UNIVERSITY OF THE
WEST *of* SCOTLAND
UWS

UWS Academic Portal

On the influence of affect in EEG-based subject identification

Arnau González, Pablo; Arevalillo-Herráez, Miguel; Katsigiannis, Stamos; Ramzan, Naeem

Link to publication on the UWS Academic Portal

On the influence of affect in EEG-based subject identification

Pablo Arnau-González, Miguel Arevalillo-Herráez, Stamos Katsigiannis, and Naeem Ramzan, *Senior Member, IEEE*

Abstract—Biometric signals have been extensively used for user identification and authentication due to their inherent characteristics that are unique to each person. The variation exhibited between the brain signals (EEG) of different people makes such signals especially suitable for biometric user identification. However, the characteristics of these signals are also influenced by the user's current condition, including his/her affective state. In this paper, we analyze the significance of the affect-related component of brain signals within the subject identification context. Consistent results are obtained across three different public datasets, suggesting that the dominant component of the signal is subject-related, but the affective state also has a contribution that affects identification accuracy. Results show that identification accuracy increases when the system has been trained with EEG recordings that refer to similar affective states as the sample that is to be identified. This improvement holds independently of the features and classification algorithm used, and it is generally above 10% under a rigorous setting, when the training and validation datasets do not share data from the same recording days. This finding emphasizes the potential benefits of considering affective information in applications that require subject identification, such as user authentication.

Index Terms—EEG, affective computing, subject identification, subject recognition, user authentication

✦

1 INTRODUCTION

Electroencephalography (EEG)-based subject identification is a relatively new biometric modality, with high robustness prospects [1]. Despite the inconveniences associated with the recording method, EEG signals are extremely hard to reproduce and cannot be furtively captured at a distance [2]. This has caused an increasing research interest in this type of biometric systems.

The inter-subject and intra-subject distances between the samples are two major aspects that influence the ability of a biometric signal to uniquely and reliably recognise an individual. These distances depend on many factors, such as the acquisition protocol and the extracted features, and are closely related to the uniqueness and permanence concepts. Uniqueness means that any two subjects in the relevant population should be sufficiently different in terms of the trait, and hence appear well separated in the particular representation space. Permanence relates to the manner in which the characteristic varies over time, imposing that the trait should be sufficiently invariant over a period of time in terms of the matching criterion [3]. In other words, two captures from the same individual at different points in time should appear sufficiently close to each other in relation to the distance between samples from different subjects.

Previous works in the field of psychology suggest that there are significant differences in the way individuals feel and express emotions [4], and have demonstrated that EEG signals are a highly individual characteristic [5], [6], [7].

However, the distance between two samples from the same individual may depend on the subject's condition and the context in which the signals were captured, in between other factors [1]. For example, the stress level of the user may produce alterations in the EEG signal [8]. These intra-subject variations may be further exploited to increase the security of the system by e.g. detecting if the user is being forced to enter the password [1], but they also contribute to reduce the relative separation between samples from different individuals.

In this paper, we analyze the impact of the user's affective state on the EEG signal, in order to determine whether this is sufficiently high to significantly alter the performance of a biometric system. To this end, we analyze the topological structure of the data in three different public EEG repositories, previously used in the field of emotion recognition. We also study the performance of a classification-based subject identification approach under different settings that consider self-reported emotional labels. Results show that there is a notable difference in performance if the data identification process contemplates the actual affective state of the individual. It must be noted that in this work, we do not attempt to build a fully functioning biometric system but rather to study and determine whether an individual's affective state may affect the performance of an EEG-based biometric system.

The rest of this paper is organized in four sections. Section II describes the related literature and provides background information in the field of EEG signals and EEG-based biometrics. Then, section III describes methodological aspects, including the datasets and the feature extraction process. After, section IV explains the experimental evaluation and presents the results obtained. Finally, section V summarizes the major conclusions that can be drawn from

---

Pablo Arnau-González, Stamos Katsigiannis and Naeem Ramzan are with the School of Engineering and Computing, University of the West of Scotland (e-mails: {pablo.arnau, stamos.katsigiannis, naeem.ramzan}@aws.ac.uk .
M. Arevalillo-Herráez is with the Departament d'Informàtica, Universitat de València. Avda. de la Universidad s/n. 46100-Burjasot, Spain (e-mail: miguel.arevalillo@uv.es).

this work.

