Stability of Features in Real-time EEG-based Emotion Recognition Algorithm

Zirui Lan[a], Olga Sourina[a], Lipo Wang[b] and Yisi Liu[a]
[a] Fraunhofer IDM @ NTU
[b] School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
LanZ0001 | EOSourina | ELPWang | LiuYS@ntu.edu.sg

***Abstract*—Stability of algorithms is very important for electroencephalogram (EEG) based applications. Stable features should exhibit consistency among repeated measurements of the same subject. Previously, power features were reported to be one of the most stable EEG features in medical application. In this paper, stability of features in emotion recognition algorithms is studied. Our hypothesis is that the most stable features give the best intra-subject accuracy across different days in real-time emotion recognition algorithm. An experiment to induce 4 emotions such as pleasant, happy, frightened, and angry is designed and carried out in 8 consecutive days (two sessions per day) on 4 subjects to record EEG data. A novel real-time subject-dependent algorithm with the most stable features is proposed and implemented. The algorithm needs just one training for each subject. The training results can be used in real-time emotion recognition applications without re-training with the adequate accuracy. The proposed algorithm is integrated with a real-time application "Emotional Avatar".**

Keywords-EEG; Emotion recognition; Fractal dimension (FD); Stability; Intra-class Correlation Coefficient (ICC)

I. INTRODUCTION

Electroencephalogram (EEG) is the continuous measure of the electric potential of human brain. The application of EEG used to be limited to medical environment, e.g. facilitating the diagnosis of brain diseases like Epileptic Seizure, Attention Deficit Hyperactive Disorder (ADHD), Alzheimer's Disease (AD) etc. However, the advancement of technology has 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, e.g. EEG-enabled games, EEG-based emotion recognition and music therapy [1][2] etc. EEG-based emotion recognition draws high attention because it is desirable that a machine can recognize human emotion and interact with us in a way we like. Current scheme for EEG-based emotion recognition is subject dependent and requires a training session prior to running a real-time emotion recognition application almost every time. During the training session, stimuli (audio/video) are presented to the subject to invoke certain targeted emotions and meanwhile the EEG of the subject is being 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 training. A stable feature is wanted, which can last long so that re-training can be omitted.

The stability issue of EEG features was firstly brought up under medical application environments. 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 [3] and [4], 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. Power feature obtained from alpha band is the most stable, followed by theta band, delta bend, and beta band. [5] recruited 19 subjects and recorded their EEG in closed-eye state in an interval of 12-16 weeks. No significant difference was found between the stability of absolute power and relative power. Peak alpha frequency and median frequency were reported to be the most stable. [6] 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 25Hz being the largest, followed by alpha mean frequency, absolute alpha and beta power, absolute delta power and alpha coherence. [7] 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.

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 emotion can be traced

978-1-4799-4677-8/14 $31.00 © 2014 IEEE
DOI 10.1109/CW.2014.27

back to 1972 [8], 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 of brain activities, and emotion recognition from EEG has the potential to assess the true inner feelings of the subject.

In the study of EEG-based emotion recognition, different features and different classifiers were employed. [9] used wavelet transform, Fast Fourier Transform (FFT) and statistics such as mean and variance as feature and employed Neural Network (NN) to classify four emotions. An accuracy of 67.7% was achieved with 3 channels. [10] 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, [11] and [12] made use of Short Time Fourier Transform (STFT) to extract feature and SVM as classifier and achieved 62.07% (16 channels) and 63% (64 channels) accuracy respectively. In another work [13], four emotions were recognized with differential asymmetry of hemispheric EEG power spectra as feature and SVM as classifier, and obtained an accuracy of 82.29% using 32 channels. [14] recognized five emotions at an accuracy of 83.04% using statistical features from different EEG bands from 62 channels and K-Nearest Neighbor (KNN) classifier. [15] used SVM to differentiate two emotions featured by logarithmic variances from 62 channels and achieved 93.5% accuracy. [16] employed fractal dimension (FD) feature together with statistical and Higher Order Crossings (HOC) features, and a 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). As can be seen, SVM is often used in EEG-based emotion recognition. [17][18][19][20][21] also support that SVM is preferred for better accuracy and effectiveness.

