Cogn Comput (2014) 6:241–252
DOI 10.1007/s12559-013-9239-7

Classification of Music-Induced Emotions Based on Information Fusion of Forehead Biosignals and Electrocardiogram

Mohsen Naji · Mohammd Firoozabadi · Parviz Azadfallah

Received: 24 November 2012 / Accepted: 16 November 2013 / Published online: 28 November 2013

Abstract Emotion recognition systems have been developed to assess human emotional states during different experiences. In this paper, an approach is proposed for recognizing music-induced emotions through the fusion of three-channel forehead biosignals (the left temporalis, frontalis, and right temporalis channels) and an electrocardiogram. The classification of four emotional states in an arousal–valence space (positive valence/low arousal, positive valence/high arousal, negative valence/high arousal, and negative valence/low arousal) was performed by employing two parallel support vector machines as arousal and valence classifiers. The inputs of the classifiers were obtained by applying a fuzzy-rough model feature evaluation criterion and sequential forward floating selection algorithm. An average classification accuracy of 88.78 % was achieved, corresponding to an average valence classification accuracy of 94.91 % and average arousal classification accuracy of 93.63 %. The proposed emotion recognition system may be useful for interactive multimedia applications or music therapy.

Keywords Emotion classification · Forehead biosignals · ECG · Arousal · Valence

M. Naji (✉)
Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
e-mail: m.naji@srbiau.ac.ir

M. Firoozabadi
Department of Medical Physics, Tarbiat Modares University, Tehran, Iran

P. Azadfallah
Department of Psychology, Tarbiat Modares University, Tehran, Iran

Introduction

From ancient times to the present day, people worldwide have listened to music for different reasons. The great Persian scientists Farabi (c. 872–950) and Avicenna (c. 980–1,037) established scientific principles concerning the musical treatment of the body and soul [1]. Today, music therapy to promote wellness is widely practiced in addition to other complementary treatments in the realm of psychotherapy. Schizophrenia, adolescent psychiatry, rehabilitation, psychosomatics, mental retardation, autism, etc., have been subjects of various studies involving applied music therapy [2]. As such, it would be useful to recognize a user's music-induced emotions without the need for self-assessment reports. Automatic emotion recognition could help music therapists as well as individual who have difficulty describing and identifying personal emotions.

The basic idea of emotion recognition arises from a large number of published works that have revealed associations between human emotions and their related neurophysiological responses. Measurements of central nervous system (CNS) responses through positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and electroencephalography (EEG) have shown that the frontal regions and auditory cortex have specific activity during musical processing [3]. Furthermore, the spectral power values of EEG bands are altered depending on the emotional type of musical stimuli [4, 5]. Activity in the peripheral nervous system (PNS) while listening to music can be obtained by measuring different signals such as galvanic skin response (GSR), blood pressure (BP), heart rate (HR), respiration rate, skin temperature (SKT), and facial expressions. Knight and Rickard showed that BP and HR decreased when listening to soothing music [6]. Bernardi et al. [7] found that ventilation, BP, and HR

increased with faster tempi and simpler structures compared with baseline measurements. Additionally, Kallinen found that mean HR values were lower during unpleasant music compared with pleasant music [8]. He concluded that low-arousal music elicited higher activation of the zygomaticus muscle (ZM) than did high-arousal music during eyes-closed listening, whereas high-arousal music elicited higher ZM activation than did low-arousal music during eyes-open listening. McFarland showed that soothing music terminated SKT decreases and perpetuated SKT increases, whereas annoying music terminated SKT increases and perpetuated SKT decreases [9]. Further, GSR is believed to be a relative indicator of emotional arousal; however, research has shown that GSR variation is dependent on listening habits [8].

In recent years, numerous researchers have examined emotion recognition during music listening. Janssen et al. [10] introduced an affective music player that was based on an estimation of probability density functions of GSR and SKT changes during music listening. Kim and André [11] investigated the potential of a surface electromyogram (EMG) of the trapezius muscle, electrocardiogram (ECG), GSR, and respiration changes for emotion recognition (positive–high arousal, negative–high arousal, negative–low arousal, and negative–low arousal) during music listening. More recently, Lin et al. [12] applied a support vector machine (SVM) to classify EEG patterns according to subjects' self-reported emotional states (i.e., joy, anger, sadness, and pleasure) during music listening with higher emotion classification accuracy that leads to higher performance for an emotion recognition system.

One of the most important factors that should be considered in designing emotion recognition systems is user comfort. In this regard, Firoozabadi et al. [13] proposed a novel biosignal acquisition method by locating three pairs of electrodes on participants' frontalis and temporalis forehead muscles. These forehead biosignals (FBS) convey the information on their adjacent standard EEG locations as well as facial expression information. In addition, these locations take the advantage of involvement in emotional processing. Rezazadeh et al. [14] applied FBS to the design of a control interface that can be adapted to a user's affective state. Further, Rad et al. [15] explored the effects of listening to pleasant and unpleasant music excerpts on the entropy of the alpha and EMG sub-bands of FBS. They concluded that the FBS provided informative data for emotion classification.

In order to improve the classification accuracy of emotion recognition systems, a fusion of the information from physiological signals could be performed. Using information from more than one signal modality could result in improved accuracy. The objective of this paper is to design an emotion recognition system to classify four music-induced emotions. This emotion recognition system should be able to serve as an alternative to questionnaires and self-reports about induced emotions. In order to design a high-accuracy emotion recognition system, we applied information fusion to FBS and ECG signals. We utilizes a feature-reduction algorithm based on a generalized fuzzy-rough model and employed support vector machines (SVMs) to classify the subjects' signals in an arousal–valence emotional space.

Materials and Methods

An emotion recognition system should be able to predict the outcome of self-reports and questionnaires for each listener on each song. Figure 1 illustrates the design stages of the proposed emotion recognition system. First, FBS as well as the main lead ECG signal are recorded during the application of appropriate musical stimuli. After this signal recording and preprocessing stage, several features from each of the biosignals are calculated. In the feature-level fusion, for different parameter values of the applied feature evaluation criterion, the most significant feature subsets involving different signal channels for arousal and valence classifications are extracted. For each feature subset, the classification accuracies are calculated. The four-class classification can be performed by juxtaposing the outputs of the most accurate valence and arousal classifiers. In the following sections, we explain this process in more detail.

