman BC. 2012. No Eminems in Asia: What's in a song? Analyzing lyrics of Top 40 songs in
553   the East & West. In: *3rd Inter-Asia Popular Music Studies (IAPMS)*. Available at
554   https://www.researchgate.net/publication/272997075_No_Eminems_in_Asia_Analyzing_ly
555   rics_of_Top_40_songs_in_the_East_West
556 Gabrielsson A. 2016. The relationship between musical structure and perceived expression. In:
557   Susan H, Ian C, Michael Thaut T, eds. *The Oxford Handbook of Music Psychology*. Oxford:
558   Oxford University Press, 215–232.
559 Gao R, Hao B, Li H, Gao Y, Zhu T. 2013. Developing Simplified Chinese Psychological
560   Linguistic Analysis Dictionary for Microblog. In: Imamura K, Usui S, Shirao T, Kasamatsu
561   T, Schwabe L, Zhong N eds. *Brain and Health Informatics*. Lecture Notes in Computer
562   Science. Cham: Springer International Publishing, 359–368. DOI: 10.1007/978-3-319-
563   02753-1_36.
564 Gendron M, Lindquist KA, Barsalou L, Barrett LF. 2012. Emotion words shape emotion
565   percepts. *Emotion* 12:314–325. DOI: 10.1037/a0026007.
566 Gordon RL, Schön D, Magne C, Astésano C, Besson M. 2010. Words and Melody Are
567   Intertwined in Perception of Sung Words: EEG and Behavioral Evidence. *PLoS ONE*
568   5:e9889. DOI: 10.1371/journal.pone.0009889.
569 Greenberg DM, Kosinski M, Stillwell DJ, Monteiro BL, Levitin DJ, Rentfrow PJ. 2016. The
570   Song Is You: Preferences for Musical Attribute Dimensions Reflect Personality. *Social
571   Psychological and Personality Science* 7:597–605. DOI: 10.1177/1948550616641473.
572 Greer T, Ma B, Sachs M, Habibi A, Narayanan S. 2019. A Multimodal View into Music's Effect
573   on Human Neural, Physiological, and Emotional Experience. In: *Proceedings of the 27th
574   ACM International Conference on Multimedia*. Nice France: ACM, 167–175. DOI:
575   10.1145/3343031.3350867.
576 Haselmayer M, Jenny M. 2017. Sentiment analysis of political communication: combining a
577   dictionary approach with crowdcoding. *Quality & Quantity* 51:2623–2646. DOI:
578   10.1007/s11135-016-0412-4.
579 Havlena W, abd Holak SL. 1991. "The good old days": Observations on nostalgia and its role in
580   consumer behavior. *Advances in Consumer Research* 18:323–329.
581 Holbrook MB, Anand P. 1990. Effects of Tempo and Situational Arousal on the Listener's
582   Perceptual and Affective Responses to Music. *Psychology of Music* 18:150–162. DOI:
583   10.1177/0305735690182004.
584 Hu X, Downie JS, Ehmann AF. 2009. Lyric text mining in music mood classification. In:
585   *Proceedings of the 10th International Society for Music Information Retrieval Conference
586   (ISMIR)*, 411-416.

