EmotionMeter: A Multimodal Framework for Recognizing Human Emotions

Wei-Long Zheng, *Student Member, IEEE*, Wei Liu, Yifei Lu, Bao-Liang Lu, *Senior Member, IEEE*, and Andrzej Cichocki, *Fellow, IEEE*

***Abstract*—In this paper, we present a multimodal emotion recognition framework called *EmotionMeter* that combines brain waves and eye movements. To increase the feasibility and wearability of *EmotionMeter* in real-world applications, we design a six-electrode placement above the ears to collect electroencephalography (EEG) signals. We combine EEG and eye movements for integrating the internal cognitive states and external subconscious behaviors of users to improve the recognition accuracy of *EmotionMeter*. The experimental results demonstrate that modality fusion with multimodal deep neural networks can significantly enhance the performance compared with a single modality, and the best mean accuracy of 85.11% is achieved for four emotions (happy, sad, fear, and neutral). We explore the complementary characteristics of EEG and eye movements for their representational capacities and identify that EEG has the advantage of classifying happy emotion, whereas eye movements outperform EEG in recognizing fear emotion. To investigate the stability of *EmotionMeter* over time, each subject performs the experiments three times on different days. *EmotionMeter* obtains a mean recognition accuracy of 72.39% across sessions with the six-electrode EEG and eye movement features. These experimental results demonstrate the effectiveness of *EmotionMeter* within and between sessions.**

***Index Terms*—Affective brain-computer interactions, deep learning, EEG, emotion recognition, eye movements, multimodal deep neural networks.**

I. Introduction

EMOTION plays an important role in human-human interactions in our everyday lives. In addition to logical intelligence, emotional intelligence is considered to be an important part of human intelligence [1]. There is an increasing focus on developing emotional artificial intelligence in human–computer interactions (HCIs) [2]. The introduction of affective factors to HCIs has rapidly been developed as an interdisciplinary research field called affective computing [3], [4]. Affective computing attempts to develop human-aware artificial intelligence that has the ability to perceive, understand, and regulate emotions. Specifically, emotion recognition is the critical phase in this affective cycle and has been a primary focus of HCI researchers. Moreover, many mental diseases are reported to be relevant to emotions, such as depression, autism, attention deficit hyperactivity disorder, and game addiction [5], [6]. However, due to the limited knowledge of the neural mechanisms underlying emotion processing, an efficient quantified measure for emotions with convenient setups to provide active feedback with evaluations for disease treatments is still lacking [2]. One of the twenty big questions about the future of humanity reported by *Scientific American* is whether we can use wearable technologies to detect human emotions.[1] Smart wearable devices have a high potential for enhancing HCI performance and treating psychiatric diseases.

Emotions are complex psycho-physiological processes that are associated with many external and internal activities. Different modalities describe different aspects of emotions and contain complementary information. Integrating this information with fusion technologies is attractive for constructing robust emotion recognition models [7], [8]. However, most studies have focused on combining auditory and visual modalities for multimodal emotion recognition [9]. In contrast, the combination of signals from the central nervous system, e.g., EEG, and external behaviors, e.g., eye movements, has been reported to be a promising approach [7], [10], [11]. Recent studies have attempted to identify emotion-specific neural markers to understand the nature of emotions [12], [13]. Affective brain-computer interfaces (aBCIs) aim to detect

Manuscript received September 4, 2017; revised November 24, 2017; accepted January 17, 2018. Date of publication February 7, 2018; date of current version February 14, 2019. The work of W.-L. Zheng, W. Liu, Y. Lu, and B.-L. Lu was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1002501, in part by the National Natural Science Foundation of China under Grant 61673266, in part by the Major Basic Research Program of Shanghai Science and Technology Committee under Grant 15JC1400103, in part by the ZBYY-MOE Joint Funding under Grant 6141A02022604, in part by the Technology Research and Development Program of China Railway Corporation under Grant 2016Z003-B, and in part by the Fundamental Research Funds for the Central Universities. The work of A. Cichocki was supported in part by the Ministry of Education and Science of the Russian Federation under Grant 14.756.31.0001, and in part by the Polish National Science Center under Grant 2016/20/W/N24/00354. This paper was recommended by Associate Editor H. A. Abbass. *(Corresponding author: Bao-Liang Lu.)*

W.-L. Zheng, W. Liu, Y. Lu, and B.-L. Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, also with the Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Brain Science and Technology Research Center, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: blu@cs.sjtu.edu.cn).

A. Cichocki is with Nicolaus Copernicus University, Torun 87-100, Poland, also with the Skolkovo Institute of Science and Technology (Skoltech), Moscow 143026, Russia, and also with RIKEN Brain Science Institute, Wako 351-0198, Japan.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2018.2797176

[1] http://www.scientificamerican.com/article/20-big-questions-about-the-future-of-humanity/

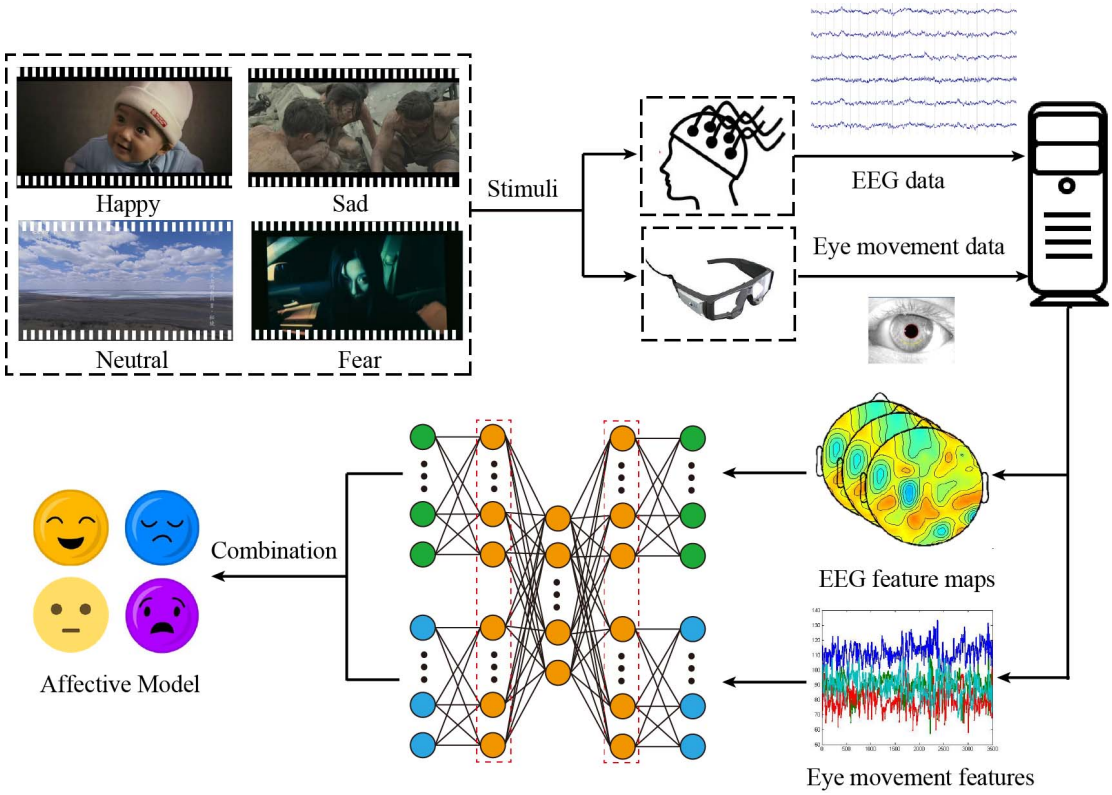


Fig. 1. Framework of our proposed approach. EEG and eye tracking data are simultaneously recorded while subjects watch emotional film clips as stimuli. The multimodal features are extracted from raw EEG and eye movement signals. The extracted features are used to feed the multimodal deep neural networks for learning high-level shared representations. The emotion predictions are given with the affective models based on the shared representations.

emotions from brain signals [14]. Compared to brain signals, eye movement signals convey important social and emotional information for context-aware environments [15], [16].

In this paper, to integrate the internal brain activities and external subconscious behaviors of users, we present a multimodal framework called *EmotionMeter* to recognize human emotions using six EEG electrodes and eye-tracking glasses. The framework of our proposed approach is shown in Fig. 1. The main contributions of this paper are as follows.

1) We developed a novel design of six symmetrical temporal electrodes that can easily be embedded in a headset or spectacle frames for implementing EEG-based emotion recognition.
2) We performed experiments for recognizing four emotions (happy, sad, fear, and neutral emotions) to evaluate the efficiency of the proposed design.
3) We revealed the complementary characteristics of EEG and eye movements for emotion recognition, and we improved the performance by using multimodal deep neural networks.
4) We investigated the stability of our proposed framework using training and test datasets from different sessions and demonstrated its stability within and between sessions.

The remainder of this paper is organized as follows. Section II briefly reports the related work in emotion recognition using EEG and eye movements. The experimental setup is presented in Section III. Section IV introduces the details of preprocessing, feature extraction, classification methods, and multimodal deep neural networks. Section V presents the experimental results and discussion. The conclusions and future work are summarized in Section VI.