Emotional Stimuli

According to cognitive theories of emotion, there is an essential cognitive basis for emotions. However, some emotions involve much less cognitive processing and structure for emotions [16]. A cognitive account of musical emotion is supported by empirical studies showing that

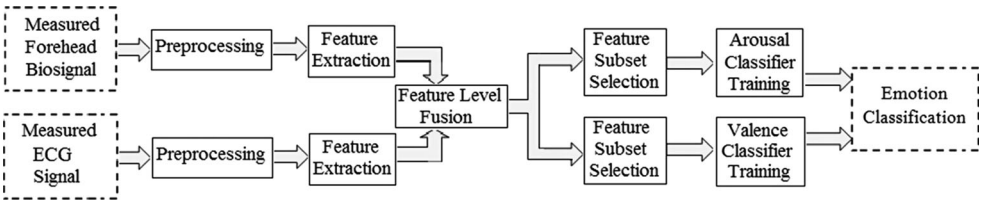

Fig. 1 Block diagram of stages involved for designing emotion classification system

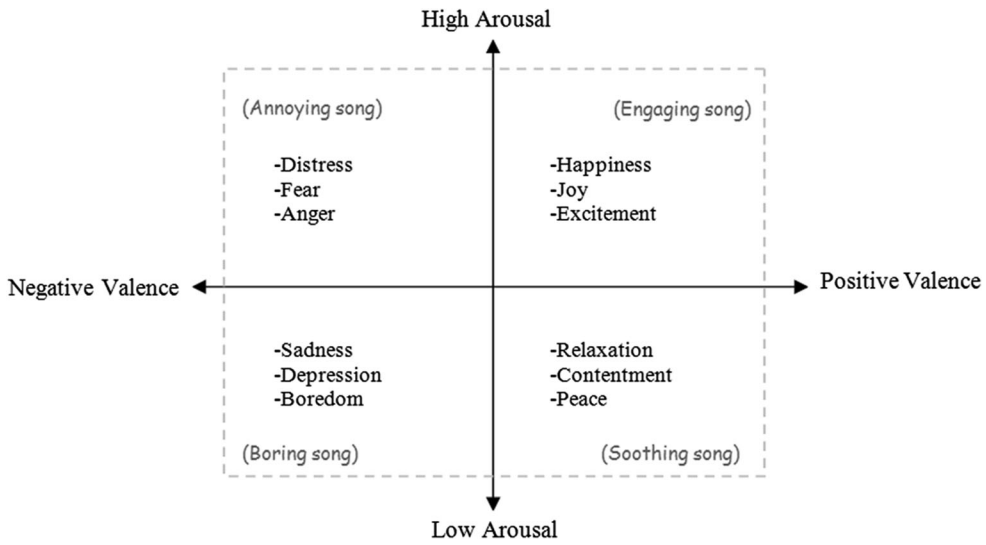

Fig. 2 Arousal–valence model of emotions with example emotions

emotional responses are systematically associated with the presence of specific musical features such as modulation, harmonic progression, and rhythm [17]. Juslin and Västfjäll argue that musical emotions must be based on the cognitive appraisal. They propose six other mechanisms that explain how musical pieces induce emotion: brain stem reflexes, evaluative conditioning, emotional contagion, visual imagery, episodic memory, and expectancies that are fulfilled or denied [18]. Konečni concluded that, however, music could lead to a variety of responses through mediators such as dance and cognitive associations to real-world events; *being moved* and *aesthetic awe* may be the most genuine and profound music-related emotional states [19].

Several-dimensional emotion categorization models have been proposed (e.g., Plutchik's model and the hourglass model [20]). According to the approach of Scholsberg [21], every emotion has a cognitive and a physiological component. That is, there are two emotion dimensions: arousal, which describes the extent of the calmness or excitation felt by people from low to high; and valence, which describes the level of pleasure or aversiveness from negative to positive. A schematic illustration of an arousal–valence model of emotion with example emotions is presented in Fig. 2. In the two-dimensional arousal–valence model, it is possible to represent emotions without using labels. The arousal–valence space has been shown to be effective for self-assessment of emotions and moods [22].

We place listeners' emotional states in the four quadrants of this arousal–valence plane. That is, emotions can be made to correspond to arousal–valence pairs using a set of musically descriptive adjectives: soothing (low arousal/positive valence), engaging (high arousal/positive valence), annoying (high arousal/negative valence), and boring (low arousal/negative valence). The music excerpts were chosen based on self-ratings in a pilot study of 50 participants with a wide range of ages and included about 15 pieces of music in different styles. The participants were asked to complete a questionnaire (Fig. 3), which consisted of questions about the subjective feelings elicited by each piece of music. A similar questionnaire had been used in a previous study [23]. After collecting the questionnaires, the selection of the pieces of music was performed. The decision about an induced emotion was extracted from values rated greater than 7 for a special feeling, while the rated values for three other feelings were less than 3. The extracted decisions were controlled by ratings on how much the subjects liked or disliked the songs, and by the self-selected feelings of the subjects (e.g., calm, sad, elated, etc.). In the selection of the pieces of music, priority was given to majority votes, less heard songs, and pieces of music that did not evoke memories. The musical excerpts were as follows: Pachelbel's *Canon in D major* (Lee Galloway), a *Persian 6/8* song, *Hi friend!* (Deadmau5 featuring MC Flipside), and *Romance* (Schumann's Symphony No. 4) for the soothing, engaging, annoying, and boring classes, respectively. It should be mentioned that the selected songs and their emotional labels were related to the culture and taste of the Iranian subjects.

Subjects and Experimental Procedure

Twenty-five healthy nonmusician volunteers participated; all were right-handed subjects (10 males, and 15 females) in the age-group of 19–28 years. None of the subjects had hearing impairment or a history of mental disorder. Because they were aware of the purpose of the experiment, they tried to concentrate as much as possible during the experiment.

- Have you heard the song before? Yes ☐ No ☐
- Does the song evoke memories for you? Yes ☐ No ☐
- Please rate your feeling of pleasure or displeasure by checking one box below:
  Totally dislike ☐--☐--☐--☐--☐--☐--☐ Totally like
- How much do you feel the song is soothing? (Please rate between 0 and 10) ----
- How much do you feel the song is engaging? (Please rate between 0 and 10) ----
- How much do you feel the song is annoying? (Please rate between 0 and 10) ----
- How much do you feel the song is boring? (Please rate between 0 and 10) ----
- Roughly speaking, how do you feel after hearing the song? (Please select from the following)

  Calm ☐ Relaxed ☐ Serene ☐ Contented ☐ Happy ☐ Elated ☐ Excited ☐ Alert ☐
  Tense ☐ Nervous ☐ Stressed ☐ Upset ☐ Sad ☐ Depressed ☐ Bored ☐ Fatigued ☐

- Please let we know if you have any comments.....

