

Real-Time Movie-Induced Discrete Emotion Recognition from EEG Signals

Yong-Jin Liu, *Senior Member, IEEE*, Minjing Yu, Guozhen Zhao, Jinjing Song, Yan Ge, and Yuanchun Shi, *Senior Member, IEEE*

Abstract—Recognition of a human's continuous emotional states in real time plays an important role in machine emotional intelligence and human-machine interaction. Existing real-time emotion recognition systems use stimuli with low ecological validity (e.g., picture, sound) to elicit emotions and to recognise only valence and arousal. To overcome these limitations, in this paper, we construct a standardised database of 16 emotional film clips that were selected from over one thousand film excerpts. Based on emotional categories that are induced by these film clips, we propose a real-time movie-induced emotion recognition system for identifying an individual's emotional states through the analysis of brain waves. Thirty participants took part in this study and watched 16 standardised film clips that characterise real-life emotional experiences and target seven discrete emotions and neutrality. Our system uses a 2-s window and a 50% overlap between two consecutive windows to segment the EEG signals. Emotional states, including not only the valence and arousal dimensions but also similar discrete emotions in the valence-arousal coordinate space, are predicted in each window. Our real-time system achieves an overall accuracy of 92.26% in recognising high-arousal and valenced emotions from neutrality and 86.63% in recognising positive from negative emotions. Moreover, our system classifies three positive emotions (joy, amusement, tenderness) with an average of 86.43% accuracy and four negative emotions (anger, disgust, fear, sadness) with an average of 65.09% accuracy. These results demonstrated the advantage over the existing state-of-the-art real-time emotion recognition systems from EEG signals in terms of classification accuracy and the ability to recognise similar discrete emotions that are close in the valence-arousal coordinate space.

Index Terms—Affective computing, EEG, emotion recognition, movie

1 INTRODUCTION

AFFECTIVE computing has recently emerged as a converging technology for recognising, interpreting, and processing that relates to, arises from, or deliberately influences emotion or other affective phenomena. One of the key elements in affective computing is emotion recognition, which identifies human emotional states from their behavioral and physiological signals [1]. This information can be used to decide upon proper actions to express emotion and design new ways for machine emotional intelligence and human-machine interaction, e.g., in chatting robot and emotion-aware intelligent recommendation. With the rapid expansion of multimedia content, characterising affective multimedia content with discriminating tags is vital for multimedia information retrieval [2]. Currently, a large portion of user-defined tags are motivated by increasing his/her popularity and reputation in social networks. Moreover, users evaluate affective media content based on subjective judgments without consistent criteria. Some might tag multimedia content with words to express their emotions while others

might use tags to describe the content. Therefore, implicit affective tagging (i.e., the effortless generation of subjective and/or emotional tags) can help recommendation and retrieval systems to improve their performance.

Some measures have been employed for emotion recognition, such as self-report, startle, behavioral responses, and physiological measures [3, 4]. Amongst them, electroencephalograph (EEG) provides a direct and comprehensive means for emotion recognition by detecting immediate responses to emotional stimuli with an excellent temporal resolution [5-7]. Compared to behavioral responses (e.g., facial expression, vocal intonation, and body posture), it is difficult to disguise an individual's real feelings by the recognition of emotions from EEG signals [8]. Compared to other physiological signals (e.g., electrocardiograph, skin temperature, and galvanic skin response), emotion recognition from EEG signals can achieve a higher classification accuracy [9].

Recently, new wireless headsets (e.g., Emotiv) have become more cost-effective, easier to use, and mobile, with increased practicability and less physical restriction [8]. EEG carries distinct advantages in practical usage and therefore has been a primary option for the development of online emotion recognition systems. In fact, there has been a growing amount of effort to recognise a person's emotion in real time from EEG signals. For example, EmoRate is a commercial product/application that detects and reacts to human emotions and will change drastically the way we interact with them in our daily lives. It is built on a 14-electrode EEG headset, created by Emotiv sys-

- Y.-J. Liu is with the CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, China and also with Department of Computer Science and Technology, Tsinghua University, Beijing, China (E-mail: liuyongjin@tsinghua.edu.cn)
- M. Yu and Y. Shi are with Tsinghua National Lab for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, China (E-mail: yumj14@mails.tsinghua.edu.cn, shiyc@tsinghua.edu.cn).
- G. Zhao, J. Song and Y. Ge are with the CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, P. R. China (E-mail: zhaogz@psych.ac.cn, songjj@psych.ac.cn, gey@psych.ac.cn).
- G. Zhao is the corresponding author.

tems, Inc. The EmoRate system was trained by comedic shows, talk shows and videos on YouTube to recognise happiness, sadness, anger and fear. However, the recognition accuracy of the EmoRate system has not been systematically studied and reported.

There are two existing models of emotion that guide researchers to understand and construct emotional space: the discrete model and the dimension model. The discrete model considers that the emotional space consists of a limited number of basic discrete emotions. Although some viewpoints are not consistent, most studies tend to agree with at least six basic emotions: joy, sadness, surprise, fear, anger, and disgust [10]. On the other hand, the dimensional model divides the emotional space into two (i.e., valence-arousal, VA) or three (i.e., pleasure-arousal-dominance, PAD) dimensions [11]. The pleasure refers to the degree of pleasure that an individual feels, namely, the positive and negative characteristics (i.e., valence). Arousal reflects the mental vigilance level of emotion and the intensity of physiological activation that an individual feels. Dominance refers to an individual's status: in control or being controlled.

From the dimensional perspective to establish the emotional model, even if it is difficult to name a particular feeling, the dimensional model can use numerical values to mark different emotions in the valence-arousal coordinate space. However, although the dimensional model is easier to operate than the discrete model, when two or more discrete emotions are close in the valence-arousal coordinate space, the recognition ability of the dimensional model will be degraded dramatically [12]. One explanation is that similar emotions (such as anxiety and anger in the valence-arousal dimension) could have their own unique representation and induce completely different behavioral responses [13].

The existing studies of real-time emotion recognition, in essence, identify different valence and arousal classes instead of discrete emotions. Although researchers combined different valence and arousal classes and attempted to map them into basic emotions, it is different from the discrete emotions that are directly elicited from target materials. In addition, the intensities of the stimuli in the same category of emotion are different. We feel happy when we see a baby smiling. We also feel happy when we see a beautiful sunset, but the feeling might not be so strong [14]. Therefore, emotions in the same category (e.g., joy) could have different levels of valence and arousal, but not vice versa.

To overcome aforementioned limitations, in this paper, we propose a real-time system to recognise an individual's emotional states (including not only the valence and arousal dimensions but also similar discrete emotions in the valence-arousal coordinate space) based on EEG signals and make the following contributions:

- To evoke target-specific emotion reliably and reproducibly, we construct a standardised database of 16 emotional film clips selected from over one thousand film excerpts, and build 8 emotional categories. The standardised database including participants' self-ratings will be made publicly available¹ to provide a useful benchmark for elicited emotion evaluation.
- Our system outperforms state-of-the-art real-time emotion recognition systems in terms of classification accuracy and the ability to recognise similar discrete emotions that are close in the valence-arousal coordinate space. In particular, our system achieved an averaged accuracy of 92.26% (50% accuracy at random) in recognising high-arousal and valenced emotions from neutrality. Based on basic discrete emotions that were elicited by standardised film clips, our system classified three positive emotions (joy, amusement, tenderness) and four negative emotions (anger, disgust, fear, sadness) with an overall accuracy of 86.43% (33.33% accuracy at random) and 65.09% (25% accuracy at random), respectively.

2 LITERATURE REVIEW

A few existing studies have implemented real-time EEG-based emotion detection systems (see Table 1).

Jatupaiboon, Pan-ngum and Israsena [15] used EEG signals to classify happy and unhappy emotions (i.e., 2 classes of valence) that were elicited by pictures and classical music in real time. They decomposed EEG signals into 5 frequency bands (i.e., delta, theta, alpha, beta, and gamma) and computed the power spectral density (PSD) of each frequency band at 2 electrode sites (T7 and T8), which led to a total of 10 features. The support vector machine (SVM) was adopted for classification. The offline analyses indicated that the average accuracies of the **subject-dependent model** and **subject-independent model** were 75.62% and 65.12%, respectively. However, the real-time classification accuracy of the two valence classes was not provided.