Though the power feature is the most stable one in the medical area, it is not necessarily fit for emotion recognition application. Liu *et al.* have demonstrated that FD feature outperform power feature in terms of accuracy [22]. In this work, we investigate the stability of various features used in the real-time EEG-based emotion recognition algorithm [16] and propose a novel real-time EEG-based emotion recognition algorithm with the most stable features. 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 intra-subject EEG data labeled with 4 emotions such as pleasant, happy, frightened and angry. The data are collected from 4 subjects during 8 consecutive days (2 sessions per day per subject).

The paper is organized as follows. In Section II, the related work including the feature extraction methods such as fractal dimension, power, statistical features and Higher Order Crossings are given. In Section III, an experiment to collect affective intra-subject EEG data is described. In Section IV, the proposed stable emotion recognition algorithm is introduced. In Section V, the data processing and analysis results and discussion are presented. In Section VI, the application of the proposed algorithm is given. Section VII concludes the paper.

II. RELATED WORK

A. Fractal Dimension Feature Extraction

FD measures the geometric complexity of objects. FD feature has been proven effective in EEG-based emotion recognition application [16]. Following work [16], a Higuchi algorithm was used to compute the FD feature of EEG.

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

$$X_k^m : X(m), X(m+k), X(m+2k),..., X(m+\left\lfloor \frac{N-m}{k} \right\rfloor k)$$
$$m = 1,2,3,...k, \tag{1}$$

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:

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

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:

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

B. 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 (4-8Hz), alpha band (8-12Hz), and beta band (12-30Hz) are computed.

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

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

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

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

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 8-12Hz range.

C. Statistical Feature Extraction

Six statistical features were adopted in [23] and used in the EEG-based emotion recognition in [16]. 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).

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

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

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

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

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

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

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

D. Higher Order Crossings (HOC) Feature Extraction

Higher Order Crossings (HOC) was proposed in [24] and used in [16] 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:

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

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:

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

E. 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 for 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 [5][6][7][25][26]. 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 [27]. Among these models, ICC(1) was often used in EEG stability study [7][25]. ICC(1) is derived from a one-way ANOVA model and defined as:

$$ICC = \frac{MS_B - MS_W}{MS_B + (k-1) MS_W} \tag{16}$$

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.

III. EXPERIMENT

A. Experiment Protocol

Since there is no available published EEG database for the analysis of stability of EEG features regarding emotion recognition, we designed and conducted an experiment to collect the affective EEG data on a small group of subjects for a period of time. This preliminary study included four subjects, three males and one female, with the age of 24-28 years old. All subjects reported no history of mental diseases and head injuries. A 14-channel Emotiv EEG device [28] was used to record the EEG data at a sampling rate of 128Hz. Sixteen sessions were recorded within eight days (two sessions per day). Each session consisted of four trials, with each trial corresponded to one induced emotion, i.e., four emotions were elicited in one session. There are standard affective stimuli libraries such as International Affective Picture System (IAPS) [29] and International Affective Digitized Sounds (IADS) [30]. In our study, the IADS was chosen for the experiment design as 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 [16]. Sound clips from the same category of the IADS were chosen and appended together to make a 76 seconds audio file, with the first 16 seconds 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 is invited to the lab and is well-instructed about the protocol of the experiment. The subject wears the Emotiv EEG device and a pair of earphones with volume properly adjusted, and he/she is required to sit still with eyes closed, and avoid muscle movements as much as possible to reduce possible artifacts from eyeballs movement, teeth clenching, neck movement etc. Following each trial, the subject is required to complete a self-assessment to describe his emotion (happy, frightened etc.). This self-assessment would be used as ground truth to assess the real emotion of the subject. The protocol of this emotion induction experiment is depicted in Fig. 1.

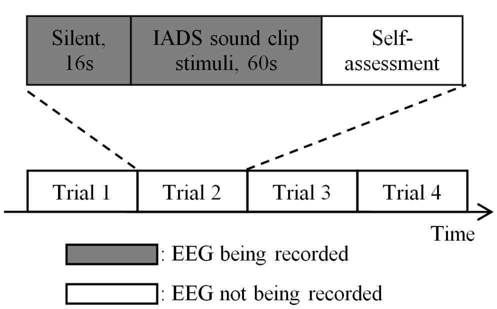

Figure 1. Protocol of emotion induction experiment.