II. Related Work

A. Emotion Recognition Systems

Various wearable emotion perception and feedback prototypes have been developed and evaluated. For example, MacLean *et al.* [17] presented a real-time biofeedback system called *MoodWings*, where a wearable butterfly can respond to users' arousal states through wing actuation. Williams *et al.* [18] designed a wearable device called *SWARM* that can react to emotions. Valtchanov *et al.* presented a system called *EnviroPulse* for automatically determining the expected affective valence of surrounding environments to individuals [19]. Recently, Hassib *et al.* [20] presented *EngageMeter* for implicit audience engagement sensing using the commercial NeuroSky MindWave headset in real-world evaluations. However, only one-channel EEG was collected from the frontal cortex (FP1) with *EngageMeter*, and the information was limited to calculating only the engagement index rather than the emotion index.

Facial expression is one of the popular modalities for emotion recognition. By studying facial expression of different cultures, Ekman and Friesen [21] proposed the concept "basic emotions" including fear, anger, surprise, disgust, joy, and

sadness that are universal across cultures. Based on this finding, many approaches to emotion recognition using facial expressions have been proposed in the past decades [4], [22]. Although emotional recognition has long focused on facial expression, there is a growing interest in many other modalities like touch and EEG. Tsalamlal *et al.* [23] presented a study of combining facial expressions and touch for evaluating emotional valence. Schirmer and Adolphs [24] compared different modalities for emotion perception, including facial expression, voice, and tactile, and these data are usually analyzed with behavioral statistics, EEG and fMRI studies. They reviewed the similarities and differences of these modalities for emotion recognition and proposed multisensory integration during different stages. Koelstra [25] proposed a multimodal approach to fusing facial expressions and EEG signals for affective tagging using feature-level and decision-level fusion strategies and presented the performance improvement. These results indicated that facial expressions and EEG contain complementary information.

Recently, Alarcao and Fonseca [26] presented a detailed survey about emotion recognition using EEG signals, including stimuli, feature extraction, and classifiers. Mühl *et al.* [14] presented the idea of aBCIs and discussed the limitations and challenges in this research field. Jenke *et al.* [27] performed a systematical comparison of feature extraction and selection methods for EEG-based emotion recognition. Daly *et al.* [28] presented the demonstration of affective brain computer music interface to detect users' affective states. Petrantonakis and Hadjileontiadis [29] proposed a new feature extraction method using hybrid adaptive filtering and higher order crossings for classifying six basic emotions with EEG. We adopted the combination of deep belief networks and hidden markov model to classify two emotions (positive and negative) using EEG signals in our previous study [30].

B. Wearable EEG Devices

Since emotions have many indicators inside and outside our body, various modalities have been adopted for constructing emotion recognition models, such as facial expressions, voices, and gestures [4]. Among these approaches, EEG-based methods are considered to be promising approaches for emotion recognition because many findings in neuroscience support the hypothesis that brain activity is associated with emotions [2], [12], [13] and that EEG allows for the direct assessment of the "inner" states of users [27]. The existing studies have shown the effectiveness and feasibility of EEG [14], [27], [31], [32]. However, most of these studies attach many wet electrodes (some with as many as 62 electrodes). In addition to the substantial time costs for mounting the electrodes, the irrelevant channels may introduce noise and artifacts in the systems, and therefore, degrade the performance. The HCI community calls for convenient setups and easy, user-friendly usage for affective brain-computer interactions.

Fortunately, with the rapid development of wearable devices and dry electrode techniques [33], [34], it is now possible to develop wearable EEG devices from laboratories to real-world applications. In fact, wearable EEG devices have many potential applications. For example, a pilot wearing such a device can adjust his/her emotions to enhance flying safety if the device detects that he or she is in an extremely emotional state. A medical rehabilitation system with emotional intelligence can adjust its rehabilitation training plan according to the emotional fluctuations of the patients. Computer games, which can change the content and scene according to the player's emotions, will have a richer user experience. The most popular commercial wearable EEG device is the Emotiv EPOC wireless headset [35]. Previous studies have assessed the feasibility of emotion recognition using the Emotiv EPOC [36], [37]. However, its design of 14 electrode placements may still be inconvenient or even inappropriate for emotion recognition. For example, the Emotiv EPOC provides only three main affective scores, namely, excitement, engagement, and frustration, which are not included in the classical emotion category. A new wearable EEG device with easy setups for emotion recognition is attractive in HCI. To achieve this goal, several open questions, such as the best electrode placements in terms of wearability and feasibility, need to be further investigated. In this paper, we utilize a considerably lower number of electrodes for EEG recordings compared with the Emotiv EPOC.

C. Eye Movement Experiments

Humans interact with their surrounding environments, and each evoked emotion has its specific context [19]. Therefore, Ptaszynski *et al.* [1] proposed the need to apply contextual analysis to emotion precessing. Eye movements have long been studied as an approach to users' behaviors and cognitive states in HCI [38]. Specifically, previous studies have reported that pupil response is associated with cognitive and emotional processes [39], [40]. Zekveld *et al.* [41] reported the neural correlates of pupil size as a measure of cognitive listening load and proposed the eye as a window to the listening brain. Moreover, other eye movements, such as fixation, saccade, and blink, provide important cues for context-aware environments [42]–[44], which reveal how a user interacts with their surroundings and what attracts a user's attention. At the Consumer Electronics Show 2016 in Las Vegas, Looxid Labs presented a prototype that combines eye tracking and two-electrode frontal EEG into one headset for device control and attention monitoring.[2] Soleymani *et al.* [7] presented a subject-independent emotion recognition approach using EEG, pupillary response and gaze distance. They achieved the best classification accuracies of 68.5% and 76.4% for three labels of valence and arousal, respectively. The combination of brain signals and eye movements has been shown to be a promising approach for modeling user cognitive states. Recently, Langer *et al.* [45] presented a multimodal dataset that combines EEG and eye tracking from 126 subjects for assessing information processing in the developing brain. Wang *et al.* [46] proposed investigating vigilance fluctuation using fMRI dynamic connectivity states. Using eye-tracking

[2]http://looxidlabs.com/

techniques, they were able to observe spontaneous eyelid closures as a nonintrusive arousal-monitoring approach. However, few studies have discussed the complementary characteristics between them and the stability across sessions.

Eye movements and facial expressions have different characteristics for emotion recognition. The popular approach using facial expressions enjoys some advantages of nonintrusive setups, low-cost hardware, and reasonable accuracy. Despite the great progress in the development of facial expression-based emotion recognition, current systems have the following limitations: 1) most systems are evaluated with posed facial expressions rather than naturalistic facial expressions, and sometimes facial expressions can be subjectively controlled due to social nature of emotion; 2) the performance of computer vision systems are usually degraded due to the illumination variations and occlusion in real-world scenarios; and 3) although emotion can never be divorced from context, most facial expression systems do not consider the contextual cues [4]. In contrast, eye movements provide an effective tool to observe users' behaviors in a natural way. Eye movements contain not only physiological signals, e.g., pupil response, but also important contextual clues for emotion recognition. In comparison with facial expressions, eye movements are also nonintrusive and accurate despite higher cost for calibration setups. Moreover, eye tracking can be well embedded in recent popular user-centered wearable technologies, e.g., virtual reality devices. Therefore, eye movements have received more and more attention in affective computing field.

D. Multimodal Approaches

In our previous studies, we applied the feature-level fusion, decision-level fusion, and bimodal deep autoencoder to classify three emotions (positive, neutral, and negative) using EEG and eye movements [10], [11], [47], [48]. Our previous experimental results indicated that eye movements contain complementary information for emotion recognition. However, these studies investigated 62 electrodes all over the entire brain areas and did not consider the wearability in real-world applications. In this paper, we dramatically simplify the EEG setup with only six electrodes placed over the ears, and we attempt to classify four emotion categories: 1) happy; 2) sad; 3) fear; and 4) neutral emotions.

The studies of internal consistency and test-retest stability of EEG can be traced back to many years ago [49], [50]. However, the stability of emotion recognition systems has received very limited attention [51]. The fluctuation in performance for emotion recognition systems over time is still unclear for the development of real-world applications. Lan *et al.* [52] presented a pilot study on the stability of features in emotion recognition algorithms. In their stability assessment, the same features derived from the same channel from the same emotion class of the same subject were grouped together to compute the correlation coefficients. In contrast to their statistical approach, we investigate the stability of our method in a machine learning framework with cross-session evaluations.

In recent years, deep neural networks have achieved substantial state-of-the-art performance in various research fields, such as object detection, speech recognition, and natural language processing, with the ability to learn different representations of data with multiple layer training [53]. Some researchers introduced deep neural networks to EEG processing, and their experimental results demonstrated the superior performance of such networks compared with the conventional shallow methods [30], [54]–[57]. To leverage the advantages of two modalities, various multimodal deep architectures have been proposed. Ngiam *et al.* [58] proposed learning effective shared representations over multiple modalities (e.g., audio and video) with multimodal deep learning. Recently, Tzirakis *et al.* [9] proposed end-to-end multimodal emotion recognition with auditory and visual modalities. They utilized a convolutional neural network and a deep residual network to construct the speech network and visual network, respectively. The outputs of the two networks were concatenated as the input of a two-layer LSTM to capture the contextual information. Despite the above promising and successful applications of multimodal deep neural networks to auditory and visual data, they remain largely unexplored in the multimodal neuroimaging domain, particularly for combining EEG and eye movements for emotion recognition.