¹ <http://enginpsych.psych.ac.cn/indexen.php>

TABLE 1
Existing real-time EEG-based emotion recognition system

Reference	Emotion	Stimulus	Device	Features	Classifier	Accuracy	Execution Time
[8]	2 valence classes ×3 arousal classes	standardised sound	Emotiv	fractal dimension value	SVM	-	-
[15]	2 valence classes	standardised picture & sound	Emotiv	power spectral density	SVM	-	5s
[16]	2 valence classes ×2 arousal classes	standardised pictures	Emotiv	fractal dimension value	SVM	70.5%	2.2s
[17]	valence	non-standardised film clip	Emotiv	power spectral density	LDA	-	-

Liu, Sourina and Minh Khoa [8] proposed a real-time fractal-dimension-based algorithm to identify 2 valence classes and 3 arousal classes. These 6 combinations were mapped into 6 basic emotions: sad, frustrated, fearful, satisfied, pleasant, and happy. The authors used music and sound stimuli from the International Affective Digitised Sounds (IADS) database to elicit the target emotions. They calculated the **fractal dimension** values at 3 electrode sites (AF3, F4, and FC6) as features. Real-time applications were proposed and implemented in 3D virtual environments (emotional avatar and music therapy).

Viet Hoang, Manh Ngo, Bang Ban and Thang Huynh [16] developed a real-time emotion recognition system to identify 2 valence classes and 2 arousal classes, which resulted in a combination of 4 basic emotions (happy, relaxed, sad, and angry) and the neutral state. The authors adopted the Higuchi fractal dimension algorithm and SVM as a classifier. EEG signals from FC6, AF3, and F4 were the input of the emotion recognition system. The average accuracy for the emotion recognition across all subjects was 70.5% (ranging from 66% to 76%).

All of these emotion recognition systems presented the standard visual or auditory stimuli to evoke emotions, such as the international affective picture system [18] and the international affective digitised sound system [19]. Standardised emotional stimuli were necessary to elicit target emotions when developing a real-time emotion recognition system. Typically, standardised stimuli are selected based on a large sample of participants' subjective feelings and emotional experiences and have been proven to better elicit a single and pure target emotion than those stimuli created relying on a smaller sample size in a single experiment. Therefore, the advantage of standardised stimuli is to label EEG signals as the ground truth regardless of a user's subjective feelings after watching the emotional content. In contrast, the majority of research on EEG-based emotion recognition uses **super-vised learning algorithms** (e.g., SVM, Naïve Bayes, K-Nearest Neighbours) to predict emotional states. Researchers collect a user's self-reported scores on the levels of arousal and valence after presenting emotional stimuli. Then, the user's ratings are separated into two (e.g., low and high) or more classes by selecting the threshold of a 9-point self-assessment rating scale [2]. Obviously, this approach does not apply to the real-time recognition of emotion. The introduction of standardised emotional stimuli is the prime condition of success.

The recent trend has been to extend traditional materials (e.g., picture and sound) to combine visual and auditory stimuli, usually film clip and music video [20]. This trend is consistent with real-life emotional experiences that rely on the presence of combined stimuli coming from different modalities. Movies or music videos are more complex and have a more realistic context for involving emotion states compared with static photos and slides [21]. In addition, the dynamic visual-and-auditory-combined stimuli could induce emotions without deception [22, 23].

Castermans, Duvinage and Riche [17] attempted to detect the valence of emotions elicited by different video

excerpts amongst several spectators seated in a realistic movie theatre, on the basis of their EEG signals. A total of 12 PSD features (3 frequency bands \times 4 electrode sites) were computed and sent to a Linear Discriminant Analysis (LDA) classifier. Preliminary results obtained with consumer-grade EEG headsets indicated that positive and negative video excerpts might be discriminated in the training data set but not on an independent data set, which indicates an over-fitting problem. In fact, this work was not a truly real-time emotion system because the authors used the participants' self-assessment to characterise their global emotional states during a movie of several minutes duration. To overcome this limitation, in this paper, we develop a standardised database of movie-induced emotional stimuli with each user's self-assessment of arousal and valence levels as well as their discrete emotional experiences. Based on this database, we propose a real-time emotion recognition system.

In addition to the lack of a standardised database of visual-and-auditory-combined stimuli, another limitation of the previous real-time emotion recognition systems was the lack of discrete emotions built beyond the traditional valence-arousal dimensional model as introduced above. In essence, all of the existing systems recognised different levels of arousal and valence, which are further mapped into discrete values/emotions. However, such mapping of the combination of arousal and valence into corresponding emotions works only when two or more discrete emotions are far from each other in the valence-arousal coordinate space.

Last but not least, real-time emotion recognition takes not only the classification accuracy but also the execution time into consideration, while emotion recognition from offline EEG signals considers only the classification accuracy. As a result, to achieve better real-time classification performance within a time limit, researchers must remove artifacts from EEG signals based on the statistical characteristics of a single trial or a few trials, extract features that are representative and sensitive to the subtle changes of brain activities due to emotional fluctuations, and implement efficient machine learning techniques to recognise emotional states. This requirement might be another reason that only a few existing studies have implemented real-time EEG-based emotion recognition systems.

3 CONSTRUCTION OF A STANDARDISED DATABASE OF CHINESE EMOTIONAL FILM CLIPS

To elicit target emotions, we construct a standardised database of film clips. Nine research assistants (4 males and 5 females) who majored in psychology watched more than one thousand Chinese film excerpts for preliminary screening. They were asked to select film excerpts that lasted 1 to 3 minutes and contained independent and integrated content to elicit a single target emotion, including four negative emotional states (sadness, anger, fear, and disgust), eight positive emotional states (joy, romance, warmth, well-being, love, mutual affection, amusement, and contentment), and a neutral state. Three specialists (2 males and 1 female) in the area of emotion

elicitation evaluated each film segment that was selected by the research assistants for its potential to successfully elicit a target emotional state. The film excerpts were decided by a consensus amongst three specialists. Finally, 111 film excerpts were selected and further examined in our experimental session.

A total of 462 Chinese-speaking undergraduate and graduate students (195 males and 267 females) whose average age was 22.2 years (range = 18-30, SD = 2.2) took part in our group film-viewing experiment. At each time, 7-10 participants constituted a group to view and rate film excerpts. They were separated by partitions and asked to use their own computers and headsets to avoid any interference from other group members (e.g., voice, behavioral response). During the experiment, the participants were advised that they felt free to withdraw consent and drop out of the experiment whenever they wanted, and all of their responses were totally anonymous.

To control the presentation of the film excerpts and record the participants' self-reported ratings, the whole experiment was implemented using the E-prime. A total of 111 film excerpts were assigned into 10 groups. In each group, the number of film clips in the same emotional category did not exceed 2. All of the film clips were viewed by at least 30 participants. The presentation order of film excerpts in each group was random. After viewing each film clip, the participants completed three questionnaires: self-assessment manikin (SAM), the positive and negative affect schedule (PANAS), and differential emotion scale (DES). The SAM was a 9-point (1 = "not at all", 9 = "extremely") self-assessment scale for arousal, valence, liking, familiarity, and dominance, which was modified from [24, 25]. The PANAS [26] consisted of two subscales: positive and negative affect. Each subscale contained 10-item words, which assessed the degree of the subject's affect while viewing the film clips based on a 5-point scale (1 = "not at all", 5 = "extremely"). The DES was on a 9-point scale (1 = "not at all", 9 = "extremely") to assess the intensity of each self-reported emotional dimension, which was modified from [23]. It consisted of the following 12 items: joy, amusement, contentment, well-being, love, mutual affection, romance, sadness, warmth, anger, fear, and disgust. The participants were encouraged to answer all of the questions based on their true emotional feelings when watching each film excerpt, instead of their expected feelings or general mood.

We conducted R-type cluster analysis with an agglomeration approach and between-group linkage, to identify the clusters of emotions. The sum of each variable was measured based on the squared Euclidean distance. We standardised the variables by computing the Z scores. The cluster analysis found that amusement was separated from the other positive emotions. Contentment and joy belonged to one cluster. Warmth, well-being, romance, mutual affection, and love were all related to tenderness. Each negative emotion was one cluster. Then a total of eight emotional categories were built: three positive emotional categories (joy, amusement, and tenderness), four negative emotional categories (anger, sadness, fear, and disgust), and the neutral state.

TABLE 2

Brief description of Chinese emotional film clips in our standardised database. More details of these film clips are presented in Supplemental Material.