587 Hu Y, Chen X, Yang D. 2009. Lyric-based Song Emotion Detection with Affective Lexicon and
588   Fuzzy Clustering Method. In: *Proceedings of the 10th International Society for Music*
589   *Information Retrieval Conference (ISMIR)*, 123-128.
590 Hunter PG, Schellenberg EG. 2010. Music and Emotion. In Mari RJ, Richard RF, Arthur NP,
591   eds. *Music Perception*. Springer, 129-164.
592 Juslin PN, Laukka P. 2004. Expression, Perception, and Induction of Musical Emotions: A
593   Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research*
594   33:217–238. DOI: 10.1080/0929821042000317813.
595 Juslin PN, Sloboda JA. 2001. Music and Emotion: Introduction. In: Juslin PN, Sloboda JA, eds.
596   *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, 3–19.
597 Kahng AB, Mantik S, Markov IL. 2002. Min-max placement for large-scale timing optimization.
598   In: *Proceedings of the 2002 international symposium on Physical design - ISPD '02*. San
599   Diego, CA, USA: ACM Press, 143. DOI: 10.1145/505388.505423.
600 Kaur R, Verma P. 2017. Sentiment analysis of movie reviews: A study of machine learning
601   algorithms with various feature selection methods. *International Journal of Computer*
602   *Sciences and Engineering* 5:113-121.
603 Kolinsky R, Lidji P, Peretz I, Besson M, Morais J. 2009. Processing interactions between
604   phonology and melody: Vowels sing but consonants speak. *Cognition* 112:1–20. DOI:
605   10.1016/j.cognition.2009.02.014.
606 Lang PJ. 1995. The emotion probe: Studies of motivation and attention. *American psychologist*
607   50: 372-385. DOI: 10.1037/0003-066X.50.5.372
608 Laurier C, Grivolla J, Herrera P. 2008. Multimodal Music Mood Classification Using Audio and
609   Lyrics. In: *2008 Seventh International Conference on Machine Learning and Applications*.
610   San Diego, CA, USA: IEEE, 688–693. DOI: 10.1109/ICMLA.2008.96.
611 Liu B. 2012. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language*
612   *Technologies* 5:1–167. DOI: 10.2200/S00416ED1V01Y201204HLT016.
613 Liu SM, Chen J-H. 2015. A multi-label classification based approach for sentiment
614   classification. *Expert Systems with Applications* 42:1083–1093. DOI:
615   10.1016/j.eswa.2014.08.036.
616 Lu K, Wu J. 2019. Sentiment Analysis of Film Review Texts Based on Sentiment Dictionary and
617   SVM. In: *Proceedings of the 2019 3rd International Conference on Innovation in Artificial*
618   *Intelligence - ICIAI 2019*. Suzhou, China: ACM Press, 73–77. DOI:
619   10.1145/3319921.3319966.
620 Malheiro R, Oliveira HG, Gomes P, Paiva RP. 2016. Keyword-based Approach for Lyrics
621   Emotion Variation Detection: In: *Proceedings of the 8th International Joint Conference on*
622   *Knowledge Discovery, Knowledge Engineering and Knowledge Management*. Porto,
623   Portugal: SCITEPRESS - Science and Technology Publications, 33–44. DOI:
624   10.5220/0006037300330044.

625   Malheiro R, Panda R, Gomes P, Paiva RP. 2018. Emotionally-Relevant Features for
626   Classification and Regression of Music Lyrics. *IEEE Transactions on Affective Computing*
627   9:240–254. DOI: 10.1109/TAFFC.2016.2598569.
628   Malheiro R, Panda R, Gomes P, Paiva R. 2016. Bi-modal music emotion recognition: Novel
629   lyrical features and dataset. In: *9th International Workshop on Music and Machine*
630   *Learning–MML'2016–in conjunction with the European Conference on Machine Learning*
631   *and Principles and Practice of Knowledge Discovery in Databases–ECML/PKDD 2016.*
632   McFee B, Raffel C, Liang D, Ellis DP, McVicar M, Battenberg E, Nieto O. 2015. librosa: Audio
633   and music signal analysis in python. In: *Proceedings of the 14th python in science*
634   *conference*, 8:18-25.
635   McVicar M, Freeman T, De Bie T. 2011. Mining the Correlation between Lyrical and Audio
636   Features and the Emergence of Mood. In: *Proceedings of the 12th International Society for*
637   *Music Information Retrieval Conference (ISMIR)*, 783-788.
638   Meyer BT, Kollmeier B. 2009. Complementarity of MFCC, PLP and Gabor features in the
639   presence of speech-intrinsic variabilities. In: *Tenth Annual Conference of the International*
640   *Speech Communication Association*, 2755-2758.
641   Morton JB, Trehub SE. 2001. Children's Understanding of Emotion in Speech. *Child*
642   *Development* 72:834–843. DOI: 10.1111/1467-8624.00318.
643   Morton JB, Trehub SE. 2007. Children's judgements of emotion in song. *Psychology of Music*
644   35:629–639. DOI: 10.1177/0305735607076445.
645   O'Callaghan C, Grocke D. 2009. Lyric analysis research in music therapy: Rationales, methods
646   and representations. *The Arts in Psychotherapy* 36:320–328. DOI:
647   10.1016/j.aip.2009.09.004.
648   Panda R, Malheiro R, Rocha B, Oliveira A, Paiva RP. 2013. Multi-modal music emotion
649   recognition: A new dataset, methodology and comparative analysis. In: *International*
650   *Symposium on Computer Music Multidisciplinary Research*, 1-13.
651   Peng H, Cambria E, Hussain A. 2017. A Review of Sentiment Analysis Research in Chinese
652   Language. *Cognitive Computation* 9:423–435. DOI: 10.1007/s12559-017-9470-8.
653   Pennebaker JW. 2004. Theories, Therapies, and Taxpayers: On the Complexities of the
654   Expressive Writing Paradigm. *Clinical Psychology: Science and Practice* 11:138–142.
655   DOI: 10.1093/clipsy.bph063.
656   Pennebaker JW, Francis ME, Booth RJ. 2001. Linguistic inquiry and word count (LIWC):
657   LIWC2001. Available at http://downloads.liwc.net.s3.amazonaws.com/
658   Petrie KJ, Pennebaker JW, Sivertsen B. 2008. Things we said today: A linguistic analysis of the
659   Beatles. *Psychology of Aesthetics, Creativity, and the Arts* 2:197–202. DOI:
660   10.1037/a0013117.
661   Pettijohn TF, Sacco DF. 2009. The Language of Lyrics: An Analysis of Popular Billboard Songs
662   Across Conditions of Social and Economic Threat. *Journal of Language and Social*
663   *Psychology* 28:297–311. DOI: 10.1177/0261927X09335259.

