

Physiological State Can Help Predict the Perceived Emotion of Music: Evidence from ECG and EDA Signals

Liang Xu¹, Jie Wang¹, Xin Wen¹, Zaoyi Sun³, Rui Sun¹, Liuchang Xu², Xiuying Qian^{1,*}

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

²College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou, China

³College of Education, Zhejiang University of Technology, Hangzhou, China

Email address:

xuliang_psy@zju.edu.cn (Liang Xu), wj_psy@zju.edu.cn (Jie Wang), wenxin0301@zju.edu.cn (Xin Wen), sunzaoyi@zju.edu.cn (Zaoyi Sun), sunrui_2019@zju.edu.cn (Rui Sun), xuliuchang@zafu.zju.edu.cn (Liuchang Xu), xqian@zju.edu.cn (Xiuying Qian)

*Corresponding author

To cite this article:

Liang Xu, Jie Wang, Xin Wen, Zaoyi Sun, Rui Sun, Liuchang Xu, Xiuying Qian. Physiological State Can Help Predict the Perceived Emotion of Music: Evidence from ECG and EDA Signals. *American Journal of Life Sciences*. Vol. 9, No. 5, 2021, pp. 105-119.
doi: 10.11648/j.ajls.20210905.12

Received: August 12, 2021; **Accepted:** September 11, 2021; **Published:** September 23, 2021

Abstract: As the soul of music, emotion information is widely used in music retrieval and recommendation systems because the pursuit of emotional experience is the main motivation for music listening. In the field of music emotion recognition, computer scientists investigated computation models to automatically detect the perceived emotion of music, but this method ignores the differences between listeners. To provide users with the most accurate music emotion information, this study investigated the effects of physiological features on personalized music emotion recognition (PMER) models, which can automatically identify an individual's perceived emotion of music. Applying machine learning methods, we formed relations among audio features, physiological features, and music emotions. First, computational modeling analysis shows that physiological features extracted from electrocardiogram and electro-dermal activity signals can predict the perception of music emotion for some individuals. Second, we compared the performance of physiological feature-based perception and feeling models and observed substantial individual differences. In addition, we found that the performance of the perception model and the feeling model is related in predicting happy, relaxed, and sad emotions. Finally, by adding physiological features to the audio-based PMER model, the prediction effect of some individuals was improved. Our work investigated the relationship between physiological state and perceived emotion of music, constructed models with practical value, and provided a reference for the optimization of PMER systems.

Keywords: Music Emotion Recognition, Physiological Signal Processing, Machine Learning, Perceived Emotion

1. Introduction

Emotion information is one of the most important attributes individuals use for music information retrieval (MIR) [57], because the pursuit of emotional experience is the main motivation for individuals to listen to music [29]. Considering the vast library of existing music, a technology for automatically identifying music emotion is urgently needed to provide individuals with accurate music emotion information [74]. In the past two decades, computer scientists have expended considerable effort investigating computation models to detect the emotion of music [6]. From the early

general music emotion recognition (MER) model [38, 55] to the recent personalized music emotion recognition (PMER) model [47, 77], the automatic recognition effect continues to improve, and the technology has attracted increasing academic attention from computer scientists, psychologists, musicologists, and so forth.

The core of traditional MER research is to form the mapping relations between music features and perceived emotions [34]; notably, Xu et al first added individual factors as additional model inputs and improved the performance of

PMER models [77]. According to our review of the literature, an individual's physiological state is rarely considered in PMER models, although it has been widely used to predict emotions evoked by music [59]. Therefore, this study investigated the effects of an individual's physiological features on recognizing perceived emotions of music. Considering the practice for real-life use, electrocardiogram (ECG) and electro-dermal activity (EDA) signals, which can be collected by simple wearable devices [64], were used here. When the PMER model is optimized by adding physiological features, the accuracy of the music emotion information that users receive will increase. This may also facilitate the construction of personalized MIR and music recommendation systems.

2. Literature Review

2.1. Music Emotion Recognition

MER is a research area that investigates computation models to detect the emotion expressed by listening to music [6]. MER solves the problem of music emotion information annotation by developing technology that automatically recognizes music emotion [74] and constitutes a process of using computers to (a) extract and analyze music features from original music, (b) use machine learning (ML) methods to construct the relationship (computation models) between music features and perceived emotions, and (c) recognize the emotional expression of the untagged music. Through these three processes, the music database can be organized and managed according to emotion [34]. Notably, early music psychology researchers have provided theoretical bases for MER research [15, 60], and the research on MER has made considerable progress in the past two decades because of the development of computer technology. With the availability and accessibility of MER toolkits, this technology has also been recently applied to psychological research [68, 77, 83].

The strong subjectivity of the ground-truth data, which reflects the perceived emotions of human beings, is one of the critical issues in MER research because different individuals listening to the same music may produce different emotion perceptions [14]. To solve this problem, in most MER studies, each musical segment is annotated by many subjects to obtain a relatively accurate emotional assessment [52]. However, this method ignores listeners' individual differences; thus, the automatically recognized emotions may be inaccurate for different individuals. Psychological research has shown that the judgment of music emotion may be influenced by, for example, age [36], music education [42], absorption [20], trait empathy [32], and personality [31]. Therefore, individuality must be considered in MER systems. Yang et al. proposed the personalized MER (PMER) approach to study the role of individuality [75], and the results of the PMER models showed that the prediction accuracy for a user may be improved if the MER system is

personalized for the user. Based on the work of [75], Xu et al. improved the prediction effect of PMER models by adding various individual features as model inputs [77]; the feature importance results illustrated that felt emotion (emotion evoked by music) plays an important role in the prediction of the perceived emotion of music, which provided a reference for this study (Section 2.2).

In summary, PMER is derived from MER research and is receiving increasing academic attention from psychologists and computer scientists. From a theoretical perspective, PMER investigates the relationship among music features, individual factors, and individual perception of music emotion. Additionally, the use of MER technology can facilitate related psychological research. From an application perspective, a continuously optimized PMER model can provide users with music emotion information that is more accurate and has been widely used in MIR [11, 71] and music recommendation systems [10, 48]. Psychological research can also provide a crucial reference to improve PMER models.

2.2. Perceived Emotion and Felt Emotion of Music

The relationship between the perceived emotion (emotion expressed by music) and felt emotion (emotion aroused by music) of music has become a firmly established part of the research agenda of music psychologists. Gabrielsson found that felt emotion (e.g., "the music makes me feel sad") is sometimes the same as ("the music is sad") and sometimes different from ("the music is happy") the perceived emotion, which reinvigorated the question of "internal locus of emotion" versus "external locus of emotion" within musical communication [14]; subsequently, many studies have investigated this relationship. For instance, Kallinen and Ravaja found that music seems to arouse emotions similar to the emotional quality perceived in music but that the relationship may change by emotional category (e.g., "fearful music was perceived as negative but felt as positive") [31]. The work of Hunter et al. showed that feeling and perception ratings are highly correlated but that perception ratings were commonly higher [21]. Schubert concluded after conducting a review that the felt emotion rating is frequently rated statistically the same or lower than the corresponding perceived emotion rating [56]. Emotional contagion theory, which holds that humans have an internal "mimicry" of the perceived voice-like emotional expression of the music [26, 27], was then considered one theoretical position for explaining the relationship between perceived emotion and felt emotion. To avoid this controversial topic, some studies have asked participants to assess the emotion of music without being explicit regarding "perceived" or "felt" emotion [68]. This method indirectly reflected the inseparability of perceived and felt emotion.

The aforementioned inseparable relationship also provided a reference for the optimization of the PMER model. As aforementioned, Xu et al. found that using felt emotion rating as input could significantly improve the prediction performance of PMER models [77]. This finding is a

reminder that when predicting an individual's perception of music emotion, the individual's emotional state should be considered. However, another problem arises, that is, no technology can detect individuals' emotional state in real time, and it is unrealistic for individuals to constantly evaluate and report on their feelings, especially in real life. However, physiological states can be detected, and many researchers have demonstrated that physiological state and emotional state are closely linked [59]. Therefore, this study regarded the felt emotion of music (emotional state) as a bridge and investigated the relationship between the physiological state and perceived emotion of music.

2.3. Human Emotion Recognition Using Physiological Signals

Human emotion recognition (HER) has attracted increasing academic attention in recent years and been widely used in many areas, for example, mental health monitoring [19], transportation safety enhancing [9], and social security [69]. Using physiological signals, the internal signals reflecting human physiological states, to predict emotions is an HER method [59]. Additionally, music is one of the stimuli often used to evoke emotions [17, 73]. Other studies have shown that felt emotion of music can be recognized by many physiological signals by using, for example, an electroencephalogram (EEG) [37], ECG [33], an electromyogram (EMG) [18], or skin conductivity (Zong, & Chetouani, 2009). Similar to the process of MER, HER based on physiological signals constitutes a process of (a) extracting and analyzing features from original physiological signals, (b) using traditional ML or deep learning methods to construct recognition models, and (c) predicting the emotion evoked by emotion stimulation [59]. We referred to the aforementioned method and investigated the relationship between physiological signals and perceived emotion of music. We assumed that physiological signals can predict or help predict the individual perception of music emotions, because many studies have proved that felt and perceived emotions are highly correlated (Section 2.2).