III. Experimental Setup

To elicit specific emotions in the experimental environment, we used carefully selected film clips as the stimuli. The reliability of film clips with audiovisual stimuli in eliciting emotions has been studied in the literature [59]–[61]. In our preliminary study, we collected film clips with highly emotional contents and ensured the integrity of the plot within the clips. The criteria for clip selection were as follows: 1) the length of the videos should not be too long to cause visual fatigue; 2) the videos should be understood without explanation; and 3) the videos should elicit a single desired target emotion.

There were 168 film clips in total for four emotions (happy, sad, fear, and neutral) in our material pool, and forty-four participants (22 females, all college students) were asked to assess their emotions when watching the film clips with keywords of emotions (happy, sad, neutral, and fear) and ratings out of ten points (from -5 to 5) for two dimensions: 1) valence and 2) arousal. The valence scale ranges from sad to happy. The arousal scale ranges from calm to excited. The mean rating distributions of the different film clips are shown in Fig. 2. As shown in this figure, there are significant differences between the conditions in terms of the ratings of valence and arousal, reflecting the successful elicitation of the targeted emotions in the laboratory environments.

Finally, 72 film clips were selected from the pool of materials that received the highest match across participants. The stimuli of these selected clips generally resulted in the elicitation of the four target emotions. The duration of each film clip was approximately two minutes. To avoid repetition, each film clip was presented only once. To investigate the stability of our model over time, we designed three different sessions

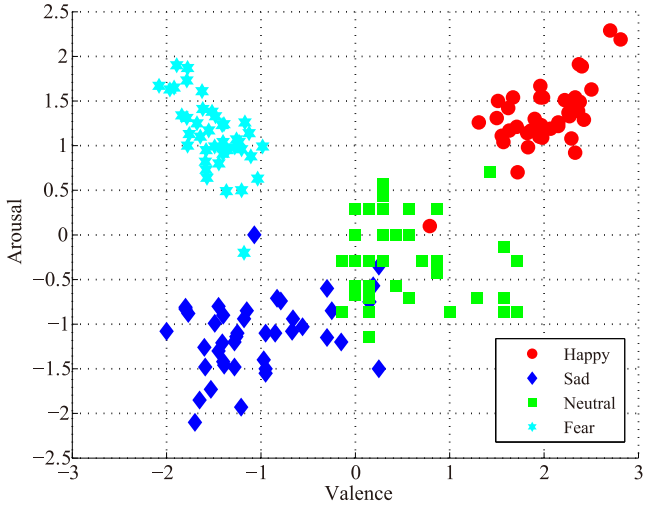


Fig. 2. Mean rating distributions of the different film clips on the arousal-valence plane for four emotions. The ratings are clustered into four classes: happy, sad, fear, and neutral emotions.

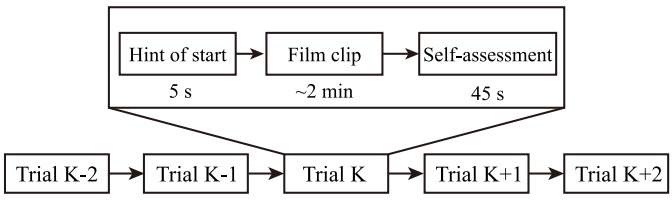


Fig. 3. Protocol of our designed emotion experiments.

for each participant on different days. Each session consisted of 24 trials (six trials per emotion), and the stimuli for the three sessions were completely different. Fig. 3 presents the detailed protocol of our designed emotion experiments. Each film clip had a 5 s hint for starting and a 45 s self-assessment with the PANAS scales [62] after each clip. The participants were asked to watch the emotional clips and elicit the corresponding emotions. The ratings of the subjects were based on how they actually felt while watching the clips rather than what they thought the film clips should be. According to the feedback, if the participants failed to elicit the correct emotions or the arousal emotions were not strong enough, the data were discarded.

The data recording experiments included a total of 15 healthy, right-handed participants (eight females) aged between 20 and 24 years. Prior to each experiment, the participants were informed of the purpose and procedure of the experiment and of the harmlessness of the equipment. Each participant participated in the experiment three times on different days, and a total of 45 experiments were evaluated. The dataset (SEED-IV) used in this paper will be freely available to the academic community as a subset of SEED.[3]

Motivated by our previous findings of critical brain areas for emotion recognition [55], [63], we selected six symmetrical temporal electrodes above the ears, which are FT7, FT8, T7, T8, TP7, and TP8 of the international 10–20 system shown in Fig. 4, as the critical channels for EEG-based emotion recognition in terms of wearability and feasibility in real-world

[3]http://bcmi.sjtu.edu.cn/~seed

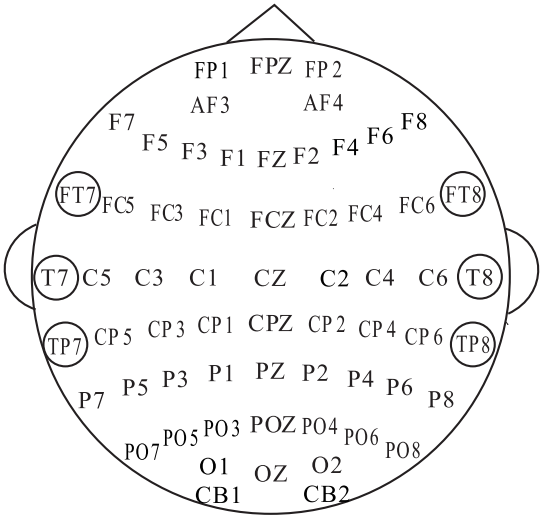


Fig. 4. EEG electrode layout of 62 channels. Six symmetrical temporal electrodes (FT7, FT8, T7, T8, TP7, and TP8) are selected in *EmotionMeter*.

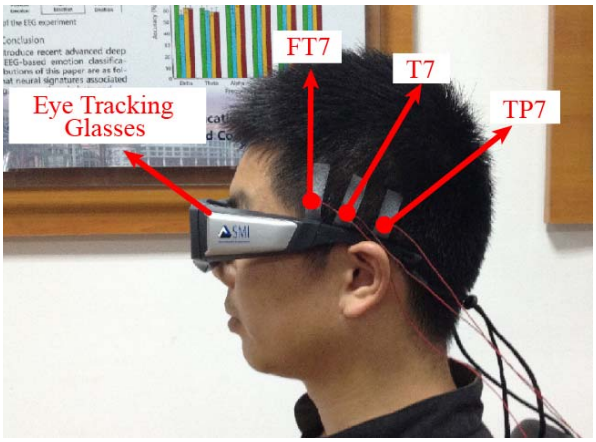


Fig. 5. Setup for the *EmotionMeter* hardware. The six symmetrical temporal electrodes above the ears are used for EEG recordings. The eye movement parameters are extracted from the wearable eye-tracking glasses.

applications. These electrodes can easily be embedded in a wearable headset or spectacle frames. Although the frontal asymmetry has been found to correlate with emotional valence in [64] and [65], we found that the frontal electrodes did not greatly contribute to enhancing classification performance with the temporal electrodes from our previous study [55]. Therefore, the frontal electrodes were not included in this paper.

For comparison, we also simultaneously recorded 62-channel EEG data according to the international 10–20 system. The raw EEG data were recorded at a 1000 Hz sampling rate using the ESI NeuroScan System.[4] Eye movements were also simultaneously recorded using SMI ETG eye-tracking glasses.[5] Fig. 5 presents the setup for the *EmotionMeter* hardware, where the left-hand side of the placement of these six electrodes is shown.

[4]http://compumedicsneuroscan.com/

[5]https://www.smivision.com/eye-tracking/product/eye-tracking-glasses/

IV. Methods

A. Preprocessing

Eye movement data from eye-tracking glasses provide various detailed parameters, such as pupil diameters, fixation details, saccade details, blink details, and event statistics. Although the pupil diameter is associated with emotional processing, it can easily be affected by the environmental luminance [40]. Based on the observations that the changes in the pupil responses of different participants to the same stimuli have similar patterns, we applied a principal component analysis (PCA)-based method to estimate the pupillary light reflex [7].

Suppose that $\mathbf{Y}$ is an $M \times N$ matrix that represents pupil diameters to the same video clip from N subjects and M samples. Then, $\mathbf{Y} = \mathbf{A} + \mathbf{B} + \mathbf{C}$, where $\mathbf{A}$ is luminance influences that are prominent, $\mathbf{B}$ is emotional influences that we want, and $\mathbf{C}$ is noise. We used PCA to decompose $\mathbf{Y}$ and computed the first principle component as the estimate of the light reflex. Let $\mathbf{Y}_{\text{rest}}$ be the emotion-relevant pupil response. We define $\mathbf{Y}_{\text{rest}} = \mathbf{Y} - \mathbf{Y}_1$. After subtracting the first principle component, the residual part contains the pupil response that is associated only with emotions. For EEG, a band-pass filter between 1 and 75 Hz was applied to filter the unrelated artifacts. We resampled the EEG and eye movement data to reduce the computational complexity and align these two modalities.

B. Feature Extraction

After preprocessing the EEG data, we extracted two types of features proposed in our previous studies, namely, power spectral density (PSD) and differential entropy (DE) [55], [66], [67], using short-term Fourier transforms with a 4 s time window without overlapping. The DE feature is defined as follows:

$$h(\mathbf{X}) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} \log\frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} dx = \frac{1}{2}\log 2\pi e\sigma^2 \quad (1)$$

where $\mathbf{X}$ submits the Gaussian distribution $N(\mu, \sigma^2)$, x is a variable, and π and e are constants. DE is equivalent to the logarithmic PSD for a fixed-length EEG sequence in a certain band. In contrast to PSD features, the DE features have the balance ability of discriminating EEG patterns between low- and high-frequency energy.

