Vis Comput (2016) 32:347–358
DOI 10.1007/s00371-015-1183-y

ORIGINAL ARTICLE

Real-time EEG-based emotion monitoring using stable features

Zirui Lan[1] · Olga Sourina[1] · Lipo Wang[2] · Yisi Liu[1]

Published online: 26 November 2015

Abstract In human–computer interaction (HCI), electroencephalogram (EEG) signals can be added as an additional input to computer. An integration of real-time EEG-based human emotion recognition algorithms in human–computer interfaces can make the users experience more complete, more engaging, less emotionally stressful or more stressful depending on the target of the applications. Currently, the most accurate EEG-based emotion recognition algorithms are subject-dependent, and a training session is needed for the user each time right before running the application. In this paper, we propose a novel real-time subject-dependent algorithm with the most stable features that gives a better accuracy than other available algorithms when it is crucial to have only one training session for the user and no re-training is allowed subsequently. The proposed algorithm is tested on an affective EEG database that contains five subjects. For each subject, four emotions (pleasant, happy, frightened and angry) are induced, and the affective EEG is recorded for two sessions per day in eight consecutive days. Testing results show that the novel algorithm can be used in real-time emotion recognition applications without re-training with the adequate accuracy. The proposed algorithm is integrated with real-time applications "Emotional Avatar" and "Twin Girls" to monitor the users emotions in real time.

Keywords EEG · Emotion recognition · Fractal dimension (FD) · Stability · Intra-class correlation coefficient (ICC)

✉ Zirui Lan
LanZ0001@e.ntu.edu.sg

Olga Sourina
EOSourina@ntu.edu.sg

Lipo Wang
ELPWang@ntu.edu.sg

Yisi Liu
LiuYS@ntu.edu.sg

[1] Fraunhofer IDM@NTU, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

[2] School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

1 Introduction

Electroencephalogram (EEG) is the time-series measure of the electric potential of human brain. Previously, the use of EEG was limited to medical applications, e.g., facilitating the diagnosis of brain diseases like epileptic seizure, Attention Deficit Hyperactive Disorder (ADHD), Alzheimer's disease, etc. However, the advancement of technology introduced to the market new EEG devices which are wearable, portable, wireless and easy to use. This has enabled the application of EEG to expand from medical use to personal entertainment use. Now, EEG-based emotion recognition draws high attention because it is desirable that a machine can recognize human emotions and interact with us in a more humanized way. EEG-enabled human–computer interfaces can be adapted to the user's internal feelings and can be driven by the user's emotions. The recognized emotions of the user can help make the user's experience more complete, more engaging, and less emotionally stressful or more stressful depending on the target of the application. It is useful to implement affective interfaces in many applications such as (1) games where the flow can be changed according to user's emotions [27]; (2) medical applications to monitor emotions of the patient who may not be able to express emotions [39]; (3) neuro-marketing to adapt online advertisement based on the recognized user's emotions, etc.

Human emotions are complex states of feelings that result in physical and psychological changes, which can be reflected by facial expressions, gestures, intonation in speech, etc. The effort to recognize human emotions can be traced back to 1972 [42], which attempted to judge the emotion based on the speech of the speaker. However, since facial expressions, gestures and intonation can be deliberately changed to hide the true emotions, emotion recognition based on such superficial features may not be reliable. EEG directly measures the changes in brain activities, and emotion recognition from EEG has the potential to assess the true inner feelings of the subject.

Current EEG-based emotion recognition algorithms are subject-dependent and require a training session prior to running the real-time emotion recognition application almost every time. During the training session, stimuli (audio/video) are presented to the subject to evoke certain targeted emotions and meanwhile the EEG of the subject is recorded. The recorded EEG data are subject to feature extraction to extract numerical feature parameters, and the extracted features are fed into a classifier for the training.

In the study of EEG-based emotion recognition, different features and different classifiers are employed. Ishino and Hagiwara [15] used wavelet transform, Fast Fourier Transform (FFT) and statistics such as mean and variance as features and employed Neural Network (NN) to classify four emotions. An accuracy of 67.7 % was achieved with 3 channels. Lin et al. [23] utilized power differences at symmetric electrode pairs and Support Vector Machine (SVM) to classify four emotions and obtained an accuracy of 90.72 % with 32 channels. Recognizing three emotions, Schaaff [36] and Chanel et al. [5] made use of Short Time Fourier Transform (STFT) to extract features, and SVM as a classifier and achieved 62.07 % (16 channels) and 63 % (64 channels) accuracy, respectively. In another work, Shaaff [37] reduced the number of channels to 4 and adopted the statistical features, power features and cross-correlation features together with an SVM classifier. The recognition accuracy was 47.11 % for 3 emotion classes. In [22], four emotions were recognized with differential asymmetry of hemispheric EEG power spectra as features and SVM as classifier, and obtained an accuracy of 82.29 % using 32 channels. Five emotions were recognized in [29] at an accuracy of 83.04 % using statistical features from different EEG bands from 62 channels and K-Nearest Neighbor (KNN) classifier. Li and Lu [21] used SVM to differentiate two emotions featured by logarithmic variances from 62 channels and achieved 93.5 % accuracy. Liu and Sourina [24] employed fractal dimension (FD) feature together with statistical and higher order crossings (HOC) features, and an SVM classifier was used. Up to eight emotions were recognized with four channels. The average accuracy obtained ranged from 53.75 % (for eight emotions) to 83.73 % (for two emotions).

In [38], Sohaib et al. evaluated the performance of different classifiers, and reported the best accuracy of 56.10 % for 3 emotions, obtained by using SVM as classifier and statistical features sourced from 6 electrodes. Similarly, in another work [30], Quadratic Discriminant Analysis (QDA) and SVM were compared when 6 emotions were to be recognized using HOC features derived from 4 channels. SVM was reported to have better accuracy, 83.3 %, as compared to 62.3 % obtained by using QDA. Wang et al. reported a similar finding in their work [41]. Adopting the statistical features and power features, Wang compared the classification performance between K-NN, SVM and multi layer perceptron (MLP). SVM, reported an accuracy of 66.51 % for identifying 4 emotion classes with 62 channels, was the highest among all. Brown et al. [4] employed the power ratio features and band power features derived from 8 channels, and 3 different classifiers (QDA, SVM and K-NN) to evaluate recognition performance for 3 emotion classes. In this study, K-NN was reported to give the best accuracy, varying from 50 to 64 % for different subjects. Frantzidis [8] exploited the Event Related Potential (ERP) and Event Related Oscillation (ERO) characteristics of EEG and proposed to use the ERP amplitude, ERP latency and ERO amplitude as features. Mahalanobis Distance (MD) classifier and SVM were chosen and compared with each other. Recognizing 4 emotion classes, SVM outperformed MD by 1.8 %, achieving 81.3 % accuracy.