Fig. 3 Questionnaire for evaluation of music-induced emotions

Before placing any electrode, the electrode placement area was abraded lightly and cleaned with alcohol to lower the electrical impedance. Three pairs of pregelled Ag–AgCl electrodes (diameter = 10 mm; *Skintact*®, Innsbruck, Austria) were placed on a subject's facial muscles in a differentiation configuration to acquire FBS. One pair was placed on a subject's frontalis muscle: above the eyebrows with about a 3-cm inter-electrode distance (frontalis channel). Two pairs were placed on the left and right temporalis muscles with about a 4-cm inter-electrode distance (left temporalis channel and right temporalis channel). Figure 4 shows the locations of these electrode pairs. The ground electrode for FBS acquisition was placed on the left earlobe. The ECG recording was performed by placing pregelled Ag–AgCl electrodes on the palmar sides of the right and left wrists (lead I configuration), and one on the right leg.

The BIOPAC MP100 system (separate amplifiers; ack100w software version; Biopac Inc., Santa Barbara, CA) was used to collect the biosignals. A sampling frequency of 256 Hz was selected.

The subjects were seated in a comfortable chair in a quiet nonsoundproofed room with minimal light for the recording of the biosignals. They were instructed to keep their eyes closed, put on the headphones, and remain seated during the music-listening experiment. Data were recorded during 60 s of silence followed by 120 s of music. All of the musical excerpts were played (with at least a 5-min rest between runs) through a high-fidelity music player (W810 Walkman™ phone, Malaysia) for each subject; the intensity level of the songs was adjusted so that the listeners felt comfortable. The musical excerpts were presented in random order. Thus, the subjects could not anticipate what the next stimulus is. To avoid interference in the natural expression of emotion during music listening, a subject's pace was not rigidly controlled. In order to validate the success of emotion induction, the subjects were requested to complete a questionnaire (Fig. 3) to report their musically induced emotions after each run.

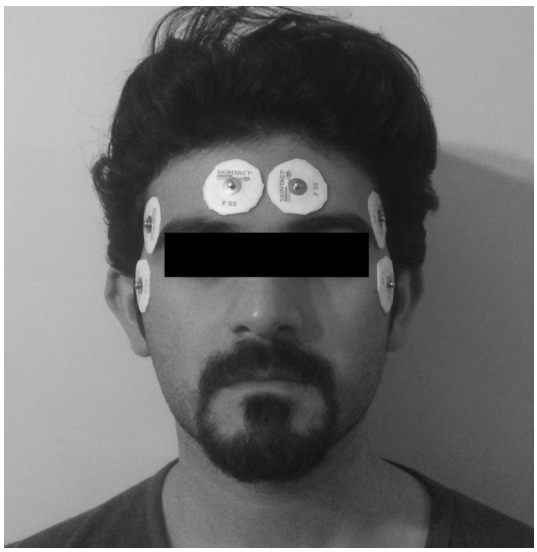

Fig. 4 Areas for electrode placement

Preprocessing and Feature Extraction

The time that each R wave of the ECG occurred was found using the AcqKnowledge software (version 3.7.3; BIOPAC systems). Afterward, the RR intervals (time intervals between consecutive R peaks) were confirmed by continual visual monitoring. In addition, the raw FBS were filtered (band-pass 1–100 Hz; band-stop 47–53 Hz).

All of the RR and filtered FBS were divided into unequal portions: a rest signal, which lasted 1 min and was acquired during the silence; and an emotional signal, which lasted 2 min and was obtained while listening to a musical excerpt. The emotional data were divided into 30-s segments [12], and in each segment, several features were computed for emotion classification purposes. In order to minimize inter-subject variability, the features were also computed for the rest of the signals to calculate the normalized features as follows:

$$F_j = \frac{F_j(\text{emotional}) - F(\text{rest})}{F(\text{rest})}, \quad j = 1, 2, 3, 4. \tag{1}$$

In (1), F (rest) is a feature value extracted from a rest signal, F_j (emotional) is a feature value extracted from the jth 30-s segment of the emotional signal, and F_j is the jth sample of the normalized feature of F during each recording. For each feature, the inter-subject variability after and before feature normalization was quantified using an analysis of variance (ANOVA). Considering the p value threshold of 0.05, there was no inter-subject variability after feature normalization. After feature normalization, the feature space was constructed by concatenation of the FBS- and ECG-based features. Because of the emotional data splitting, 25 subjects, and four excerpts, a total of 400 samples were obtained.

FBS Feature Extraction

The following features were extracted from the FBS:

- Relative powers

As mentioned in the "Introduction," listening to music affects CNS activation. Thus, the relative powers (RP) [23] of each EEG band could be used to assess emotional status. To estimate the FBS power spectral density (PSD) function, 256-point windowed epochs (Hanning windowing) were extended to 512 points by zero padding. Then, a 512-point short-time Fourier transform (STFT) was applied to compute PSD. The RP were calculated for the following frequency bands: theta (θ: 4–7 Hz), slow alpha (α1: 8–10 Hz), fast alpha (α2: 11–13 Hz), alpha (α: 8–13 Hz), slow beta (β1: 13–19 Hz), fast beta (β2: 20–30 Hz), beta (β 13–30 Hz), and gamma (γ 31–50 Hz). The relative powers were calculated using the total power of a frequency band divided by the signal power.

- Mean frequency

The mean frequency (MF) [24] in the range $f_1 - f_2$ of a FBS is calculated as

$$\text{MF} = \frac{\sum_{f_i=f_1}^{f_2} f_i \cdot P(f_i)}{\sum_{f_i=f_1}^{f_2} P(f_i)}, \tag{2}$$

where $P(f_i)$ is PSD at frequency f_i. We selected the range of 4–35 Hz (a frequency range including theta, alpha, and beta bands) for MF calculation. This feature has been used to quantify the level of mental arousal.

- Average nonlinear energy

For the FBS $x_{\text{FBS}}(n)$, the nonlinear energy (NE) operator [25] is represented by

$$\text{NE}[n] = x_{\text{FBS}}^2(n) - x_{\text{FBS}}(n-1)x_{\text{FBS}}(n+1) \tag{3}$$

It was found that this operator improves SNR and measures the instantaneous energy changes of signals. Because FBS may include facial expressions, the authors decided to use this operator for FBS processing. After NE is obtained, the feature is weighted with a Hanning window. Then, the mean of the windowed data, the average nonlinear energy (ANE), is calculated.

- Higher-order crossings

Higher-order crossings (HOC) [26] are obtained by counting the number of zero crossings in the filtered time series. The HOC of order m, HOC_m, for a zero-mean time series of $x_{\text{FBS}}(n)$ can be calculated as

$$\text{HOC}_m = \text{NZC}\{\nabla^{m-1}(x_{\text{FBS}}(n))\}, \quad m = 1, 2, 3, \ldots, 10 \tag{4}$$

where ∇ is a backward difference operator and NZC$\{\cdot\}$ denotes the number of zero crossings. The HOC feature was first used for an EEG-based emotion recognition system. In the present work, HOC_1 to HOC_{10} were calculated and divided by the duration of the FBS time series.