We computed the PSD and DE features in five frequency bands for each channel: 1) delta: 1–4 Hz; 2) theta: 4–8 Hz; 3) alpha: 8–14 Hz; 4) beta: 14–31 Hz; and 5) gamma: 31–50 Hz. The dimensions of the PSD and DE features are 10, 20 and 30 for two electrodes (T7 and T8), four electrodes (T7, T8, FT7, and FT8), and six electrodes (T7, T8, FT7, FT8, TP7, and TP8), respectively. We applied the linear dynamic system approach to filter out noise and artifacts that were unrelated to the EEG features [68].

For eye movements, we extracted various features from different detailed parameters used in the literature, such as pupil diameter, fixation, saccade, and blink [7], [11]. The details of the features extracted from eye movements are shown in Table I. The total number of dimensions of the eye movement features is 33. Fig. 6 illustrates five eye movement parameters.

TABLE I
Details of the Extracted Eye Movement Features

Eye movement parameters	Extracted features
Pupil diameter (X and Y)	Mean, standard deviation and DE features in four bands: 0-0.2 Hz, 0.2-0.4 Hz, 0.4-0.6 Hz, and 0.6-1 Hz
Dispersion (X and Y)	Mean, standard deviation
Fixation duration (ms)	Mean, standard deviation
Blink duration (ms)	Mean, standard deviation
Saccade	Mean, standard deviation of saccade duration (ms) and saccade amplitude (°)
Event statistics	Blink frequency, fixation frequency, maximum fixation duration, total fixation dispersion, maximum fixation dispersion, saccade frequency, average saccade duration, average saccade amplitude, and average saccade latency.

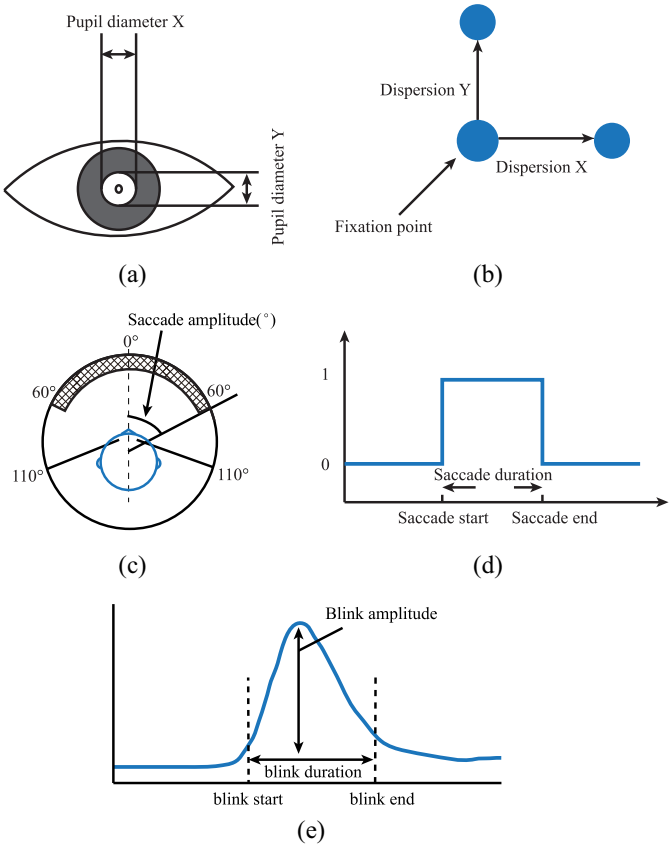


Fig. 6. Illustration of various eye movement parameters: pupil diameter, fixation dispersion, saccade amplitude, saccade duration, and blink.

C. Classification Methods

As a baseline classification method, we used support vector machine (SVM) with a linear kernel as the classifier. We used

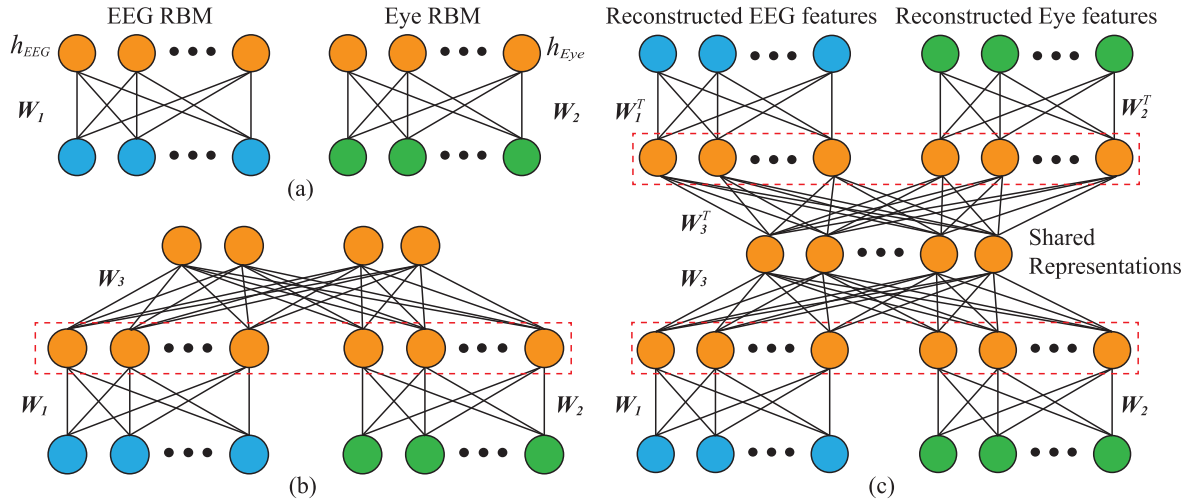


Fig. 7. Deep neural network architectures adopted in this paper. (a) Two RBMs were constructed using EEG and eye movement features as input. (b) Two hidden layers of EEG RBM and eye RBM were concatenated, and an upper RBM was trained above them. (c) Stacked RBMs were unfolded into a BDAE, and the shared representations of both modalities were learned from the neural networks.

the Liblinear toolbox to implement SVM [69]. For training, we searched the parameter space $2^{[-10:10]}$ for C to find the optimal value. The feature extraction and SVM used in this paper are implemented in MATLAB. We used both the accuracy and standard deviation for evaluation. We adopted feature-level fusion as a baseline for modality fusion. The feature vectors of EEG and eye movements were directly concatenated into a larger feature vector as the inputs of the classifiers. For the experimental evaluations of a single modality and multiple modalities, we separated the data from one experiment into training data and test data, where the first 16 trials are the training data, and the last eight trials containing all emotions (each emotion with two trials) are the test data. To analyze the performance consistency across sessions, the data of one session are used as the training data, and the data of another session are used as the test data.

D. Multimodal Deep Learning

To enhance the recognition performance, we adopted a bimodal deep auto-encoder (BDAE) [58] to extract the shared representations of both EEG and eye movements. Fig. 7 depicts the deep neural network architectures designed for our proposed approach. In contrast to the conventional approaches, where the direct concatenation of both feature vectors from different modalities are simply fed to a neural network, we train individual networks for different modalities. Two restricted Boltzmann machines (RBMs) called EEG RBM and eye RBM were constructed using EEG and eye movement data, respectively. An RBM is an undirected graph model with a visible layer and a hidden layer. There are no visible-visible connections and no hidden-hidden connections.

The visible and hidden layers each have a bias vector, $\mathbf{a}$ and $\mathbf{b}$, respectively. In an RBM, the joint distribution $p(\mathbf{v}, \mathbf{h}; \theta)$ over the visible units $\mathbf{v}$ and hidden units $\mathbf{h}$, given the model parameters θ, is defined in terms of an energy function $E(\mathbf{v}, \mathbf{h}; \theta)$ of

$$P(\mathbf{v}, \mathbf{h}; \theta) = \frac{\exp(-E(\mathbf{v}, \mathbf{h}; \theta))}{Z} \tag{2}$$

where $Z = \sum_v \sum_h \exp(-E(\mathbf{v}, \mathbf{h}; \theta))$ is a normalization factor, and the marginal probability that the model assigns to a visible vector $\mathbf{v}$ is

$$P(\mathbf{v}; \theta) = \frac{\sum_h \exp(-E(\mathbf{v}, \mathbf{h}; \theta))}{Z}. \tag{3}$$

For a Bernoulli (visible)—Bernoulli (visible) RBM, the energy function is defined as

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij} v_i h_j - \sum_{i=1}^{I} b_i v_i - \sum_{j=1}^{J} b_j h_j \tag{4}$$

where w_{ij} is the symmetric interaction term between visible unit v_i and hidden unit h_j, b_i, and a_j are the bias terms, and I and J are the numbers of visible and hidden units, respectively. $\mathbf{W}$ denotes the weights between visible and hidden layers. The conditional probabilities can efficiently be calculated as follows:

$$P\left(h_j = 1 | \mathbf{v}; \theta\right) = \sigma\left(\sum_{i=1}^{I} w_{ij} v_i + a_j\right) \tag{5}$$

$$P\left(v_j = 1 | \mathbf{h}; \theta\right) = \sigma\left(\sum_{j=1}^{J} w_{ij} h_j + b_i\right) \tag{6}$$

where $\sigma(x) = 1/(1 + \exp(x))$.