It has to be pointed out that a direct comparison between different algorithms is not appropriate, as the reported accuracies were obtained under different experimental settings. Nevertheless, some conclusions can be drawn without overgeneralization. The accuracies are generally higher when more EEG channels are involved. SVM, as a popular classifier, has been extensively used in these studies [4,5,8,14,21–24,30–32,36–38,41]. Moreover, controlled experiments have been conducted in [5,8,22,30,38,41] in order to evaluate the performance of different classifiers. SVM was more preferable than other classifiers for its effectiveness and better accuracies.

The stability issue of EEG features was firstly brought up under medical application. A feature must demonstrate high stability in order to be accepted for clinical use. A stable feature should exhibit consistency among repeated EEG measurements of the same subject. Stability of several common EEG features such as band power, coherence, and entropy has been studied. In [9,10], 26 subjects were involved in a 10-month experiment. Absolute power feature and relative power feature were reported to have similar stability while coherence was less stable than the former two. The power feature true obtained from alpha band was the most stable, followed by theta band, delta band, and beta band power features. Salinsky et al. [34] recruited 19 subjects and recorded their EEG in closed-eye state in an interval of 12–16 weeks. Peak alpha frequency and median frequency were reported to be the most stable. No significant difference was found between the stability of absolute power and relative power. Kondacs and Szabó [18] investigated power spectral features and coherence features of the resting, closed-eye EEG of 45 subject in 25–62 months interval. The stability was reported as total power of frequency range from 1.5 to 25 Hz being the largest, followed by alpha mean frequency, absolute alpha and beta power, absolute delta power and alpha coherence. Gudmundsson et al. [11] studied the power spectral parameters, entropy and coherence features. EEG data were from 15 elderly subjects, each recorded 10 sessions within 2 months. Power spectral parameters were reported to be more stable than entropy, coherence being the least stable. Among the power features, theta band was the most stable, followed by alpha, beta, delta and gamma band. Admittedly, parallels cannot be drawn easily between these studies, as subjects, features, data processing techniques, test–retest interval were all different. However, some common findings can be drawn: absolute power features and relative power features have similar stability performance; power features are more stable than coherence feature.

In EEG-based emotion recognition, stable EEG features are also needed, so that re-training can be omitted. A stable EEG feature should ideally give consistent measurements of the same emotion of the same subject over time. Though the power features are the most stable one in the medical applications, further analysis for emotion recognition application is needed. Liu et al. have demonstrated that the fractal dimension feature outperforms power features in terms of accuracy in valence levels recognition [25]. In [19], we investigated the stability of various features used in the real-time EEG-based emotion recognition algorithm [24]. In this paper, we further investigate stability of more features and propose a novel real-time EEG-based emotion recognition algorithm with the most stable features. The implemented algorithm is integrated with two applications for user's emotions monitoring such as "Emotional Avatar" and "Twin Girls". The proposed algorithm allows having just one training session for the subject, and this training can be used in the applications without re-training for each new session. We design and implement experiment to collect EEG data labeled with four emotions such as pleasant, happy, frightened and angry. The data are collected from five subjects during eight consecutive days (two sessions per day per subject).

The paper is organized as follows. In Sect. 2, the related work including the feature extraction methods such as fractal dimension, power, statistical features and higher order crossings are given. In Sect. 3, an experiment to collect affective intra-subject EEG data is described. In Sect. 4, the proposed stable emotion recognition algorithm is introduced. In Sect. 5, the data processing and analysis results and discussion are presented. In Sect. 6, the application of the proposed algorithm is given. Section 7 concludes the paper.

2 Related work

2.1 Fractal dimension feature extraction

FD measures the geometric complexity of objects. FD feature has been proven effective in EEG-based emotion recognition application [24]. Following work [24], a Higuchi algorithm [13] was used to compute the FD feature of EEG.

Let $X(1), X(2), \ldots, X(N)$ denote time series samples (similarly hereinafter), construct new time series by picking up one sample from every k samples:

$$X_k^m : X(m), X(m+k), \ldots, X\left(m + \left[\frac{N-m}{k}\right] \cdot t\right)$$
$$m = 1, 2, 3 \ldots, k, \tag{1}$$

where m is the initial time and k is the interval time.

Then, for each of the k new time series, compute $L_m(k)$ as:

$$L_m(k) = \frac{1}{k} \cdot \left[\frac{\left(\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} |X(m+ik) - X(m+(i-1)k|\right)(N-1)}{\lfloor \frac{N-m}{k} \rfloor k}\right]. \tag{2}$$

Let $\langle L(k)\rangle$ denote the average of $L_m(k)$, i.e., $\langle L(k)\rangle = \frac{1}{k}\sum_{m=1}^{k} L_m(k)$, the following proportionality exists:

$$\langle L(k)\rangle \propto k^{-FD}, \tag{3}$$

where FD is the fractal dimension value, which can be calculated as:

$$FD = -\lim_{k\to\infty} \frac{\log\langle L(k)\rangle}{\log k}. \tag{4}$$

2.2 Power feature extraction

In EEG study, there is common agreement on partitioning the EEG power spectrum into several sub-bands (though the frequency range may slightly differ from case to case): alpha band, theta band, beta band, etc. In our study, the EEG power features from theta band (4–8 Hz), alpha band (8–12 Hz), and beta band (12–30 Hz) are computed.

The power features are obtained by first performing DFT on the EEG signals:

$$X(\mathrm{e}^{j\omega}) = \sum_{n=0}^{N-1} X(n)\mathrm{e}^{-j\omega n}, \tag{5}$$

where N is the number of input samples, $\omega = \frac{2\pi}{N}$. Then the power spectrum density is computed as:

$$\hat{S}_{NX}(\omega) = \frac{1}{N}|X(e^{j\omega})|^2. \tag{6}$$

At last, the power features are obtained by averaging the power spectrum density over the targeted sub-band, e.g., the alpha power parameter is computed by averaging $\hat{S}_{NX}(\omega)$ over 8–12 Hz range.

2.3 Statistical feature extraction

Six statistical features were adopted in [33] and used in the EEG-based emotion recognition in [24]. They were mean (7), standard deviation (8), mean of absolute values of the first differences (9), mean of absolute values of the first differences of normalized EEG (10), mean of absolute values of the second differences (11), mean of the absolute values of the second differences of the normalized EEG (12), as formulated in (7)–(12).

$$\mu_X = \frac{1}{N}\sum_{n=1}^{N} X(n). \tag{7}$$

$$\sigma_X = \sqrt{\frac{1}{N}\sum\nolimits_{n=1}^{N}(X(n) - \mu_X)^2}. \tag{8}$$

$$\delta_X = \frac{1}{N-1}\sum_{n=1}^{N-1}|X(n+1) - X(n)|. \tag{9}$$

$$\overline{\delta_X} = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|\overline{X}(n+1) - \overline{X}(n)\right| = \frac{\delta_X}{\sigma_X}. \tag{10}$$

$$\gamma_X = \frac{1}{N-2}\sum_{n=1}^{N-2}|X(n+2) - X(n)|. \tag{11}$$

$$\overline{\gamma_X} = \frac{1}{N-2}\sum_{n=1}^{N-2}\left|\overline{X}(n+2) - \overline{X}(n)\right| = \frac{\gamma_X}{\sigma_X}, \tag{12}$$

where n and N is the running index and total number of samples, respectively. $\overline{X}(n)$ is the normalized EEG signal $\overline{X}(n) = \frac{X(n)-\mu_X}{\sigma_x}$.