In addition, considering the practice for real-life use, we use ECG and EDA as target signals because they can be collected by simple wearable devices without compromising comfort and privacy [64]. ECG is one of the most sensitive markers for emotional arousal. Heart rate (HR) and heart rate variability (HRV) extracted from ECG signals have been widely used for HER [44, 72]. Additionally, many studies have shown that music can produce specific physiological reactions of change in HR and HRV, which are associated with different emotions [43, 67]. Hsu *et al.* showed that using only an ECG signal can also predict the emotions evoked by music [22]. Similarly, EDA, which modulates the sweat amount from skin pores [66], is a sensitive marker for emotional arousal. Wu *et al.* (2010) proposed a method for recognizing human emotions based on only galvanic skin response (GSR; the same as EDA) signals, and many other studies have combined more signals. For instance, Monajati *et al.* employed fuzzy-adaptive resonance theory to

automatically recognize human emotions by combining GSR, HR, and respiration rate [41]. Das *et al.* combined ECG and GSR signals to recognize emotions among happy, sad, and neutral [8]. Song *et al.* designed and built a multimodal physiological emotion database, which collected EEG, GSR, ECG, and respiration signals, to explore human emotions [62]. As for this study, we referred to the processing methods of ECG and EDA signals in other HER studies to investigate the effect of individual physiological state on perceived emotion of music.

2.4. Theoretical Model of Music Emotion

For MER, emotions should be defined and accessed quantitatively. In most MER studies, two types of naturalistic emotion models have usually been applied to emotion evaluation. One model is the discrete emotion model, which divides emotions into discrete categories [23]. This type of model has been widely used in music emotion classification (MEC) studies, a sub-domain of MER, whose goal is to obtain one or more emotion labels corresponding to a music segment [74]. Several discrete emotions were chosen as target emotions from discrete emotion models [13, 24], and music segments with these emotion labels were used as the ground truth for modeling. For example, in [7], angry, happy, sad, and peaceful were selected as target emotion classes, and participants were asked to annotate emotion classes for each music segment. The researchers then applied support vector machines (SVMs) to construct MEC models for each emotion class, which showed good performance on MER. Notably, multidimensional emotion space models, which use multiple dimensions to label emotions (e.g., valence and arousal) [53], are more frequently used in MER studies [74]. Additionally, these studies have usually required that the participants annotate dimension values in the dimensional models (e.g., arousal and valence values in the VA model) directly while listening to music [25]. However, most MER studies have ignored an important controversy: whether musical and naturalistic emotions map onto one another in a 1-1 fashion (Allen, Walsh, & Zangwill, 2013). Many researchers posited that music emotion was not “naturalistic” [35, 78], whereas others believed in considerable overlap between musical and naturalistic emotions [54, 79].

Considering the aforementioned situation, we had three reasons for using four representative emotions—happy, relaxed, sad, and angry—to conduct the evaluation. First, considering the “naturalistic” emotion models, they cover the four parts of the 2D model of emotion [63] and are related to basic emotions [29]. Second, regarding the “musical” emotion models, they are related to the dimensions in the 9-factorial model of music-induced emotions proposed by [79]. Third, recognition models based on these four emotions may be directly applied in real life, which is not possible with dimensional models (e.g., arousal and valence). For example, in a MIR system, appropriate user information is “this is a relaxing song,” but “this is a low-arousal song” may be inappropriate.

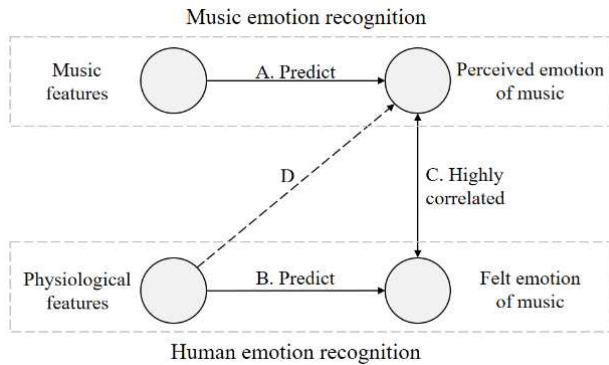


Figure 1. Relationship among music feature, physiological feature, perceived emotion, and felt emotion.

2.5. The Present Study

According to our review of the literature (Figure 1), we found that (A) music features have been widely used to predict perceived emotion of music in MER studies; (B) features extracted from physiological signals generated when individuals listen to music can predict the felt emotion of music; and (C) the perceived emotion and felt emotion of music are highly correlated. Additionally, the goal of this study is to investigate (D) the effect of physiological state on the perceived emotion of music. From a theoretical perspective, we wanted to explore whether the physiological arousal caused by music can affect or predict the “objective” part of music emotion. Using ML methods, we investigated the relationship between them from a computational perspective. In addition, our work is of practical value. Physiological signals that can be collected through wearable devices (ECG and EDA in this study) were considered to improve the predictive effect of the PMER model, which can increase the accuracy of the music emotion information provided to users in real life. This can also provide a reference for personalized MIR and music recommendation systems.

Considering the aforementioned goals, our three research questions are as follows. First, humans have an internal “mimicry” of the perceived voice-like emotional expression of the music [26], which is one of the reasons why music evokes emotion. Additionally, emotional arousal is often accompanied by physiological arousal. Conversely, physiological arousal may also identify the perceived emotion of music. Therefore, we assumed that physiological signals generated when listening to music can directly predict the perceived emotion of music (Hypothesis 1). Second, most studies have investigated the relationship between physiological arousal and emotional arousal [59] but rarely that of perceived emotion and physiological state. The relationship between the physiological state and felt emotion may be closer; thus, we assumed that when only physiological features were used as the model input, recognition models of felt emotion (feeling models) would perform better than perception models (Hypothesis 2). Third, we attempted to improve the prediction effect of the PMER models, and the physiological state may provide more

information for models. Therefore, we assumed that the PMER models based on music features would perform better after adding physiological features (Hypothesis 3).

3. Methods

To construct the PMER models, we first performed a listening task for each participant, to collect the ground-truth data (perceived and felt emotions of music) and physiological data (Section 3.1). The processing of audio and physiological signals was then conducted to extract model inputs, respectively, in Sections 3.2 and 3.3. Finally, the ML methods used for modeling were introduced in Section 3.4.

3.1. Experimental Design

For stimuli, sixty famous popular songs were collected from Chinese albums. Because the length of the segment for popular music is usually 25–30 seconds in MER studies [71], the collected excerpts were first trimmed to 25 seconds. Next, the trimmed excerpts were converted to a uniform format: 22,050 Hz, 16 bits, and mono channel PCM WAV (see [75]). These 60 processed music excerpts were used as experimental stimuli and for subsequent music feature extraction.

Referring to the method of [51], we constructed separate PMER models for each individual. Therefore, we needed each participant to complete the emotional annotation experiment of all the music excerpts. After assessing the behavioral and physiological data, this study finally obtained ten complete datasets. On average, the final ten participants (five females, five males) were aged 23.40 years ($SD = 2.32$) and were all undergraduate or graduate students recruited from their campus. In addition, none of them had received professional music training, and all of them often listened to music.

In the experiments, each participant listened to all the excerpts, and each excerpt was heard by ten unique participants. After a brief description of the experiment, participants received a listening order that was independently randomized to minimize the influence of presentation order. Each excerpt was preceded by 30 seconds of silence and followed by self-report questionnaires. Additionally, the participants were asked to concentrate on the music and listen to the music with their eyes closed. Kallinen and Ravaja advocated that the subjective measurements must be performed first because they may dilute faster than the more objective evaluation of the perceived emotion [31]. Additionally, the felt emotion may have included physiological responses that decrease as a function of time. Therefore, in the self-report phase, participants were first asked to evaluate the emotions that the music aroused in them while listening (i.e., emotion felt; “Did you feel happy when you listened to the music?”) on a scale from 1 (*not at all*) to 5 (*very much*). After the subjective measures, the participants were asked to evaluate the perception of the emotions expressed by the music (i.e., emotion perceived; “Do you think the music express a happy emotion?”), also

evaluated on a scale from 1 (*not at all*) to 5 (*very much*).

Physiological data were collected synchronously in the listening task. Participants used their left hand to provide emotion ratings while their right hand was connected to the Biopac MP160 data acquisition system for measurement of physiological responses. Using wireless receivers, the participants wore ECG100C and EDA100C amplifiers, simulating wearable devices, to collect physiological signals. ECG signals were recorded at 2000Hz by attaching electrodes, connected to the ECG100C amplifier, to the right wrist, and right and left ankles. EDA signals were collected by attaching two TSD203 Ag–AgCl electrodes to the distal phalanges of the index and middle fingers of the right hand [12], using Velcro straps. Additionally, EDA signals were resampled at 31.25Hz. Physiological data would be subjected to feature analysis to extract features that may be associated with emotions (Section 3.3).