Taking the gradient of the log likelihood $\log p(\mathbf{v}; \theta)$, we can derive the update rule for the RBM weights as

$$\Delta w_{ij} = E_{\text{data}}\left(v_i h_j\right) - E_{\text{model}}\left(v_i h_j\right) \tag{7}$$

where $E_{\text{data}}(v_i h_j)$ is the expectation observed in the training set and $E_{\text{model}}(v_i h_j)$ is the same expectation under the distribution defined by the model. The contrastive divergence algorithm [70] is adopted to train RBMs using Gibbs sampling since $E_{\text{model}}(v_i h_j)$ is intractable.

As shown in Fig. 7, two hidden layers of EEG RBM and eye RBM were further concatenated as the input of the upper RBM. The stacked RBMs were unfolded into a BDAE, and an unsupervised back-propagation algorithm was used to fine

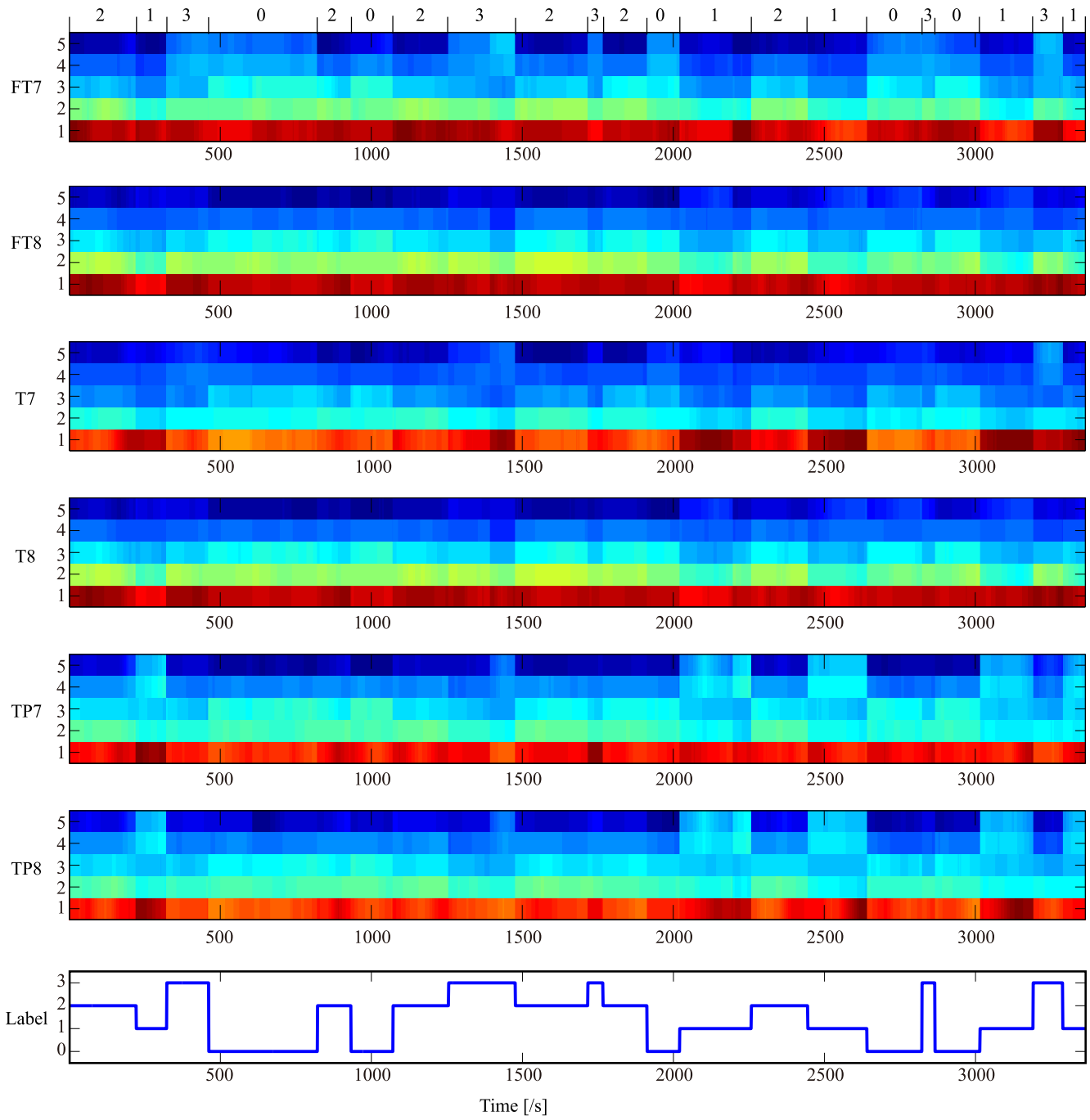


Fig. 8. Visualization of the DE features and the corresponding labels in one experiment. Here, labels with 0, 1, 2, and 3 denote the ground truth, neutral, sad, fear, and happy emotions, respectively. The numbers 1, 2, 3, 4, and 5 in the vertical coordinates, respectively, denote the five frequency bands: 1) δ; 2) θ; 3) α; 4) β; and 5) γ. The dynamic neural patterns in high frequency bands (α, β, and γ) have consistent changes with the emotional labels during the whole experiment.

tune the weights. Using this approach, the shared representations of both modalities were extracted, and linear SVMs were trained using the new shared representations as the inputs. The input features of EEG and eye movements for the RBMs were normalized to the range from zero to one. The numbers of neurons in the hidden layers were fixed to be the same for the three RBMs when training and were tuned in [200, 150, 100, 90, 70, 50, 30, 20, 15, 10] units using cross-validation. The learning rate was set to 0.001. The mini-batch size was 100. The multimodal deep models were implemented in Python using the deep learning libraries Keras[6] and Tensorflow.[7]

V. Experimental Results

A. EEG-Based Emotion Recognition

First, we evaluated the performance of *EmotionMeter* regarding accuracy for different setups of EEG recordings. Our

[6]https://keras.io/
[7]https://www.tensorflow.org/

objective was to investigate how the performance varies with the number of attached electrodes. We designed three setups for EEG recordings: 1) T7 and T8; 2) *T7, T8, FT7*, and FT8; and 3) *T7, T8, FT7, FT8, TP7*, and TP8. The mean accuracies and the standard deviations of all 45 experiments for different features obtained from separated and total frequency bands are presented in Table II. "Total" denotes the direct concatenation of five frequency bands. We compared the performance of the PSD and DE features in recognizing four emotions. As shown in this table, the DE features outperformed the PSD features with higher accuracies and lower standard deviations in most cases. The beta and gamma bands performed slightly better than the other frequency bands in general. These results gave a further verification on our previous work [71]. The visualization of the DE features and the labels in one experiment are shown in Fig. 8, which presents the dynamic neural patterns in high-frequency bands. In Fig. 8, the DE features of the delta band does not show significant changes, whereas the gamma and beta responses have consistent changes with the emotional labels. These results indicate that the alpha, beta, and gamma bands contain the most discriminative information. For neutral and happy emotions, the neural patterns have significantly higher beta and gamma responses than for the sad and fear emotions, whereas the neural patterns of neutral emotions have higher alpha responses compared to the other emotions.

Moreover, the electrode placements with two, four, and six electrodes can achieve relatively good performance for the four emotions. As shown in Table II, the best mean accuracies and their standard deviations of two, four, and six electrodes are 64.24%/15.39%, 67.02%/15.87%, and 70.33%/14.45%, respectively. The setup with only six electrodes can achieve comparable performance with a slightly lower mean accuracy compared with 62 electrodes (70.33% versus 70.58%). Although the system can achieve slightly higher accuracies with more electrodes as expected, the computational complexity and calibration time used are also considerably increased. In real-world applications, considering the feasibility and comfort, fewer electrodes will be preferred. These results demonstrate the efficiency of our design using only six EEG electrodes.

TABLE II
MEAN ACCURACY RATES (%) OF DIFFERENT SETUPS (TWO ELECTRODES: T7 AND T8; FOUR ELECTRODES: FT7, FT8, T7, AND T8; SIX ELECTRODES: FT7, FT8, T7, T8, TP7, AND TP8; AND 62 ELECTRODES) FOR THE TWO DIFFERENT FEATURES OBTAINED FROM THE SEPARATE AND TOTAL FREQUENCY BANDS. HERE, SVMS WITH LINEAR KERNELS WERE USED AS CLASSIFIERS

#	Feat.	Stat.	δ	θ	α	β	γ	Total
2	PSD	Mean	39.23	39.19	38.77	39.90	41.22	**60.22**
		Std.	8.36	8.00	7.35	9.15	10.77	16.18
	DE	Mean	36.84	37.18	39.25	39.23	41.21	**64.24**
		Std.	6.76	7.01	9.63	9.61	9.59	15.39
4	PSD	Mean	47.30	47.57	50.73	52.83	51.26	**58.98**
		Std.	12.29	12.22	15.65	13.26	15.04	17.85
	DE	Mean	45.21	47.99	50.65	54.41	57.19	**67.02**
		Std.	14.64	12.69	13.68	14.44	17.59	15.87
6	PSD	Mean	49.74	49.52	48.95	57.58	54.84	**57.70**
		Std.	17.48	13.47	15.66	20.92	18.46	16.94
	DE	Mean	50.47	48.88	51.99	60.70	60.57	**70.33**
		Std.	17.62	14.61	15.10	19.88	16.89	14.45
62	PSD	Mean	53.68	57.13	60.57	**63.60**	58.47	56.34
		Std.	13.40	14.23	15.54	18.78	18.60	14.54
	DE	Mean	57.58	57.98	61.22	66.66	66.34	**70.58**
		Std.	12.64	12.30	16.46	18.80	17.49	17.01