664 Pieschl S, Fegers S. 2016. Violent Lyrics = Aggressive Listeners?: Effects of Song Lyrics and
665   Tempo on Cognition, Affect, and Self-Reported Arousal. *Journal of Media Psychology*
666   28:32–41. DOI: 10.1027/1864-1105/a000144.
667 Poulin-Charronnat B, Bigand E, Madurell F, Peereman R. 2005. Musical structure modulates
668   semantic priming in vocal music. *Cognition* 94:B67–B78. DOI:
669   10.1016/j.cognition.2004.05.003.
670 Proverbio AM, Benedetto F, Guazzone M. 2020. Shared neural mechanisms for processing
671   emotions in music and vocalizations. *European Journal of Neuroscience* 51:1987–2007.
672   DOI: 10.1111/ejn.14650.
673 Rachman FH, Samo R, Fatichah C. 2019. Song Emotion Detection Based on Arousal-Valence
674   from Audio and Lyrics Using Rule Based Method. In: *2019 3rd International Conference*
675   *on Informatics and Computational Sciences (ICICoS)*. Semarang, Indonesia: IEEE, 1–5.
676   DOI: 10.1109/ICICoS48119.2019.8982519.
677 Schmidt EM, Turnbull D, Kim YE. 2010. Feature selection for content-based, time-varying
678   musical emotion regression. In: *Proceedings of the international conference on Multimedia*
679   *information retrieval - MIR '10*. Philadelphia, Pennsylvania, USA: ACM Press, 267. DOI:
680   10.1145/1743384.1743431.
681 Schwartz JW, Gouzoules H. 2019. Decoding human screams: perception of emotional arousal
682   from pitch and duration. *Behaviour* 156:1283–1307. DOI: 10.1163/1568539X-00003566.
683 Sen A, Srivastava M. 2012. Multiple regression. In: Sen A, Srivastava M, eds. *Regression*
684   *analysis: theory, methods, and applications*. Springer Science & Business Media: 28-49.
685 Serafine M. 1984. Integration of melody and text in memory for songs. *Cognition* 16:285–303.
686   DOI: 10.1016/0010-0277(84)90031-3.
687 Serafine ML, Davidson J, Crowder RG, Repp BH. 1986. On the nature of melody-text
688   integration in memory for songs. *Journal of Memory and Language* 25:123–135. DOI:
689   10.1016/0749-596X(86)90025-2.
690 Silverman MJ. 2020. Therapeutic songwriting to address distress tolerance for adults on an acute
691   care mental health unit: A pilot study. *The Arts in Psychotherapy* 71:101716. DOI:
692   10.1016/j.aip.2020.101716.
693 Slatcher RB, Pennebaker JW. 2006. How Do I Love Thee? Let Me Count the Words: The Social
694   Effects of Expressive Writing. *Psychological Science* 17:660–664. DOI: 10.1111/j.1467-
695   9280.2006.01762.x.
696 Sousou SD. 1997. Effects of Melody and Lyrics on Mood and Memory. *Perceptual and Motor*
697   *Skills* 85:31–40. DOI: 10.2466/pms.1997.85.1.31.
698 Speck JA, Schmidt EM, Morton BG, Kim YE. 2011. A Comparative Study of Collaborative vs.
699   Traditional Musical Mood Annotation. In: *Proceedings of the 12th International Society for*
700   *Music Information Retrieval Conference (ISMIR)*, 549-554.
701 Stratton VN, Zalanowski AH. 1994. Affective Impact of Music Vs. Lyrics. *Empirical Studies of*
702   *the Arts* 12:173–184. DOI: 10.2190/35T0-U4DT-N09Q-LQHW.