Emotion	Film Name	Length (sec)	Clip Content
Joy	Singing When We Are Young	60	Young people are talking about their future
	Better And Better	85	The Spring Festival party of a village
Amusement	Just Another Pandora's Box	67	Humorous battle scenes
	Love on Delivery	83	The dogshit is patted to a bad man's face
Tenderness	Hear Me	60	The hero rides the motorcycle and carries the heroine
	A Simple Life	99	Recall the master's happy childhood together
Anger	City of Life and Death	73	Nanjing massacre
	YiP Man I	60	A Chinese martial artist is shot by Japanese
Sadness	Bodyguards and Assassins	114	A father finds his son is dead
	Changjiang qihao	139	A little boy is sad because of his father's death
Fear	The Chrysalis	68	The heroine encounters a ghost in the toilet
	Inner Senses	92	The hero sees a spirit
Disgust	Farewell My Concubine	84	The hero suffers sexual assault by an old eunuch
	Farewell My Concubine	60	
Neutrality	Raise the Red Lantern	62	Hang the lanterns
	Black Coal, Thin Ice	65	The hero takes his coat to a laundry

We selected film clips for each emotional category based on two objective scores: 1) Hit rate is defined as the proportion of users who rate the emotional target score at least one point higher than the non-target emotion; and 2) Success index represents the sum of two z-scores derived from normalising the hit rate and the target emotional score [23, 27]. For neutral films, the mean ratings on each emotional dimension were less than 3. For the remaining seven emotional categories, there were seven **hit rate scores** and seven success indices for each film clip. The selected film clip met two criteria at the same time: the hit rate was more than 20%, and the mean rating on the target score was at least 1 larger than that on the non-target scores. Eventually, we selected 16 film clips according to their success indices and constructed a standardised database of Chinese emotional film clips, in which all of the subtitles and advertisements were removed. Each emotional category consisted of two film clips (see Table 2). Note that two disgusting film clips came from the same film.

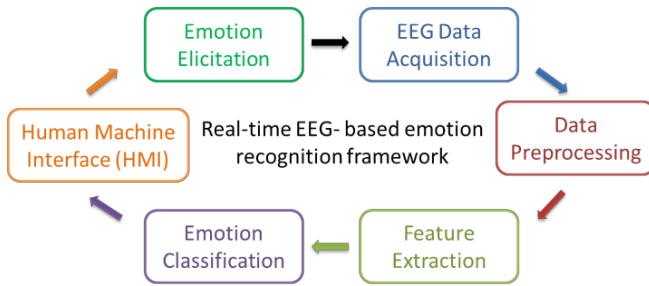


Fig. 1. The process of real-time emotion recognition system.

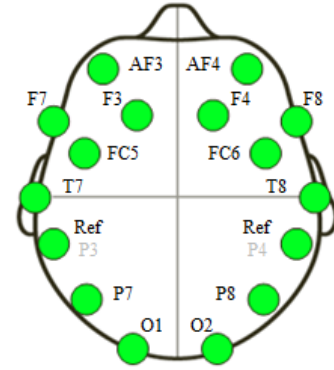


Fig. 2. Electrode placement for the Emotiv EPOC system.

All of the selected film clips were expected to elicit high-arousal emotions except for the neutral clips. The standardised database including 30 participants' self-ratings will be made publicly available.

In the following section, we propose a real-time discrete emotion recognition system that uses these standardised film excerpts to induce emotions.

4 REAL-TIME MOVIE-INDUCED DISCRETE EMOTION RECOGNITION

We propose a real-time EEG-based emotion recognition framework (see Fig. 1). This framework consists of six modules: (1) emotion elicitation, (2) data acquisition, (3) data preprocessing, (4) feature extraction, (5) emotion learning and pattern classification, and (6) a human machine interface (HMI).

The extracted EEG features in module 4 and the feature-based classification techniques used in module 5 are deliberately selected from the literature. The contribution of our framework is to develop a working system with high performance because such a good model did not exist for discrete emotion recognition using realistic movie clips with high ecological validity. The performance of our recognition system can be regarded as a benchmark, such that more advanced techniques for feature extraction, feature selection, and classification deserve further study in their own right and will be reported in future work. Emotion elicitation is presented in section 3. Modules 2-6 are presented in the following subsections.

4.1 EEG Data Acquisition

In our work, motivated by the success of Emotiv EPOC system (see Fig. 2) [28] in the application of emotion recognition [8, 15-17, 29-31], we chose this system to collect EEG signals.

Emotiv EPOC is a 14-electrode wireless EEG headset with a 128 Hz sampling rate. We used it to record EEG signals when users watch film clips. It monitors the user's brain activities and enables a real-time display of the Emotiv headset data stream, including the EEG, signal quality, Fast Fourier Transform (FFT), gyro (if applicable), wireless packet acquisition/loss display, marker events, and headset battery level.

Emotiv EPOC provides only some basic EEG data processing functions (e.g., filters and FFT with adjustable sampling window sizes and window functions). To incorporate more powerful signal processing, feature extraction and machine learning algorithms, we build a buffer to store the EEG data stream from the Emotiv EPOC and then load the EEG data section by section in EEGLAB, which is an open source Matlab toolbox for physiological signal processing [32]. Specifically, when a reading command is triggered, all of the samples in the buffer are taken out, and the buffer is cleared. Then, we use a queue to buffer the EEG data from Emotiv to EEGLAB. The queue is refreshed by the current number of samples in the Emotiv's buffer every time that the reading command is triggered.

EEGLAB provides alternative and powerful algorithms to process the EEG data, extract the features, and recognise the emotional states. Moreover, the Matlab environment allows us to further combine Emotiv EPOC with other wearable devices (e.g., eye tracking glasses, smart bracelet) to provide real-time multi-channel physiological feedback from the entire set of user experiences for emotional applications.

4.2 EEG Data Preprocessing

EEG signals in the acquisition process can easily be contaminated with noise signals. We preprocess the data by the removal of artifacts in the process of EEG recording to maintain signal stability and retain the effective data segments. Continuous EEG data are filtered with a band-pass (1-45Hz) filter to remove linear trends and minimise the introduction of artifacts.

To remove electrooculography (EOG) artifacts, we perform independent component analysis (ICA) to decompose the whole training dataset of the EEG signals into 14 independent components (ICs). The 14 computed ICs are characterized by their topographies and PSDs which are visually inspected and rated by two experts as either an EOG artifact or EEG signal component. The SVM with a RBF kernel function is trained and used for identification of EOG artifacts in the testing dataset of the EEG signals. We implement the CORRMAP in EEGLAB to remove EOG artifacts. After automated classification, the components classified as artifacts are discarded from the subsequent process. The remaining ones, classified as signal

components, are back projected to reconstruct artifact-free EEG signals [33, 34].

4.3 Feature Extraction and Normalization

To recognise a discrete emotion in real time, we use the short-time Fourier transform (STFT) with a sliding time-window approach for feature extraction and normalisation based on time-frequency (TF) analysis. TF analysis represents the EEG signals in a two-dimensional spectral domain and provides insights into both the frequency and temporal evolution of the energy- and power-related features that can be associated with brain activities [35, 36]. The TF features of the EEG signals are estimated based on the concept of event-related desynchronisation (ERD) and synchronisation (ERS) [37-40]. The ERD/ERS theory assumes that the event-related activity represents changes of the power in specific frequency bands. The ERD/ERS phenomena could be considered to be changes in one or more parameters that control oscillations in neuronal networks.

STFT is a linear decomposition of the EEG signals into elementary components. The STFT involves a sliding window of the EEG signal $x(m)$ around a time instant t from an experimental trial and the subsequent calculation of the Fourier transform for each t .

$$X(t, f) = \int_{-\infty}^{\infty} w(m - t) x(m) e^{-j2\pi fm} dm \quad (1)$$

where $w(m - t)$ represents the short-time analysis window, and $X(t, \omega)$ represents the power distribution of the signal in the TF plane. In essence, STFT extracts several frames of the signal to be analysed by using a window that moves with time. If the time window is sufficiently narrow, then each frame that is extracted can be viewed as stationary, which enables the Fourier transform to be used. By moving the window along the time axis, the relation between the variance of the frequency and time can be identified.

Fig. 3 shows the schematic representation of the sliding time-window approach for the TF feature extraction. The power spectral features of the EEG signals are computed in a time window w_n and in the frequency band f_b of interest. The power spectral features that are recorded during the process of watching a film clip $X_{f_b}^{w_n}$ are averaged over both time and frequency, i.e.,

$$X_{f_b}^{w_n} = \frac{1}{N_{w_n}} \sum_t \frac{1}{N_{f_b}} \sum_f X(t, f) \quad (2)$$

where N_{w_n} and N_{f_b} denote the number of samples in each time window w_n and the number of samples in each frequency band f_b , respectively [35]. Similarly, the power spectral features recorded during a resting period X_{f_b} are averaged over the same frequency band and time. Here, a resting period is used to correct the emotional baseline to exclude the confounding effects of other factors.

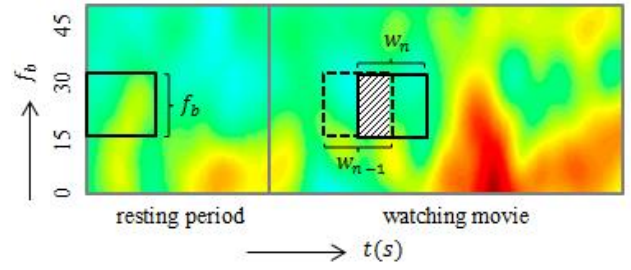


Fig. 3. Schematic representation of our sliding time-window approach for time-frequency feature extraction.