2 STATE-OF-THE-ART

EEG signals have been commonly used in medicine to diagnose a diversity of pathological conditions and disorders [9], [10], [11], [12], [13], but their use has recently been extended to other fields of research. At the beginning of this century, EEG correlates of emotions reported in a number of neuropsychological studies [14] also motivated their use in the emotion (affect) recognition field [15], [16], [17], [18]. In emotion recognition, feature selection has been proved to significantly affect the classification performance [14], [19]. This justifies that many previous works have focused on finding the set of features that best correlates with specific emotions. Features in the time domain, frequency domain, time-frequency domain and electrode combinations have extensively been used in this context [19], [15].

More recently, the large quantity of information that EEG signals encode about the subject has been exploited in biometrics [20], [21], [22], [23]. The theoretical aspects and future perspectives of EEG-biometric systems have been reviewed in [3] and [24]. Another survey, more focused on practicality and usability issues has been presented in [25]. This includes signal acquisition devices, with medical-grade sensors and low-cost devices been used in this context. The former produce more reliable signals by using a large number of electrodes for data capture, which need to be moistened by electrolytes (e.g. saline solution). The latter use a smaller number of electrodes and commonly use dry sensors to reduce cost and improve usability.

Independently from the type of recording device, the identity content carried in brain signals may critically depend upon many factors, which include the type of task in which the user is engaged. Thus, a major consideration in biometric recognition from EEG signals relates to the particular setting which is employed. In some cases, it may be appropriate to involve the user into a concrete task, or to use a predefined stimulus to evoke the desired brain activity. In this regard, several approaches have been reported in the literature, mainly covering three types of brain stimulation: resting state, sensory (audio/visual) stimuli, and cognitive tasks—(verbal instructions) [25]. The first two have predominantly been used for subject identification, where the goal is to match the identity of a subject out of a closed pool of people. Cognitive tasks have also been used for person authentication, with the aim to accept or reject an individual who claims a concrete identity.

Another line of research in EEG biometrics is related to the analysis of the small voltages generated in the brain structures in response to a specific event or stimuli. These are called event-related potentials (ERPs) [26] and are subject-dependent. Many works in this direction have focused on the study of P300 (P3) waves, a waveform that can be detected within 250 and 500 ms after a visual stimulus. High subject identification rates have also been obtained under this setting. For example, a recognition accuracy of 95% was initially reported in [27] and later improved to a peak classification accuracy of 98.12% in [28]; whereas 100% identification accuracy in a pool of 50 users was reported in [21].

Some approaches in the literature have also used brain activity for subject identification and/or authentication purposes [29], [30], [31]. In this case, the EEG signal is acquired while the subject performs some kind of intentional task, such as mathematical calculations or a concrete physical movement. Imagined speech was used for user identification in [31], reaching 99.76% accuracy out of 6 different subjects. A 99% authentication accuracy has been reported in [30] using single-channel signals from consumer-grade EEG sensor technology, by choosing custom tasks and custom acceptance thresholds for each subject. More recently, and with the aim to ease acquisition and better fulfill the collectability requirement, the use of a wearable in-ear EEG sensor has been proposed [32]. Results using Power Spectral Density (PSD) features and Support Vector Machines (SVM) with different kernels showed an average accuracy of 95.7% at a user verification task, without mixing the training and validation data from the same recording days.

A current topic of interest refers to the study of the longitudinal behavior of EEG signals, to analyze their permanence across time and the influence of aging effects on the discriminative capabilities of EEG signals over long-term periods. In [33], authors used event-related potential (ERP) biometrics, and addressed permanence by asking a subset of the participants to return to the lab between a week and six months after their first session. They concluded that the features used were stable over time, on the basis of a continued accurate identification of individuals from the ERPs when they returned. Also a good level of stability was reported in [34], from the analysis of the performance achieved with several EEG-based biometric system configurations on a database that contained EEG recordings from 50 users, which were acquired along three distinct sessions spanning a period of one month and a half, under eyes-closed and eyes-open resting conditions. A more rigorous study over a considerably longer time span has been recently presented in [35]. In this case, EEG signals from 45 subjects were recorded in 5 to 6 sessions with an interval between sessions no shorter than a month, covering an overall period of about 3 years. This data was used to provide evidence that aging actually affects EEG biometric traits, and propose several strategies to mitigate aging effects and achieve Equal Error Rates (EERs) below 2% when comparing samples taken at temporal distances in the order of years.