703 Swaminathan S, Schellenberg EG. 2015. Current Emotion Research in Music Psychology.
704   *Emotion Review* 7:189–197. DOI: 10.1177/1754073914558282.

705 van Besouw RM, Howard DM, Ternström S. 2005. Towards an understanding of speech and
706   song perception. *Logopedics Phoniatrics Vocology* 30:129–135. DOI:
707   10.1080/14015430500262160.

708 van Waesberghe JS. 1955. *A Textbook of Melody: A course in functional melodic analysis*.
709   American Institute of Musicology Press.

710 Vempala NN, Russo FA. 2018. Modeling Music Emotion Judgments Using Machine Learning
711   Methods. *Frontiers in Psychology* 8:2239. DOI: 10.3389/fpsyg.2017.02239.

712 Vidas D, Calligeros R, Nelson NL, Dingle GA. 2020. Development of emotion recognition in
713   popular music and vocal bursts. *Cognition and Emotion* 34:906–919. DOI:
714   10.1080/02699931.2019.1700482.

715 Viega M, Baker FA. 2017. What's in a song? Combining analytical and arts-based analysis for
716   songs created by songwriters with neurodisabilities. *Nordic Journal of Music Therapy*
717   26:235–255. DOI: 10.1080/08098131.2016.1205651.

718 Wang J-C, Yang Y-H, Wang H-M, Jeng S-K. 2012. The acoustic emotion gaussians model for
719   emotion-based music annotation and retrieval. In: *Proceedings of the 20th ACM*
720   *international conference on Multimedia - MM '12*. Nara, Japan: ACM Press, 89. DOI:
721   10.1145/2393347.2393367.

722 Xu L, Li L, Jiang Z, Sun Z, Wen X, Shi J, Sun R, Qian X. 2021. A Novel Emotion Lexicon for
723   Chinese Emotional Expression Analysis on Weibo: Using Grounded Theory and Semi-
724   Automatic Methods. *IEEE Access* 9:92757–92768. DOI: 10.1109/ACCESS.2020.3009292.

725 Xu L, Wen X, Shi J, Li S, Xiao Y, Wan Q, Qian X. 2020. Effects of individual factors on
726   perceived emotion and felt emotion of music: Based on machine learning methods.
727   *Psychology of Music*:030573562092842. DOI: 10.1177/0305735620928422.

728 Xu L, Yun Z, Sun Z, Wen X, Wang J, Qian X. 2020. PSIC3839: Predicting the Overall Emotion
729   and Depth of Entire Songs. Available at https://github.com/xl2218066/PSIC3839.

730 Xu L, Zheng Y, Xu D, Xu L. 2021. Predicting the Preference for Sad Music: The Role of
731   Gender, Personality, and Audio Features. *IEEE Access* 9:92952–92963. DOI:
732   10.1109/ACCESS.2021.3090940.

733 Yang X, Dong Y, Li J. 2018. Review of data features-based music emotion recognition methods.
734   *Multimedia Systems* 24:365–389. DOI: 10.1007/s00530-017-0559-4.