In this work, EEG signals are segmented into a number of time windows with a 50% overlap between two successive time windows (see Fig. 4). The power spectral features of the EEG signals on 14 channels are extracted in 5 frequency bands: delta (1-4Hz), theta (4-8Hz), alpha (8-12Hz), beta (13-30Hz), and gamma (31-45Hz). In addition to the power spectral features, the difference between the spectral power of all of the 7 symmetrical pairs of electrodes on the right and the left hemispheres in 5 frequency bands is extracted to measure the possible asymmetry in the brain activity due to the emotional stimuli. These 7 pairs of electrodes are: AF3-AF4, F3-F4, F7-F8, FC5-FC6, T7-T8, P7-P8, and O1-O2. As a result, a total number of 105 (70 PSD and 35 ASM) EEG features are used.

To reduce the individual variability and the lasting effects of the previous elicited emotions on the current target emotion, in our experimental protocol, each trial begins with a 60-s Go/NoGo task such that the participant is kept in a neural state (see Section 5.2 for more details). Furthermore, after feature extraction, we normalise the segmented data when watching each film clip with a corresponding resting period with eyes open. Then, we subtract the baseline power of each resting period from all of the corresponding segments of the extracted features, in such a way that the change in the power is based only on the current film clip and is not affected by the previous clips [41].

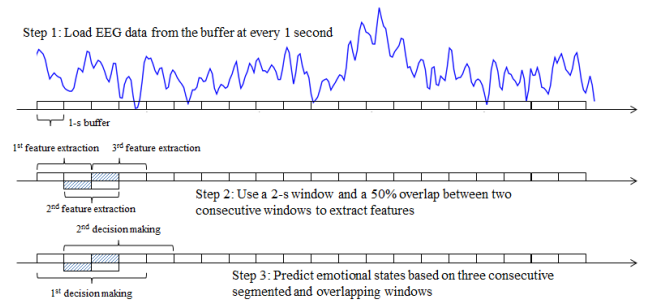


Fig. 4. EEG data are loaded every 1 second. Then, a 2-s window with a 50% overlap between two consecutive windows is used to segment the filtered and artifact-free EEG data. 105 EEG features are computed in each 2-s window to predict emotional states. We make a decision upon the three consecutive 2-s windows according to the majority rule. In other words, we make the first decision after EEG data are continuously imported for 4 seconds. After that, emotional states are predicted every 1 second.

4.4 Emotion Classification with Feature Selection

105 EEG features are extracted from EEG signals on 14 electrodes. We observed that not all of these features are necessary to be used for emotion classification; that is, some features may be irrelevant to emotion classification, and some features may be highly correlated and thus redundant in emotion classification. In our work, we applied a classifier that performed feature selection; that is, a classifier that intelligently used a subset of all 105 EEG features.

We choose Linear Discriminant Analysis (LDA) [42] for feature selection, which projects high-dimensional data with labels into a low-dimensional space with good class-separability by maximizing Fisher separation criterion. Feature selection is achieved by finding a subset of features that have the largest Fisher separation value. However, when the number of selected features is still larger than the number of observations, the within-class covariance matrix of the features becomes singular. To solve this problem, we apply the Sparse LDA (SLDA) [43], which overcomes this limitation by performing LDA with a sparseness criterion imposed such that classification and feature selection are performed simultaneously. Accordingly, the selected features entered as input features to a Support Vector Machine (SVM) for classification. In our implementation using Matlab, we apply the SpaSM toolbox for SLDA and LIBSVM toolbox for SVM. We optimize the parameter STOP in SpaSM toolbox by searching the range $[-105, -1]$ with a step size 15. In LIBSVM toolbox, we choose the RBF kernel function and optimize the cost parameter c (in the range $[-8, 8]$) and the gamma parameter g (in the range $[-8, 8]$) using SVMcgForClass.

We use SVM to classify eight discrete emotions; that is, three positive emotions (joy, amusement and tenderness), four negative emotions (anger, sadness, fear and disgust) and neutrality. Note that the original SVM only solves a two-class problem. To achieve a multiclass classification, LIBSVM constructs $k(k-1)/2$ classifiers with each classifier trained for two classes, where k is the number of classes [44]. In addition to directly apply eight-class classification in LIBSVM, we further improve the classification accuracy by incorporating a priori knowledge: in eight emotions, three positive emotions including joy, amusement and tenderness are similar and four negative emotions including anger, sadness, fear and disgust are similar. Based on this priori knowledge, we propose a three-level classification scheme. At the first level, neural and non-neural emotions are classified. At the second level, non-neural emotions are classified into positive and negative emotions. At the third level, positive emotions are classified into joy, amusement and tenderness, and negative emotions are classified into anger, sadness, fear and disgust. The experimental results summarised in Section 5 demonstrate that our three-level classification scheme significantly improves the accuracy of the default eight-class classification in LIBSVM.



Fig. 5. The human machine interface of the proposed real-time movie-induced emotion recognition system from EEG signals. In this screenshot, our system successfully recognises joy (target emotion) from amusement and tenderness (both are high-arousal and positive emotions).

4.5 Human Machine Interface and Prototype System

We developed a prototype system to implement the framework illustrated in Fig. 1. The human machine interface of this proposed system provides five pieces of information and is illustrated in Fig. 5. First, a film clip that is selected to elicit a user's target emotion is presented in the middle of the HMI. When a user wearing the Emotiv EPOC headset watches this film clip (upper right), this suite continuously records his/her EEG signals (lower right). This system recognises not only the arousal and valence levels (lower left) but also discrete emotional states in terms of 3 positive emotional categories, 4 negative emotional categories, and neutrality (upper left). The number of colour bars reflects the classification posterior probability or the degree of confidence when classifying an emotional state into a specific class.

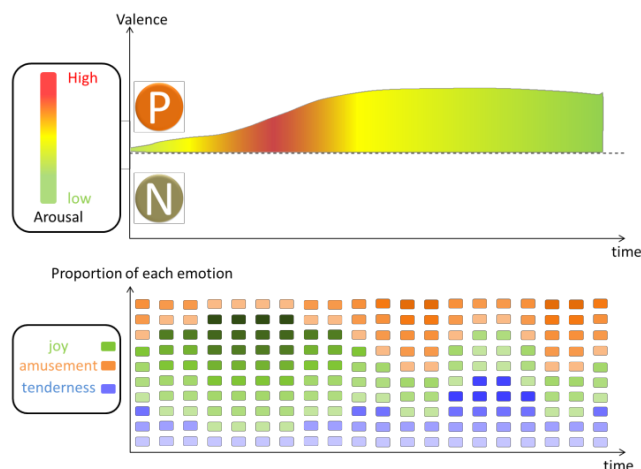


Fig. 6. The proposed system provides real-time recognition of discrete emotions and the levels of arousal and valence.

In the screenshot shown in Fig. 5, a film clip aims to elicit a positive (i.e., more yellow bars in the right side) and high-arousal (i.e., the number of red bars is more than half) emotion, namely, joy. The proposed system successfully recognises joy from amusement and tenderness (i.e., the number of green bars is more than the number of orange and blue bars), which are two similar emotions in the valence-arousal coordinate space. Due to their rich content and dynamic characteristics, real-world emotions might not be singular and unique but instead could be a mix of components. In this case, our system can predict a distribution of different emotions depending on the degree of physiological activation of each emotion instead of classifying into a single emotion.

In addition, after watching a film clip, the user can view the summary of his/her personal emotional experiences. As illustrated in Fig. 6, the HMI can present the levels of arousal and valence that change over time (upper panel). The redder area indicates a higher level of mental vigilance of the target emotion that an individual feels. Meanwhile, the HMI can also present the portion of each discrete emotion within its valence category, which changes over time (lower panel). The larger portion indicates that the corresponding emotion was elicited more than the other emotions that were similar in the valence-arousal coordinate space. The deeper the colour is, the higher the intensity of physiological activation that an individual feels. By viewing this summary, the user can realize when and where a specific emotion (e.g., joy) is evoked and how long it lasts.

5 EXPERIMENT

5.1 Participants

30 healthy, male, right-handed undergraduate or graduate students (age range: 19 - 26, mean = 23, SD = 1.73 years) without neurological illness or psychiatric disorder took part in the experiment. All of the participants were required to not take any tobacco or caffeine for 24 hours before the experiment.