3.2. Audio Signal Processing

Audio signal processing is a vital part of MER studies and can extract effective information from original music excerpts as input for later models [74]. In this study, we first applied librosa [40], a Python package for audio and music signal processing, to extract low-level audio features. Fourteen types of features were considered: mel-frequency cepstrum coefficients, root mean square energy, spectral centroid, spectral bandwidth, spectral contrast, spectral flatness, spectral roll-off frequency, short-time Fourier transform of chromagram, constant-Q transform of chromagram, chroma energy normalized (CENS), tonal centroid features (tonnetz), zero crossing rate, beat, and tempo. These features have demonstrated good performance in MER studies [50, 61, 75].

After feature extraction, each excerpt was represented in a subspace of high dimensionality. Therefore, feature reduction was then conducted to reduce the storage and computational space [68]. In this study, we used principal components analysis to reduce the dimensionality of the data. By forming a linear combination set of the new features that retain the variation of the original features in some fashion, the multidimensional data were mapped into a low-dimensional subspace [76]. As a result, 54 new features, explaining 95% of the variation, were retained and utilized as basic inputs of our final PMER models.

3.3. Physiological Signal Processing

Signal preprocessing was conducted to remove interference signals before feature extraction. Filtering was first performed by using AcqKnowledge 5.0 software [1], and the following high-pass (HP) and/or low-pass (LP) filters were applied to the original physiological data: ECG (LP = 35Hz; HP = 0.5Hz), and EDA (LP = 1Hz). Second, baseline correction was conducted by subtracting the equivalent signal obtained in the final 20 seconds of the silence that preceded the excerpt. Third, hot deck imputation was applied to handle outliers caused by the participant's movement or breathing

[2]. The preprocessed physiological signals were then used for feature extraction.

3.3.1. ECG Feature Extraction

Twenty-four features were extracted from the time-domain, frequency-domain, and nonlinear analyses of ECG signals for each excerpt [22, 65]. In the time-domain analysis, 11 features were calculated as follows:

(1) The standard deviation of R-R time intervals (RR intervals; *ECG_SDNN*):

$$ECG_SDNN = \sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 / n} \quad (1)$$

where R_i is the i th RR interval, \bar{R} is the average of the RR intervals, and n is the number of the RR intervals;

(2) The root mean square of differences between adjacent RR intervals (*ECG_RMSSD*):

$$ECG_RMSSD = \sqrt{\sum_{i=2}^n (R_i - R_{i-1})^2 / (n - 1)} \quad (2)$$

(3) The percentage of successive RR intervals that differ more than 50 ms (*ECG_pNN50*):

$$ECG_pNN50 = \text{count}(|t_{i+1} - t_i| > 50\text{ms}) / (n - 1) \quad (3)$$

where t_i is the time when the i th R-wave appears, and $\text{count}(\dots)$ means calculate the quantity in brackets;

(4) The standard deviation of differences between adjacent RR intervals (*ECG_SDSD*):

$$ECG_SDSD = \sqrt{\sum_{i=2}^n \frac{(R_i - R_{i-1} - ECG_RMSSD)^2}{n - 1}} \quad (4)$$

(5) The number of R-waves within one epoch divided by 1 min (*ECG_BPM*);

(6) The median value of RR intervals (*ECG_MED_RRI*);

(7) The interquartile range of RR intervals (*ECG_IQR_RRI*);

(8) The mean absolute deviation of RR intervals (*ECG_MAD_RRI*):

$$ECG_MAD_RRI = \frac{1}{n} \sum_{i=1}^n |R_i - \bar{R}| \quad (5)$$

(9) The mean of the difference between adjacent RR intervals (*ECG_DIFF_RRI*):

$$ECG_DIFF_RRI = \frac{1}{n-1} \sum_{i=2}^n |R_i - R_{i-1}| \quad (6)$$

(10) The coefficient of variation of RR intervals (*ECG_CV_RRI*):

$$ECG_CV_RRI = \frac{1}{\bar{R}} \sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 / n} \quad (7)$$

(11) The difference between the maximum and the minimum RR interval (*ECG_RANGE*).

In the frequency-domain analysis, we used the fast Fourier transform (FFT)-based method to calculate the power spectral density (PSD) of the resampled RR intervals [22]. The PSD analysis could be used to calculate the power of

specific frequency ranges and the peak frequencies for three different frequency bands: very-low-frequency range (VLF; 0.0033–0.04 Hz), low-frequency range (LF; 0.04–0.15 Hz), and high-frequency range (HF; 0.15–0.4 Hz). In addition to the power calculated in the VLF, LF, and HF bands (1, ECG_VLF ; 2, ECG_LF ; 3, ECG_HF), another seven features were calculated as follows:

(4) The total power in the full frequency range (ECG_TP):

$$ECG_TP = ECG_VLF + ECG_LF + ECG_HF \quad (8)$$

(5) The ratio of power calculated within the LF band to that calculated within the HF band (ECG_LF/HF);

(6) LF power normalized to the sum of the LF and HF power (ECG_LF_{norm}):

$$ECG_LF_{norm} = ECG_LF / (ECG_TP - ECG_VLF) \quad (9)$$

(7) HF power normalized to the sum of the LF and HF power (ECG_HF_{norm}):

$$ECG_HF_{norm} = ECG_HF / (ECG_TP - ECG_VLF) \quad (10)$$

(8) VLF power expressed as percentage of the total power (ECG_pVLF):

$$ECG_pVLF = (ECG_VLF / ECG_TP) \times 100 \quad (11)$$

(9) LF power expressed as percentage of the total power (ECG_pLF):

$$ECG_pLF = (ECG_LF / ECG_TP) \times 100 \quad (12)$$

(10) HF power expressed as percentage of the total power (ECG_pHF):

$$ECG_pHF = (ECG_HF / ECG_TP) \times 100 \quad (13)$$

In the nonlinear analysis, we only considered Poincaré plot analysis, which measured the quantitative beat-to-beat correlation between adjacent RR intervals [82]. Additionally, three features were calculated as follows:

(1) The standard deviation of the instantaneous beat-to-beat RR interval variability (ECG_SD_1):

$$ECG_SD_1 = \sqrt{0.5(std(t_{i+2} - 2t_{i+1} + t_i))^2} \quad (14)$$

where t_i is the time when the i th R-wave appears, and $std(\dots)$ means calculate the standard deviation in brackets;

(2) The standard deviation of the continuous long-term beat-to-beat RR interval variability (ECG_SD_2):

$$ECG_SD_2 = \sqrt{2(std(t_{i+1} - t_i))^2 - 0.5(std(t_{i+2} - 2t_{i+1} + t_i))^2} \quad (15)$$

(3) The ratio of ECG_SD_1 to ECG_SD_2 (ECG_SD_{12}).

3.3.2. EDA Feature Extraction

Referring to the methods of [64] and [70], 12 statistical features and three skin conductance level (SCL)-related features were extracted from the original EDA signals for each excerpt as follows:

- (1) The average value of EDA signals (EDA_MEAN);
- (2) The standard deviation of EDA signals (EDA_STD);
- (3) The maximum and (4) minimum values of EDA signals (EDA_MAX , and EDA_MIN);
- (5) The average value of the first-order difference of EDA signals (EDA_1D_MEAN):

$$EDA_1D_MEAN = \frac{1}{n-1} \sqrt{\sum_{i=1}^{n-1} (x_{i+1} - x_i)} \quad (16)$$

where x_i is the value of the i th resampled EDA signal;

(6) The standard deviation of the first-order difference of EDA signals (EDA_1D_STD):

$$EDA_1D_STD = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n-2} [(x_{i+1} - x_i) - EDA_1D_MEAN]^2} \quad (17)$$

(7) The maximum and (8) minimum value of the first-order difference of EDA signals (EDA_1D_MAX , and EDA_1D_MIN);

(9) The average value of the second-order difference of EDA signals (EDA_2D_MEAN):

$$EDA_2D_MEAN = \frac{1}{n-2} \sqrt{\sum_{i=1}^{n-2} (x_{i+2} - 2x_{i+1} + x_i)} \quad (18)$$

(10) The standard deviation of the second-order difference of EDA signals (EDA_2D_STD):

$$EDA_2D_STD = \sqrt{\frac{1}{n-3} \sum_{i=1}^{n-3} [(x_{i+2} - 2x_{i+1} + x_i) - EDA_2D_MEAN]^2} \quad (19)$$

(11) The maximum and (12) minimum values of the second-order difference of EDA signals (EDA_2D_MAX , and EDA_2D_MIN).

Skin electricity can be separated into two parts: *tonic*, which corresponds to the slowly changing SCL, and *phasic*, which corresponds to rapid skin electrical fluctuations. Therefore, the first-order polynomial can be used to fit the skin electrical signal as follows:

$$X_{tonic} = k(X_{tonic} + X_{phasic}) + b \quad (20)$$

where X_{tonic} and X_{phasic} represent the values of the *tonic* and *phasic* parts, respectively. Additionally, the first-order fitting coefficients, k and b , were used as SCL-related features (EDA_SCL_COEFK , and EDA_SCL_COEFB). The average value of SCL (EDA_SCL_MEAN) was also extracted:

$$EDA_SCL_MEAN = \frac{1}{N} \sum_{n=1}^N X_{tonic_n} \quad (21)$$

In summary, 39 physiological features were extracted from ECG and EDA signals for each music excerpt (Table 1). These features were utilized as inputs in our recognition models.