IV. METHODS

A. Feature Extraction

Prior to feature extractions, all raw EEG data were centralized (shifted to have zero-mean). Then, a 2-42Hz band-pass filter was applied, since the major EEG waves (alpha, theta, beta, delta, and gamma) all lie within this bandwidth [31]. The FD feature, alpha power, theta power, beta power, 6 statistical features, HOC features of order up to 36 as introduced in Section II were calculated from the EEG of the four emotion states. Discarding the first 16-sec silent part, the first 5-sec and the last 6-sec audio elicited parts, EEG from the 22nd-sec to the 70th-sec were used in data processing. Sixteen 49-sec EEG epochs per subject per emotion were obtained. The five channels bearing the highest channel selection scores as was justified in [32] were chosen, namely: channel FC5, F4, F7, AF3, and T7. 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 [16]. Following work [6], log-transform was applied to the power features.

B. Stability Assessment

Four kinds of EEG features (FD, power, statistics and HOC) were computed from the EEG data of each of the four subjects. The same features derived from the same channel from the same emotion class from the same subject were grouped together to compute the ICCs. 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 4 emotion classes and the 5 channels.

C. Classification

The SVM classifier implemented by LIBSVM [33] 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 tuned parameters g=1, d=5, r=1 and c=1. We also did a within session classification to compare with the accuracy across different days, which means the EEG data from the same session are partitioned into training and testing data. For the within session classification, 5-fold cross validation was used. The 5-fold 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.

V. RESULT AND DISCUSSION

The average ICC results for each subject and each feature are shown in Fig. 2. Also shown are the average ICC results across the four 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 three band power features. The stability of HOC features tends to decrease when the order increases. The 1st statistical

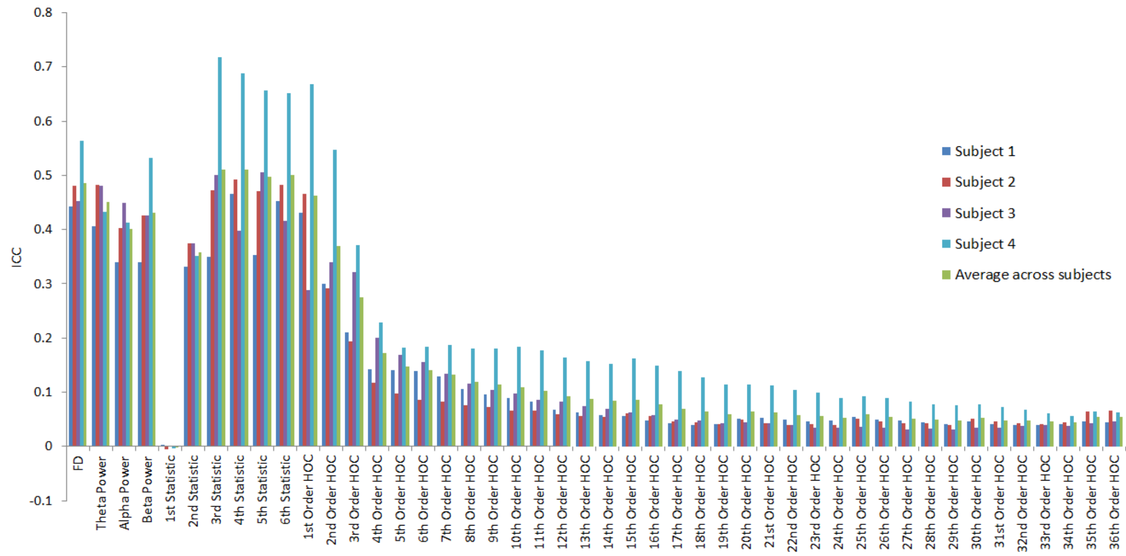

Figure 2. The average ICC for each subject and each feature.

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.

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 three band power features 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 I. 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 II. To calculate the within session classification accuracy, 5-fold cross validation was performed to 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 [16]; FC2 denotes the proposed novel stable feature combination in this paper, i.e. FD, 2nd-6th statistics, 1st order HOC and three band powers.