735 Yu Y, Tang S, Raposo F, Chen L. 2019. Deep Cross-Modal Correlation Learning for Audio and
736   Lyrics in Music Retrieval. *ACM Transactions on Multimedia Computing, Communications,*
737   *and Applications* 15:1–16. DOI: 10.1145/3281746.

738 Yu Y, Wu D, Zhang J, Fang P. 2019. Lyrics only or lyrics with music? The effect of different
739   lyric conditions on prosocial-related outcomes. *PsyCh Journal* 8:503–512. DOI:
740   10.1002/pchj.269.

741 Zacharopoulou K, Kyriakidou A. 2009. A cross-cultural comparative study of the role of musical
742   structural features in the perception of emotion in Greek traditional music. *Journal of*
743   *Interdisciplinary Music Studies* 3:1-15.
744 Zhang C, Tong T, Bu Y. 2019. Examining differences among book reviews from various online
745   platforms. *Online Information Review* 43:1169–1187. DOI: 10.1108/OIR-01-2019-0037.
746 Zhao N, Jiao D, Bai S, Zhu T. 2016. Evaluating the Validity of Simplified Chinese Version of
747   LIWC in Detecting Psychological Expressions in Short Texts on Social Network Services.
748   *PLOS ONE* 11:e0157947. DOI: 10.1371/journal.pone.0157947.
749 Zhou J, Chen X, Yang D. 2019. Multimodel Music Emotion Recognition Using Unsupervised
750   Deep Neural Networks. In: Li W, Li S, Shao X, Li Z eds. *Proceedings of the 6th*
751   *Conference on Sound and Music Technology (CSMT)*. Lecture Notes in Electrical
752   Engineering. Singapore: Springer Singapore, 27–39. DOI: 10.1007/978-981-13-8707-4_3.
753

Correlation between lyric features and perceived arousal in music

PastM: proportion of past tense markers; *Insight*: proportion of words related to insight; *Time*: proportion of words related to time*; Achieve*: proportion of words related to achievement; *tPast*: proportion of words related to the past; *WordCount*: the total number of words; *WordPerSentence*: average number of words per sentence; *RateLatinWord*: the ratio of Latin words. **Correlation is significant at the 0.01 level (2-tailed).

1   **Table 1** Correlation between lyric features and perceived arousal in music

	1	2	3	4	5	6	7	8	9
1. Arousal	1								
2. *PastM*	-.121**	1							
3. *Insight*	-.122**	.168**	1						
4. *Time*	-.115**	.299**	.139**	1					
5. *Achieve*	.111**	0.023	.146**	-0.006	1				
6. *tPast*	-.124**	.517**	.153**	.322**	0.014	1			
7. *WordCount*	.206**	.056**	.080**	0.004	.073**	0.014	1		
8. *WordPerSentence*	.179**	.062**	.105**	0.021	.088**	0.025	.873**	**1**	
9. *RateLatinWord*	.183**	-0.029	0.02	-0.032	.103**	-0.031	.174**	.114**	1

2   **Notes:** **Correlation is significant at the 0.01 level (2-tailed).

3   **Abbreviations:** *PastM*: proportion of past tense markers; *Insight*: proportion of words related to insight; *Time*:
4   proportion of words related to time*; Achieve*: proportion of words related to achievement; *tPast*: proportion of words
5   related to the past; *WordCount*: the total number of words; *WordPerSentence*: average number of words per sentence;
6   *RateLatinWord*: the ratio of Latin words.

Correlation between lyric features and perceived valence in music

Adverb: proportion of adverbs; *TenseM*: proportion of tense markers; *PastM*: proportion of past tense markers; *NegEmo*: proportion of negative emotion words; *Sad*: proportion of words related to sadness; *CogMech*: proportion of words related to cognition; *Tentat*: proportion of words related to tentativeness; *tPast*: proportion of words related to the past. **Correlation is significant at the 0.01 level (2-tailed).