2.4 Higher order crossings (HOC) feature extraction

Higher order crossings (HOC) was proposed by Kedem [16] and used in [24,30,32] as features to recognize human emotion from EEG signals. The HOC is computed as follows.

First, the input raw EEG data has to be centralized: $Z(n) = X(n) - \mu_X$.

Then, filters of order k are applied to the centralized EEG data:

$$\nabla^{k-1}Z(n) \equiv \sum_{j=1}^{k}\frac{(k-1)!}{(j-1)!(k-j)!}(-1)^{(j-1)}Z(n-j+1). \tag{13}$$

The crossings for order k are counted as:

$$D_k = \sum_{n=2}^{N}\left[X_n(k) - X_{n-1}(k)\right]^2, \tag{14}$$

where $X_n(k)$ is the characteristic function:

$$X_n(k) = \begin{cases} 1 & \text{if } \nabla^{k-1}Z(n) \geq 0, \\ 0 & \text{if } \nabla^{k-1}Z(n) < 0. \end{cases} \tag{15}$$

2.5 Intra-class correlation coefficient

The stability of feature parameters was quantified by the intra-class correlation coefficient (ICC). Unlike the Pearson correlation coefficient, which is based on pairwise comparison, the ICC allows assessment of similarity in grouped data. It describes how well the data from the same group resemble each other. Both Pearson correlation coefficient and ICC have been used to examine the stability of EEG parameters [1,11,18,34,40]. However, when examining the stability of EEG parameters coming from multiple sessions, ICC was preferred. Multiple ICC models such as ICC(1), ICC(C,1), ICC(A,1) are available [28]. Among these models, ICC(1) was often used in EEG stability study [1,11]. ICC(1) is derived from a one-way ANOVA model and defined as:

$$\mathrm{ICC} = \frac{\mathrm{MS}_B - \mathrm{MS}_W}{\mathrm{MS}_B + (k-1)\mathrm{MS}_W}, \tag{16}$$

where MS_B, MS_W and k represent the mean square error between subjects, the mean square error within subjects, and the number of sessions, respectively. When $\mathrm{MS}_W = 0$, ICC value becomes 1, indicating the highest similarity. A smaller ICC value indicates a lower similarity level. ICC value can drop below zero in the case when MS_W is larger than MS_B, accounting for non-similarity among the grouped data.

3 Experiment

3.1 Experiment protocol

The stability of affective EEG features is of our interest of investigation. In contrast to current EEG benchmark dataset such as DEAP dataset [17], which includes a large number of subjects but only one single EEG recording session for each subject, we designed and conducted an experiment to collect the affective EEG data from multiple sessions on a

Fig. 1 Emotiv EEG device [7]

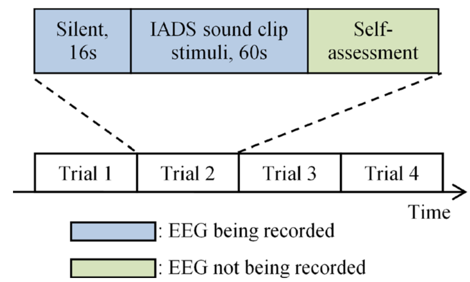

Fig. 2 Protocol of emotion induction experiment

small group of subjects. This preliminary study included five subjects, four males and one female, aged 24–28. All subjects reported no history of mental diseases or head injuries. Two sessions were recorded per day for each subject for eight consecutive days, i.e., 16 sessions were recorded for each subject. A 14-channel Emotiv EEG device (see Fig. 1) was used to record the EEG data at a sampling rate of 128 Hz. Each session consisted of four trials, with each trial corresponding to one induced emotion, i.e., four emotions were elicited in one session, so totally each subject has $4 \times 2 \times 8 = 64$ trials. There are standard affective stimuli libraries such as International Affective Picture System (IAPS) [20] and International Affective Digitized Sounds (IADS) [3]. In our study, the IADS was chosen for the experiment design because during the exposure of the subjects to the audio stimuli, the subjects can keep their eyes closed and hence avoid possible ocular movements which could contaminate the EEG signals. The emotion induction experiment protocol followed work [24]. Sound clips from the same category of the IADS were chosen and appended to make a 76-s audio file, with the first 16 s silent to calm the subject down. Four audio files were used as stimuli to evoke four different emotions, namely pleasant, happy, angry and frightened. During each session of the experiment only one subject was invited to the lab and was well-instructed about the protocol of the experiment. The subject wore the Emotiv EEG device and a pair of earphones with volume properly adjusted, and he/she was required to sit still with eyes closed and avoided muscle movements as much as possible to reduce possible artifacts from eyeballs movement, teeth clenching, neck movement, etc. Following each trial, the subject was required to complete a self-assessment to describe his emotion (happy, frightened, etc.). This self-assessment was used as a ground truth to assess the real emotion of the subject. The protocol of this emotion induction experiment is depicted in Fig. 2.

4 Methods

4.1 Feature extraction

Prior to feature extractions, all raw EEG data were centralized (zero-mean). Then, a 2–42 Hz band-pass filter was applied,

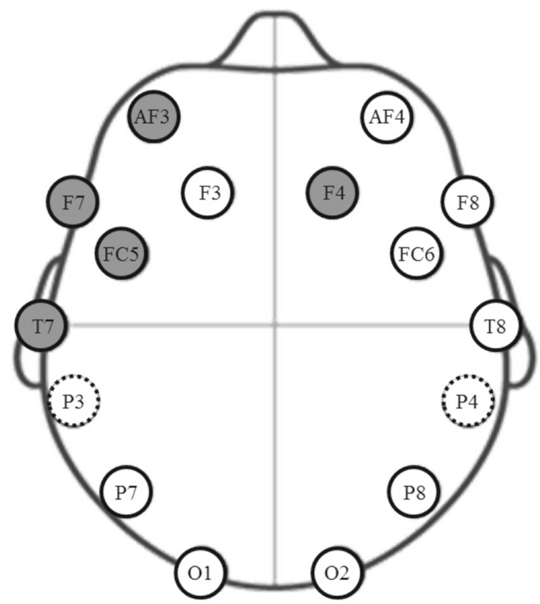

Fig. 3 Emotiv electrode position (*shaded circles* electrodes used in the experiment; *dashed circles* reference electrodes defaulted by Emotiv)

since the major EEG waves (alpha, theta, beta, delta, and gamma) all lie within this bandwidth [35]. The FD feature, alpha power, theta power, beta power, theta/beta ratio, 6 statistical features, HOC features of order up to 36 as introduced in Sect. 2 were calculated from the EEG of the four emotion states. Discarding the first 16-s silent part, the first 5-s and the last 6-s audio elicited parts, EEG from the 22nd sec to the 70th sec were used in data processing. Sixteen 49-s EEG epochs per subject per emotion were obtained. The five channels bearing the highest channel selection scores as was justified in [26] were chosen, namely: channel FC5, F4, F7, AF3, and T7 (see Fig. 3). Channels were referenced to the average of two mastoids, as defaulted by Emotiv. All features were calculated from the centralized, filtered EEG data with a sliding window of size 512 and 75 % overlap (shift forwards by 128 sample points each time) as was proposed in [24]. Following work [18], log-transform was applied to the power features.