ECG Feature Extraction

After extracting the RR time series from the ECG signals, several features were calculated. Some of these features such as statistical features (e.g., mean and standard deviation), nonlinear features (e.g., sample entropy), and triangular phase space mapping have previously appeared in the ECG-based emotion assessment literature [7, 11, 27, 28]. In the present work, the following features were extracted ($x_{\text{RR}}(t)$, $t=t_1, \ldots, t_N$, is the N-sample RR signal):

- The mean of the RR signal

$$\text{RR}_{\text{mean}} = \frac{1}{N}\sum_{n=1}^{N} x_{\text{RR}}(t_n) \tag{5}$$

- The standard deviation of the RR signal

$$\text{RR}_{\text{std}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left(x_{\text{RR}}(t_n) - \text{RR}_{\text{mean}}\right)^2} \tag{6}$$

- The average waveform length

$$\mathrm{RR_{WL}} = \frac{1}{N-1}\sum_{n=1}^{N-1} |x_{\mathrm{RR}}(t_{n+1}) - x_{\mathrm{RR}}(t_n)| \tag{7}$$

- The slope of the regression curve of the RR signal

For the regression equation of $x = \mathrm{RR_{slp}}t + b$, the slope $\mathrm{RR_{slp}}$ is calculated as

$$\mathrm{RR_{slp}} = \left(N\sum_{n=1}^{N} t_n x_{\mathrm{RR}}(t_n) - \left(\sum_{n=1}^{N} t_n\right)\left(\sum_{n=1}^{N} x_{\mathrm{RR}}(t_n)\right)\right) \Big/ \left(N\sum_{n=1}^{N} t_n^2 - \left(\sum_{n=1}^{N} t_n\right)^2\right) \tag{8}$$

- Katz's fractal dimension

$$D^{\mathrm{Katz}} = \frac{\log_{10} L}{\log_{10} d} \tag{9}$$

In (9), L is the total waveform length (or $\mathrm{RR_{WL}}\,(N-1)$), and d is the largest distance between $x_{\mathrm{RR}}(t_1)$ and $x_{\mathrm{RR}}(t_i)$, $i = 2,\ldots N$ [27].

- Poincaré geometry

The Poincaré geometry is a feature that is extracted from a Poincaré plot. In this plot, each RR interval is plotted as a function of the previous RR interval. The level of the long-term HRV, called SD2, is assessed by computing the standard deviation of the RR points along a 45° axis, while the level of the short-term HRV, called SD1, is assessed by computing the standard deviation of the distances of the RR points to the 45° axis. The Poincaré geometry is defined as the ratio of SD1 to SD2 [27].

- Sample entropy

The procedure for estimating the SampEn algorithm consists of the following steps [27]:

1. Create m vectors defined as $\mathrm{X}(i) = [x_{\mathrm{RR}}(t_i)\, x_{\mathrm{RR}}(t_{i+1}) \cdots x_{\mathrm{RR}}(t_{i+m-1})]$ for $i = 1,\ldots,N-m+1$.
2. Calculate the Euclidian distance between two vectors: $|\mathrm{X}(i) - \mathrm{X}(j)|$.
3. Calculate the similarity measures of C_i^m and C_i^{m-1} for each $l \le i \le N-m+1$ as $C_i^m = (N-m+1)^{-1}\sum_{j=1,i\neq j}^{N-m+1} \Theta\{r - |X(i) - X(j)|\}$, where $\Theta\{\cdot\}$ denotes a Heaviside step function.
4. From C_i^m and C_i^{m-1}, calculate C^m and C^{m-1} as $C^m = (N-m+1)^{-1}\sum_{i=1}^{N-m+1} C_i^m$.
5. Estimate SampEn: $\mathrm{SampEn}(m,r) = -\ln\frac{C^m}{C^{m-1}}$.

We selected $m = 2$, 3, and $\mathrm{r} = 0.15\,\mathrm{std}(x_{\mathrm{RR}}(t))$, where std(.) denotes standard deviation. This parameter selection was similar to that in [11].

- Triangular phase space mapping (TPSM) features

A triangle is obtained by plotting the absolute value of the zero-mean normalized RR values as a function of the RR values. From this mapping, some geometric features such as the angles, area, and peripheral of the triangle could be extracted. The interested reader is referred to [28], in which the formulations of TPSM features were explicitly introduced. In this paper, the left-side angle, area, peripheral, and quality of TPSM were extracted.

Feature Evaluation Criterion

To evaluate the extracted features, we used a novel feature significance measure, presented by Hu et al. [29], which was derived from a generalized fuzzy-rough model. After normalizing the observations of each dimension in the feature space, the significance measure will be obtained after the calculations given below.

In a feature space of FS, the fuzzy equivalence class $[x_i]_{\mathrm{FS}}$ of observation x_i is defined as

$$[x_i]_{\mathrm{FS}} = \frac{r_{i1}}{x_1} + \frac{r_{i2}}{x_2} + \cdots + \frac{r_{iN}}{x_N}, \tag{10}$$

where N is the total number of observations, "+" means the union, and r_{ij} is the output of a symmetrical membership function that measures the value of the fuzzy similarity degree between x_i and x_j. That is, $r_{ij} = f(|x_i - x_j|)$, where $|x_i - x_j|$ represents an Euclidean distance between x_i and x_j. In this paper, a Gaussian similarity relation function is adopted:

$$r_{ij} = \exp\left(-|x_i - x_j|^2 \Big/ 2\sigma^2\right), \quad \sigma = 0.25 \tag{11.}$$

We can define the lower approximation of decision X as

$$\underline{\mathrm{FS}_k}X = \left\{x_i | I([x_i]_{\mathrm{FS}}, X) \ge k, x_i \in U\right\}, \quad 1 \ge k \ge 0.5, \tag{12}$$

where $I(A,B) = \frac{\sum_{x\in U}\mu_{A\cap B}(x)}{\sum_{x\in U}\mu_A(x)}$, and $\mu_A(x)$ is the *membership* degree of x in fuzzy set A. k is a parameter that reflects a user's tolerance to the degree of noise, with a smaller value for k, indicating a greater tolerance to noise. For a two-class problem (X_1 and X_2), the lower approximation of classification D is defined as

$$\underline{\mathrm{FS}_k}D = \left\{\underline{\mathrm{FS}_k}X_1, \underline{\mathrm{FS}_k}X_2\right\}. \tag{13}$$

Finally, the feature significance measure of feature space FS for classification D is calculated as

$$\gamma = \frac{|\mathrm{FS}_k D|}{N} \tag{14}$$

where |.| is the cardinality of a set. Obviously, $0 \leq \gamma \leq 1$. A higher value of γ indicates a higher class separability.