5.2 Experimental Protocol and Procedure

All of the film clips were adjusted to the same resolution (720 × 576). They were presented randomly on a 15-inch LCD screen. The film clips that targeted the same valence were presented no more than twice consecutively [23]. Two speakers were used, and the volume was adjusted to an appropriate level to keep the participants comfortable. The participants were seated in a soundproofed room and were required to turn off all wireless and Bluetooth devices to isolate possible interferences. Their eyes were approximately 0.6 meter from the screen centre.

Each trial began with a 60-s Go/NoGo task to keep all of the participants in a neutral state. In general, Go/NoGo testing refers to a pass/fail principle that uses two boundary conditions. The test is passed only when the Go condition is met and also the NoGo condition fails. The test was used to “distract” the participants from the influence of the last film clip. Then, the participants were instructed

to rest with their eyes open for 40-s and their eyes closed for another 40-s. After two resting baselines, the participants watched a film clip and then completed two questionnaires: SAM and DES, as introduced in section 3. All of the participants were instructed to rate all of the scales according to their real feelings while watching the film clips instead of meeting the expected standards or their daily mood.

In the training session, we played 8 film clips to elicit 7 discrete emotions and one neutral state. The play order was randomised: the film clips with the same valence or those that targeted the same emotional category were not presented consecutively more than twice. After each trial of the training session, the EEG signals were analysed immediately, to train the system for each discrete emotion. The training session was followed by a 5-min break. Afterward, the testing session had another 8 film clips played for real-time recognition of emotion. The play order was randomised following the same rules.

Before recording the EEG, we put EMOTIV on the participant's head for a while to prevent undesired emotions that can arise from unfamiliar or uncomfortable feelings. Then, we described the process of recording and advised the participant to stay as still as possible to prevent artifacts from movements of the body. The 14-channel wireless EMOTIV EPOC consists of AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, and O2. The sampling rate is 128 Hz.

We collected the participants' true feelings and responses after watching each film clip. Because our system used the standardised database of movie-induced emotions as the ground truth to label the EEG signals, we compared every participant's self-assessments with the standardised database and ignored the cases where he or she did not rate the target emotion at least one point higher than the other non-target emotions (i.e., inconsistent with the target emotions [8]).

5.3 Experimental Results

We used the standardised database of Chinese film clips to elicit seven high-arousal and valenced discrete emotions and neutrality (Section 5.3.1-5.3.3). Furthermore, our system applied the three-level classification scheme proposed in Section 4.4 to recognise eight discrete emotions at the same time, which outperformed the default eight-class classification in LIBSVM (Section 5.3.4).

5.3.1 Assessment of valence and arousal

The participants' assessments for valence and arousal were consistent with our expectations (see Fig. 7). Self-reported scores on the valence dimension were the highest for positive emotions ($M = 6.12$, $SD = 1.57$), which were higher than the scores for neutrality ($M = 3.03$, $SD = 1.86$) and negative emotions ($M = 1.66$, $SD = 0.79$). The higher score on the valence dimension is, the more positive the characteristics of an emotion. On the other hand, both positive ($M = 5.58$, $SD = 1.66$) and negative ($M = 6.26$, $SD = 1.62$) film clips successfully elicited higher arousal states than neutrality ($M = 2.74$, $SD = 1.61$).

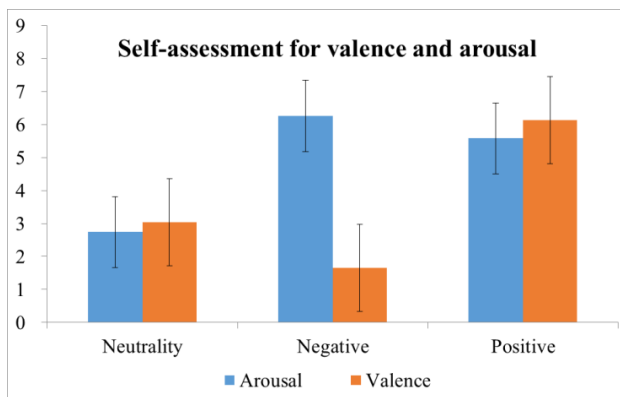


Fig. 7. Self-reported valence and arousal of positive, negative, and neutral emotions. Error bars indicate ± 1 standard error.

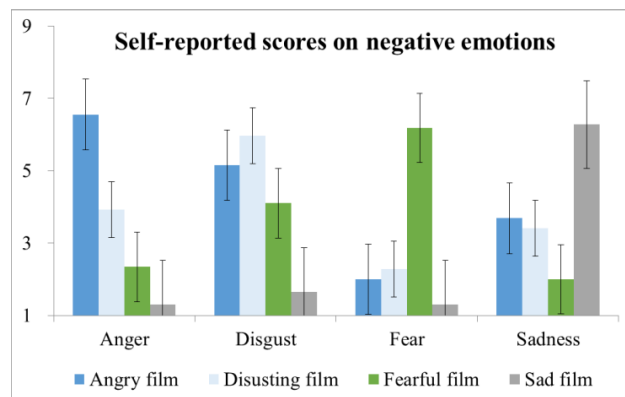


Fig. 9. Self-reported scores on negative emotions (anger, disgust, fear, and sadness) elicited by four negative film clips (angry, disgusting, fearful, and sad). Error bars indicate ± 1 standard error.

5.3.2 Assessment of positive emotions

The participants' self-reported scores on three positive emotions: joy, amusement, and tenderness, were shown in Fig. 8. On average, their ratings on differential emotion scales met the criterion of the standardised database of emotional film clips. After watching a joyful film clip, the participants rated the joy ($M = 6.93$, $SD = 1.46$) at least one point higher than amusement ($M = 5.31$, $SD = 2.36$) and tenderness ($M = 5.72$, $SD = 1.41$). Similarly, an amusing film clip elicited higher subjective feelings of amusement ($M = 7.76$, $SD = 1.27$) than joy ($M = 6.19$, $SD = 1.86$) and tenderness ($M = 2.34$, $SD = 1.82$). Tenderness was successfully elicited with a higher score on the target dimension ($M = 7.03$, $SD = 1.76$) compared to joy ($M = 4.78$, $SD = 1.95$) and amusement ($M = 1.59$, $SD = 1.15$).

The participants' self-reported scores on the arousal and valence dimensions of three positive emotions were similar. Individual subjective ratings on the arousal: joy ($M = 5.83$, $SD = 1.61$), amusement ($M = 6$, $SD = 1.69$), and tenderness ($M = 5.38$, $SD = 1.59$). On the other hand, the participants rated the valence: joy ($M = 6.52$, $SD = 1.24$), amusement ($M = 6.76$, $SD = 1.55$), and tenderness ($M = 5.72$, $SD = 1.69$). The difference in the self-reported scores on the arousal and valence between three positive emotions was less than 1.

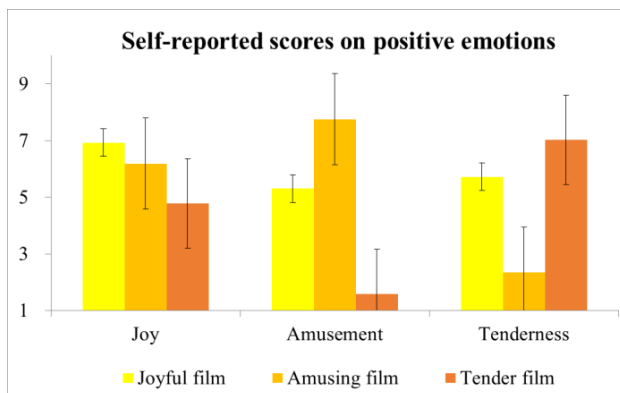


Fig. 8. Self-reported scores on positive emotions (joy, amusement, and tenderness) elicited by three positive film clips (joyful, amusing, and tender). Error bars indicate ± 1 standard error.

5.5.3 Assessment of negative emotions

The participants' self-reported scores on four negative emotions: anger, disgust, fear, and sadness were illustrated in Fig. 9. Their averaged ratings on the target emotion scale were at least one point greater than the other negative emotion scales. Specifically, the participants rated the anger ($M = 6.55$, $SD = 1.64$) higher than disgust ($M = 5.15$, $SD = 2.15$), fear ($M = 2$, $SD = 1.49$), and sadness ($M = 3.69$, $SD = 2.48$) after watching an angry film clip. A disgusting film clip elicited higher subjective feelings of disgust ($M = 5.96$, $SD = 2.1$) than anger ($M = 3.93$, $SD = 2.05$), fear ($M = 2.28$, $SD = 1.69$), and sadness ($M = 3.41$, $SD = 2.01$). Fear was successfully elicited with a higher score on the target dimension ($M = 6.19$, $SD = 1.64$) compared to anger ($M = 2.34$, $SD = 1.88$), disgust ($M = 4.1$, $SD = 2.38$), and sadness ($M = 2$, $SD = 1.6$). The selected sad film clip acquired an optimal result: self-reported scores on the sadness dimension ($M = 6.28$, $SD = 2.19$) were evidently greater than those on the anger ($M = 1.31$, $SD = 0.81$), disgust ($M = 1.66$, $SD = 1.54$), and fear dimension ($M = 1.31$, $SD = 0.71$).