In Table I, it can be seen that the accuracy across sessions would fluctuate instead of constantly declining or rising. From the results it shows that FC2 on average outperforms FC1 in three out of four subjects. FC2 constantly outperforms FC1 in every session for Subject 4 and in all sessions except session 12 for Subject 1. For Subject 2 and 3, FC2 and FC1 achieve similar accuracy. This is 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)×5=215 and (1+5+1+3)×5=50 respectively).

TABLE I. FOUR EMOTION RECOGNITION ACCURACY ACROSS DIFFERENT SESSIONS (%)

Subject	Feature	Session Number				
		2	*3*	*4*	*5*	*6*
S1	FC1	42.35	35.20	30.61	33.16	26.53
	FC2	47.45	52.55	37.76	37.76	34.69
S2	FC1	22.96	38.78	25.00	21.43	14.80
	FC2	29.08	27.55	28.57	18.88	22.45
S3	FC1	39.29	24.49	26.02	35.20	32.14
	FC2	32.14	31.12	23.47	33.16	28.57
S4	FC1	40.82	26.02	24.49	45.41	34.69
	FC2	50.51	31.12	52.55	69.39	60.71

Subject	Feature	Session Number					
		7	*8*	*9*	*10*	*11*	*12*
S1	FC1	21.43	43.88	22.45	32.65	21.43	51.53
	FC2	22.45	46.94	34.69	35.71	32.14	48.98
S2	FC1	29.59	38.27	15.82	28.57	32.14	22.96
	FC2	20.92	37.76	25.00	26.02	30.10	30.61
S3	FC1	25.51	25.00	27.04	19.39	34.18	25.51
	FC2	27.04	16.84	28.57	18.37	33.67	26.53
S4	FC1	44.90	10.20	12.76	11.73	17.86	23.47
	FC2	57.14	50.00	48.47	44.90	42.35	41.84

Subject	Feature	Session Number				
		13	*14*	*15*	*16*	*Average*
S1	FC1	40.82	42.35	26.53	36.22	33.81
	FC2	42.35	48.47	36.73	40.31	**39.93**
S2	FC1	20.92	30.61	17.86	23.47	25.54
	FC2	29.59	14.80	27.55	26.53	**26.36**
S3	FC1	18.37	51.02	29.59	26.02	29.25
	FC2	23.98	48.47	30.10	20.92	**28.20**
S4	FC1	10.71	46.94	19.39	5.10	24.97
	FC2	32.14	65.82	30.10	43.37	**48.03**

TABLE II. FOUR EMOTION RECOGNITION ACCURACY WITHIN EACH SESSION (%)

Subject	Feature	Session Number					
		1	*2*	*3*	*4*	*5*	*6*
S1	**FC1**	63.33	60.83	75.00	58.33	56.67	55.83
	FC2	57.50	58.33	65.00	59.17	60.00	58.33
S2	**FC1**	56.67	59.17	53.33	33.33	48.33	80.83
	FC2	55.83	46.67	48.33	31.67	62.50	75.00
S3	**FC1**	37.50	52.50	59.17	28.33	36.67	49.17
	FC2	42.50	62.50	38.33	44.17	27.50	40.00
S4	**FC1**	88.33	85.00	88.33	88.33	84.17	85.83
	FC2	86.67	91.67	68.33	86.67	75.00	90.83
Subject	**Feature**	**Session Number**					
		7	*8*	*9*	*10*	*11*	*12*
S1	**FC1**	65.83	76.67	40.83	74.17	60.00	65.83
	FC2	47.50	63.33	32.50	67.50	65.83	64.17
S2	**FC1**	45.83	65.83	67.50	30.00	41.67	55.83
	FC2	44.17	56.67	57.50	43.33	41.67	56.67
S3	**FC1**	49.17	55.00	64.17	68.33	52.50	49.17
	FC2	26.67	53.33	56.67	50.83	42.50	48.33
S4	**FC1**	72.50	86.67	79.17	82.50	65.83	71.67
	FC2	83.33	83.33	65.83	75.00	58.33	59.17
Subject	**Feature**	**Session Number**					
		13	*14*	*15*	*16*	*Average*	
S1	**FC1**	65.83	67.50	69.17	75.00	**64.43**	
	FC2	61.67	70.83	57.50	78.33	60.47	
S2	**FC1**	30.00	46.67	50.83	49.17	**50.94**	
	FC2	39.17	26.67	50.83	42.50	48.70	
S3	**FC1**	40.00	71.67	42.50	40.83	**49.79**	
	FC2	33.33	68.33	51.67	51.67	46.15	
S4	**FC1**	91.67	85.00	88.33	94.17	**83.59**	
	FC2	97.50	72.50	87.50	91.67	79.58	