Feature Selection Strategy

For each parameter k of the described feature evaluation criterion, we used the sequential forward floating selection (SFFS) approach [30] to select informative feature subsets, *FSs*. Thereafter, these feature subsets were imposed on classifiers, and classification rates were calculated. The selection procedure consists of the following steps:

Step 1 Initialization: $\mathrm{FS}^0 = \emptyset; \quad i = 0;$

Step 2 Select the feature with the maximum γ ratio: $F^+ = \arg \max_{F^+ \notin \mathrm{FS}^i} [\gamma(\mathrm{FS}^i + F^+)]$.

Step 3 Inclusion: If $\gamma(\mathrm{FS}^i + F^+) > \gamma(\mathrm{FS}^i)$ then $\mathrm{FS}^{i+1} = \mathrm{FS}^i + F^+; \quad i = i + 1.$

Step 4 Find the least significant feature in FS^i: $F^- = \arg \max_{F^- \in \mathrm{FS}^i} [\gamma(\mathrm{FS}^i - F^-)];$

Step 5 Exclusion: If $\gamma(\mathrm{FS}^i - F^-) > \gamma(\mathrm{FS}^i)$ then $\mathrm{FS}^{i+1} = \mathrm{FS}^i - F^-; \quad i = i + 1;$ Go to Step 4, else go to Step 2

Step 6 End

Classifiers

This study employed binary support vector machines (SVMs) [30] for pattern classification. In an SVM, a kernel function is applied to map the input feature space to a high-dimensional feature space. In the learning phase of the SVM, the goal is to maximize the separation margin between two classes. After testing different kernels (radial basis, quadratic, and polynomial), this study employed a third-order polynomial kernel function for data projection.

To design an emotion recognition system that classifies music-induced emotions into four classes, positive valence/low arousal, positive valence/high arousal, negative valence/high arousal, and negative valence/low arousal, we performed the following steps:

1. Place the data into the two classes of low arousal and high arousal based on their labels.
2. Change threshold k of the feature evaluation criterion from 0.7 to 0.95 in 0.05 steps; select the optimal features for the high arousal–low arousal classification problem; and determine the average arousal classification accuracy for each k.
3. Determine threshold k_A for the most accurate arousal classifier.
4. Place the data into the two classes of negative valence and positive valence based on their labels.
5. Change threshold k of the feature evaluation criterion from 0.7 to 0.95 in 0.05 steps; select the optimal features for the positive valence–negative valence classification problem; and determine the average valence classification accuracy for each k.
6. Determine threshold k_V for the most accurate valence classifier.
7. Combine the outputs of the k_A-based arousal classifier and k_V-based valence classifier to develop the emotion recognition system.

In order to estimate the true power of the arousal or valence classifiers (in the second, fifth, and seventh above-mentioned steps), a hundred repetitions of the fourfold cross-validation technique were used. It should be noted that the same index data, with different features, were fed to the valence and arousal classifiers. The four-class classification results from the total emotion recognition system were determined by the binary outputs of the valence and arousal classifiers (Fig. 1). For example, when the valence classifier had an output of 1 and the arousal classifier had an output of 0, the class of the induced emotion was regarded as positive valence/low arousal. The rationale of fusing the valence and arousal classifiers is that the SVM classifiers are binary by nature. An alternative scheme for a four-class classification problem using SVMs is to train four independent SVMs to classify samples for one class in relation to the three other classes. This alternative method is referred to as "one-against-all" (OAA). Obviously, the proposed classifier fusion method requires fewer SVMs than OAA does.

Effect of Discarding Signal Modality

To evaluate the emotion recognition accuracy after discarding one of the signal modalities, the average classification rates were recalculated by applying FBS- and ECG-discarded selected feature subsets. The *t* test was performed for different subsets of features to determine statistical significance.

Results

Arousal Classification

Table 1 shows that the optimal selected feature subsets and corresponding arousal classification accuracies varied with the specified threshold of the feature evaluation criterion. Significantly, two sets of features had the maximum mean

arousal classification accuracy (93.63 and 93.36 % corresponding to $k_A = 0.85$ and $k_A = 0.9$). The best-selected feature subsets, labeled as arousal-discriminant feature subset (AFS), for classifying high-arousal and low-arousal emotional states are as follows:

- AFS1: Concatenation of RP_θ and FMN of the left temporalis channel; $RP_{\alpha 1}$, $RP_{\beta 2}$, and FMN of the frontalis channel; RP_θ, FMN, HOC_5, and ANE of the right temporalis channel; and SampEn(3, r) of the ECG channel.
- AFS2: Concatenation of RP_θ and $RP_{\alpha 1}$ of the left temporalis channel; $RP_{\alpha 1}$ and $RP_{\beta 2}$ of the frontalis channel; RP_θ, FMN, HOC_5, and ANE of the right temporalis channel; and SampEn(3,r) and the TPSM angle of the ECG channel.

The mean false-positive (FP) and false-negative (FN) rates of the selected feature subsets show that there is no notable classification bias toward high arousal or low arousal classes.

Valence Classification

Table 2 shows that the optimal valence-discriminant feature subsets and corresponding classification accuracies varied with the specified threshold of the feature evaluation criterion. Significantly, two sets of features had the maximum mean valence classification accuracy (94.88 and 94.91 % corresponding to $k_V = 0.85$ and $k_V = 0.9$). The best-selected feature subsets, labeled as valence-discriminant feature subset (VFS), for classifying the negative-valence and positive-valence emotions are as follows:

VFS1 Concatenation of RP_γ of the left temporalis channel; RP_θ, $RP_{\beta 1}$, and HOC_5 of the frontalis channel; $RP_{\alpha 1}$ and HOC_6 of the right temporalis channel; and RR_{mean} and the SampEn(2, r) feature of the ECG channel

VFS2 Concatenation of RP_γ of the left temporalis channel; RP_θ, $RP_{\beta 1}$, and HOC_5 of the frontalis channel; $RP_{\alpha 2}$, HOC_8, and FMN of the right temporalis channel; and RR_{mean} and the SampEn(2, r) feature of the ECG channel

The mean FP and FN rates of the selected feature subsets show that negative-valence samples have more classification error than positive-valence samples.

Final Results

As described in Sect. 2, in the final stage of the emotion recognition system design, the valence classifier and arousal classifier outputs were combined. By applying an input pattern, a binary output of 0 or 1 is obtained for each classifier. Therefore, by juxtaposing the outputs of the arousal and valence classifiers, a four-class emotion recognition system can be designed.