B. Analysis of Complementary Characteristics

For emotion recognition using only eye movements, we obtained an average accuracy and standard deviation of 67.82%/18.04%, which was slightly lower than that obtained using only EEG signals (70.33%/14.45%). For modality fusion, we compare two approaches: 1) feature-level fusion and 2) multimodal deep learning. For feature-level fusion, the feature vectors of EEG and eye movements are directly concatenated into a larger feature vector as the inputs of SVMs. Table III shows the performance of each single modality (eye movements and EEG) and of the two modality fusion approaches, and Fig. 9 presents the box plot of the accuracies using different modalities. The average accuracies and standard deviations of the feature-level fusion and multimodal deep learning were 75.88%/16.44% and 85.11%/11.79%, respectively, for all the experiments. We used one-way analysis of variance (ANOVA) to determine the statistical significance. The performance with modality fusion is significantly greater than that with only a single modality ($p < 0.01$), which indicates that modality fusion with multimodal deep learning can combine the complementary information in each single modality and effectively enhance the performance. These results demonstrate the efficiency of *EmotionMeter* combining EEG and eye movements for emotion recognition.

In comparison with the feature-level fusion, multimodal deep learning can learn the high-level shared representations between two modalities. Through the processing of multiple layers in deep neural networks, the effective shared representations are automatically extracted. In the feature-level fusion, it is very difficult to relate the original features in one modality to features in other modality and this method usually learns unimodal features [58]. Moreover, the relations across various modalities are deep instead of shallow. Multimodal deep learning can capture these relations across various modalities with deep architectures and improve the performance.

To further investigate the complementary characteristics of EEG and eye movements, we analyzed the confusion matrices of each modality, which reveals the strength and weakness of each modality. Figs. 10 and 11 present the confusion graph and the confusion matrices of eye movements and EEG, respectively. As indicated by these results, EEG and eye movements have important complementary characteristics. We observe that EEG has the advantage of classifying happy emotion (80%) compared to eye movements (67%), whereas eye movements outperform EEG in recognizing fear emotion (67% versus 65%). It is difficult to recognize fear emotion using only EEG and happy emotion using only eye movements. Sad emotion has the lowest classification accuracies for both modalities. However, the misclassifications of these two modalities are different. EEG misclassifies more sad emotion as neutral emotion (23%), whereas eye movements

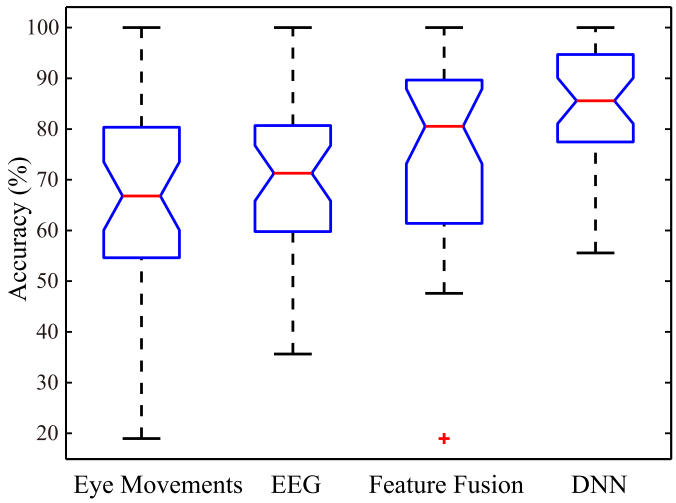


Fig. 9. Box plot of the accuracies with each single modality (eye movements and EEG) and the two modality fusion approaches. The performance with modality fusion is significantly greater than that with only a single modality ($p < 0.01$, ANOVA). The red lines indicate the median accuracies.

TABLE III
PERFORMANCE OF EACH SINGLE MODALITY (EYE MOVEMENTS AND EEG) AND THE TWO MODALITY FUSION APPROACHES

Method	Eye	EEG	Feature Fusion	DNN
Mean	67.82	70.33	75.88	**85.11**
Std.	18.04	14.45	16.44	**11.79**

misclassify more sad emotion as fear emotion (23%). Both EEG and eye movements can achieve relatively high accuracies of 78% and 80% for neutral emotion, respectively. These results indicate that EEG and eye movements have different discriminative powers for emotion recognition. Combining the complementary information of these two modalities, modality fusion can significantly improve the classification accuracies (85.11%).

As shown by the confusion matrices of the multimodal fusion methods presented in Fig. 10, the feature fusion method can significantly enhance the performance of classifying sad and fear emotions with 6% and 12% improvements in accuracies, respectively. Moreover, multimodal DNN provides even better improvements for sad, fear, and neutral emotions with increases in accuracies of 22%, 20%, and 12%, respectively, in comparison with a single modality, particularly for sad emotion. The single EEG modality provides a relatively high classification accuracy for happy emotion. Both fusion methods do not improve the classification accuracy of happy emotion compared to the single EEG modality. These experimental results reveal why the combination of both modalities can enhance the performance of emotion recognition. The fusion method integrates the advantages of EEG for recognizing happy emotion and the advantages of eye movements for recognizing fear emotion while simultaneously improving the classification accuracies of sad emotion. Moreover, the performance of classifying neutral emotion is also improved.

Humans convey and interpret emotional states through several modalities jointly, including audio-visual (facial expression, voice, and so on), physiological (respiration, skin temperature, and so forth), and contextual information (environment, social situation, and so on) [72]. Researchers have reached a consensus for constructing multimodal emotion recognition while concerning the fusion architecture of these multimodal information. However, most studies simply feed all multiple modalities into the machine learning models and do not investigate or interpret the underlying mechanisms of the improvement, even for the popular audio and visual modalities. In this paper, we utilize the attractive modalities six-channel EEG and eye movements and study the interactions between both modalities for multimodal emotion recognition. Eye tracking using wearable techniques has received considerable attention in recent years due to its natural observations and informative features of users' nonverbal behaviors [38], [73], [74]. The previous study of Ding *et al.* [75] showed that eye contact contained reliable information for speaker identification in three-party conversations. Through eye tracking, more qualitative indices could be included to enhance HCIs. Compared to other modalities, eye tracking has the advantage of providing contextual information.

TABLE IV
AVERAGE ACCURACIES AND STANDARD DEVIATIONS (%) OF EEG WITH DIFFERENT NUMBERS OF ELECTRODES AND EYE MOVEMENTS ACROSS SESSIONS. ("1ST," "2ND," AND "3RD"' DENOTE THE DATA OBTAINED FROM THE FIRST, SECOND, AND THIRD EXPERIMENTS WITH ONE PARTICIPANT, RESPECTIVELY)

#Elec	Train	Test		
		1st	2nd	3rd
2	1st	**80.76** (7.97)	70.88 (7.61)	66.47 (7.77)
	2nd	63.33 (7.64)	**89.14** (10.07)	65.59 (7.43)
	3rd	66.62 (10.12)	69.22 (9.10)	**85.27**(9.83)
4	1st	**85.13** (11.72)	69.68 (8.64)	62.97 (7.42)
	2nd	63.72 (6.84)	**85.72** (12.66)	64.49 (11.08)
	3rd	64.16 (7.55)	68.33 (11.73)	**84.24** (11.49)
6	1st	**83.26** (13.40)	70.42 (6.42)	66.24 (9.10)
	2nd	62.02 (9.54)	**86.41** (13.80)	62.65 (8.90)
	3rd	64.97 (8.48)	70.57 (9.07)	**84.97** (8.84)

C. Analysis of Stability Across Sessions

The systematic evaluation of a robust emotion recognition system involves not only the accuracy but also the stability over time. The novelty of our dataset compared with other datasets is that it consists of three sessions for each participant to investigate the stability of *EmotionMeter* across sessions.

We select the DE features of the total frequency bands and eye movement features from different sessions with the same participants as the training and test datasets. The average accuracies and standard deviations for two, four, and six electrodes are shown in Table IV. A mean classification accuracy of 72.39% was achieved across sessions with the six-electrode EEG and eye movement features using multimodal deep learning, whereas for feature-level fusion with SVM, the mean accuracy was 59.52%. The multimodal deep learning approach achieves significantly better performance than feature-level fusion. An interesting finding is that the performance was better with training and test data obtained from sessions

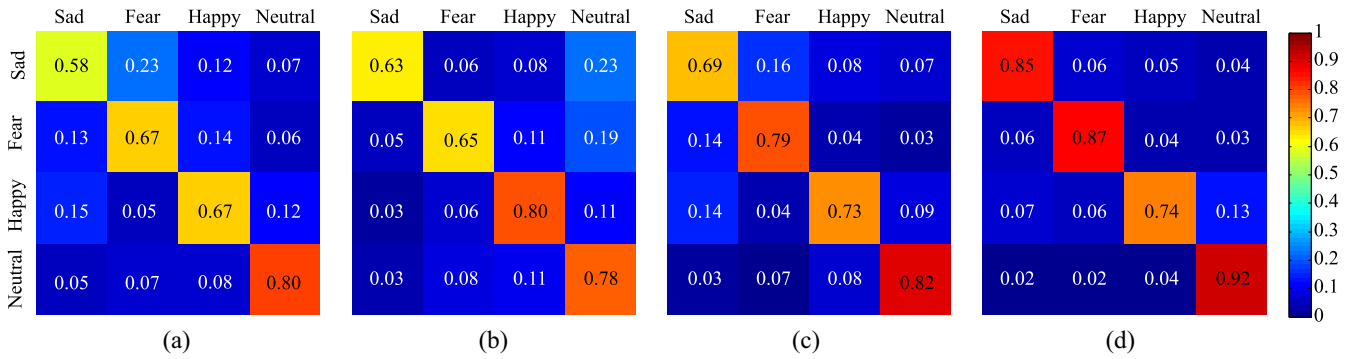


Fig. 10. Confusion matrices of single modality and multimodal fusion methods: feature-level fusion and multimodal deep neural networks. Each row of the confusion matrices represents the target class and each column represents the predicted class. The element (i, j) is the percentage of samples in class i that is classified as class j. (a) Eye. (b) EEG. (c) Feature Fusion. (d) DNN.