4.2 Stability assessment

Four kinds of EEG features (FD, power and power ratio, statistics and HOC) were computed from the EEG data of each of the five subjects. Then, the same feature parameters derived from the same channel from the same emotion class from the same subject were grouped together to compute the ICC. In this way, for each subject, each emotion class, each feature and each channel, we had one ICC assessment. The ICCs were then averaged across the five channels and four emotion classes.

4.3 Classification

The SVM classifier implemented by LIBSVM [6] was used in our work. For classification across different days, the training used the EEG data recorded from the 1st session, and testing data were the EEG from each of the rest 15 sessions (session 2 to session 16). The polynomial kernel was chosen for the SVM with parameters $g = 1, d = 5, r = 1$ and $c = 1$, given by a grid search approach. We also did a within-session classification to compare with the accuracy obtained across different days, which means the EEG data from the same session are partitioned into training and testing data. For the within-session classification, fivefold cross-validation was used. The fivefold cross-validation was done by first dividing the EEG session to five non-overlapping epochs, then using four epochs to train the SVM and one epoch to test the classification accuracy. The average accuracy across five runs was reported.

5 Results and discussion

The average ICC results for each feature from each subject are shown in Fig. 4. Also shown in Fig. 4 are the average ICC results across the five subjects. It can be seen that on average, the 2nd to 6th statistical features have the highest ICC and hence the most stable, followed by FD, HOC of 1st order, and the four band power features (theta, alpha, beta power and theta/beta ratio). The stability of HOC features tends to decrease when the order increases. The 1st statistical feature (i.e., mean value) has an ICC close to zero, which means the feature is highly unstable and tends to change drastically in each measure.

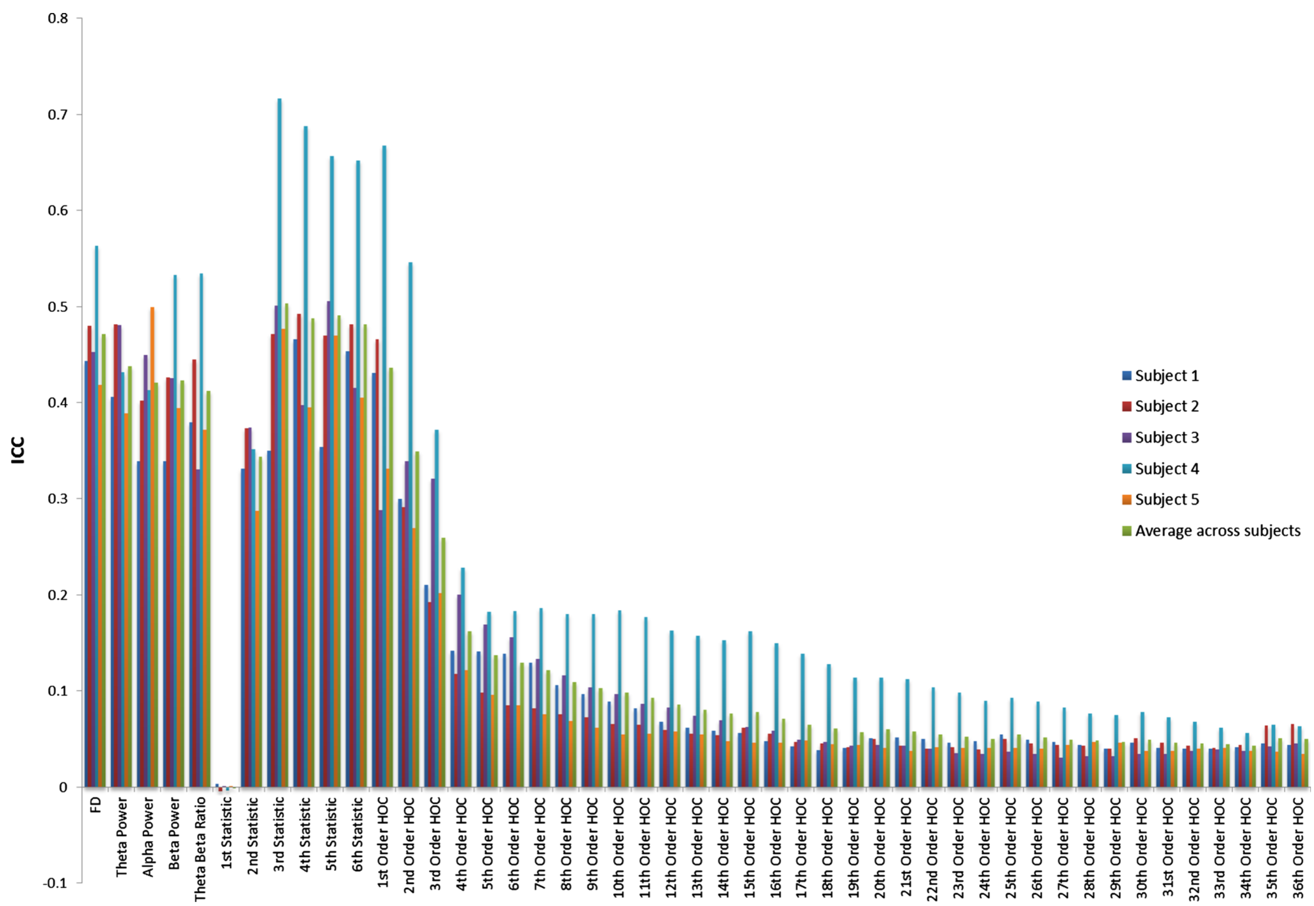

Fig. 4 The average ICC for each subject and each feature

Table 1 Four emotion recognition accuracy across different sessions (%)