3.4. Machine Learning Methods

During the construction of the PMER models, the ML method links the variables. A multitude of ML methods is

available for supervised problems (i.e., classification and regression). Additionally, this study formulated PMER as a regression problem, to predict a real value from some observed features [58]. We considered three ML methods to train regressors. First, we used multiple linear regression (MLR) as a baseline algorithm because of its relatively low computational complexity and its effectiveness [61, 74]. Second, SVM has been found superior to the existing ML methods [75]. Therefore, we adopted support vector

regression (SVR), an extension of SVM to regression problems, to construct PMER models as a reference. Third, random forest regression (RFR), a widely used ML method in MER studies [3, 34], was also adopted for modeling. Additionally, scikit-learn [49], a Python module integrating a wide range of ML algorithms for medium-scale supervised and unsupervised problems, was applied for model construction and training. More details on the ML methods are presented along with the results in Section 4.

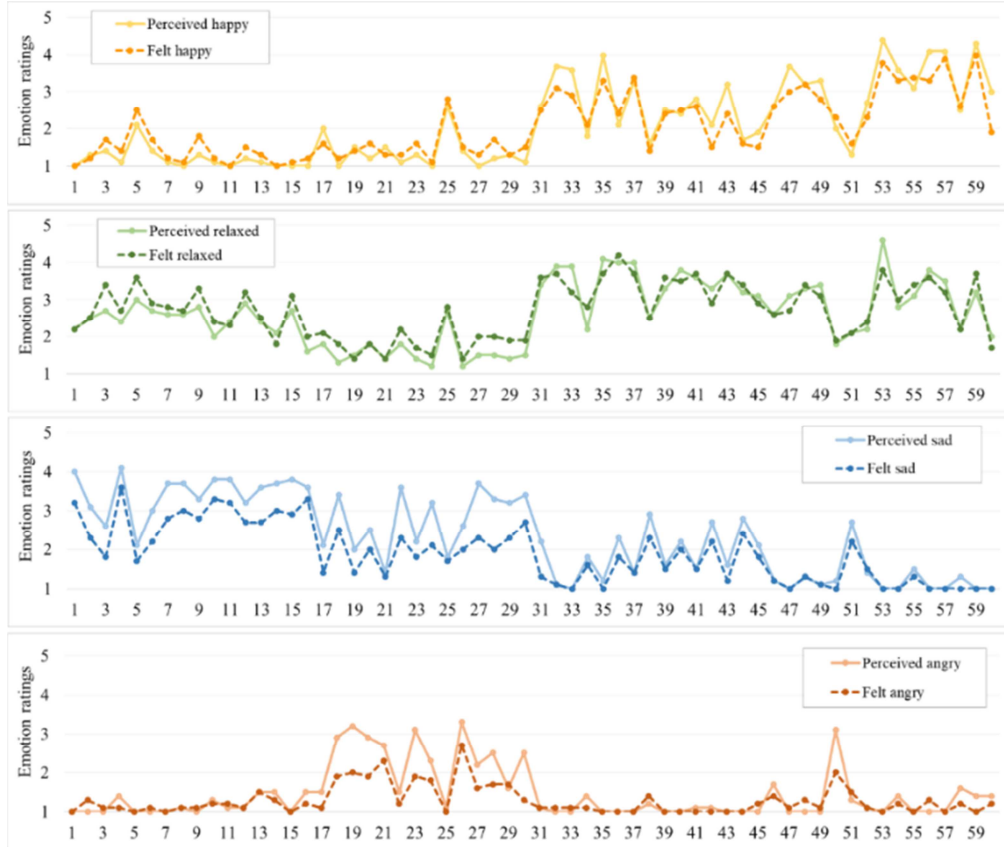


Figure 2. Distribution of perceived and felt emotion ratings for each music excerpt.

4. Results

4.1. Relationship Between Perceived and Felt Emotions of Music

As the first step of data exploration, we investigated the correlations between perceived and felt emotions. Figure 2 shows the mean ratings for perceived and felt emotions of each excerpt. We observed that perceived emotion ratings were positively correlated with felt emotion ratings (Happy: $r(59) = 0.95$, $p < 0.01$; Relaxed: $r(59) = 0.93$, $p < 0.01$; Sad: $r(59) = 0.96$, $p < 0.01$; Angry: $r(59) = 0.90$, $p < 0.01$). This result is similar to that of other studies [21, 31], which reflects the inseparability of perceived and felt emotions.

In addition, we found that perception ratings of negative emotions were higher than feeling ratings. Perceived sad ratings ($M = 2.34$, $SD = 1.03$) were significantly higher than felt sad ratings ($M = 1.90$, $SD = 0.75$), $t = 8.93$, $p < 0.01$, $d =$

0.50. Additionally, perceived angry ratings ($M = 1.46$, $SD = 0.68$) were significantly higher than felt sad ratings ($M = 1.28$, $SD = 0.37$), $t = 3.65$, $p < 0.01$, $d = 0.33$. By contrast, we observed that perceived relaxed ratings ($M = 2.62$, $SD = 0.87$) were significantly lower than felt relaxed ratings ($M = 2.72$, $SD = 0.75$), $t = 2.34$, $p < 0.05$, $d = 0.12$, which is the opposite of the result in [21], that is, perception ratings were commonly higher. One possible explanation is that the participants may feel relaxed in a quiet experimental environment. Combined with their relaxed state, felt relaxed ratings may be evaluated as higher. This hypothesis was indirectly confirmed by the feelings reported after the experiment, that is, most of the participants felt “relaxed,” “comfortable,” and “a little bit sleepy” during the listening task.

In summary, the results show that the perceived and felt emotions of music were highly correlated, although there were some differences. These findings provide a vital basis for our subsequent analysis.

4.2. Recognition Models of Perceived Emotions Using Physiological Features

In this section, we built the physiological features-based recognition models of perceived emotions and attempted to verify Hypothesis 1. We used physiological features as inputs, perceived emotion ratings as the ground truth, and three ML methods (MLR, SVR, and RFR) to construct recognition models (also called regressors). For each participant, we built four types of personalized recognition models separately to predict the perception of happy, relaxed, sad, and angry. For each type of model, three models using different ML algorithms were constructed. Therefore, twelve personalized recognition models were built for each participant. In this manner, we investigated the predictive effect of physiological features on perceived emotions, compared the effect of the different ML algorithms, and analyzed the differences when predicting different emotions.

To maximize ML, all the physiological inputs were scaled to a value between 0 and 1 for each feature [68]. The ground-truth values were also scaled to a range between 0 and 1. After data preprocessing, we applied scikit-learn to train regressors for each individual, respectively. Additionally, a grid parameter search was applied to find the best parameters for each regressor [75]. The performances of our models were then evaluated by the tenfold cross-validation technique, which uses 10% of the data as the testing data and uses the remaining instances as training data to train the regressor. Additionally, the prediction accuracy of a regressor was evaluated in terms of the correlation (r) between the actual and predicted scores [16].

Table 1. Mean correlation of recognition models of different participants.

	Happy	Relaxed	Sad	Angry
MLR	-0.04	0.01	0.02	0.03
SVR	0.14	0.07	0.10	0.01
RFR	0.17	0.22	0.17	0.15

Table 1 shows the average performance of each recognition model for different emotions. In general, physiological features can predict the perceived emotion ratings of happy ($r = 0.17$), relaxed ($r = 0.22$), sad ($r = 0.17$), and angry ($r = 0.15$), but the predictive effect is poor. This result indicates that there is a weak connection between perceived emotion and physiological features and proves Hypothesis 1. Notably, models using the RFR method performed significantly better than models using the SVR and MLR methods (Happy: $\chi^2 = 26.54$, $p < 0.01$; Relaxed: $\chi^2 = 18.96$, $p < 0.01$; Sad: $\chi^2 = 11.06$, $p < 0.01$; Angry: $\chi^2 = 8.667$, $p < 0.05$). The poor performance of the MLR method reflects that the physiological features do not have a purely linear relationship with the perceived emotions of music. Additionally, the RFR method, which creates an ensemble of decision trees to predict the perceived emotions, may be more suitable for explaining the relationship between the variables. In addition, we observed that when predicting different emotion ratings, there is no significant difference in model prediction results ($\chi^2 = 3.48$, $p = 0.32$), although the average correlation is the highest when predicting perceived relaxed ratings. This shows that the physiological features considered in this study do not have a significant gap in predicting different perceived emotions.