The participants' subjective ratings on the arousal between four negative emotions were close to each other: anger ($M = 6.38$, $SD = 1.72$), fear ($M = 6.69$, $SD = 1.44$), and sadness ($M = 6.41$, $SD = 1.62$) except disgust ($M = 5.55$, $SD = 1.53$). On the other hand, the participants rated the valence: anger ($M = 1.55$, $SD = 0.74$), disgust ($M = 1.55$, $SD = 0.74$), fear ($M = 1.66$, $SD = 0.77$), and sadness ($M = 1.86$, $SD = 0.92$). The difference in the self-reported scores on the valence between four negative emotions was less than 0.5.

5.3.4 Recognition of discrete emotions

In Section 4.4, we proposed a three-level classification scheme based on prior knowledge to differentiate eight discrete emotions at the same time. At the first level, we separated high-arousal and valenced emotions from neutrality with an average accuracy of 92.26% ($SD = 0.17\%$) (see Table 3). At the second level, high-arousal and valenced emotions were further classified into positive or negative categories with an accuracy of 86.63% ($SD = 0.27\%$). The decrement in the classification accuracy could be attributed to a lower explanatory ability of the valence dimension compared to the valence-arousal coordinate

space. From a traditional valence-arousal perspective, the differences between high-arousal and valenced emotions and neutrality exist in both valence and arousal spaces, while discrimination between positive and negative emotions relies on valenced characteristics exclusively.

At the third level, positive emotions were classified into joy, amusement and tenderness with an accuracy of 86.43% (SD = 1.03%) (33.33% accuracy at random). Negative emotions were classified into anger, sadness, fear and disgust with an accuracy of 65.09% (SD = 0.66%) (25% accuracy at random). Applying this three-level classification scheme based on prior knowledge, our proposed system recognised eight discrete emotions at the same time with an overall accuracy of 60.55%, which was computed by $0.9226 \times 0.8663 \times (0.8643 + 0.6509) / 2$. In contrast, a baseline approach was to directly use the default eight-class classification in LIBSVM, which constructed 28 two-class classifiers (computed by $8 \times (8-1) / 2$). This eight-class classification resulted in an average accuracy of 32.31% (SD = 0.46%) (12.5% accuracy at random). Obviously, our proposed three-level classification scheme with prior knowledge outperformed an eight-class classification to differentiate eight discrete emotions at the same time.

To validate the integrity of the trained classification models, we repeat the classification task but with the shuffled emotion labels (i.e., randomly assigned labels) of testing dataset. In this scenario, our system achieved an overall accuracy of 47.67% in recognising high-arousal and valenced emotions from neutrality and 47.35% in recognising positive from negative emotions. Moreover, our system classified three positive emotions (joy, amusement, tenderness) with an average of 31.39% accuracy and four negative emotions (anger, disgust, fear, sadness) with an average of 20.77% accuracy. These accuracy ratios were much lower than those with unshuffled emotion labels, demonstrating that there was no systematic bias in the classification procedure.

TABLE 3

Comparison of classification accuracy between default eight-class classification in LIBSVM and three-level classification scheme with prior knowledge (average and standard deviation across participants)

Classification (number of two-class classifiers)	Accuracy (%)
Default eight-class classification in LIBSVM	
(28 two-class classifiers)	32.31 (0.46)
Our three-level classification scheme	
I. Non-neutrality vs. neutrality (2-classes: 1 two-class classifiers)	92.26 (0.17)
II. Positive vs. negative emotions (2-classes: 1 two-class classifiers)	86.63 (0.27)
III. Similar emotions in the valence-arousal coordinate space	
amusement vs. joy vs. tenderness (3-classes: 3 two-class classifiers)	86.43 (1.03)
anger vs. disgust vs. fear vs. sadness (4-classes: 6 two-class classifiers)	65.09 (0.66)

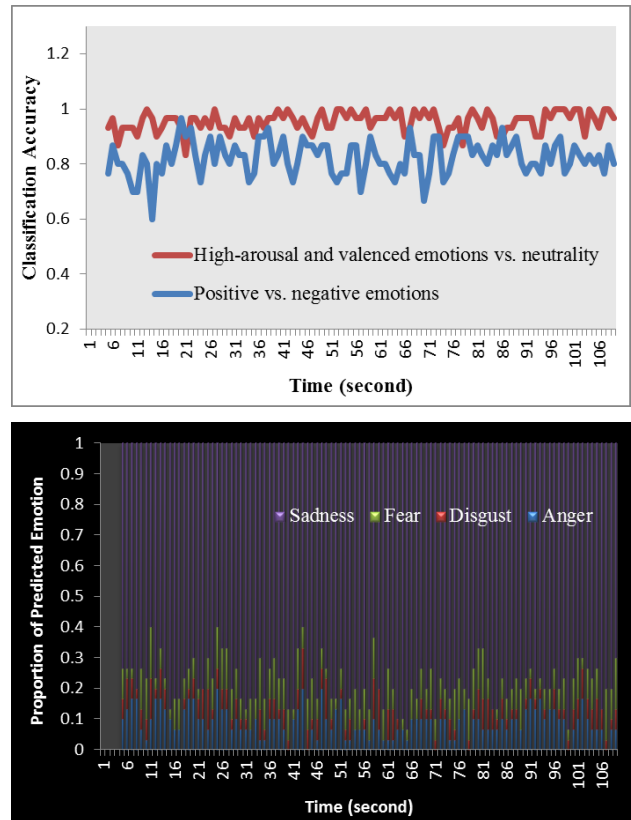


Fig. 10. Continuous recognition of valence and arousal states (upper panel) and four negative emotions similar in the valence-arousal coordinate space elicited by a sad film clip (lower panel).

More results are presented in Supplemental Material.

In addition, our system can be further used to recognise discrete emotions in a long-duration trial. Fig. 10 and Supplemental Material present the assessment of system performance as a function of time during movie watching. As illustrated in Fig. 10, our system predicted continuous classification accuracies of valence and arousal states and continuous recognition of four negative emotions similar in the valence-arousal coordinate space when participants watched a sad film clip. The highest proportion (average = 77.78%, range = 60%-93.33%) of negative emotions was sadness and this proportion of sadness reached the maximum at the time from 63 to 65 seconds (film content: the father found this son was died).

5.3.5 SLDA-based selected features

We found that frontal alpha asymmetry (FAA) features played an important role to recognize different emotional states. As shown in Table 4, F3-F4 distinguished between positive and negative emotions, while AF3-AF4 differentiated a positive emotion from similar emotions in both valence and arousal dimensions. Human brain reflect the affective and motivational processing of emotion states asymmetrically [45]. FAA refers to the patterns of asymmetrical anterior EEG activity in the alpha band, which was proved to be a typical indicator of asymmetric brain activity in the frontal cortex [46].

TABLE 4

SLDA-based top 5 PSD and asymmetry features. See Figure 2 for the notations of electrode placement.

Emotional states	Frequency bands				
	Delta	Theta	Alpha	Beta	Gamma
Positive vs. negative emotions	F4		FC5, F3-F4	AF4, F4	
Non-neutrality vs. neutrality		T7, T8, FC5	P7	F4	
amusement vs. joy vs. tenderness	F8		AF3- AF4, F7, P3	P4	
anger vs. disgust vs. fear vs. sadness	F8	P3, P4, P3-P4			T7

Past research have indicated that positive emotions correlates with approach motivation and negative emotions always correlates with withdrawal motivation [47]. The valence model of FAA states that greater left hemisphere activity (lower alpha power) is associated with positive emotion, whereas greater right hemisphere activity is associated with negative emotion [48]. Both trait positive (or negative) emotions and state positive (or negative) emotions are associated with greater left than right frontal cortical activity, which are manifested in greater right than left power values of alpha band [49]. However, anger is a typical negatively valenced emotion which evokes behavioral tendencies of approach [50-52]. The fact that higher scores on a trait anger scale associated with higher resting levels of left frontal activity and lower right frontal activity has been reported by several studies [53, 54]. Similar results are also found in the state-induced anger [47, 55].

We also found that delta and beta frequency bands in the right frontal and parietal lobe, theta frequency band in the temporal and parietal lobe, and gamma frequency band in the temporal lobe were sensitive to recognize different emotional states. Previous studies observed a positive correlation of valence with beta and gamma bands, emanating from anterior temporal cerebral sources [56]. A highly significant increase of temporal gamma power was found in [2], which might result from the impact of muscle activity. Theta frequency band in the frontal middle line [57, 58] and parietotemporal region [59] was another source of dominant features to recognize positive and negative emotions.