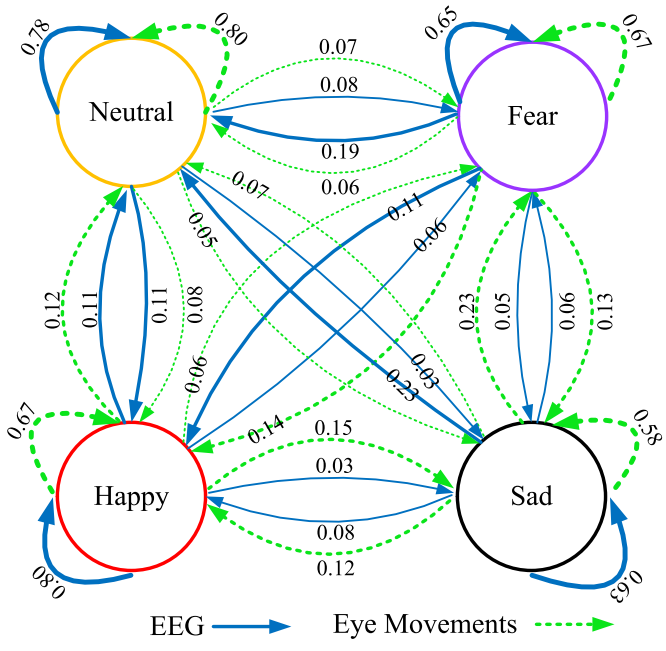


Fig. 11. Confusion graph of EEG and eye movements, which shows their complementary characteristics for emotion recognition. [The numbers denote the percentage values of samples in the class (arrow tail) classified as the class (arrow head). Bolder lines indicate higher values.]

performed in nearer time. These results demonstrate the comparative stability of our proposed *EmotionMeter* framework.

For real-world applications, the intuitive approach of the training and calibration phase is to use the past labeled data as the training data and make inferences on the new data. However, there are some differences in feature distributions across sessions, and these differences may be due to the non-stationary characteristics of EEG and changing environments such as noise, impedance variability, and the relative position of the electrodes. As time passes, the performance of the emotion recognition system may deteriorate. Therefore, adapting emotion recognition models should be further studied in the future [76]–[80]. To overcome the across-day variability in neural recording conditions and make the brain-machine interfaces robust to future neural variability, Sussillo *et al.* [81] exploited the previously collected data to construct a robust decoder using a multiplicative recurrent neural network.

VI. Conclusion

Emotions are manifested via internal physiological responses and external behaviors. Signals from different modalities provide different aspects of emotions, and complementary information from different modalities can be integrated to construct a more robust emotion recognition system compared to unimodal approaches. In this paper, we have presented *EmotionMeter*, which is a multimodal framework to recognize human emotions with EEG and eye movements. Considering its wearability and feasibility, we have designed a six-electrode placement above the ears, which is suitable for attachment in a wearable headset or headband. We have demonstrated that modality fusion combining EEG and eye movements with multimodal deep learning can significantly enhance the emotion recognition accuracy (85.11%) compared with a single modality (eye movements: 67.82% and EEG: 70.33%). Moreover, we have also investigated the complementary characteristics of EEG and eye movements for emotion recognition and evaluated the stability of our proposed framework across sessions. The quantitative evaluation results have indicated the effectiveness of our proposed *EmotionMeter* framework.

In the future, we plan to improve the accuracy of our emotion recognition system with the adaptation over time and the integration of user-specific profiles. We are also working on implementing wearable prototypes with hardware and software in more real social interaction environments.

References

[1] M. Ptaszynski, P. Dybala, W. Shi, R. Rzepka, and K. Araki, "Towards context aware emotional intelligence in machines: Computing contextual appropriateness of affective states," in *Proc. Int. Joint Conf. Artif. Intell.*, 2009, pp. 1469–1474.

[2] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cogn. Emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[3] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

[4] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.

[5] A. M. Al-Kaysi *et al.*, "Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification," *J. Affect. Disorders*, vol. 208, pp. 597–603, Jan. 2017.

[6] A. V. Bocharov, G. G. Knyazev, and A. N. Savostyanov, "Depression and implicit emotion processing: An EEG study," *Clin. Neurophysiol.*, vol. 47, no. 3, pp. 225–230, 2017.

[7] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 211–223, Apr./Jun. 2012.

[8] S. K. D'Mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Comput. Surveys*, vol. 47, no. 3, p. 43, 2015.

[9] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 8, pp. 1301–1309, Dec. 2017.

[10] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 5040–5043.

[11] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition," in *Proc. Int. Joint Conf. Artif. Intell.*, 2015, pp. 1170–1176.

[12] P. A. Kragel and K. S. LaBar, "Advancing emotion theory with multivariate pattern classification," *Emotion Rev.*, vol. 6, no. 2, pp. 160–174, 2014.

[13] H. Saarimäki *et al.*, "Discrete neural signatures of basic emotions," *Cerebral Cortex*, vol. 26, no. 6, pp. 2563–2573, 2016.

[14] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: Principles, state-of-the-art, and challenges," *Brain-Comput. Interfaces*, vol. 1, no. 2, pp. 66–84, 2014.

[15] C. Aracena, S. Basterrech, V. Snáel, and J. Velásquez, "Neural networks for emotion recognition based on eye tracking data," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2015, pp. 2632–2637.

[16] D. H. Lee and A. K. Anderson, "Reading what the mind thinks from how the eye sees," *Psychol. Sci.*, vol. 28, no. 4, pp. 494–503, 2017.

[17] D. MacLean, A. Roseway, and M. Czerwinski, "MoodWings: A wearable biofeedback device for real-time stress intervention," in *Proc. 6th Int. Conf. Pervasive Technol. Related Assistive Environ.*, 2013, p. 66.

[18] M. A. Williams, A. Roseway, C. O'Dowd, M. Czerwinski, and M. R. Morris, "SWARM: An actuated wearable for mediating affect," in *Proc. 9th ACM Int. Conf. Tangible Embedded Embodied Interaction*, 2015, pp. 293–300.

[19] D. Valtchanov and M. Hancock, "EnviroPulse: Providing feedback about the expected affective valence of the environment," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 2073–2082.

[20] M. Hassib *et al.*, "EngageMeter: A system for implicit audience engagement sensing using electroencephalography," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, 2017, pp. 5114–5119.

[21] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Pers. Soc. Psychol.*, vol. 17, no. 2, pp. 124–129, 1971.

[22] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[23] M. Y. Tsalamlal, M.-A. Amorim, J.-C. Martin, and M. Ammi, "Combining facial expression and touch for perceiving emotional valence," *IEEE Trans. Affect. Comput.*, to be published, doi: 10.1109/TAFFC.2016.2631469.

[24] A. Schirmer and R. Adolphs, "Emotion perception from face, voice, and touch: Comparisons and convergence," *Trends Cogn. Sci.*, vol. 21, no. 3, pp. 216–228, 2017.

[25] R. A. L. Koelstra, "Affective and implicit tagging using facial expressions and electroencephalography," Ph.D. dissertation, School Electron. Eng. Comput. Sci., Queen Mary Univ. at London, London, U.K., 2012.

[26] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, to be published, doi: 10.1109/TAFFC.2017.2714671.

[27] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul./Sep. 2014.

[28] I. Daly *et al.*, "Affective brain–computer music interfacing," *J. Neural Eng.*, vol. 13, no. 4, 2016, Art. no. 046022.

[29] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 81–97, Jul./Dec. 2010.

[30] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2014, pp. 1–6.

[31] Y.-P. Lin *et al.*, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.

[32] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, Apr. 2014.

[33] Y.-J. Huang, C.-Y. Wu, A. M.-K. Wong, and B.-S. Lin, "Novel active comb-shaped dry electrode for EEG measurement in hairy site," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 256–263, Jan. 2015.

[34] C. Grozea, C. D. Voinescu, and S. Fazli, "Bristle-sensors–low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications," *J. Neural Eng.*, vol. 8, no. 2, 2011, Art. no. 025008.

[35] W. D. Hairston *et al.*, "Usability of four commercially-oriented EEG systems," *J. Neural Eng.*, vol. 11, no. 4, 2014, Art. no. 046018.