Despite the extensive literature on the use of EEG signals in biometrics and emotion recognition, the potential effect of the user's emotional state in biometric identification tasks has received little attention. Correlates of emotions provide evidence that the subject's affective state has an influence in the EEG signal. This allows exploiting specific emotional responses to concrete events in order to characterize users and help subject identification or authentication. However, affect may also have a negative effect and hinder the user identification task. When a subject interacts with a biometric system, the individual is in an emotional state which is generally unknown and most biometric systems have disregarded. This emotional state may vary between subsequent operations, introducing a form of noise that may obstruct pattern matching. In this paper, we analyze the implications of this effect, to determine whether the influence of the emo-

Table 1: Overview of the datasets used in this study

Dataset	Number of participants	Number of videos	Number of sessions	Video content	Video duration	Recording device	Number of channels	Sampling frequency	Number of features
DEAP [36]	32	40	1	Music videos	60 s	Biosemi Active II	32	512 Hz ($\downarrow$ to 128 Hz)	230
MANHOB [37]	27	20	1	Excerpts from movies	34.9-117 s ($\mu = 81$ s)	Biosemi Active II	32	512 Hz ($\downarrow$ to 128 Hz)	230
SEED [38]	15	15	3	Excerpts from movies	240 s	ESI NeuroScan	62 (32 used)	1000 Hz ($\downarrow$ to 200 Hz)	230

tion on the EEG signal can potentially cause a significant decrease of performance in a typical biometric system for subject identification.

3 EXPERIMENTAL METHODOLOGY

3.1 Datasets

The growing interest in emotion recognition through EEG signals has led to the development of public datasets specifically aimed to the analysis of human affective states. We have chosen three of the most widely used datasets in the field to carry out our study, namely DEAP [36], MAHNOB-HCI [37], and SEED [38]. These datasets contain EEG recordings acquired from subjects exposed to video stimuli, specifically selected in order to elicit emotions in each of the 4 quadrants of Russell's Circumplex Model [39]. In DEAP and MAHNOB-HCI, the stimuli were presented in random order to each participant and emotional labels for valence and arousal were assigned according to a self-report using self-assessment manikins (SAM) [40]. In SEED, however, the recordings were annotated according to the emotional response that the stimulus was expected to elicit.

DEAP [36] contains EEG data for 32 subjects and the video stimulus consisted of a set of 40 music videos of 63 seconds each. EEG signals were recorded at a sampling rate of 512 Hz using the Biosemi Active II system, with 32 active $AgCl$ electrodes placed according to the international 10-20 system. EEG data were common average referenced, down-sampled to 128 Hz, and high-pass filtered with a 2 Hz cutoff frequency. MANHOB-HCI [37] contains data from 27 subjects exposed to video stimuli consisting of a set of 20 film clips. EEG signals were recorded by using the same equipment as in DEAP.

The SEED dataset contains EEG recordings from 15 different individuals exposed to audiovisual stimuli during three separate sessions. EEG signals were captured using an ESI NeuroScan System with a sampling rate of 1 000 Hz and a 62-channel electrode cap, and were down-sampled to 200 Hz for further processing. The presented stimuli consisted of 15 film clips designed to elicit specific emotions as follows: 5 for positive, 5 for negative, and 5 for neutral emotion. Each of the three sessions replicated exactly the same experimental procedure, using the same video sequences in the same order [38]. It must be noted that the recorded signals in the SEED dataset are not annotated in terms of a continuous Arousal/Valence scale, thus the samples of the SEED dataset were annotated with the emotional labels of the associated video stimuli. An overview of the three datasets used in this work is given in Table 1.