1   **Table 2** Correlation between lyric features and perceived valence in music

	1	2	3	4	5	6	7	8	9
1. Arousal	1								
2. *PastM*	-.121**	1							
3. *Insight*	-.122**	.168**	1						
4. *Time*	-.115**	.299**	.139**	1					
5. *Achieve*	.111**	0.023	.146**	-0.006	1				
6. *tPast*	-.124**	.517**	.153**	.322**	0.014	1			
7. *WordCount*	.206**	.056**	.080**	0.004	.073**	0.014	1		
8. *WordPerSentence*	.179**	.062**	.105**	0.021	.088**	0.025	.873**	1	
9. *RateLatinWord*	.183**	-0.029	0.02	-0.032	.103**	-0.031	.174**	.114**	1

2   **Notes:** **Correlation is significant at the 0.01 level (2-tailed).

3   **Abbreviations:** *Adverb*: proportion of adverbs; *TenseM*: proportion of tense markers; *PastM*: proportion of past tense

4   markers; *NegEmo*: proportion of negative emotion words; *Sad*: proportion of words related to sadness; *CogMech*:

5   proportion of words related to cognition; *Tentat*: proportion of words related to tentativeness; *tPast*: proportion of words

6   related to the past.

The best performing parameters for each random forest regression

AF indicates audio features; LF indicates lyric features; and CF indicates combined features.

1   **Table 3** The best performing parameters for each random forest regression

Ground truth	Inputs	Parameters				
		n_estimators	*max_depth*	*min_samples_leaf*	*min_samples_split*	*max_features*
Arousal	AF	156	10	8	18	0.2
	LF	196	50	5	8	0.8
	CF	136	27	3	12	0.4
Valence	AF	179	15	13	22	0.5
	LF	193	38	3	8	0.8
	CF	191	43	5	25	0.6

2   **Abbreviations:** AF indicates audio features; LF indicates lyric features; and CF indicates combined features.