[36] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-time EEG-based human emotion recognition and visualization," in *Proc. IEEE Int. Conf. Cyberworlds*, 2010, pp. 262–269.

[37] M. Hassib, M. Pfeiffer, S. Schneegass, M. Rohs, and F. Alt, "Emotion actuator: Embodied emotional feedback through electroencephalography and electrical muscle stimulation," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, 2017, pp. 6133–6146.

[38] A. Duchowski, *Eye Tracking Methodology: Theory and Practice*, vol. 373. London, U.K.: Springer, 2007.

[39] E. Granholm and S. R. Steinhauer, "Pupillometric measures of cognitive and emotional processes," *Int. J. Psychophysiol.*, vol. 52, no. 1, pp. 1–6, 2004.

[40] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," *Psychophysiology*, vol. 45, no. 4, pp. 602–607, 2008.

[41] A. A. Zekveld, D. J. Heslenfeld, I. S. Johnsrude, N. J. Versfeld, and S. E. Kramer, "The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load," *Neuroimage*, vol. 101, pp. 76–86, Nov. 2014.

[42] A. Bulling, D. Roggen, and G. Troester, "What's in the eyes for context-awareness?" *IEEE Pervasive Comput.*, vol. 10, no. 2, pp. 48–57, Apr./Jun. 2011.

[43] A. Bulling, J. A. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 741–753, Apr. 2011.

[44] J. Simola, K. Le Fevre, J. Torniainen, and T. Baccino, "Affective processing in natural scene viewing: Valence and arousal interactions in eye-fixation-related potentials," *NeuroImage*, vol. 106, pp. 21–33, Feb. 2015.

[45] N. Langer *et al.*, "A resource for assessing information processing in the developing brain using EEG and eye tracking," *Sci. Data*, vol. 4, Apr. 2017, Art. no. 170040.

[46] C. Wang, J. L. Ong, A. Patanaik, J. Zhou, and M. W. L. Chee, "Spontaneous eyelid closures link vigilance fluctuation with fMRI dynamic connectivity states," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 34, pp. 9653–9658, 2016.

[47] W. Liu, W.-L. Zheng, and B.-L. Lu, "Emotion recognition using multimodal deep learning," in *Proc. Int. Conf. Neural Inf. Process.*, 2016, pp. 521–529.

[48] H. Tang, W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using deep neural networks," in *Proc. Int. Conf. Neural Inf. Process.*, 2017, pp. 811–819.

[49] S. Gudmundsson, T. P. Runarsson, S. Sigurdsson, G. Eiriksdottir, and K. Johnsen, "Reliability of quantitative EEG features," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2162–2171, 2007.

[50] L. K. McEvoy, M. Smith, and A. Gevins, "Test–retest reliability of cognitive EEG," *Clin. Neurophysiol.*, vol. 111, no. 3, pp. 457–463, 2000.

[51] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, to be published, doi: 10.1109/TAFFC.2017.2712143.

[52] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Stability of features in real-time EEG-based emotion recognition algorithm," in *Proc. IEEE Int. Conf. Cyberworlds*, 2014, pp. 137–144.

[53] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[54] D. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: Fast classification and anomaly measurement," *J. Neural Eng.*, vol. 8, no. 3, 2011, Art. no. 036015.

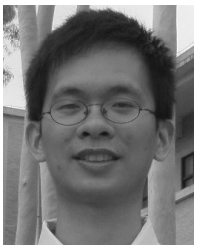
[55] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[56] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[57] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Deep feature learning for EEG recordings," *arXiv preprint arXiv:1511.04306*, 2015.

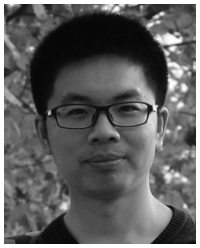
[58] J. Ngiam *et al.*, "Multimodal deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 689–696.

[59] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cogn. Emotion*, vol. 9, no. 1, pp. 87–108, 1995.
[60] 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," *Cogn. Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.
[61] Y. Baveye, E. Dellandréa, C. Chamaret, and L. Chen, "LIRIS-ACCEDE: A video database for affective content analysis," *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 43–55, Jan./Mar. 2015.
[62] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *J. Pers. Soc. Psychol.*, vol. 54, no. 6, pp. 1063–1070, 1988.
[63] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," in *Proc. 5th Int. IEEE/EMBS Conf. Neural Eng.*, 2011, pp. 667–670.
[64] R. E. Wheeler, R. J. Davidson, and A. J. Tomarken, "Frontal brain asymmetry and emotional reactivity: A biological substrate of affective style," *Psychophysiology*, vol. 30, no. 1, pp. 82–89, 1993.
[65] J. A. Coan and J. J. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biol. Psychol.*, vol. 67, no. 1, pp. 7–49, 2004.
[66] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 6627–6630.
[67] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng.*, 2013, pp. 81–84.
[68] L.-C. Shi and B.-L. Lu, "Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2010, pp. 6587–6590.
[69] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jan. 2008.
[70] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.
[71] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 1223–1226.
[72] M. Pantic *et al.*, "Multimodal emotion recognition from low-level cues," in *Emotion-Oriented Systems*. Heidelberg, Germany: Springer, 2011, pp. 115–132.
[73] K. Krafka *et al.*, "Eye tracking for everyone," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2176–2184.
[74] A. M. Feit *et al.*, "Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, 2017, pp. 1118–1130.
[75] Y. Ding, Y. Zhang, M. Xiao, and Z. Deng, "A multifaceted study on eye contact based speaker identification in three-party conversations," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, 2017, pp. 3011–3021.
[76] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[77] X. Li, C. Guan, H. Zhang, K. K. Ang, and S. H. Ong, "Adaptation of motor imagery EEG classification model based on tensor decomposition," *J. Neural Eng.*, vol. 11, no. 5, 2014, Art. no. 056020.
[78] W.-L. Zheng, Y.-Q. Zhang, J.-Y. Zhu, and B.-L. Lu, "Transfer components between subjects for EEG-based emotion recognition," in *Proc. IEEE Int. Conf. Affect. Comput. Intell. Interaction*, 2015, pp. 917–922.
[79] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 20–31, Feb. 2016.
[80] W.-L. Zheng and B.-L. Lu, "Personalizing EEG-based affective models with transfer learning," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 2732–2738.
[81] D. Sussillo, S. D. Stavisky, J. C. Kao, S. I. Ryu, and K. V. Shenoy, "Making brain–machine interfaces robust to future neural variability," *Nature Commun.*, vol. 7, Dec. 2016, Art. no. 13749.



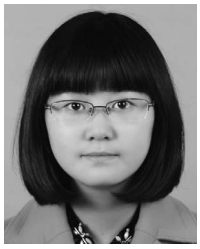
Wei-Long Zheng (S'14) received the bachelor's degree in information engineering with the Department of Electronic and Information Engineering, South China University of Technology, Guangzhou, China, in 2012. He is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China.

His current research interests include affective computing, brain-computer interaction, machine learning, and pattern recognition.



Wei Liu received the bachelor's degree in automation science from the School of Advanced Engineering, Beihang University, Beijing, China, in 2014. He is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China.

His current research interests include affective computing, brain-computer interface, and machine learning.



Yifei Lu received the bachelor's and master's degrees from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2014 and 2016, respectively, both in computer science and technology.

Her current research interests include emotion recognition and brain-computer interaction.

Bao-Liang Lu (M'94–SM'01) received the B.S. degree in instrument and control engineering from the Qingdao University of Science and Technology, Qingdao, China, in 1982, the M.S. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 1989, and the Dr.Eng. degree in electrical engineering from Kyoto University, Kyoto, Japan, in 1994.

He was with the Qingdao University of Science and Technology from 1982 to 1986. From 1994 to 1999, he was a Frontier Researcher with the Bio-Mimetic Control Research Center, Institute of Physical and Chemical Research (RIKEN), Nagoya, Japan, and a Research Scientist with the RIKEN Brain Science Institute, Wako, Japan, from 1999 to 2002. Since 2002, he has been a Full Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include brain-like computing, neural network, machine learning, brain-computer interaction, and affective computing.

Prof. Lu was the President of the Asia Pacific Neural Network Assembly (APNNA) and the General Chair of the 18th International Conference on Neural Information Processing in 2011. He is currently the Steering Committee Member of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, the Associate Editor of the IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENT SYSTEMS, and a Board Member of the Asia Pacific Neural Network Society (APNNS, previously APNNA).

Andrzej Cichocki (SM'06–F'13) received the M.Sc. (Hons.), Ph.D., and Dr.Sc. (Habilitation) degrees from the Warsaw University of Technology, Warsaw, Poland, all in electrical engineering.

He spent several years with University Erlangen-Nuerenberg, Erlangen, Germany, as an Alexander-von-Humboldt Research Fellow and a Guest Professor. From 1995 to 1997, he was a Team Leader with the Brain Information Processing Group, Laboratory for Open Information Systems, Frontier Research Program RIKEN, Wako, Japan, directed by Prof. S.-I. Amari. He was recently a Senior Team Leader and the Head of the Cichocki Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute. He has authored and co-authored over 600 technical scientific papers and six internationally recognized monographs (two of them translated to Chinese). He has over 34 000 Google Scholar citations and an H-index of over 80. The new Laboratory Tensor Networks and Deep Learning for Applications in Data Mining is established at SKOLTECH, Moscow, Russia, under his guidance.