3.2 Preprocessing and Feature Extraction

In all datasets, the subject label for the identification task was set by using the identifier assigned to each subject. This is an integer between 1 and 32 in DEAP, in the range 1 to 27 in MAHNOB and in between 1 and 15 in SEED. However, there were relevant differences about the labeling used to encode emotional responses in the three datasets. In SEED, three different values were used, namely -1 (negative), 0 (neutral) and 1 (positive), according to the emotion that the stimulus was expected to elicit. However, in DEAP and MAHNOB, they were self-reported in a continuous scale between 1 and 9, for each dimension in the Russell's circumplex model. In order to unify the experiments across all different databases and yet maintain consistency with other previous works in the literature that use DEAP and MANHOB, e.g. [16], [36], [41], [42], the arousal and valence labels in these two datasets were discretized into two levels by applying a threshold to the originally reported values. The threshold was set to 5, in order to be coherent with other previous studies [16], [36], [41], [42]. For consistency with the DEAP and MAHNOB datasets, the samples of the SEED dataset that referred to neutral emotion were discarded, and only those referring to positive or negative emotion were used in this study.

With regard to feature extraction, DEAP and MANHOB only use a subset of 32 channels out of the 62 used in SEED. To also make the data uniform across the three databases, we discarded the information from the additional 30 channels that were included in SEED but not in the other two datasets. Then, we computed the *Power Spectral Density* (PSD) in each of the 32 channels. For consistency reasons and to ease the future comparison of results, this was done as in the original publications [36], [37], [38], down-sampling the signals to 128 Hz in the case of MANHOB, and using Welch's method with a Hamming window of 1 sec, with 50% overlapping. The spectral power was averaged over the θ (4-8 Hz), $\tilde{\alpha}$ (8-10 Hz), α (8-12 Hz), β (12-30 Hz), and γ (30+ Hz) bands for all electrodes. In addition, the difference between the spectral power of all the symmetrical pairs of electrodes on the right and left hemisphere in the same bands was also computed in order to measure the possible asymmetry in the brain activities due to emotional stimuli. This yielded a vector of 230 features per recording (32 electrodes $\times$ 5 bands + 14 pairs $\times$ 5 bands).

As a result, after data preparation, each recording in DEAP and MANHOB had 230 features and three labels associated with it, namely: the user identifier, the arousal (high/low) and the valence (negative/positive). In SEED, we used the same features, but the labels were the user identifier, the session identifier (1 to 3) and the polarity of the emotion (positive/negative).

3.3 Experimental Protocols

It should be noted that we do not attempt to build an EEG-based biometric system, which has been the subject of many other previous works in the literature [43], [44], [45]. Rather, we aim to determine whether the user's affective state may alter the user identification performance of an EEG-based biometric system. Hence, the evaluation of the identification performance in comparison to the state-of-the-art lies outside the scope of this research.

In order to analyze the effect of emotions in a subject identification context, we conducted a series of studies that are consistent with the purpose of the research. In all cases, we considered a set of samples composed of the EEG features extracted for each entire recording and the label of the corresponding individual as provided in each dataset.

The first study was a qualitative analysis of the data topology, aimed to assess the influence of the emotion on the distance between data samples. In it, we first show the suitability of the EEG signals for biometric subject identification, by analyzing intra- and inter-subject distances between data samples. Then, we independently evaluate the samples from each subject, to determine whether the label has an implicit influence on the distance between samples.

The second study aims to evaluate whether the evidence found in the first study is sufficient to alter the classification results and significantly affect the subject identification accuracy. To this end, we use two substantially different protocols, which are described next.

3.3.1 Protocol P1: Biased scenario

Many previous studies on EEG biometric identification have claimed high accuracy results [46] by using single session datasets, e.g. [47], or by randomly selecting training and validation samples regardless of the data acquisition days [48]. This was called the biased scenario in [32], and can be affected by session-specific exogenous conditions (e.g. capacitative coupling of electrodes and cables with other devices, induction loops created between the employed equipment and the body, power supply artifacts) [35] or other noise-dependent features related to contamination artefacts from subjects' movements (e.g. eye blinks, chewing) [32]. This approach benefits from the recording-day-dependent EEG characteristic in the classification, it is hence biased in favour of high classification rate and it cannot evaluate the repeatability/reproducibility of the application [32]. Still, recordings for the DEAP and MAHNOB datasets were acquired during a single session per subject and do not allow for a cross-session-validation study.