In Table II, 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 [16]. 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 48.70% to 79.58%. Hence, if training is permitted every time prior to real-time emotion recognition, the FC1 feature combination proposed in [16] 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, 5-fold cross validation was used to get the accuracy of each session and each pair of emotions. The average accuracies over all 16 sessions across 6 pairs of emotion combinations are reported in Table III under the column "within sessions". For the across session recognition, each time we selected one pair from the aforementioned six emotion pairs. A 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 III under the column "across sessions". From Table III, we can see that the FC1 feature combination proposed in [16] always achieves better accuracy than FC2 in within session recognition, while FC2, the proposed stable features in this paper, outperforms FC1 in across session recognition in three out of four subjects.

In addition, we also combine together the EEG data labeled with positive emotion, namely pleasant and happy, and those with negative emotion, namely frightened and angry, to classify the positive and negative emotions in valence emotion dimension. The results are shown in Table IV. From Table IV, 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 scheme 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.

TABLE III. COMPARISON OF TWO OUT OF FOUR EMOTION RECOGNITION BETWEEN WITHIN SESSION ACCURACY AND ACROSS SESSIONS ACCURACY (%)

Subject	Within Sessions		Across Sessions	
	FC1	*FC2*	*FC1*	*FC2*
S1	82.48	80.68	62.98	60.51
S2	72.95	72.34	50.78	52.95
S3	72.92	70.14	53.24	54.04
S4	93.37	91.15	60.81	70.67

TABLE IV. COMPARISON OF POSITIVE AND NEGATIVE EMOTION RECOGNITION BETWEEN WITHIN SESSION ACCURACY AND ACROSS SESSIONS ACCURACY (%)

Subject	Within Sessions		Across Sessions	
	FC1	*FC2*	*FC1*	*FC2*
S1	85.05	80.83	65.27	66.57
S2	70.31	68.49	51.36	53.67
S3	72.81	68.02	53.47	55.14
S4	91.09	91.67	69.63	71.80

VI. APPLICATION

The proposed algorithm can be integrated with different applications for stable real-time emotion recognition with one training session. For example, an application called "Emotional Avatar" is implemented. This application enables the real-time assessment of human emotion. As the proposed algorithm is a subject-dependent one, a training session is needed. Fig. 3 shows the screenshot of the classifier training menu. During the training session, the subject is exposed to affective stimuli to evoke certain emotions and the EEG data are recorded simultaneously. After each stimulus, a self-assessment dialogue is prompted to the subject to evaluate his/her real elicited emotion. The feedback from subject includes the arousal level, dominance level, valence level, likeness level and familiarity level (all are discrete and scaled from one to nine, with one being the weakest level and nine the strongest level) and a word describing his/her emotion. Then, a SVM classifier is trained using the recorded EEG data and the emotion labels from the subject's self-assessment. After training, the emotion recognition application is ready for real-time recognition. In the real-time recognition phase, the EEG signals are collected by Emotiv device and passed to the emotion recognition algorithm where feature extraction and classification are done. As a result, an emotion label is assigned as the current emotional state of the subject. The recognized emotions of the subject from EEG are visualized and animated as the facial expressions of a 3D Haptek avatar [34]. In Fig. 4, the current recognized emotional state of the subject is pleasant, and the avatar shows a smile on her face to visualize the pleasant emotion.

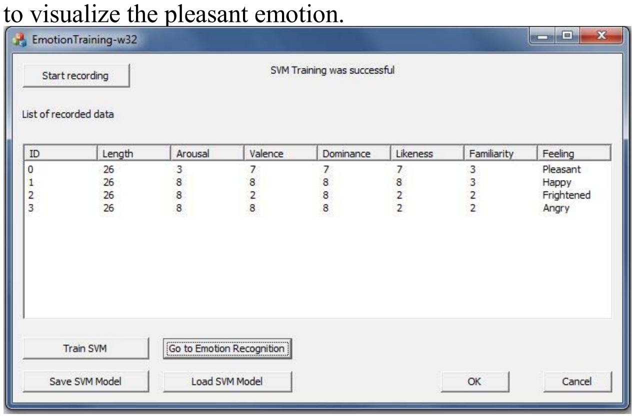

Figure 3. Screenshot of the classifier training menu.