6 DISCUSSION

In this experiment, our system achieved an average accuracy of 92.26% in recognising high-arousal and valenced emotions from neutrality and 86.63% in recognising two valence classes. Due to having a limited number of existing real-time emotion recognition systems, we did not find any result in the literature that was comparable to our findings. Viet Hoang, Manh Ngo, Bang Ban and Thang Huynh [16] developed a real-time emotion recog-

nition system to identify 2 valence classes and 2 arousal classes, which resulted in 4 basic emotions (happy, relaxed, sad, and angry) and the neutral state. The average accuracy in recognising each emotion was 70.5% across all subjects. However, the existing work used standardised pictures or sounds to elicit the target emotions. Predictions of emotional states were made upon a number of independent pictures or sounds. In contrast, our system used standardised film clips to elicit target emotions. The EEG signals recorded from a film clip were segmented into a number of consecutive windows, and predictions were made on each segment. Moreover, the definition of the classification accuracy was different. The existing work (e.g., [16]) computed the proportion of correct classification for each emotion and the average accuracy over 5 emotional states. In contrast, we defined the accuracy as the number of correct predictions from all of the predictions made.

Interestingly, the average accuracy for recognising high-arousal and valenced emotions from neutrality (i.e., 92.26%) was higher than that of classifying two valence classes (i.e., 86.63%). One possible reason was that the trained classifier captured more information to enlarge its ability to perform pattern recognition in the valence-arousal coordinate space than the model that had a single valence dimension. Compared to the neutral state (i.e., low-arousal and non-valenced), a discrete emotion (i.e., high-arousal and valenced) in this work was different in both the arousal and valence dimensions. As a result, it was reasonable for our system to achieve a higher classification accuracy to recognise a discrete emotion from neutrality. Jatupaiboon, Pan-ngum and Israsena [15] used real-time EEG signals to classify happy and unhappy emotions (i.e., 2 classes of valence) that were elicited by pictures and classical music. The average accuracy of the subject-dependent model was 75.62%, which was lower than the performance of our system in recognising two valence classes.

In addition, our system successfully classified three positive emotions and four negative emotions with an overall accuracy of 86.43% and 65.09%, respectively. To the best of our knowledge, this work was the first attempt to classify similar discrete emotions that were close in the valence-arousal coordinate space. For example, in our experiment, the participants' subjective ratings on arousal were 5.83 for joy and 6.0 for amusement. Their self-reported scores on valence were 6.52 for joy and 6.76 for amusement. From the dimensional perspective to establish the emotional model, it was almost impossible to classify two or more similar discrete emotions that are close in the valence-arousal coordinate space. Even if we set up a cut-off point to divide joy and amusement into two classes, the recognition ability of the dimensional model would decline dramatically [12]. Similar emotions (such as joy and amusement in the valence-arousal dimension) could have their own unique representation and induce completely different behavioral responses [13]. In this case, our system demonstrated an advantage over the existing state-of-the-art real-time emotion recognition systems from EEG signals in terms of accuracy and

the ability to recognise similar emotional states in both valence and arousal dimensions.

7 CONCLUSIONS

In this paper, we established a standardised database of 16 Chinese film clips, and proposed a real-time EEG-based emotion recognition system for identifying an individual's emotional states through the analysis of brain waves. The system consisted of six modules: emotion elicitation, EEG data acquisition, data preprocessing, feature extraction, emotion classification, and human machine interface. We conducted an experiment to validate the efficiency and effectiveness of the system. Thirty participants took part in this experiment and watched a few emotional film clips that targeted seven basic discrete emotions (joy, amusement, tenderness, anger, disgust, fear, sadness) and neutrality. Our system achieved an average accuracy of 92.26% in recognising high-arousal and valenced emotions from neutrality and 86.63% in recognising two valence classes. In addition, the system successfully classified three positive emotions (joy, amusement, tenderness) and four negative emotions (anger, disgust, fear, sadness) with an overall accuracy of 86.43% and 65.09%, respectively. These results demonstrated an advantage over the existing state-of-the-art real-time emotion recognition systems from EEG signals in terms of the accuracy and the ability to recognise several similar discrete emotions that are close in the valence-arousal coordinate space.

ACKNOWLEDGMENTS

The authors thank the editor and reviewers for their valuable comments, which help improve this paper. This paper was done when Yong-Jin Liu visited the Institute of Psychology, Chinese Academy of Sciences, Beijing, China. This work was supported in part by the National Key Research and Development Plan (2016YFB1001200), the National Natural Science Foundation of China (Grant No. 31300852, 61661130156, 61521002) and the Royal Society-Newton Advanced Fellowship.

REFERENCES

- [1] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, "Toward Emotion Aware Computing: An Integrated Approach Using Multichannel Neurophysiological Recordings and Affective Visual Stimuli," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 589-597, May, 2010.
- [2] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A Database for Emotion Analysis Using Physiological Signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18-31, Jan-Mar, 2012.
- [3] I. B. Mauss, and M. D. Robinson, "Measures of emotion: A review," *Cognition and emotion*, vol. 23, no. 2, pp. 209-237, 2009.
- [4] R. A. Calvo, and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18-37, 2010.
- [5] M. Balconi, and C. Lucchiari, "EEG correlates (event-related desynchronization) of emotional face elaboration: a temporal analysis," *Neuroscience letters*, vol. 392, no. 1, pp. 118-123, 2006.
- [6] M. Y. Bekkedal, J. Rossi, and J. Panksepp, "Human brain EEG indices of emotions: delineating responses to affective vocalizations by measuring frontal theta event-related synchronization," *Neuroscience & Biobehavioral Reviews*, vol. 35, no. 9, pp. 1959-1970, 2011.
- [7] P. R. Davidson, R. D. Jones, and M. T. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 832-839, 2007.
- [8] Y. Liu, O. Sourina, and N. Minh Khoa, "Real-Time EEG-Based Emotion Recognition and Its Applications," *Transactions on Computational Science Xii: Special Issue on Cyberworlds*, Lecture Notes in Computer Science M. L. Gavrilova, C. J. K. Tan, A. Sourin and O. Sourina, eds., pp. 256-277, 2011.
- [9] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," pp. 667-670.
- [10] E. L. Van den Broek, "Ubiquitous emotion-aware computing," *Personal and Ubiquitous Computing*, vol. 17, no. 1, pp. 53-67, 2013.
- [11] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261-292, 1996.
- [12] J. R. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional," *Psychological science*, vol. 18, no. 12, pp. 1050-1057, 2007.
- [13] D. Yin, S. Bond, and H. Zhang, "Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews," *Mis Quarterly*, vol. 38, no. 2, pp. 539-560, 2014.
- [14] S. Hamann, "Mapping discrete and dimensional emotions onto the brain: controversies and consensus," *Trends in Cognitive Sciences*, vol. 16, no. 9, pp. 458-466, Sep, 2012.
- [15] N. Jatupaiboon, S. Pan-ngum, and P. Israsena, "Real-Time EEG-Based Happiness Detection System," *Scientific World Journal*, 2013, 2013.
- [16] A. Viet Hoang, V. Manh Ngo, H. Bang Ban, and Q. Thang Huynh, "A Real-Time Model Based Support Vector Machine for Emotion Recognition Through EEG," pp. 191-196.
- [17] T. Castermans, M. Duvinage, and N. Riche, "Online Emotion Classification from Electroencephalographic Signals: A First Study Conducted in a Realistic Movie Theater," *Artificial Intelligence and Soft Computing, Pt II*, Lecture Notes in Artificial Intelligence L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh and J. M. Zurada, eds., pp. 251-262, 2013.
- [18] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*: NIMH, Center for the Study of Emotion & Attention, 2005.
- [19] M. M. Bradley, and P. J. Lang, "The International Affective Digitized Sounds (; IADS-2): Affective ratings of sounds and