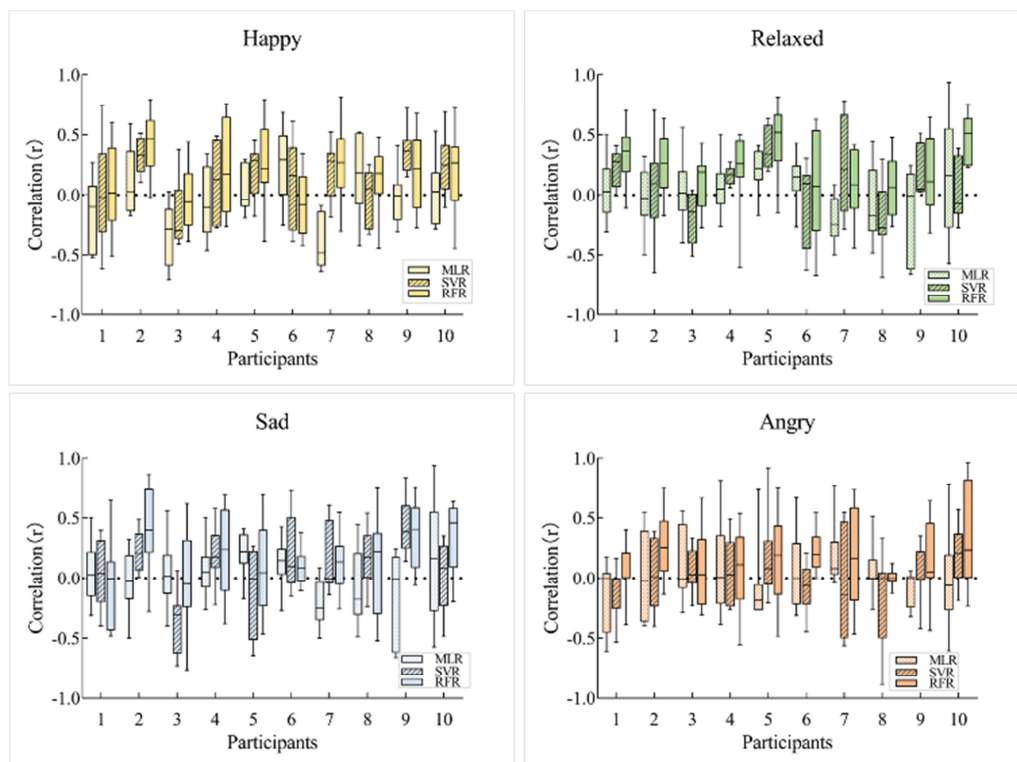


Figure 3. Prediction accuracy of each recognition model for each participant. Each compact boxplot indicates the prediction accuracy of each model by showing the correlations (r) in different cross-validation folds. The black line in the middle of each box indicates the median.

However, the aforementioned findings are not consistent with those of every individual. Figure 3 shows the prediction accuracy (r) of each recognition model for each participant. We observed that the model prediction effects of some individuals, such as the RFR-based recognition models of participant 2 (Happy: $M(r) = 0.42$, $SD = 0.26$; Relaxed: $M(r) = 0.26$, $SD = 0.25$; Sad: $M(r) = 0.41$, $SD = 0.35$; Angry: $M(r) = 0.28$, $SD = 0.28$) and participant 10 (Happy: $M(r) = 0.18$, $SD = 0.35$; Relaxed: $M(r) = 0.48$, $SD = 0.20$; Sad: $M(r) = 0.34$, $SD = 0.28$; Angry: $M(r) = 0.35$, $SD = 0.42$), which are much better than those of other participants (Happy: $\chi^2 = 16.91$, $p = 0.05$; Relaxed: $\chi^2 = 22.08$, $p < 0.01$; Sad: $\chi^2 = 15.34$, $p = 0.08$; Angry: $\chi^2 = 9.17$, $p = 0.42$). Additionally, some participants' recognition models are almost ineffective, such as participant 3 (Happy: $M(r) = -0.03$, $SD = 0.26$; Relaxed: $M(r) = 0.09$, $SD = 0.23$; Sad: $M(r) = -0.02$, $SD = 0.42$; Angry: $M(r) = 0.07$, $SD = 0.32$). These results are a reminder that when using only physiological features to construct MER models, individual differences affect the model effect.

Analyzing what traits caused the aforementioned differences

may become an important further research direction. However, individuality is too subtle to be captured by each individual factor [75], and it is difficult to analyze the impact of all individual factors. Additionally, by using only individual data to build a personalized recognition model, it is possible to avoid the deviations caused by differences in some individual features. For instance, personality differences may affect the assessment of music emotions [31], but personality differences can be disregarded when constructing a model using only individual data, because differences are directly reflected in the ground-truth values (perceived emotion ratings in this study). This reason is also one of the reasons why personalized models have been widely promoted and applied in recent years [4, 47]. In addition, from a practical perspective, the use of physiological features alone cannot effectively recognize the perceived emotions of some individuals. Therefore, in the subsequent modeling, we used physiological features as an additional input to the PMER model based on audio features and compared the changes in model recognition effects (Section 4.4).

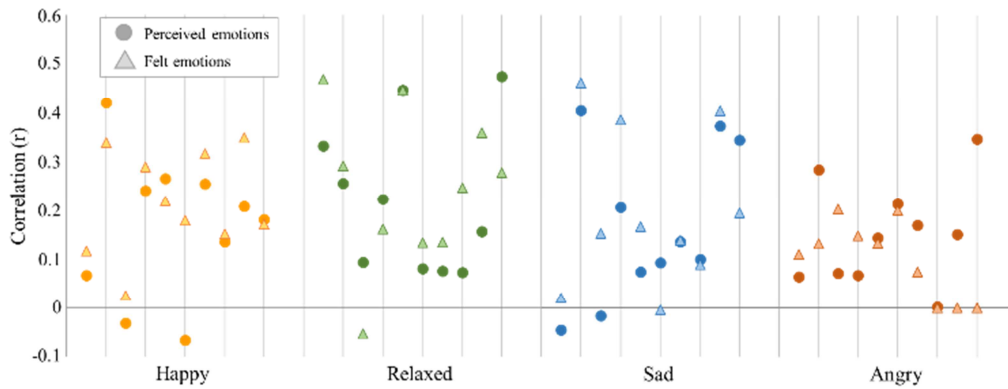


Figure 4. Mean prediction accuracy for each participant when using physiological features to predict perceived and felt emotions.

4.3. Comparison of Physiological Features-Based Perception and Feeling Models

We then attempted to verify Hypothesis 2 by comparing the prediction effects of physiological features-based perception and feeling models. We used physiological features as inputs—perceived and felt emotion ratings as ground truth, respectively—to construct recognition models. In addition, because RFR performed best in Section 4.2 and was widely applied in HER studies [72], we only used the RFR method in the subsequent modeling. Similarly, for each participant, we built eight types of personalized recognition models separately to predict the felt and perceived emotion ratings (happy, relaxed, sad, and angry). The performances of our models were also evaluated by a tenfold cross-validation technique, and the mean correlation (r) between actual and predicted scores is presented in Figure 4.

The result shows that the performance of the perception model and the feeling model is related in predicting happy, relaxed, and sad emotions (Happy: $r(9) = 0.78$, $p < 0.01$;

Relaxed: $r(9) = 0.65$, $p < 0.05$; Sad: $r(9) = 0.79$, $p < 0.01$). This discovery once again illustrates the close connection between felt emotion and perceived emotion of music. Next, by comparing the prediction effects of perception and feeling models, we found that generally, physiological features did not perform significantly better in predicting felt emotions (Happy: $Z = 1.58$, $p = 0.11$; Relaxed: $Z = 0.56$, $p = 0.58$; Sad: $Z = 1.07$, $p = 0.29$; Angry: $Z = 1.17$, $p = 0.24$), which does not support Hypothesis 2. Additionally, we observed substantial individual differences in prediction accuracy and in the difference between the accuracy of the perception and feeling models; in particular, the latter may imply two different internal mechanisms. When the individual's perception model performed better, felt emotion might be caused by the internal “mimicry” of the perceived emotion [26]. Therefore, the physiological features performed better in predicting perceived emotions. Conversely, when the individual's feeling model performed better, felt emotion might be caused by visual imagery [46], episodic memory [5], or other reasons [28]. Additionally, physiological arousal was associated with emotional arousal,

which indirectly influenced the perceived emotions. As a result, the physiological features performed better in

predicting felt emotions for these individuals. Of course, the aforementioned guess requires further verification.