Subject	Feature	Session number								
		2	3	4	5	6	7	8	9	
S1	FC1	42.35	35.20	30.61	33.16	26.53	21.43	43.88	22.45	
	FC2	52.55	56.63	33.16	36.73	32.14	28.06	56.12	35.20	
S2	FC1	22.96	38.78	25.00	21.43	14.80	29.59	38.27	15.82	
	FC2	34.18	30.61	25.51	20.92	22.45	27.04	38.78	22.45	
S3	FC1	39.29	24.49	26.02	35.20	32.14	25.51	25.00	27.04	
	FC2	28.57	28.57	18.37	34.18	36.73	26.53	21.43	30.61	
S4	FC1	40.82	26.02	24.49	45.41	34.69	44.90	10.20	12.76	
	FC2	49.49	34.69	54.08	71.43	60.20	50.00	50.00	52.55	
S5	FC1	11.73	34.18	14.29	35.71	32.65	25.51	20.92	28.57	
	FC2	25.00	27.55	30.10	41.84	42.86	32.65	29.59	31.12	
Subject	Feature	Session number								
		10	11	12	13	14	15	16	Average (std)	p value
S1	FC1	32.65	21.43	51.53	40.82	42.35	26.53	36.22	33.81 (9.15)	
	FC2	40.31	32.14	47.45	44.90	44.39	37.76	42.35	41.33 (8.96)	2.41e−04*
S2	FC1	28.57	32.14	22.96	20.92	30.61	17.86	23.47	25.54 (7.35)	
	FC2	25.00	31.12	34.69	28.57	17.35	22.96	24.49	27.07 (5.87)	4.10e−01
S3	FC1	19.39	34.18	25.51	18.37	51.02	29.59	26.02	29.25 (8.25)	
	FC2	21.43	34.69	24.49	25.00	47.96	29.08	23.47	28.74 (7.47)	6.76e−01
S4	FC1	11.73	17.86	23.47	10.71	46.94	19.39	5.10	24.97 (14.32)	
	FC2	45.92	44.39	46.94	37.76	68.88	31.12	46.94	49.63 (11.18)	1.20e−06*
S5	FC1	31.12	21.43	29.59	29.08	32.14	23.98	30.61	26.77 (7.08)	
	FC2	31.12	26.53	41.33	26.53	33.67	31.63	29.08	32.04 (5.69)	6.10e−03*

According to the ICC assessment, we proposed to combine features with high stability (i.e., large ICC values), namely FD, 2nd–6th statistics, 1st order HOC and the four band power features (theta, alpha, beta power and theta/beta ratio) and evaluated the performance of such feature combination. Our hypothesis is that the most stable features give the best intra-subject accuracy across different sessions in real-time emotion recognition algorithm. The accuracy across different sessions is reported in Table 1. As we used the first session as training data and each of the rest 15 sessions as testing data, in total 15 accuracies plus an average accuracy are obtained.

We also calculated the accuracy within each session for comparison and the results are given in Table 2. To calculate the within-session classification accuracy, fivefold cross-validation was performed on each of the 16 sessions for each subject. In both tables, FC1 represents the feature combination of FD, 6 statistics and HOC of order from 1st to 36th, which gives the best accuracy as it was proposed and reported in [24]; FC2 denotes the proposed novel stable feature combination in this paper, i.e., FD, 2nd–6th statistics, 1st order HOC and four band powers.

In Table 1, it can be seen that the accuracy across sessions would fluctuate instead of constantly declining. The results show that FC2 on average outperforms FC1 in four out of five subjects. The accuracies are improved by 7.52, 1.53, 24.66 and 5.27 % for Subject 1, 2, 4 and 5, respectively, as compared to FC1. The improvement is significant for Subject 1, 4 and 5 ($p \leq 0.05$, marked by *). In comparison with our previous work [19], the averaged accuracies obtained by FC2 in Table 1 are improved by 1.4, 0.71, 0.54 and 1.6 % for Subject 1, 2, 3 and 4, respectively. The standard deviations of FC2 are always smaller than FC1 for the five subjects, suggesting that the accuracies given by FC2 fluctuate less than FC1, hence more stable. For Subject 4, FC2 constantly outperforms FC1 in all sessions. For Subject 1, FC2 outperforms FC1 in all sessions except session 12. For Subject 5, FC2 gives better accuracies in all sessions but session 3, 13 and 16. For Subject 2 and 3, FC2 and FC1 achieve similar accuracy. Interestingly, it may be worth pointing out that both Subject 2 and 3 had participated in similar emotion induction experiment before our experiment and expressed that they were too familiar with the affective stimuli (IADS) and hence the stimuli may not be effective in inducing targeted emotions on them, due

Table 2 Four emotion recognition accuracy within each session (%)

Subject	Feature	Session number								
		1	2	3	4	5	6	7	8	9
S1	FC1	63.33	60.83	75.00	58.33	56.67	55.83	65.83	76.67	40.83
	FC2	58.33	55.83	66.67	65.00	53.33	59.17	47.50	62.50	40.83
S2	FC1	56.67	59.17	53.33	33.33	48.33	80.83	45.83	65.83	67.50
	FC2	61.67	52.50	41.67	32.50	67.50	73.33	37.50	60.83	54.17
S3	FC1	37.50	52.50	59.17	28.33	36.67	49.17	49.17	55.00	64.17
	FC2	37.50	61.67	41.67	41.67	29.17	49.17	30.00	54.17	57.50
S4	FC1	88.33	85.00	88.33	88.33	84.17	85.83	72.50	86.67	79.17
	FC2	87.50	92.50	75.83	91.67	75.83	90.00	84.17	80.83	67.50
S5	FC1	54.17	80.83	52.50	50.00	55.00	48.33	50.83	50.83	36.67
	FC2	63.33	75.83	50.00	61.67	53.33	40.00	39.17	63.33	44.17
Subject	Feature	Session number								
		10	11	12	13	14	15	16	Average (std)	p value
S1	FC1	74.17	60.00	65.83	65.83	67.50	69.17	75.00	64.43 (9.19)	
	FC2	69.17	64.17	64.17	61.67	68.33	60.00	77.50	60.89 (8.75)	5.21e−02
S2	FC1	30.00	41.67	55.83	30.00	46.67	50.83	49.17	50.94 (13.78)	
	FC2	39.17	46.67	56.67	35.00	31.67	60.00	48.33	49.95 (12.90)	6.81e−01
S3	FC1	68.33	52.50	49.17	40.00	71.67	42.50	40.83	49.79 (12.04)	
	FC2	52.50	45.00	41.67	37.50	65.83	55.00	45.00	46.56 (10.67)	2.08e−01
S4	FC1	82.50	65.83	71.67	91.67	85.00	88.33	94.17	83.59 (7.69)	
	FC2	77.50	60.00	63.33	97.50	72.50	90.83	95.00	81.41 (11.59)	2.68e−01
S5	FC1	49.17	60.83	51.67	38.33	55.00	40.83	61.67	52.29 (10.33)	
	FC2	44.17	55.00	45.83	41.67	37.50	53.33	45.83	50.89 (10.77)	5.77e−01

to habituation effect [2]. Notwithstanding, for the other three subjects, it is still clear that FC2 in most cases outperforms FC1. This could be owing to the fact that FC2 omits the features that change drastically throughout days (i.e., have a lower ICC). Such features may be useful in representing the transient states of the brain, but including such features will also increase the intra-subject variance, hence decreasing the accuracy when training is done once and testing is done throughout days. Hence, if training is limited to one-time only, we suggest using FC2 for its better accuracy throughout days and much smaller feature vector dimensionality (for FC1 and FC2, the dimensionality of the feature vectors are $(1 + 6 + 36) \times 5 = 215$ and $(1 + 5 + 1 + 4) \times 5 = 55$, respectively).