The overall classification accuracies, sensitivities, and specificities corresponding to the application of optimal AFSs and VFSs are tabulated in Table 3. As shown in this table, the maximum average classification accuracy of 88.78 % when classifying four-class emotional states is obtained by juxtaposing the outputs of the AFS1-input arousal classifier and VFS2-input valence classifier. The average sensitivities when classifying the four emotional states related to the maximum mean classification rate are 84.22, 91.84, 88.45, and 89.89 % for the negative valence/low arousal (boring), positive valence/low arousal (soothing), negative valence/high arousal (annoying), and positive valence/high arousal (engaging) classes, respectively. The corresponding average specificities are 98.01, 92.92, 97.57, and 95.84 % for the boring, soothing, annoying, and engaging classes, respectively. The sensitivity values are

Table 1 Selected features and corresponding arousal classification rates versus parameter of feature evaluation criterion

k	Selected features for left temporalis channel	Selected features for frontalis channel	Selected features for right temporalis channel	Selected features for ECG channel	Classification accuracy	Mean false positive	Mean false negative
0.7	RP_θ, HOC_2	HOC_7	FMN	TPSM angle, TPSM quality	89.41 (4.59)	10.62	10.94
0.75	RP_θ, HOC_2	RP_α, $RP_{\beta 2}$, $RP_{\alpha 1}$	FMN. HOC_5	Slope, SampEn2, TPSM angle	90.92 (4.66)	9.48	10.12
0.8	RP_θ, SE	RP_2, $RP_{\alpha 1}$, FMN	RP_θ, FMN, HOC_5, ANE	Std SampEn3	91.05 (4.59)	9.44	10.06
0.85	**RP_θ, FMN**	**$RP_{\alpha 1}$, $RP_{\beta 2}$, FMN**	**RP_θ, FMN, HOC_5, ANE**	**SampEn3**	**93.63 (4.1)**	6.28	6.46
0.9	**$RP_{\alpha 1}$, RP_θ**	**$RP_{\alpha 1}$, $RP_{\beta 2}$**	**RP_θ, FMN, HOC_5, ANE**	**SampEn3 TPSM angle**	**93.36 (4.3)**	6.30	6.52
0.95	RP_α	HOC_4	RP_θ	SampEn4 TPSM quality	80.29 (5.94)	18.26	20.94

k the parameter of fuzzy-rough model (feature evaluation criterion)

The bolded rates indicate the most classification accuracies among all the feature subsets ($p < 0.05$)

Table 2 Selected features and corresponding valence classification rates versus parameter of feature evaluation criterion

k	Selected features for left temporalis channel	Selected features for frontalis channel	Selected features for right temporalis channel	Selected features for ECG channel	Classification accuracy	Mean false positive	Mean false negative
0.7	$RP_{\alpha 1}$				70.51 (6.85)	28.36	31.25
0.75	HOC_1	$RP_{\beta 1}$, RP_{θ}, $RP_{\alpha 1}$	HOC_2	SampEn2, Dkatz	91.74 (4.42)	7.68	8.86
0.8		RP_{β}, $RP_{\alpha 1}$	HOC_3, HOC_6, $RP_{\alpha 1}$	SampEn2, Dkatz	91.91 (4.33)	7.55	8.78
0.85	**RP_{γ}**	**HOC_5, $RP_{\beta 1}$, RP_{θ}**	**HOC_6, $RP_{\alpha 1}$**	**SampEn2, mean**	**94.88 (3.37)**	4.52	5.75
0.9	**RP_{γ}**	**HOC_5, $RP_{\beta 1}$, RP_{θ}**	**HOC_8, $RP_{\alpha 2}$, FMN**	**SampEn2, mean**	**94.91 (3.71)**	4.50	5.72
0.95	HOC_7, FMN	HOC_1, RP_{θ}, $RP_{\alpha 1}$, $RP_{\beta 1}$	HOC_1, FMN	SampEn2, mean	93.08 (4.34)	6.48	7.26

k the parameter of fuzzy-rough model (feature evaluation criterion)

The bolded rates indicate the most classification accuracies among all the feature subsets ($p < 0.05$)

Table 3 Overall classification rates, sensitivity, and specificity values for selected feature subsets

Feature type of arousal classifier	Feature type of valence classifier	Classification accuracy (%)	Average sensitivity (%)				Average specificity (%)			
			Soothing	Engaging	Annoying	Boring	Soothing	Engaging	Annoying	Boring
AFS1	VFS1	88.18 ± 5.92	91.01	89.77	88.74	83.85	92.98	95.85	97.52	97.57
AFS1	VFS2	88.78 ± 5.88	91.84	89.89	88.45	84.22	92.92	95.84	97.57	98.01
AFS2	VFS2	86.9 ± 5.86	89.56	88.25	88.38	83.44	92.08	95.01	97.74	97.24
AFS2	VFS2	87.01 ± 5.75	90.24	88.31	87.69	83.82	92.07	94.83	97.83	97.59

obtained by calculating the average confusion matrix (a table with the true classes in rows and the predicted classes in columns). Table 4 presents the average confusion matrix obtained by the maximum-accuracy classifiers. The first row of the matrix shows that the maximum confusion of the proposed emotion recognition system occurred for a boring musical stimulus, with an average confusion of 9.65 % for the soothing class. Furthermore, Table 3 shows that the minimum average specificity rate is 92.92 % for the soothing class.

Table 5 presents the effect of discarding the signal modalities on the classification rates. As seen, discarding FBS-based features from the optimal feature subsets greatly reduces the classification rates. According to the classification rates after discarding the ECG-based features from the optimal feature subset, the overall classification accuracy, the valence classification rate, and the classification sensitivities to all classes except the annoying class are significantly reduced. Obviously, FBS have more impact than ECG signals on the classification rates.

Table 4 Average confusion matrix of final emotion recognition system using AFS1 and VFS2

Input music	Output: listener's recognized affective states			
	Boring (%)	Soothing (%)	Annoying (%)	Engaging (%)
Boring	84.22	9.65	3.28	2.85
Soothing	2.03	91.84	1.59	4.54
Annoying	3.67	4.11	88.45	3.77
Engaging	0.7	6.97	2.44	89.89

Discussion

Nowadays, emotion recognition and sentiment analysis systems are in the focus of opinion mining research [31, 32]. As mentioned in the "Introduction," the recognition of induced emotions during music therapy is important. To establish an appropriate replacement for self-reports of induced emotions, emotion recognition systems that utilize neurophysiological signal processing have been proposed. The present study was performed to demonstrate the feasibility of fusing ECG and FBS information to classify music-induced emotions. Nowadays, these biosignals are fairly widely used in affective computing and HMI applications.