For this biased scenario, we created 20 000 sub-sets of samples for each affective label available in all datasets (DEAP-Arousal, DEAP-Valence, MAHNOB-Arousal, MAHNOB-Valence, SEED-Polarity), by randomly

Table 2: Rejected subjects, minimum number of training samples per subject, and number of samples per sub-set for each dataset and emotional dimension

Dataset	Rejected subjects	M	n
DEAP-Arousal	None	32	10
DEAP-Valence	3, 12, 13, 20	28	10
MAHNOB-Arousal	3, 5, 7, 9, 10, 11, 13, 14, 15, 17, 18, 20, 21, 25, 26	12	8
MAHNOB-Valence	1, 3, 8, 9, 10, 13, 14, 15, 17, 18, 21, 25, 26, 27	13	8
SEED-Polarity	None	15	15

selecting exactly one sample per subject with a given label. 10 000 of these sub-sets contained only samples labeled as "positive" (or "high" for arousal) and the remaining 10 000 contained samples labeled as "negative" (or "low" for arousal). Each of these sub-sets was used as a test set, in order to compare various classification models at the task of subject identification under two different settings. In the *SAME* setting, the corresponding training set was composed of the remaining samples in the dataset that shared the same label. In the *DIFF* setting, the training set contained all the remaining samples from the same database with a different label.

Furthermore, in order to be able to do a fair paired comparison between the classification performance for the two approaches (*SAME* and *DIFF*), subjects that did not have at least n samples for each emotional label were discarded, ensuring that at least $n - 1$ samples for each subject and emotional label were included in the training set, plus 1 sample in the test set. The number of samples n was manually selected for each dataset, so as to reach a reasonable balance between the amount of rejected subjects and the size of the training set ($n_{DEAP} = 10$, $n_{MAHNOB} = 8$, and $n_{SEED} = 15$). Furthermore, n remained the same at each dataset for both the *SAME* and *DIFF* approaches. As a result, the final number of samples, M, in each of the 20 000 sub-sets for each dataset and emotional dimension was adjusted depending on the number of rejected subjects in each case. The identifiers of the rejected subjects, the minimum number of samples per subject n, and the number of samples M for each combination of emotional dimension and dataset can be found in Table 2.

In order to ensure that detected differences in performance do not occur due to the particularities of a specific classification algorithm, the aforementioned experiments were conducted using seven different classifiers: Support Vector Machines with Linear and Radial kernels (LSVM and RSVM, respectively), k-Nearest Neighbor ($k = 1, 3, 5$), Multilayer Perceptron (MLP) with one hidden layer of 10 hidden units using cross-entropy as the cost function and stochastic gradient descent for training, and the AdaBoost ensemble method using decision trees as the weak learner [49].
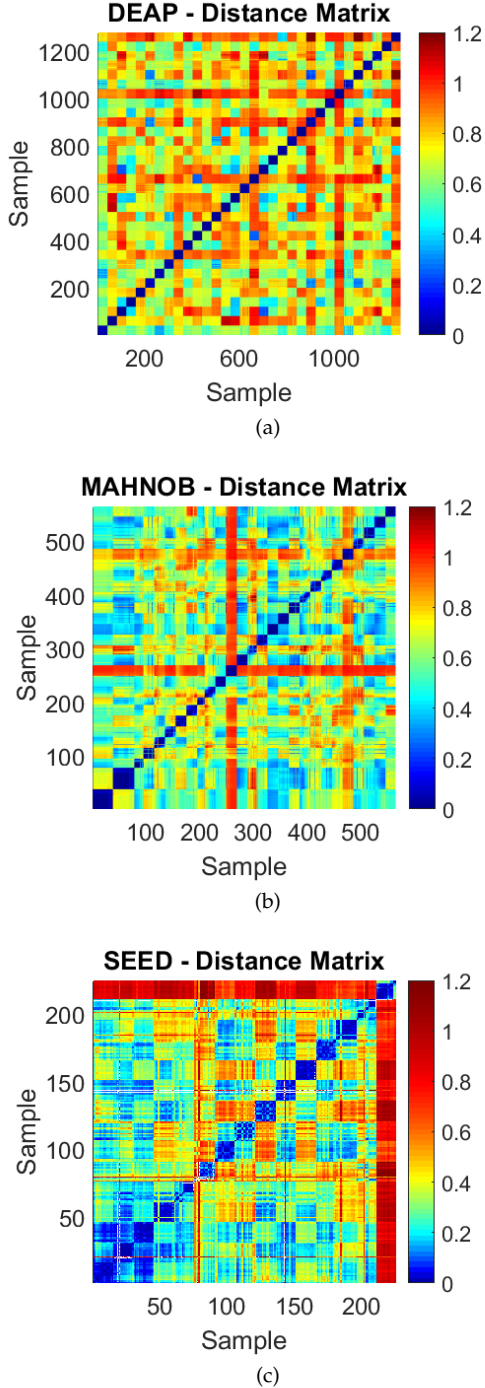
(a)