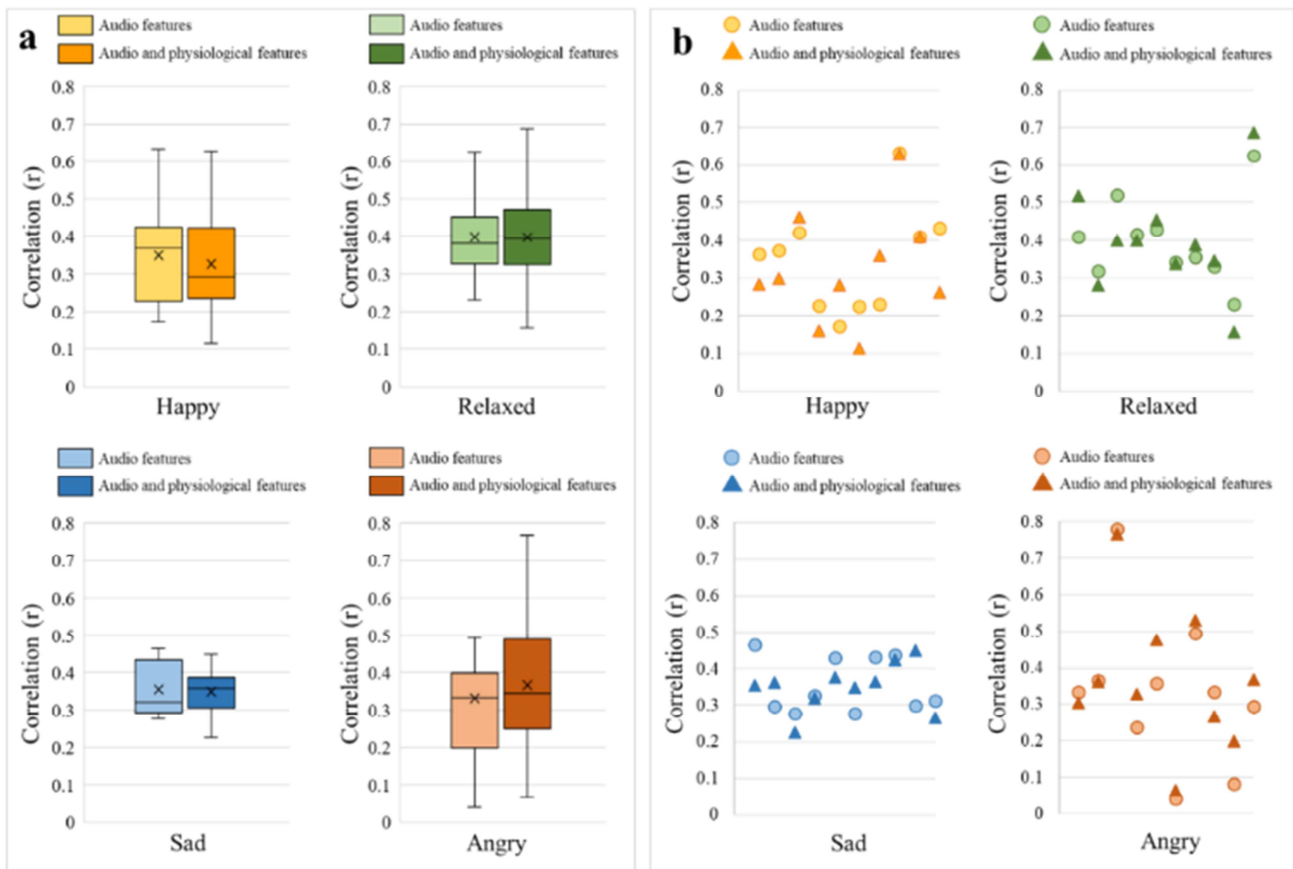


Figure 5. Comparison of models using audio features only and models using audio and physiological features. (a) overall distribution of the prediction accuracy (r) of all the participants. The black line in the middle of each box indicates the median; “ \times ” indicates the mean value. (b) mean prediction accuracy (r) for each participant.

4.4. Effects of Physiological Features on PMER Models

According to the result in Sections 4.2 and 4.3, we used physiological features to improve the performance of PMER models based on audio features and attempted to verify Hypothesis 3. First, we used audio features as input and perceived emotion ratings as the ground truth and applied RFR to construct basic PMER models for each individual. Tenfold cross-validation was also used to evaluate the performance of our models. Additionally, the basic PMER models achieved a mean correlation value of 0.35 for perceived happy ratings, 0.40 for perceived relaxed ratings, 0.36 for perceived sad ratings, and 0.33 for perceived angry ratings. Next, we added physiological features as additional inputs to the aforementioned PMER models. The result shows that the model effect was not significantly improved (Figure 5a; Happy: $Z = -0.76$, $p = 0.45$; Relaxed: $Z = -0.05$, $p = 0.96$; Sad: $Z = -0.36$, $p = 0.72$; Angry: $Z = -1.48$, $p = 0.14$), and this does not support Hypothesis 3.

Thus, on the whole, physiological features cannot effectively improve the effect of the PMER model, which may cause difficulties in further applications. In addition, considering the individuality, Figure 5b depicts the predictive effect (r) of each

PMER model. We observed substantial individual differences in the improvement rate of the current model (adding physiological features) compared with the basic model. For example, the improvement rate of the PMER models of participant 5 reached 64.49% for happy, which was significantly better than that of other participants. This is a reminder that physiological features can help predict the perceived emotion of music for certain individuals, and the characteristics of these individuals requires further investigation.

To better understand which features played a major role, we then explained the model by examining the information gain of features. Because the final PMER models were built by RFR, they can be interpreted by calculating feature importance [51]. In Figures 6a–d, we observed that for the final models, the feature importance values of the audio features accounted for 74.36% of happy, 75.93% of relaxed, 75.19% of sad, and 74.57% of angry. Thus, in general, audio features dominate the prediction of perceived emotions, which is similar to the results of [77]. Individual differences were also observed. For instance, Figures 6e and 6f show the distribution of feature importance values for the happy emotion recognition models of participants 6 and 8, respectively. We observed that audio features accounted for

77.29% of participant 6 and 69.88% of participant 8, and the same physiological features showed different feature importance (e.g., *EDA_SCL_COEFB* is crucial to participant 8 but not to participant 6).

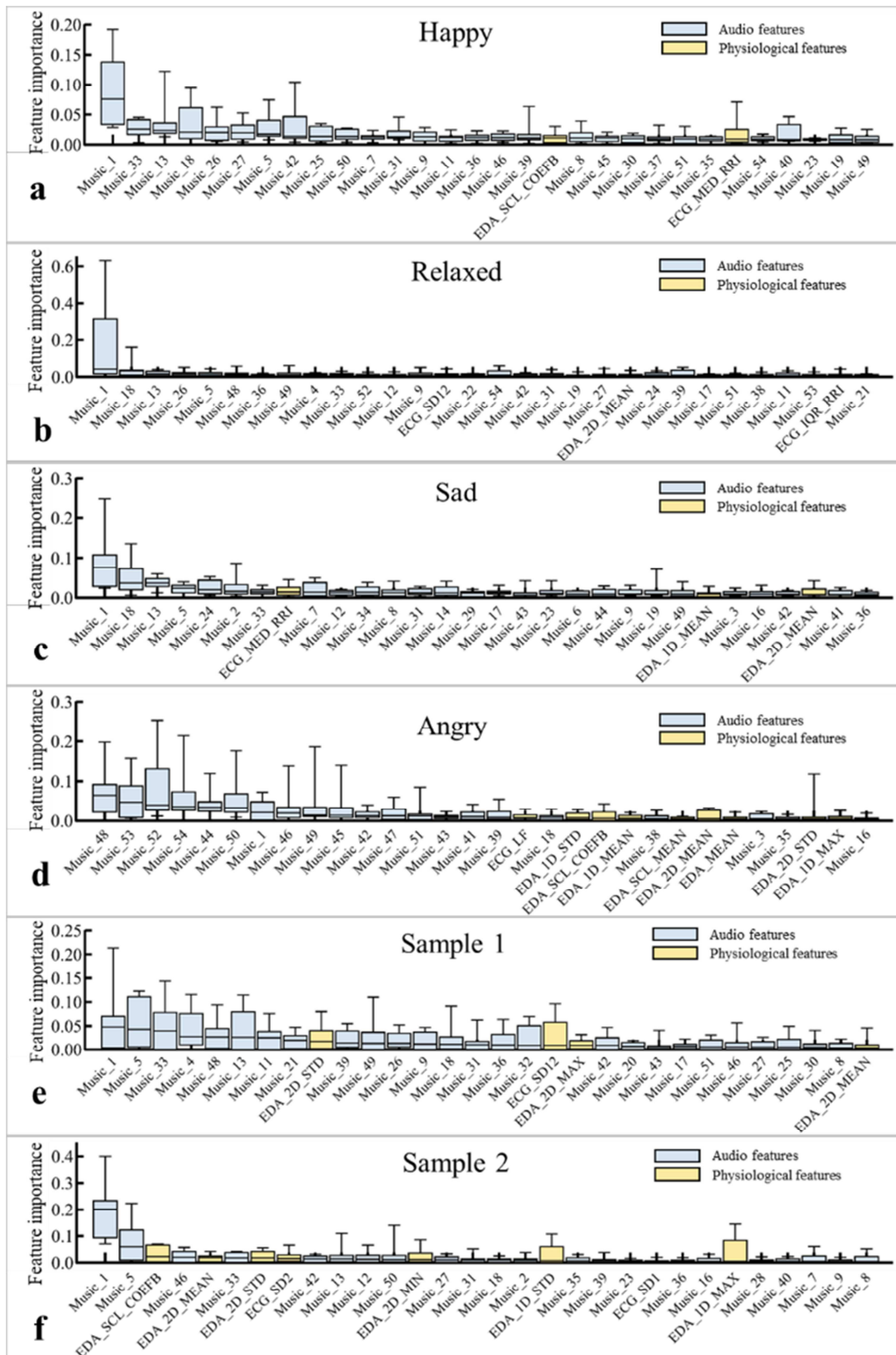


Figure 6. Distribution of feature importance of each feature for the final PMER models. Boxplots are sorted by median, and only the top 30 features were included for visibility with the trend of the remaining features being approximately the same. a–d, the overall distribution of feature importance values for each feature across all participants. e–f, the distribution of feature importance values for the happy emotion recognition models of participants 6 and 8.