In Table 2, we can see that for within-session emotion recognition, FC1 outperforms FC2 in most cases and on average. This result is consistent with the work [24]. This is reasonable as FC1 has a much larger feature vector than FC2. FC1 contains more information that reflects the transient states of the brain during emotional moment, while FC2 preserves less such information. Recognizing four emotion classes, FC1 on average achieves accuracy from 49.79 to 83.59 %, and FC2 achieves 46.56–81.41 %. Therefore, if training is permitted every time prior to real-time emotion recognition, the FC1 feature combination proposed in [24] is still preferred.

The performance of the feature combinations for classifying any two emotions within each session and across different sessions was also investigated. Totally there were six pairs of emotion combinations, i.e., happy–pleasant, happy–angry, happy–frightened, pleasant–angry, pleasant–frightened and frightened–angry. For the within-session recognition, five-fold cross-validation was used to get the accuracy of each session and each pair of emotions. The average accuracies over all 16 sessions cross-validation accuracies across 6 pairs of emotion combinations are reported in Table 3 under the column within sessions. For the across-session recognition, each time we selected one pair from the aforementioned six emotion pairs. An SVM was trained with the first session and tested with the rest 15 sessions. The average accuracies across 15 sessions and 6 emotion pairs are reported in Table 3 under the column across sessions. From Table 3, we can see that the FC1 feature combination proposed in [24] always achieves better accuracy than FC2 in within-session recog-

Table 3 Comparison of two out of four emotion recognition between within-session accuracy and across-session accuracy (%)

Subject	Within sessions		Across sessions	
	FC1	FC2	FC1	FC2
S1	82.48	82.03	62.97	61.30
S2	72.95	72.14	50.78	53.54
S3	72.92	71.48	53.24	54.47
S4	93.37	92.19	60.80	71.75
S5	73.72	73.70	50.09	56.16

Table 4 Comparison of positive and negative emotion recognition between within-session accuracy and across-session accuracy (%)

Subject	Within sessions		Across sessions	
	FC1	FC2	FC1	FC2
S1	85.05	81.88	65.27	69.29
S2	70.31	69.32	51.36	52.93
S3	72.81	69.48	53.47	55.68
S4	91.09	92.03	69.63	73.10
S5	73.59	70.31	54.83	57.07

nition, while FC2, the proposed stable features in this paper, outperforms FC1 in across-session recognition in four out of five subjects.

In addition, we also combine together the EEG data labeled with positive emotion, namely happy and pleasant, and those with negative emotion, namely frightened and angry, to classify the positive and negative emotions in valence dimension. The results are shown in Table 4. From Table 4, it can be seen that FC2 always outperforms FC1 in across-session tests, while FC1 performs better in within-session cross-validation in all subjects but Subject 4. This again demonstrates that FC1 is fit for the scene that training is allowed each time prior to emotion recognition, while the proposed stable feature FC2 is preferred when only one-time training is permitted.

6 Applications for emotion monitoring

The proposed algorithm can be integrated with different applications for stable real-time emotion recognition with one training session for a new user. As the proposed algorithm is a subject-dependent one, a training session is needed for a new subject. In the training session, the user listens to sound clips labeled with emotions which are supposed to be elicited. After listening to the clips, the user is asked to assess arousal, valence and dominance levels of his/her feelings by moving the bar on a scale of 1–9. In Fig. 5, the screenshot of the menu of the training session is shown. The top left corner

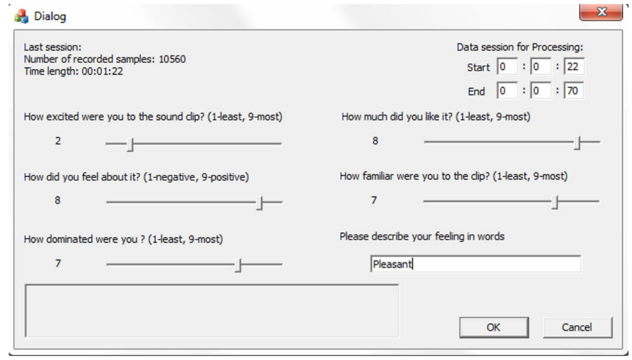

Fig. 5 Screenshot of the training session

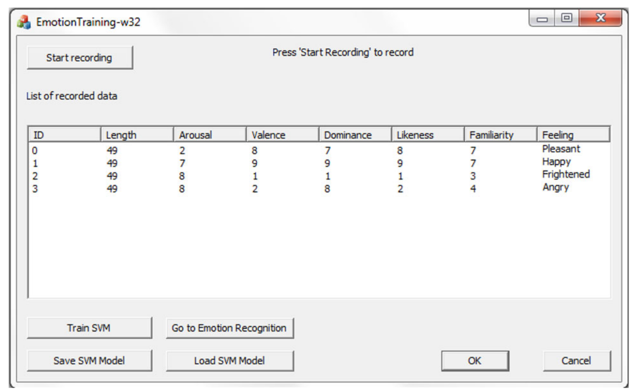

Fig. 6 Screenshot of the classifier training menu

of the screen shows the number of recorded samples of EEG data and the recorded length of time for the training session. The top right corner allows the user to choose the duration of the recorded data for training. With the arousal, valence and dominance levels entered by the user to label the recorded EEG data, the SVM model is trained. The results are saved and later are used to classify new EEG data samples in the applications. Figure 6 shows the screenshot of the classifier training menu. During the training, subjects are exposed to affective stimuli from IADS database to evoke certain emotions, and the EEG data are recorded simultaneously.

Then, an SVM classifier is trained using the recorded EEG with the corresponding emotion label from the subjects self-assessment. The proposed algorithm can be integrated with different applications for real-time emotion recognition with one training session. For example, an application called Emotional Avatar is implemented. This application enables the real-time monitoring of human emotions. The recognized emotions of the subject from EEG are visualized and animated as the facial expressions of a 3D Haptek avatar [12]. In Fig. 7, the current recognized emotional state of the subject is angry, and the avatar shows rage face to visualize the angry emotion.