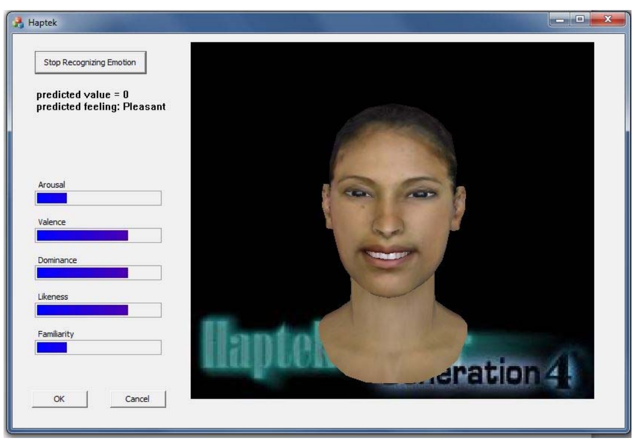

Figure 4. Screenshot of the real-time emotion recognition application.

VII. CONCLUSION

In this paper, stability of different EEG features for real-time emotion recognition was analyzed. An experiment to induce 4 emotions such as pleasant, happy, frightened and angry was designed and carried out in 8 consecutive days (two sessions per day) on 4 subjects to record EEG data. A novel real-time emotion recognition algorithm was proposed based on the most stable features such as FD, 5 statistics features, 1st order HOC and 3 band power features 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 confirmed. 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. The proposed algorithm was integrated with the real-time application "Emotional Avatar". In the future, we are planning to conduct our experiment with more subjects and establish a database for the research of stability of affective EEG signals.

ACKNOWLEDGMENT

This research was done for Fraunhofer IDM@NTU, which is funded by the National Research Foundation (NRF) and managed through the multi-agency Interactive & Digital Media Programme Office (IDMPO) hosted by the Media Development Authority of Singapore (MDA).

REFERENCES

[1] Y. Liu, O. Sourina, and M. Nguyen, "Real-time EEG-based emotion recognition and its applications," Transactions on computational science XII, 2011.

[2] O. Sourina, Y. Liu, and M. K. Nguyen, "Real-time EEG-based emotion recognition for music therapy," Journal on Multimodal User Interfaces, vol. 5, no. 1–2, pp. 27–35, Dec. 2011.

[3] T. Gasser, P. Bächer, and H. Steinberg, "Test-retest reliability of spectral parameters of the EEG.," Electroencephalography and clinical neurophysiology, vol. 60, no. 4, pp. 312–9, Apr. 1985.

[4] T. Gasser, C. Jennen-Steinmetz, and R. Verleger, "EEG coherence at rest and during a visual task in two groups of children.," Electroencephalography and clinical neurophysiology, vol. 67, no. 2, pp. 151–8, Aug. 1987.

[5] M. C. Salinsky, B. S. Oken, and L. Morehead, "Test-retest reliability in EEG frequency analysis.," Electroencephalography and clinical neurophysiology, vol. 79, no. 5, pp. 382–92, Nov. 1991.

[6] A. Kondacs and M. Szabó, "Long-term intra-individual variability of the background EEG in normals.," Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology, vol. 110, no. 10, pp. 1708–16, Oct. 1999.

[7] S. Gudmundsson, T. P. Runarsson, S. Sigurdsson, G. Eiriksdottir, and K. Johnsen, "Reliability of quantitative EEG features.," Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology, vol. 118, no. 10, pp. 2162–71, Oct. 2007.

[8] C. E. Williams and K. N. Stevens, "Emotions and speech: Some acoustical correlates." The Journal of the Acoustical Society of America 52, no. 4B, 1238-1250, 1972

[9] K. Ishino and M. Hagiwara, "A feeling estimation system using a simple electroencephalograph," in SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics, vol. 5, pp. 4204–4209.