The selection of musical excerpts and the signal recordings were performed using different sets of subjects. It should be stressed that the system is able to predict the outcome of the questionnaire even if an induced emotion is contrary to those of the majority. The comparison between questionnaire outcomes and stimulus labels after signal

Table 5 Effects of discarding signal modalities on classification rates

	Optimal feature subsets	FBS-discarded optimal feature subsets	ECG-discarded optimal feature subsets
Total mean accuracy (%)	88.78	**47.2**	**86.63**
Mean arousal classification accuracy (%)	93.63	**56.43**	93.07
Mean valence classification accuracy (%)	94.91	**82.86**	**92.96**
Output sensitivity of classification			
Boring (%)	84.22	**49.27**	**82.63**
Soothing (%)	91.84	**41.65**	**88.64**
Annoying (%)	88.45	**36.05**	88.13
Engaging (%)	89.89	**49.92**	**88.75**

Bold values indicate significant decreases ($p < 0.05$) in classification rates in comparison with the condition of applying optimal feature subsets

recording aimed to ensure the exact input–output relationship for the classifiers and an equal sample size for each class.

After the acquisition and preprocessing of the biosignals, several linear and nonlinear features were extracted. Because the SVM classifier has been successfully applied to emotion recognition applications [12, 33] in the past, it was also applied in the present work. To classify the emotional data to arousal–valence quadrants, two parallel SVM classifiers were designed: a valence classifier and an arousal classifier. The final emotion recognition system was designed by juxtaposing the outputs of these valence and arousal classifiers. The inputs of the classifiers varied according to the inclusion parameter of the optimal feature evaluation algorithm. Tables 1 and 2 show that the feature subsets are obtained by applying a fuzzy-rough model (feature evaluation algorithm) and SFFS. Although the SFFS algorithm may not be the optimal search method, it proceeds dynamically, including and excluding features until the optimal feature subset is obtained. As shown in the tables, most of the features are selected from the frontalis and right temporalis channels. However, the optimum feature subsets are obtained by fusing informative features of all of the signal channels. The major selection of the RP features reflects the interrelation of emotional states and power spectrum variations in the EEG subbands, as previously shown [3–5].

The best average valence classification rate, the best average arousal classification rate, and the corresponding total classification rate were 93.63, 94.91, and 88.78 %, respectively. To the best of our knowledge, this was the first time that the fusion of FBS and electrocardiogram data was applied to emotion classification in an arousal–valence plane. The results proved the hypothesis about the ability of FBS to classify music-induced emotions. Furthermore, it was shown that the inclusion of ECG-based features yields improvements in the valence and total classification rates. This confirms previous studies that have suggested that ECG activation is sensitive to emotional valence [8, 33]. The maximum classification sensitivity was obtained for the recognition of the positive valence/low arousal (soothing) class, and the lowest sensitivity was obtained for the recognition of the negative valence/low arousal (boring) class. To provide a complete picture of the system performance, the FP and FN rates of each classifier as well as the average confusion matrix (Table 4) were included. The mean FP and FN rates of the valence classifier beside the first row of the confusion matrix reveal that the weakest point of the proposed system is the confusion of low-arousal/negative-valence samples with low-arousal/positive-valence samples (9.65 %). However, the classification results, sensitivity, and specificity values are generally satisfying (Table 3). Because the highest confusions are with soothing class, it yielded the lowest specificity of 92.92 %.

Currently, researchers are working on designing and developing emotion recognition systems by applying various biomedical data. Using four-channel biosignals (ECG, respiration, GSR, and EMG), Kim and André [11] presented an average classification rate of 70 % for the subject-independent recognition of music-induced emotions over four subjects. Lin et al. [12] proposed an EEG-based emotion recognition system for distinguishing four music-induced emotions with a maximum average accuracy of 82.29 % over 26 subjects. In [34], Liu et al. proposed a three-channel EEG-based emotion recognition algorithm and reported arousal and valence classification accuracies of 84.9 and 90 %, respectively, over ten subjects. This research is not limited to the application of musical stimuli. In [20], the HOC features of EEG channels and SVMs were used to classify six visually induced emotions (happiness, surprise, anger, fear, disgust, and sadness); they reported a mean classification rate of 83.33 % across 16 subjects. Soleymani et al. [35] used EEG, pupillary responses, and gaze distance to classify the affective states induced by video stimuli; they achieved the best user-independent classification accuracies of 68.5 and 76.4 % (over 24 participants) for three valence labels and three arousal labels,

respectively. In addition, Khosrowabadi et al. [36] reported a mean classification rate of 84.5 % (across 26 subjects) by applying eight EEG channels for classifying four emotions elicited by audiovisual stimuli. It is clear that the classification rates of our proposed system are considerably better than those of the previous related works. However, the lack of public musical stimuli and factors such as data acquisition conditions and decision-making schemes might affect the classification rates. Furthermore, in some cases, the difference in the number of emotional states does not provide the same conditions for comparison of the published works.

The proposed emotion recognition system has the advantages of subject independence and user comfort. However, it should be noted that, if desired, the recognition of more than four emotional classes in the arousal–valence space would require modifications, including the selection of additional musical stimuli (emotional states), changes in the optimal subject-independent AFSs and VFSs, and retraining of the classifiers. In future work, the performance of the proposed system will be evaluated using subjects who suffer from alexithymia (the phenomenon of being unable to express emotions).

Conclusions

This paper introduced and evaluated the application of information fusion of forehead and ECG biosignals to classify users' music-induced emotions in an arousal–valence emotional space. The results of this study show that the optimal selection of features of the FBS and ECG signals for arousal and valence classifiers is an effective technique for the classification of induced emotions. The best result for arousal classification (93.63 % mean classification rate) was obtained by the concatenation of RP_θ and FMN of the left temporalis channel; $RP_{\alpha 1}$, $RP_{\beta 2}$, and FMN of the frontalis channel; RP_θ, FMN, HOC_5, and ANE of the right temporalis channel; and SampEn(3, r) of the ECG channel. The best result for valence classification (94.91 % mean classification rate) was obtained by the concatenation of RP_γ of the left temporalis channel; RP_θ, $RP_{\beta 1}$, and HOC_5 of the frontalis channel; $RP_{\alpha 2}$, HOC_8, and *FMN* of the right temporalis channel; and RR_{mean} and the SampEn(2, r) feature of the ECG channel. Finally, the best mean accuracy for classifying the low arousal/positive valence, low arousal/negative valence, high arousal/positive valence, and high arousal/negative valence classes was 88.78 %. By using the proposed emotion classification system, we hope to see greater progress in the fields of music therapy, affective computing, and interactive multimedia systems.