5. Conclusion

In this article, the effects of physiological features on PMER models were investigated by (a) constructing recognition models of perceived emotions using physiological features only, (b) comparing the performance of physiological features-based perception and feeling models, and (c) adding physiological features as additional inputs to audio features-based PMER models. Substantial individual differences were observed in all the three steps, which verified the advocacy of other studies [75, 77], that is, MER studies should consider individuality. From a theoretical perspective, by applying ML methods, this study formed relations among audio features, physiological features, and perception of music emotions, which investigated their relationship by using computational modeling [68]. From an application perspective, we attempted to optimize the PMER models by adding physiological features, which can provide individuals with music emotion information that is more accurate. This may also improve the traditional MER system and indirectly contribute to the MIR [11] and music recommendation systems [48].

We first directly investigated the relationship between physiological features and perceived emotions by using only physiological features to construct perceived emotion recognition models. The results show that physiological features can slightly predict perceived emotions, and the prediction effect of different individuals is different. However, the relationship between physiological features and perceived emotion has no causal explanation. We speculated that musical stimuli first evoke individual emotions, affecting the assessment of perceived emotions. Therefore, we then compared the difference in the results of predicting perceived and felt emotions, respectively, through physiological features. If the aforementioned assumptions are true, the performance of feeling models should be improved. However, the results show that some individuals' feeling models performed better than the perception models, and others did not. The aforementioned difference might be related to the mechanism by which music evokes emotions. One of the mechanisms in the BRECVEMA framework [27]; Juslin et al. indicates that individuals have an internal "mimicry" of the perceived emotion of music (Contagion) [26, 30]. In this case, the individual's physiological state might be more effective in predicting perceived emotions. Conversely, for individuals whose emotions are evoked by other mechanisms, the perception models may perform better. Notably, the aforementioned speculation requires further verification.

Considering real-life applications, we then added physiological features as additional model inputs to the PMER models based on audio features for exploration. The results show that the model effect of some individuals was significantly improved; thus, the models have obtained more information from physiological features. However, the model effect of other individuals declined, which is a reminder that

the increase of redundant information may reduce the model effect. Notably, music features, external clues directly collected by the human auditory system, are usually used as the only input for MER research [74]. However, additional input of individual features can also improve the effect of the PMER model [77]. Starting from causality, this study attempted to optimize the model effect by adding physiological features, to increase the accuracy of the music emotion information provided to individuals. However, we found that the physiological feature is a double-edged sword and is only effective for certain individuals. Therefore, in subsequent applications, we must pre-test each individual to determine whether to input physiological features into the individual's PMER model.

This study has several notable limitations. First, the sample size of this study is relatively small and may be insufficient robust to support the general results. In further research, we can collect more samples to strengthen the reliability of the general results. Notably, it is difficult for participants to annotate a large number of music excerpts while collecting physiological data, and the quality may decline because of fatigue or sudden movements [68]. Thus, a better approach for collecting annotation is necessary. Second, only traditional ML methods were considered in this study. If a sufficiently large dataset is obtained, more flexible methods, such as deep neural networks [45] and recurrent neural networks [39], can be applied to pursue model effects. Finally, this study used computational modeling methods to investigate the relationship between variables, which may be less effective in explaining samples than traditional statistical analysis. Psychological research using ML methods tends to focus on prediction to explore trends and laws from data to achieve the generalization of results [81]. Therefore, in the process of data analysis, trade-offs must continue to be made between (a) constructing a theoretically supported, simple, and interpretable model with limited repeatability and (b) constructing a model with strong predictive power but an insufficient understanding of the internal mechanism of the current data.

References

- [1] Almond, A. (2017). Acqknowledge (Software). The International Encyclopedia of Communication Research Methods, 1-2.
- [2] Andridge, R. R., & Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International statistical review*, 78 (1), 40-64.
- [3] Bai, J., Peng, J., Shi, J., Tang, D., Wu, Y., Li, J., & Luo, K. (2016, August). Dimensional music emotion recognition by valence-arousal regression. In *2016 IEEE 15th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)* (pp. 42-49). IEEE.
- [4] Barros, P., Parisi, G., & Wermter, S. (2019, May). A Personalized Affective Memory Model for Improving Emotion Recognition. In *International Conference on Machine Learning* (pp. 485-494).

- [5] Baumgartner, H. (1992). Remembrance of things past: Music, autobiographical memory, and emotion. *Advances in Consumer Research*, 19, 613–620.
- [6] Chen, S. H., Lee, Y. S., Hsieh, W. C., & Wang, J. C. (2015, December). Music emotion recognition using deep Gaussian process. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)* (pp. 495–498). IEEE.
- [7] Chin, Y. H., Lin, C. H., Siahaan, E., Wang, I. C., & Wang, J. C. (2013, March). Music emotion classification using double-layer support vector machines. In *2013 1st International Conference on Orange Technologies (ICOT)* (pp. 193–196). IEEE.
- [8] Das, P., Khasnobish, A., & Tibarewala, D. N. (2016, June). Emotion recognition employing ECG and GSR signals as markers of ANS. In *2016 Conference on Advances in Signal Processing (CASP)* (pp. 37–42). IEEE.
- [9] De Nadai, S., D'Inca, M., Parodi, F., Benza, M., Trotta, A., Zero, E.,... & Sacile, R. (2016, June). Enhancing safety of transport by road by on-line monitoring of driver emotions. In *2016 11th System of Systems Engineering Conference (SoSE)* (pp. 1–4). IEEE.
- [10] Deng, J. J., Leung, C. H., Milani, A., & Chen, L. (2015). Emotional states associated with music: Classification, prediction of changes, and consideration in recommendation. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5 (1), 4.
- [11] Downie, J. S. (2008). The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology*, 29 (4), 247–255.
- [12] Edelberg, R. (1967). Electrical properties of the skin. *Methods in psychophysiology*, 1–53.
- [13] Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6 (3–4), 169–200.
- [14] Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different?. *Musicae scientiae*, 5 (1_suppl), 123–147.
- [15] Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Series in affective science. Music and emotion: Theory and research* (pp. 223–248). Oxford University Press.
- [16] Gladstone, J. J., Matz, S. C., & Lemaire, A. (2019). Can psychological traits be inferred from spending? Evidence from transaction data. *Psychological science*, 30 (7), 1087–1096.
- [17] Gong, P., Ma, H. T., & Wang, Y. (2016, June). Emotion recognition based on the multiple physiological signals. In *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)* (pp. 140–143). IEEE.
- [18] Guendil, Z., Lachiri, Z., Maaoui, C., & Pruski, A. (2015, December). Emotion recognition from physiological signals using fusion of wavelet based features. In *2015 7th International Conference on Modelling, Identification and Control (ICMIC)* (pp. 1–6). IEEE.
- [19] Guo, R., Li, S., He, L., Gao, W., Qi, H., & Owens, G. (2013, May). Pervasive and unobtrusive emotion sensing for human mental health. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops* (pp. 436–439). IEEE.
- [20] Herbert, R. (2013). An empirical study of normative dissociation in musical and non-musical everyday life experiences. *Psychology of Music*, 41 (3), 372–394.
- [21] Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: similarities, differences, and mixed emotions. *Psychology of Aesthetics Creativity & the Arts*, 4 (1), 47–56.
- [22] Hsu, Y. L., Wang, J. S., Chiang, W. C., & Hung, C. H. (2017). Automatic ecg-based emotion recognition in music listening. *IEEE Transactions on Affective Computing* (Vol. 11, pp. 85–99).
- [23] Izard, C. E. (1977). Differential emotions theory. In *Human emotions* (pp. 43–66). Springer, Boston, MA.
- [24] Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on psychological science*, 2 (3), 260–280.
- [25] Jun, S., Rho, S., Han, B. J., & Hwang, E. (2008). A fuzzy inference-based music emotion recognition system. In *5th International Conference on Visual Information Engineering (VIE)* (pp. 673–677).
- [26] Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin, & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). Oxford, UK: Oxford University Press.
- [27] Juslin, P. N., & Daniel Västfjäll. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behavioral & Brain Sciences*, 31 (6), 751–751.
- [28] Juslin, P. N., Harmat, L., & Eerola, T. (2013). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, 42 (4), 599–623.
- [29] Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research*, 33 (3), 217–238.
- [30] Juslin, P. N., Liljeström, S., Västfjäll, D., & Lundqvist, L. O. (2010). How does music evoke emotions? Exploring the underlying mechanisms. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 605–642). Oxford, UK: Oxford University Press.
- [31] Kallinen, K., & Ravaja, N. (2006). Emotion perceived and emotion felt: Same and different. *Musicae Scientiae*, 10 (2), 191–213.
- [32] Kawakami, A., & Katahira, K. (2015). Influence of trait empathy on the emotion evoked by sad music and on the preference for it. *Frontiers in psychology*, 6, 1541.
- [33] Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (12), 2067–2083.
- [34] Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J.,... & Turnbull, D. (2010, August). Music emotion recognition: A state of the art review. In *Proceedings of International Society for Music Information Retrieval (IMSIR)* (Vol. 86, pp. 937–952).