Figure 1

The proposed modeling method.
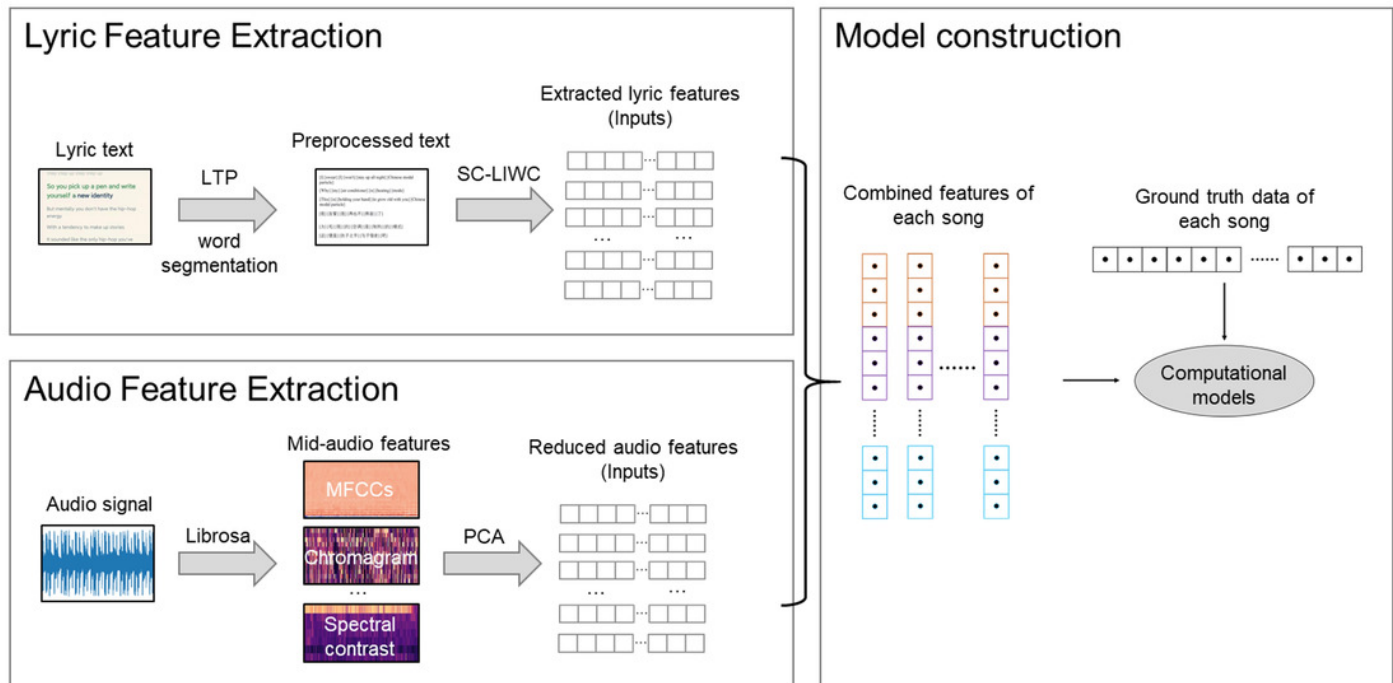
Figure 2

Data distribution of the 2372 Chinese songs in this study.

(a) Distribution of annotated emotions in the valence-arousal emotion space. (b) Word clouds of the top words used in each quadrant of the valence-arousal emotion space. The font size depends on the usage frequency of the word (positive correlation).
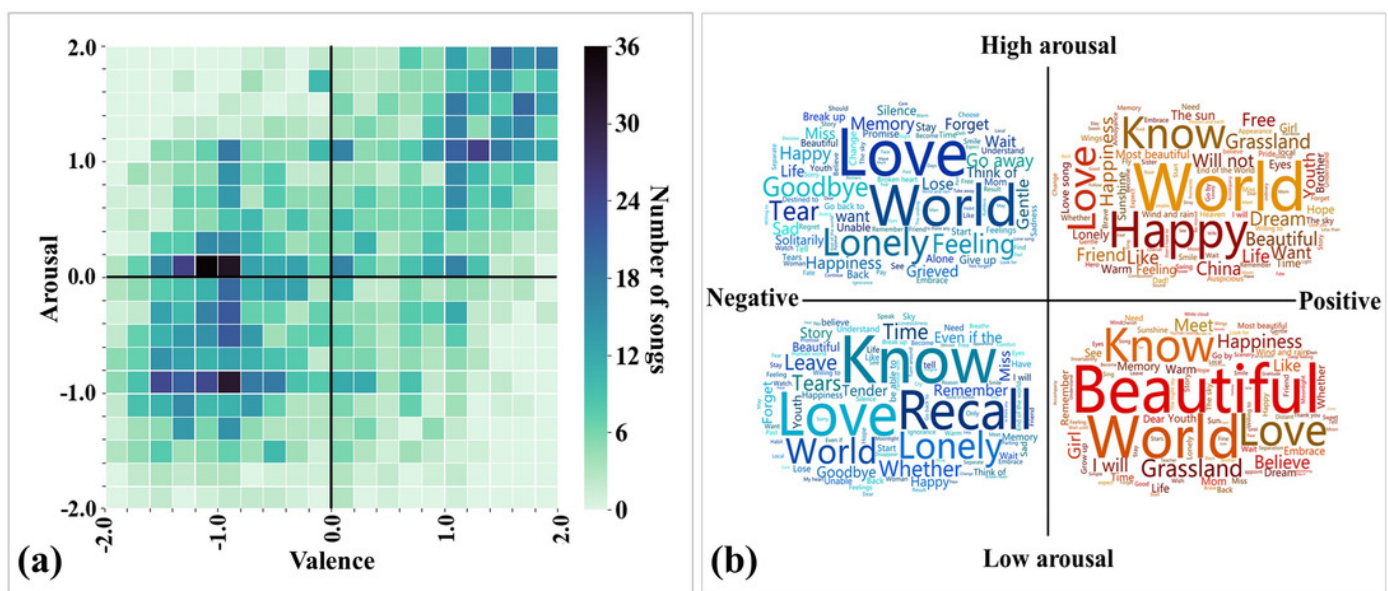
Figure 3

Performance of constructed MER models with different inputs and algorithms.