(b)

(c)

Figure 1: Distance matrices between all the samples in (a) DEAP, (b) MAHNOB, and (c) a representative session in SEED. The matrices contain the value of the cosine distances encoded as a color according to the legend shown at the right-hand side. Please, note that the first two subjects in MAHNOB have almost double the samples than the rest

3.3.2   Protocol P2: Rigorous scenario

The importance of using data recorded across multiple sessions per user to validate the robustness of an EEG-based biometrics application has been remarked in a number of previous works [32], [35]. It is widely accepted that a sound experimental setting for biometric purposes should include recordings for each user over multiple days. In addition, the experimental methodology should ensure that the training and validation datasets do not share data from the same recording days. This has been called the rigorous scenario in [32], and has been used in multiple longitudinal studies that evaluated EEG signals as biometric identifiers [29], [33], [34]. In our work, this setting could only be attempted in SEED, as this is the only database of the ones considered that contains labeled recordings from multiple separate sessions per subject.

A similar experimental procedure as in Protocol P1 was followed. The 20 000 test sub-sets previously created for the SEED dataset were used in order to repeat the *SAME* and *DIFF* experiments as previously described, with the additional constraint that the training sets for the classification models could only contain samples referring to a different session than the ones contained in the test set. Furthermore, apart from the PSD features, we also examined the performance of the *Mel Frequency Cepstral Coefficients* (MFCC) features and the *Auto Regression Reflection Coefficients* (ARRC) features, as proposed by Piciucco et al. [45] for EEG-based subject identification. As also proposed in [45], the samples were divided into frames with a length of 5 *sec* and 75% overlapping, leading to a total of 46 new samples generated from each of the original recordings in the dataset. Then, the 20 000 sub-sets were computed again from the newly created samples of the SEED dataset and the *SAME* and *DIFF* experiments were conducted again for the MFCC and the ARRC features.

4   EXPERIMENTAL RESULTS

4.1   Distance based analysis

Along with other properties, uniqueness is essential to any biometric recognition system. This translates into the necessity that samples that belong to the same individual appear close to each other, according to a concrete distance and in relation to samples coming from different subjects. To test if this condition holds for our data, we have computed the cosine distance (D_C) between all pairs of samples in each dataset, in line with other research works that have remarked on the benefits of this distance in specific contexts [50], [51]. This has been computed as:

$$D_C = 1 - S_C \tag{1}$$

where S_C is the cosine similarity between two feature vectors $\boldsymbol{F_a}$ and $\boldsymbol{F_b}$ and is computed as:

$$S_C = \cos(\theta_{F_a,F_b}) = \frac{\boldsymbol{F_a} \cdot \boldsymbol{F_b}}{||\boldsymbol{F_a}||_2 \, ||\boldsymbol{F_b}||_2} \tag{2}$$