Another example of the application of human emotion monitoring is the Twin Girls emotional companion applica-

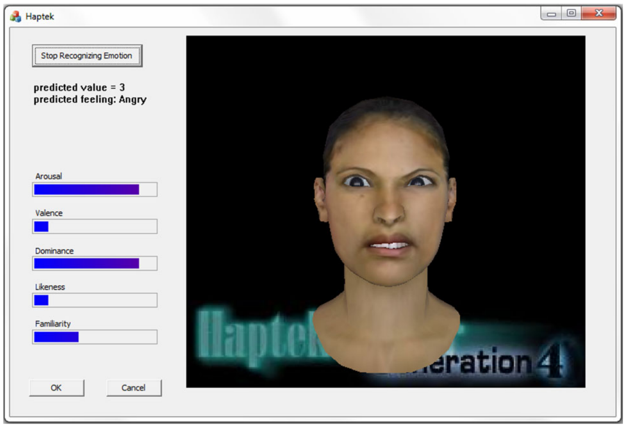

Fig. 7 Screenshot of the real-time emotion recognition application Emotional Avatar

Fig. 8 Screenshot of the real-time emotion recognition application Twin Girls

tion. In this application, the twins' behaviors are changed according to the recognized emotions from EEG. For example, when the recognized emotion is happy (Fig. 8), the twins caper; when the recognized emotion is frightened, the twins curl up. Additionally, one girl's dress and hair color and the other girl's facial expression are changed. For example, when the recognized emotion is happy (Fig. 8), the hair color of the girl on the left-hand side changes to red and the dress color changes to pink; the facial expression of the girl on the right-hand side is smiling happily; when the recognized emotion is frightened, both the hair color and dress color of the girl on the left-hand side change to black; the facial expression of the girl on the right-hand side is terrified. The proposed algorithm can be further integrated with other applications. For example, the recognized emotion results can be sent to other applications, and other applications may execute different commands based on different emotion states of the user. In music therapy [39], the music player receives the emotion state recognized by our emotion-monitoring algorithm, and select proper music to play to the user. If the music player receives a negative emotion from the user, the music player will choose the positive music to play in order to cheer up the user.

7 Conclusion

In this paper, stability of different EEG features for real-time emotion recognition was analyzed. An experiment to induce four emotions such as pleasant, happy, frightened and angry was designed and carried out in eight consecutive days (two sessions per day) on five subjects to record EEG data. A novel real-time emotion recognition algorithm was proposed based on the most stable features such as FD, five statistics features, 1st order HOC and four band power features (alpha power, theta power, beta power, theta/beta ratio) and it was compared with the previous algorithms. Our hypothesis that the most stable features give the best intra-subject accuracy across different days in real-time emotion recognition algorithm is validated. The proposed algorithm is a subject-dependent one which needs just one training for the subject. The training results can be used in real-time emotion recognition applications without re-training with the adequate accuracy (up to 49.63 % accuracy for 4 emotions classification, 71.75 % for any two emotions classification, and 73.10 % for positive negative emotions classification). A real-time emotion-monitoring application employing the proposed stable features is implemented. The proposed algorithm is integrated in Emotional Avatar and Twin Girls applications. The Emotional Avatar application can monitor and visualize the current emotion of the user assessed from his/her EEG during the human–machine interaction. The Twin Girls application monitors the users current emotion and can be used as the users companion that can show the same or opposite to the users emotion depending on the application task.

The proposed algorithm can be further integrated with other applications seamlessly, by sending the recognized emotion results to other applications which execute different commands accordingly. In this way, the human–machine interaction is made more adaptive to user's feeling and the user may feel more engaged. In the future, we are planning to establish a dataset for the research of stability of affective EEG signals.

Acknowledgments The work is supported by Fraunhofer IDM@NTU, which is funded by the National Research Foundation (NRF) and managed through the multi-agency Interactive and Digital Media Programme Office (IDMPO).

References

1. Allen, J.J., Urry, H.L., Hitt, S.K., Coan, J.A.: The stability of resting frontal electroencephalographic asymmetry in depression. Psychophysiology **41**(2), 269–280 (2004)
2. Bouton, M.E.: Learning and Behavior: A Contemporary Synthesis. Sinauer Associates, Sunderland (2007)
3. Bradley, M.M., Lang, P.J.: The international affective digitized sounds (; iads-2): affective ratings of sounds and instruction manual. Technical report B-3, University of Florida, Gainesville (2007)