- [35] Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychology of Aesthetics, Creativity, and the Arts*, 2 (2), 115.
- [36] Laukka, P. (2007). Uses of music and psychological well-being among the elderly. *Journal of happiness studies*, 8 (2), 215.
- [37] Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T. L., Jeng, S. K., Duann, J. R., & Chen, J. H. (2010). EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57 (7), 1798-1806.
- [38] Liu, D., Lu, L., & Zhang, H. J. (2003). Automatic mood detection from acoustic music data. In *Proceedings of International Society for Music Information Retrieval (IMSIR)* (pp. 81-87).
- [39] Malik, M., Adavanne, S., Drossos, K., Virtanen, T., Ticha, D., & Jarina, R. (2017). Stacked convolutional and recurrent neural networks for music emotion recognition. *arXiv preprint arXiv: 1706.02292*.
- [40] McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015, July). librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference* (Vol. 8).
- [41] Monajati, M., Abbasi, S. H., Shabaninia, F., & Shamekhi, S. (2012). Emotions states recognition based on physiological parameters by employing of fuzzy-adaptive resonance theory. *International Journal of Intelligence Science*, 2, 166-175.
- [42] Mualem, O., & Lavidor, M. (2015). Music education intervention improves vocal emotion recognition. *International Journal of Music Education*, 33 (4), 413-425.
- [43] Naji, M., Firoozabadi, M., & Azadfallah, P. (2014). Classification of music-induced emotions based on information fusion of forehead biosignals and electrocardiogram. *Cognitive Computation*, 6 (2), 241-252.
- [44] Nardelli, M., Valenza, G., Greco, A., Lanata, A., & Scilingo, E. P. (2015). Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, 6 (4), 385-394.
- [45] Orjeseck, R., Jarina, R., Chmulik, M., & Kuba, M. (2019). DNN-Based music emotion recognition from raw audio signal. In *2019 29th International Conference Radioelektronika* (pp. 1-4). New York, NY: IEEE.
- [46] Osborne, J. W. (1980). The mapping of thoughts, emotions, sensations, and images as responses to music. *Journal of Mental Imagery*, 5, 133-136.
- [47] Panwar, S., Rad, P., Choo, K. K. R., & Roopaci, M. (2019). Are you emotional or depressed? Learning about your emotional state from your music using machine learning. *Journal of Supercomputing*, 75 (6), 2986-3009.
- [48] Park, S. H., Ihm, S. Y., Jang, W. I., Nasridinov, A., & Park, Y. H. (2015). A music recommendation method with emotion recognition using ranked attributes. In *Computer Science and its Applications* (pp. 1065-1070). Springer, Berlin, Heidelberg.
- [49] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12 (Oct), 2825-2830.
- [50] Pouyanfar, S., & Sameti, H. (2014, February). Music emotion recognition using two level classification. In *2014 Iranian Conference on Intelligent Systems (ICIS)* (pp. 1-6). IEEE.
- [51] Quiroz, J. C., Geangu, E., & Yong, M. H. (2018). Emotion Recognition Using Smart Watch Sensor Data: Mixed-Design Study. *JMIR mental health*, 5 (3), e10153.
- [52] Raykar, V. C., Yu, S., Zhao, L. H., Valadez, G. H., Florin, C., Bogoni, L., & Moy, L. (2010). Learning from crowds. *Journal of Machine Learning Research*, 11 (Apr), 1297-1322.
- [53] Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39 (6), 1161.
- [54] Scherer, K., & Zentner, M. (2008). Music evoked emotions are different—more often aesthetic than utilitarian. *Behavioral and Brain Sciences*, 31 (5), 595-596.
- [55] Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception: An Interdisciplinary Journal*, 21 (4), 561-585.
- [56] Schubert, E. (2013). Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Frontiers in psychology*, 4, 837.
- [57] Sease, R., & McDonald, D. W. (2011). The organization of home media. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18 (2), 1-20.
- [58] Sen, A., & Srivastava, M. (2012). Multiple regression. In *Ashish Sen, & Muni Srivastava (Eds.), Regression analysis: theory, methods, and applications* (pp. 28-49). Springer Science & Business Media.
- [59] Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D.,... & Yang, X. (2018). A review of emotion recognition using physiological signals. *Sensors*, 18 (7), 2074.
- [60] Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In P. N. Juslin & J. A. Sloboda (Eds.), *Series in affective science. Music and emotion: Theory and research* (p. 71-104). Oxford University Press.
- [61] Soleymani, M., Caro, M. N., Schmidt, E. M., Sha, C. Y., & Yang, Y. H. (2013, October). 1000 songs for emotional analysis of music. In *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia* (pp. 1-6). ACM.
- [62] Song, T., Zheng, W., Lu, C., Zong, Y., Zhang, X., & Cui, Z. (2019). MPED: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access*, 7, 12177-12191.
- [63] Thayer, R. E. (1989). *The Biopsychology of Mood and Arousal*, New York, Oxford University Press.
- [64] Udovičić, G., Đerek, J., Russo, M., & Sikora, M. (2017, October). Wearable emotion recognition system based on GSR and PPG signals. In *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care* (pp. 53-59).
- [65] Valenza, G., Citi, L., Lanata, A., Scilingo, E. P., & Barbieri, R. (2014). Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. *Scientific reports*, 4, 4998.
- [66] Van Dooren, M., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & behavior*, 106 (2), 298-304.

- [67] Vanderlei, F. M., de Abreu, L. C., Garner, D. M., & Valenti, V. E. (2016). Symbolic analysis of heart rate variability during exposure to musical auditory stimulation. *Altern. Ther. Health Med*, 22, 24-31.
- [68] Vempala, N. N., & Russo, F. A. (2018). Modeling Music Emotion Judgments Using Machine Learning Methods. *Frontiers in psychology*, 8, 2239.
- [69] Verschuere, B., Crombez, G., Koster, E., & Uzieblo, K. (2006). Psychopathy and physiological detection of concealed information: A review. *Psychologica Belgica*, 46 (1-2).
- [70] Wagner, J., Kim, J., & André, E. (2005, July). From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In 2005 IEEE international conference on multimedia and expo (pp. 940-943). IEEE.
- [71] Wang, J. C., Yang, Y. H., Wang, H. M., & Jeng, S. K. (2012, October). The acoustic emotion Gaussians model for emotion-based music annotation and retrieval. In Proceedings of the 20th ACM international conference on Multimedia (pp. 89-98). ACM.
- [72] Wen, W., Liu, G., Cheng, N., Wei, J., Shangguan, P., & Huang, W. (2014). Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Transactions on Affective Computing*, 5 (2), 126-140.
- [73] Wong, W. M., Tan, A. W., Loo, C. K., & Liew, W. S. (2010, December). PSO optimization of synergetic neural classifier for multichannel emotion recognition. In 2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC) (pp. 316-321). IEEE.
- [74] Yang, X., Dong, Y., & Li, J. (2018). Review of data features-based music emotion recognition methods. *Multimedia Systems*, 24 (4), 365-389.
- [75] Yang, Y. H., Su, Y. F., Lin, Y. C., & Chen, H. H. (2007, September). Music emotion recognition: The role of individuality. In Proceedings of the international workshop on Human-centered multimedia (pp. 13-22). ACM.
- [76] Yazdani, A., Kappeler, K., & Ebrahimi, T. (2011, November). Affective content analysis of music video clips. In Proceedings of the 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies (pp. 7-12).
- [77] Xu, L., Wen, X., Shi, J., Li, S., Xiao, Y., Wan, Q., & Qian, X. (2020). Effects of individual factors on perceived emotion and felt emotion of music: Based on machine learning methods. *Psychology of Music*. Advance online publication.
- [78] Zangwill, N. (2011). Music, essential metaphor, and private language. *American Philosophical Quarterly*, 48 (1), 1-16.
- [79] Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, 8 (4), 494.
- [80] Zong, C., & Chetouani, M. (2009, December). Hilbert-Huang transform based physiological signals analysis for emotion recognition. In 2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) (pp. 334-339). IEEE.
- [81] Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12 (6), 1100-1122.
- [82] Tulppo, M. P., Makikallio, T. H., Takala, T. E., Seppanen, T. H. H. V., & Huikuri, H. V. (1996). Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *American journal of physiology-heart and circulatory physiology*, 271 (1), H244-H252.
- [83] Xu, L., Zheng, Y., Xu, D., & Xu, L. (2021). Predicting the Preference for Sad Music: The Role of Gender, Personality, and Audio Features. *IEEE Access*, 9, 92952-92963.