- instruction manual," *University of Florida, Gainesville, FL, Tech. Rep. B-3*, 2007.
- [20] Y. Baveye, E. Dellandrea, C. Chamaret, and L. Chen, "LIRIS-ACCEDE: A Video Database for Affective Content Analysis," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 43-55, 2015.
- [21] L. M. Jenkins, and D. G. Andrewes, "A New Set of Standardised Verbal and Non-verbal Contemporary Film Stimuli for the Elicitation of Emotions," *Brain Impairment*, vol. 13, no. 02, pp. 212-227, 2012.
- [22] P. Philippot, "Inducing and assessing differentiated emotion-feeling states in the laboratory," *Cognition & Emotion*, vol. 7, no. 2, pp. 171-193, 1993.
- [23] J. J. Gross, and R. W. Levenson, "Emotion elicitation using films," *Cognition & Emotion*, vol. 9, no. 1, pp. 87-108, 1995.
- [24] P. J. Lang, "Behavioral treatment and bio-behavioral assessment: computer applications," *Technology in mental health care delivery systems*, J. B. Sidowski, J. H. Johnson and T. A. Williams, eds., pp. 119-137: Ablex, 1980.
- [25] M. M. Bradley, and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49-59, 1994.
- [26] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales," *Journal of personality and social psychology*, vol. 54, no. 6, pp. 1063, 1988.
- [27] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cognition and Emotion*, vol. 24, no. 7, pp. 1153-1172, 2010.
- [28] EMOTIV. "Brain Computer Interface & Scientific Contextual EEG (<https://emotiv.com/product-specs/Emotiv%20EPOC%20Specifications%202014.pdf>)."
- [29] M. Soleymani, S. Asghari Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17-28, 2016.
- [30] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42-55, Jan-Mar, 2012.
- [31] M. Soleymani, M. Pantic, and T. Pun, "Multimodal Emotion Recognition in Response to Videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211-223, Apr-Jun, 2012.
- [32] A. Delorme, and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9-21, 2004.
- [33] T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert, "EEG artifact elimination by extraction of ICA-component features using image processing algorithms," *Journal of neuroscience methods*, vol. 243, pp. 84-93, 2015.
- [34] J.-C. Liao, and W.-C. Fang, "An ICA-based automatic eye blink artifact eliminator for real-time multi-channel EEG applications." pp. 532-535.
- [35] S. K. Hadjilidimitriou, and L. J. Hadjileontiadis, "EEG-Based Classification of Music Appraisal Responses Using Time-Frequency Analysis and Familiarity Ratings," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 161-172, Apr-Jun, 2013.
- [36] P. C. Petrantonakis, and L. J. Hadjileontiadis, "Adaptive Emotional Information Retrieval From EEG Signals in the Time-Frequency Domain," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2604-2616, May, 2012.
- [37] G. Pfurtscheller, and F. L. Da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842-1857, 1999.
- [38] S. Makeig, S. Debener, J. Onton, and A. Delorme, "Mining event-related brain dynamics," *Trends in cognitive sciences*, vol. 8, no. 5, pp. 204-210, 2004.
- [39] G. Pfurtscheller, "Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest," *Electroencephalography and clinical neurophysiology*, vol. 83, no. 1, pp. 62-69, 1992.
- [40] G. Pfurtscheller, and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG," *Electroencephalography and clinical neurophysiology*, vol. 42, no. 6, pp. 817-826, 1977.
- [41] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation," *Scientific World Journal*, 2014, 2014.
- [42] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification (Second Edition)*: John Wiley & Sons, 2012.
- [43] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, "Sparse discriminant analysis," *Technometrics*, vol. 53, no. 4, 2011.
- [44] C.-C. Chang, and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.
- [45] J. Wacker, M. Heldmann, and G. Stemmler, "Separating Emotion and Motivational Direction in Fear and Anger: Effects on Frontal Asymmetry," *Emotion*, vol. 3, no. 2, pp. 167-193, Jun, 2003.
- [46] B. B. Briesemeister, S. Tamm, A. Heine, and A. M. Jacobs, "Approach the Good, Withdraw from the Bad—A Review on Frontal Alpha Asymmetry Measures in Applied Psychological Research," *Psychology*, vol. 04, no. 03, pp. 261-267, 2013.
- [47] E. Harmon-Jones, and J. Sigelman, "State anger and prefrontal brain activity: Evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression," *Journal of Personality and Social Psychology*, vol. 80, no. 5, pp. 797-803, 2001.
- [48] B. D. Poole, and P. A. Gable, "Affective motivational direction drives asymmetric frontal hemisphere activation," *Experimental Brain Research*, vol. 232, no. 7, pp. 2121-2130, Jul, 2014.
- [49] E. Harmon-Jones, "Clarifying the emotive functions of asymmetrical frontal cortical activity," *Psychophysiology*, vol. 40, no. 6, pp. 838-848, Nov, 2003.
- [50] M. Lewis, S. M. Alessandri, and M. W. Sullivan, "Violation of expectancy, loss of control, and anger expressions in young infants," *Developmental Psychology*, vol. 26, no. 5, pp. 745-751, 1990.
- [51] M. Lewis, M. W. Sullivan, D. S. Ramsay, and S. M. Alessandri, "Individual differences in anger and sad expressions during

extinction: antecedents and consequences," *Infant Behavior and Development*, vol. 15, no. 4, pp. 443-452, 1992.

- [52] E. Harmon-Jones, "Contributions from research on anger and cognitive dissonance to understanding the motivational functions of asymmetrical frontal brain activity," *Biological Psychology*, vol. 67, no. 1-2, pp. 51-76, Oct, 2004.
- [53] J. Hewig, D. Hagemann, J. Seifert, E. Naumann, and D. Bartussek, "On the selective relation of frontal cortical asymmetry and anger-out versus anger-control," *Journal of Personality and Social Psychology*, vol. 87, no. 6, pp. 926-939, Dec, 2004.
- [54] M. Rybak, J. W. Crayton, I. J. Young, E. Herba, and L. M. Konopka, "Frontal alpha power asymmetry in aggressive children and adolescents with mood and disruptive behavior disorders," *Clinical Eeg and Neuroscience*, vol. 37, no. 1, pp. 16-24, Jan, 2006.
- [55] E. Harmon-Jones, "Trait anger predicts relative left frontal cortical activation to anger-inducing stimuli," *International Journal of Psychophysiology*, vol. 66, no. 2, pp. 154-160, Nov, 2007.
- [56] J. A. Onton, and S. Makeig, "High-frequency broadband modulation of electroencephalographic spectra," *Frontiers in Human Neuroscience*, vol. 3, pp. 61, 2009.
- [57] L. I. Aftanas, and S. A. Golocheikine, "Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalized attention: high-resolution EEG investigation of meditation," *Neuroscience Letters*, vol. 310, no. 1, pp. 57-60, Sep, 2001.
- [58] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, no. 2, pp. 293-304, Mar, 2007.
- [59] L. Aftanas, A. Varlamov, S. Pavlov, V. Makhnev, and N. Reva, "Affective picture processing: event-related synchronization within individually defined human theta band is modulated by valence dimension," *Neuroscience Letters*, vol. 303, no. 2, pp. 115-118, 2001.



Guozhen Zhao received the B.S. degree in industrial engineering, from Tianjin University, China, in 2007, and the M.S. and Ph.D. degrees in industrial and systems engineering from the State University of New York, Buffalo, in 2009 and 2011, respectively. He is an Assistant Professor in the Institute of Psychology, Chinese Academy of Sciences. His current research interest is the mathematical modeling of human cognition and performance, human computer interaction, emotion recognition, and augmented cognition.



Jinjing Song received the B.S. degree in psychology, from Wuhan University, China, in 2013, and M.S. degree from the Institute of Psychology, Chinese Academy of Sciences, China, in 2016. Her research interest is emotion recognition based on EEG and physiological signals.



Yan Ge received the B.S. degree from Beijing in psychology, from Normal University, China, in 2003, and the Ph.D. degree from Institute of Psychology, Chinese Academy of Sciences in 2008. She is an Associate Professor at Institute of Psychology, Chinese Academy of Sciences. Her main research interest is human-computer interaction.



Yuanchun Shi is a Changjiang distinguished professor at the Department of Computer Science, Tsinghua University. She received her PhD, MS and BS in Computer Science from Tsinghua University. She was a senior visiting scholar at MIT AI lab during 2001-2002. She had chaired several conferences including ACM Ubicomp2011. She serves as the Area Editor of Pervasive and Mobile

Computing (PMC, Elsevier), Editor of Interacting with Computer (IwC, Oxford University Press) and Vice Editor-in-Chief of Communications of China Computer Federation (CCF). Prof. Shi has published more than one hundred papers in IJHCS, IEEE TPDS, TKDE, ACM CHI, MM, UIST, etc. Her research interests include human computer interaction, pervasive computing and multimedia communication.



Yong-Jin Liu received the B. Eng. Degree in mechano-electronic engineering from Tianjin University, China, in 1998, and the Ph.D. degree from the Hong Kong University of Science and Technology, Hong Kong, in 2003. He is an Associate Professor with the Department of Computer Science and Technology, Tsinghua University, China. He is a senior member of IEEE. His research interests include pattern analysis, human-computer interaction and computer graphics.



Min-Jing Yu is a Ph.D student with TNList, Department of Computer Science and Technology, Tsinghua University. She received her B.Eng. degree from Wuhan University, China, in 2014. Her research interests include cognitive computation, computer graphics and computer vision.