[10] Y. P. Lin, C. H. Wang, T. L. Wu, S. K. Jeng, and J. H. Chen, "EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2009, pp. 489–492.

[11] K. Schaaff, "EEG-based Emotion Recognition," Universitat Karlsruhe (TH), 2008.

[12] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," International Journal of Human-Computer Studies, vol. 67, no. 8, pp. 607–627, Aug. 2009.

[13] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen, “EEG-based emotion recognition in music listening.,” IEEE transactions on bio-medical engineering, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.

[14] M. Murugappan, R. Nagarajan, and S. Yaacob, “Combining spatial filtering and wavelet transform for classifying human emotions using EEG Signals,” Journal of Medical and Biological Engineering, vol. 31, no. 1, pp. 45–51, 2011.

[15] M. Li and B. L. Lu. "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.

[16] Y. Liu and O. Sourina, “EEG Databases for Emotion Recognition,” in Proc. 2013 Int. Conf. on Cyberworlds, 2013, pp. 302–309.

[17] P. Rani and N. Sarkar, “An empirical study of machine learning techniques for affect recognition in human-robot interaction,” in International Conference on Intelligent Robots and Systems, 2005, pp. 2662–2667.

[18] G. Chen and R. Hou, “A New Machine Double-Layer Learning Method and Its Application in Non-Linear Time Series Forecasting,” in 2007 International Conference on Mechatronics and Automation, 2007, pp. 795–799.

[19] Y. H. Yu, P. C. Lai, L. W. Ko, C. H. Chuang, B. C. Kuo, and C. T. Lin, “An EEG-based classification system of Passenger’s motion sickness level by using feature extraction/selection technologies,” in The 2010 International Joint Conference on Neural Networks (IJCNN), 2010, pp. 1–6.

[20] W. Y Huang, X. Q. Shen, and Q. Wu, “Classify the number of EEG current sources using support vector machines,” in Proceedings. International Conference on Machine Learning and Cybernetics, 2002, vol. 4, pp. 1793–1795.

[21] M. Tangermann, I. Winkler, S. Haufe, and B. Blankertz. “Classification of artifactual ICA components.” Int J Bioelectromagnetism Volume 11, no. 2 (2009): 110-114.

[22] Y. Liu and O. Sourina, “Real-Time Fractal-Based Valence Level Recognition from EEG,” Transactions on Computational Science XVIII, vol. 7848, pp. 101–120, 2013.

[23] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: Analysis of affective physiological state,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175–1191, 2001.

[24] K. Takahashi, “Remarks on emotion recognition from multi-modal bio-potential signals,” in IEEE International Conference on Industrial Technology, 2004, vol. 3, pp. 1138–1143 Vol. 3.

[25] J. J. B. Allen, H. L. Urry, S. K. Hitt, and J. A. Coan, “The stability of resting frontal electroencephalographic asymmetry in depression.,” Psychophysiology, vol. 41, no. 2, pp. 269–80, Mar. 2004.

[26] A. J. Tomarken, R. J. Davidson, R. E. Wheeler, and L. Kinney, “Psychometric properties of resting anterior EEG asymmetry: temporal stability and internal consistency.,” Psychophysiology, vol. 29, no. 5, pp. 576–92, Sep. 1992.

[27] K. O. McGraw and S. P. Wong, “Forming inferences about some intraclass correlation coefficients.” Psychological methods 1, no. 1 pp. 30-46, 1996.

[28] Emotiv, [Online]. Available: http://www.emotive.com

[29] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. NIMH, Center for the Study of Emotion & Attention, 2005.

[30] M. M. Bradley and P. J. Lang, “The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual,” University of Florida, Gainesville, 2007.

[31] S. Sanei and J. A. Chambers, EEG Signal Processing. John Wiley & Sons, 2008, p. 312.

[32] Y. Liu and O. Sourina, “Real-time Subject-dependent EEG-based Emotion Recognition Algorithm,” Transactions on Computational Science XII, Lecture Notes in Computer Science XII, in print, 2014.

[33] C. C. Chang and C. J. Lin, “LIBSVM: a library for support vector machines,” 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[34] Haptek, [Online]. Available: http://www.haptek.com.