(a) Prediction results of perceived arousal recognition models, measured by R^2 statistics. (b) Prediction results of perceived arousal recognition models, measured by $RMSE$. (c) Prediction results of perceived valence recognition models, measured by R^2 statistics. (d) Prediction results of perceived valence recognition models, measured by $RMSE$. Error bars indicate the standard deviations.
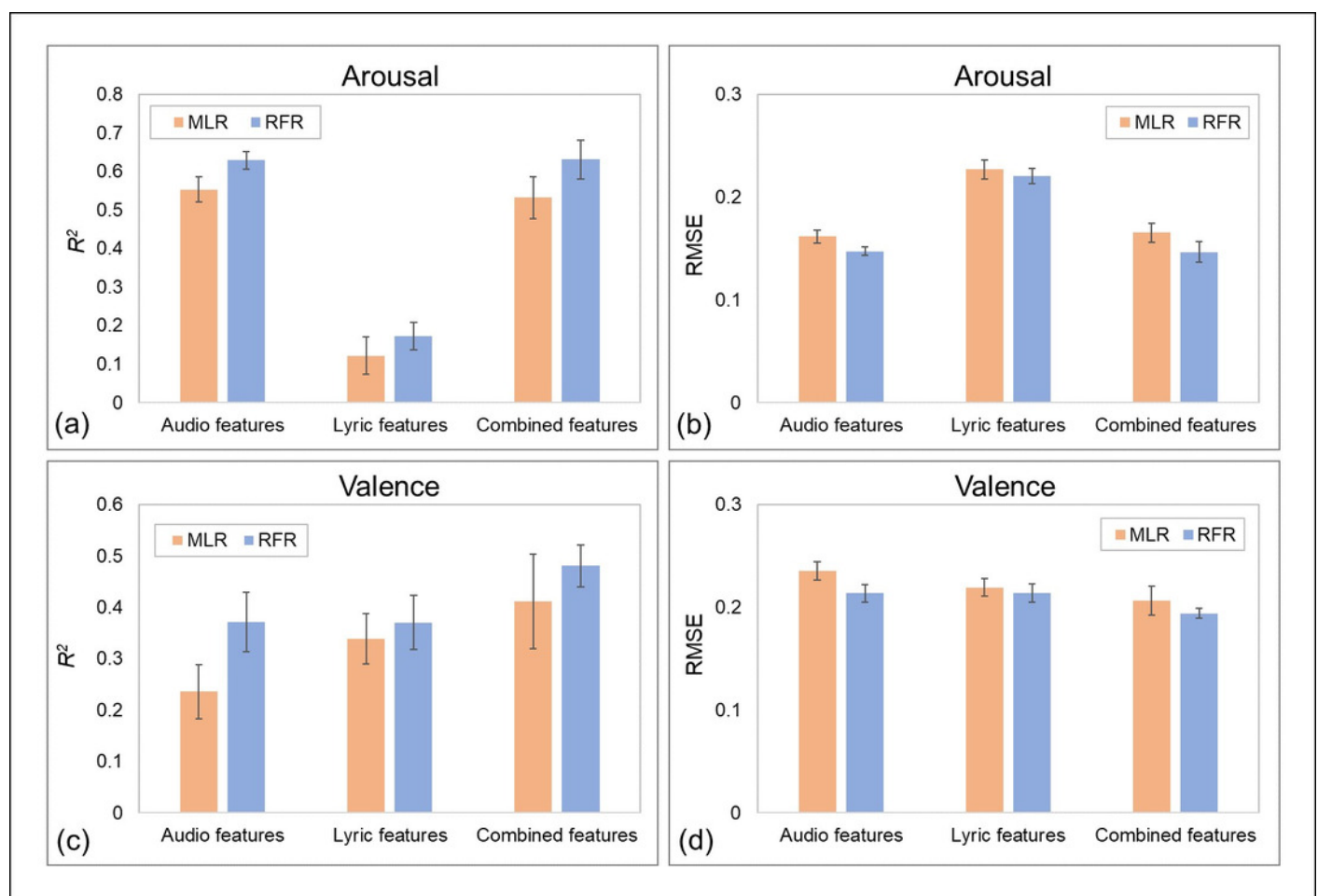
Figure 4

Distribution of feature importance for RFR-based recognition models of perceived arousal and perceived valence.

Arranged in descending order of the mean value, the top 30 features were included for visibility, and the trend of the remaining features was approximately the same. Error bars indicate the standard deviations.