Acknowledgments We gratefully acknowledge the assistance of Ms. Atena Bajoulvand for her help with the collection of the data of female subjects. The authors would also like to thank the anonymous reviewers for their insightful comments.

References

1. Barkišli M. Les idées scientifiques de Farabi dans la musique. Pažūhišgāh-i Mūsīqī-šināsī-i Īrān; 1978.
2. Aldridge D. An overview of music therapy research. Complementary Ther Med. 1994;2:204–16.
3. Trainor LJ, Schmidt LA. Processing emotions induced by music. In: Peretz I, Zatorre R, editors. The cognitive neuroscience of music. Oxford: Oxford University Press; 2003. p. 310–24.
4. Sammler D, Grigutsch M, Fritz T, Koelsch S. Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. Psychophysiology. 2007;44:293–304.
5. Pavlygina RA, Sakharov DS, Davydov VI. Spectral analysis of the human EEG during listening to musical compositions. Hum Physiol. 2004;30:54–60.
6. Knight WEJ, Rickard NS. Relaxing music prevents stress-induced increases in subjective anxiety, systolic blood pressure, and heart rate in healthy males and females. J Music Ther. 2001;38:254–72.
7. Bernardi L, Porta C, Sleight C. Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. Heart. 2006;92:459–70.
8. Kallinen K. Emotion related psychological responses to listening to music with eyes-open versus eyes-closed: electrodermal (EDA), electrocardiac (ECG), and electromyographic (EMG) measures. In: Proceedings of 8th international conference on music perception and cognition. 2004. p. 299–301.
9. McFarland RA. Relationship of skin temperature changes to the emotions accompanying music. Biofeedback Self Regul. 1985;10:255–67.
10. Janssen JH, Van den Broek EL, Westerink JHDM. Personalized affective music player. In: Proceedings of IEEE 3rd international conference on affective computing and intelligent interaction. Eindhoven. 2009. p. 1–6.
11. Kim J, André E. Emotion recognition based on physiological changes in music listening. IEEE Trans Pattern Anal Mach Intell. 2008;30:2067–83.
12. Lin YP, Wang CH, Jung TP, Wu TL, Jeng SK, Duann JR, Chen JH. EEG-based emotion recognition in music listening. IEEE Trans Biomed Eng. 2010;57:1798–806.
13. Firoozabadi SMP, Oskoei MRA, Hu H. A Human–Computer interface based on forehead Multi-Channel bio-signals to control a virtual wheelchair. In: Proceedings of 14th Iranian conference on biomedical engineering, Tehran. 2008. p. 108–113.
14. Rezazadeh IM, Wang X, Firoozabadi M, Golpayegani MRH. Using affective human–machine interface to increase the operation performance in virtual construction crane training system: a novel approach. Autom Constr. 2011;20:289–98.
15. Rad RH, Firoozabadi M, Rezazadeh IM. Discriminating affective states in music induction environment using forehead bioelectric signals. In: Proceedings of 1st middle east conference on biomedical engineering, Sharjah. 2011. p. 343–346.
16. Ortony A, Clore GL, Collins A. The cognitive structures of emotions. Cambridge: Cambridge University Press; 1990.
17. Beigand E, Viellard S, Madurell F, Marozeau J, Dacquet A. Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts. Cogn Emot. 2005;19:1113–39.

18. Juslin PN, Västfjäll D. Emotional responses to music: the need to consider underlying mechanisms. Behav Brain Sci. 2008;31:559–621.
19. Konečni VJ. Does music induce emotions? A theoretical and methodological analysis. Psychol Aesthet Creat Arts. 2008;2: 115–29.
20. Cambria E, Livingstone A, Hussain A. The hourglass of emotions. In: Esposito A, Esposito AM, Vinciareli A, Hoffmann R, Muller VC, editors. Cognitive behavioural systems. Berlin: Springer; 2012. p. 144–57.
21. Schlosberg H. Three dimensions of emotion. Psychol Rev. 1954;61:81–8.
22. Russel JA. A circumplex model of affect. J Pers Soc Psychol. 1980;39:1161–78.
23. Flores-Gutiérrez EO, Díaz JL, Barrios FA, Favila-Humara R, Guevara MA, Del Río-Portilla Y, Corsi-Cabrea M. Metabolic and electric brain patterns during pleasant and unpleasant emotions induced by music masterpieces. Int J Psychophysiol. 2007;65:69–84.
24. Pop-Jordanova N, Pop-Jordanova J. Spectrum-weighted EEG frequency ("brain-rate") as a quantitative indicator of mental arousal. Prilozi. 2005;26:35–42.
25. Kaiser JF. On a simple algorithm to calculate the 'energy' of a signal. In: Proceedings of IEEE ICASSP'90, New Mexico. 1990. p. 381–384.
26. Petrantonakis PC, Hadjileontiadis LJ. Emotion recognition from EEG using higher order crossings. IEEE Trans Inf Technol Biomed. 2010;14:186–97.
27. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Eng Comput. 2006;44: 1031–51.
28. Dabanloo NJ, Moharreri S, Parvaneh S, Nasrabadi AM. Application of novel mapping for heart rate phase space and its role in cardiac arrhythmia diagnosis. In: Computers in cardiology, Belfast. 2010. p. 209–212.
29. Hu Q, Xie Z, Yu D. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. Pattern Recogn. 2007;40:3509–21.
30. Theodoridis S, Koutroumbas K. Pattern recognition. 3rd ed. San Diego: Academic Press; 2006.
31. Grassi M, Cambria E, Hussain A, Piazza F. Sentic web: a new paradigm for managing social media affective information. Cognit Comput. 2011;3:480–9.
32. Poria S, Gelbukh A, Hussain A, Howard N, Das D, Bandyopadhyay S. Enhanced senticNet with affective labels for concept-based opinion mining. IEEE Intell Syst. 2013;28:31–8.
33. Lang PJ, Greenwald MK, Bradley MM, Hamm AO. Looking at pictures: affective, facial, visceral, and behavioral reactions. Psychophysiology. 1993;30:261–73.
34. Liu Y, Sourina O, Nguyen MK. Real-time EEG-based human emotion recognition and visualization. In: Proceedings of international conference on cyberworlds, Singapore. 2010. p. 262–9.
35. Soleymani M, Pantic M, Pun T. Emotion recognition in response to videos. IEEE Trans Affect Comput. 2012;3:211–23.
36. Khosrowabadi R, Heijnen M, Wahab A, Quek HC. The dynamic emotion recognition system based on functional connectivity of brain regions. In: Proceedings of IEEE intelligent vehicles symposium, San Diego. 2010. p. 377–381.