4. Brown, L., Grundlehner, B., Penders, J.: Towards wireless emotional valence detection from EEG. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp. 2188–2191. IEEE (2011)
5. Chanel, G., Kierkels, J.J., Soleymani, M., Pun, T.: Short-term emotion assessment in a recall paradigm. Int. J. Hum.–Comput. Stud. **67**(8), 607–627 (2009)
6. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) **2**(3), 27 (2011)
7. Emotiv. http://www.emotiv.com. Accessed 26 Feb 2015
8. Frantzidis, C., Bratsas, C., Papadelis, C.L., Konstantinidis, E., Pappas, C., Bamidis, P.D., et al.: Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. Inf. Technol. Biomed. IEEE Trans. **14**(3), 589–597 (2010)
9. Gasser, T., Bächer, P., Steinberg, H.: Test-retest reliability of spectral parameters of the EEG. Electroencephalogr. Clin. Neurophysiol. **60**(4), 312–319 (1985)
10. Gasser, T., Jennen-Steinmetz, C., Verleger, R.: EEG coherence at rest and during a visual task in two groups of children. Electroencephalogr. Clin. Neurophysiol. **67**(2), 151–158 (1987)
11. Gudmundsson, S., Runarsson, T.P., Sigurdsson, S., Eiriksdottir, G., Johnsen, K.: Reliability of quantitative EEG features. Clin. Neurophysiol. **118**(10), 2162–2171 (2007)
12. Haptek. http://www.haptek.com. Accessed 28 Feb 2015
13. Higuchi, T.: Approach to an irregular time series on the basis of the fractal theory. Phys. D: Nonlinear Phenom. **31**(2), 277–283 (1988)
14. Hosseini, S.A., Khalilzadeh, M.A., Naghibi-Sistani, M.B., Niazmand, V.: Higher order spectra analysis of EEG signals in emotional stress states. In: 2010 Second International Conference on Information Technology and Computer Science (ITCS), pp. 60–63. IEEE (2010)
15. Ishino, K., Hagiwara, M.: A feeling estimation system using a simple electroencephalograph. In: IEEE International Conference on Systems, Man and Cybernetics, 2003. vol. 5, pp. 4204–4209. IEEE (2003)
16. Kedem, B., Yakowitz, S.: Time Series Analysis by Higher Order Crossings. IEEE Press, Piscataway (1994)
17. Koelstra, S., Mühl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: Deap: a database for emotion analysis; using physiological signals. Affect. Comput. IEEE Trans. **3**(1), 18–31 (2012)
18. Kondacs, A., Szabó, M.: Long-term intra-individual variability of the background EEG in normals. Clin. Neurophysiol. **110**(10), 1708–1716 (1999)
19. Lan, Z., Sourina, O., Wang, L., Liu, Y.: Stability of features in real-time EEG-based emotion recognition algorithm. In: 2014 International Conference on Cyberworlds (CW), pp. 137–144. IEEE (2014)
20. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (iaps): affective ratings of pictures and instruction manual. Technical report A-8 (2008)
21. Li, M., Lu, B.L.: Emotion classification based on gamma-band EEG. In: Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, pp. 1223–1226. IEEE (2009)
22. Lin, Y.P., Wang, C.H., Jung, T.P., Wu, T.L., Jeng, S.K., Duann, J.R., Chen, J.H.: EEG-based emotion recognition in music listening. Biomed. Eng. IEEE Trans. **57**(7), 1798–1806 (2010)
23. Lin, Y.P., Wang, C.H., Wu, T.L., Jeng, S.K., Chen, J.H.: EEG-based emotion recognition in music listening: a comparison of schemes for multiclass support vector machine. In: IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009, pp. 489–492. IEEE (2009)
24. Liu, Y., Sourina, O.: EEG databases for emotion recognition. In: 2013 International Conference on Cyberworlds (CW), pp. 302–309. IEEE (2013)
25. Liu, Y., Sourina, O.: Real-time fractal-based valence level recognition from EEG. In: Transactions on Computational Science XVIII, pp. 101–120. Springer (2013)
26. Liu, Y., Sourina, O.: Real-time subject-dependent EEG-based emotion recognition algorithm. In: Transactions on Computational Science XXIII, pp. 199–223. Springer (2014)
27. Liu, Y., Sourina, O., Nguyen, M.K.: Real-time EEG-based emotion recognition and its applications. In: Transactions on computational science XII, pp. 256–277. Springer (2011)
28. McGraw, K.O., Wong, S.P.: Forming inferences about some intraclass correlation coefficients. Psychol. Methods **1**(1), 30 (1996)
29. Murugappan, M., Nagarajan, R., Yaacob, S.: Combining spatial filtering and wavelet transform for classifying human emotions using EEG signals. J. Med. Biol. Eng. **31**(1), 45–51 (2011)
30. Petrantonakis, P.C., Hadjileontiadis, L.J.: Emotion recognition from EEG using higher order crossings. Inf. Technol. Biomed. IEEE Trans. **14**(2), 186–197 (2010)
31. Petrantonakis, P.C., Hadjileontiadis, L.J.: A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition. Inf. Technol. Biomed. IEEE Trans. **15**(5), 737–746 (2011)
32. Petrantonakis, P.C., Hadjileontiadis, L.J.: Adaptive emotional information retrieval from EEG signals in the time-frequency domain. Signal Process. IEEE Trans. **60**(5), 2604–2616 (2012)
33. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: analysis of affective physiological state. Pattern Anal. Mach. Intell. IEEE Trans. **23**(10), 1175–1191 (2001)
34. Salinsky, M., Oken, B., Morehead, L.: Test-retest reliability in EEG frequency analysis. Electroencephalogr. Clin. Neurophysiol. **79**(5), 382–392 (1991)
35. Sanei, S., Chambers, J.A.: EEG Signal Processing. Wiley, New York (2013)
36. Schaaff, K.: EEG-Based Emotion Recognition. Universitat Karlsruhe (TH), Karlsruhe (2008)
37. Schaaff, K., Schultz, T.: Towards an EEG-based emotion recognizer for humanoid robots. In: The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009. RO-MAN 2009, pp. 792–796. IEEE (2009)
38. Sohaib, A.T., Qureshi, S., Hagelbäck, J., Hilborn, O., Jerčić, P.: Evaluating classifiers for emotion recognition using EEG. In: Foundations of Augmented Cognition, pp. 492–501. Springer (2013)
39. Sourina, O., Liu, Y., Nguyen, M.K.: Real-time EEG-based emotion recognition for music therapy. J. Multimodal User Interf. **5**(1–2), 27–35 (2012)
40. Tomarken, A.J., Davidson, R.J., Wheeler, R.E., Kinney, L.: Psychometric properties of resting anterior EEG asymmetry: temporal stability and internal consistency. Psychophysiology **29**(5), 576–592 (1992)
41. Wang, X.W., Nie, D., Lu, B.L.: EEG-based emotion recognition using frequency domain features and support vector machines. In: Neural Information Processing, pp. 734–743. Springer (2011)
42. Williams, C.E., Stevens, K.N.: Emotions and speech: some acoustical correlates. J. Acoust. Soc. Am. **52**(4B), 1238–1250 (1972)

Zirui Lan received his B.Sc. in Computer Science from Sun Yat-sen University in 2013. He is currently a research student in Fraunhofer IDM@NTU, School of Electrical and Electronic Engineering, NTU, and Faculty of Computer Science and Biomedical Engineering, TU Graz, under a joint Ph.D. program between Nanyang Technological University and Technische Universität Graz. His research interests are in EEG signal processing, EEG-based brain state classification and EEG-based emotion recognition.

Olga Sourina received her M.Sc. in Computer Engineering from Moscow Engineering Physics Institute (MEPhI) in 1983, and her Ph.D. in Computer Science from NTU in 1998. Dr. Sourina is a Principal Research Scientist and head of Cognitive Human–Computer Interaction lab in FraunhoferIDM@NTU. Her research interests are in human–computer interfaces, brain–computer interaction and biosignals. Dr. Sourina has more than 130 publications including more than 40 research papers in journals and 3 books. She is a senior member of the IEEE, member of the International Organization of Psychophysiology and a member of the Biomedical Engineering Society.

Lipo Wang received the Bachelor degree from National University of Defense Technology (China) and Ph.D. from Louisiana State University (USA). His research interest is intelligent techniques with applications to communications, image/video processing, biomedical engineering, and data mining. He is (co-)author of over 270 papers, of which more than 90 are in journals. He holds a U.S. patent in neural networks and a Chinese patent in VLSI. He has co-authored 2 monographs and (co-)edited 15 books. He was/will be keynote/panel speaker for 16 international conferences. He is/was Associate Editor/Editorial Board Member of 30 international journals, including 3 IEEE Transactions, and guest editor for 10 journal special issues. He is an AdCom member (2010–2015) of the IEEE Computational Intelligence Society (CIS) and served as CIS Vice President for Technical Activities and Chair of Emergent Technologies Technical Committee. He is a member of the Board of Governors of the International Neural Network Society (2011–2016) and was an AdCom member of the IEEE Biometrics Council. He served as Chair of Education Committee, IEEE Engineering in Medicine and Biology Society (EMBS). He was President of the Asia-Pacific Neural Network Assembly (APNNA) and received the APNNA Excellent Service Award. He was founding Chair of both the EMBS Singapore Chapter and CIS Singapore Chapter. He serves/served as chair/committee members of over 200 international conferences.

Yisi Liu is a research fellow of Fraunhofer IDM@NTU. She received her B.Eng. degree in 2008 from University of Science and Technology, Beijing, China, M.Sc. degree in 2009 and Ph.D. degree in 2014 in School of Electrical and Electronic Engineering from Nanyang Technological University, Singapore. Her current research interests include human factor, human machine interaction, brain computer interface, EEG-based brain states recognition, biosignal/EEG processing, serious games, and neuroscience.