In order to visualize the data in a form that can easily be interpreted, we first sorted the samples in the datasets by subject identifier, so that all elements for the same subject appeared in a consecutive order. Then, the distance matrix
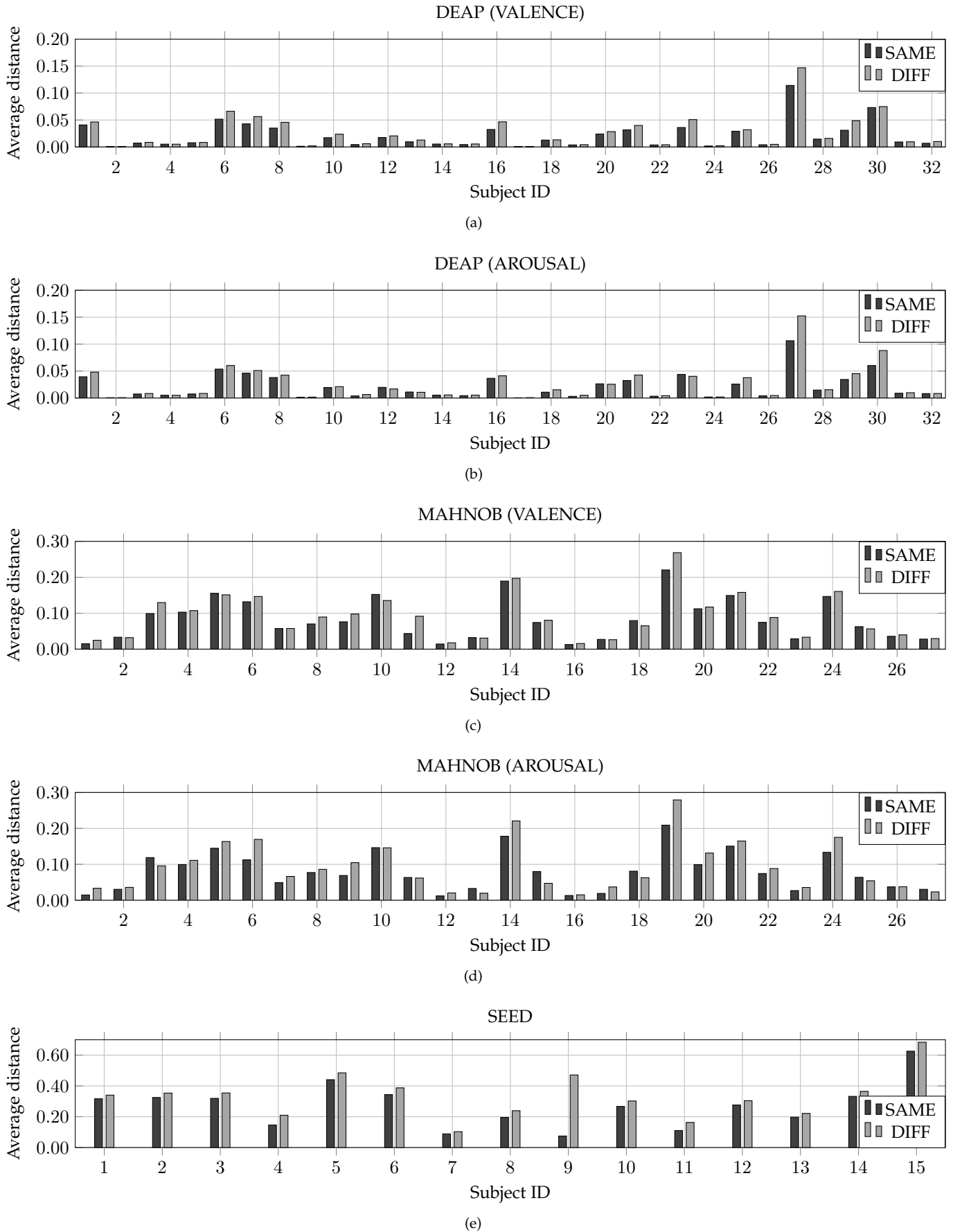
Figure 2: Bar plots containing the *per-subject* difference of distance between samples with the same emotional state (SAME) and with different emotional state (DIFF) for (a) DEAP - Valence, (b) DEAP - Arousal, (c) MAHNOB - Valence, (d) MAHNOB - Arousal, and (e) SEED.

Table 3: Classification accuracy for each dataset, emotional dimension, and setting

| | DEAP | | | | MAHNOB | | | | SEED | |
| | VALENCE | | AROUSAL | | VALENCE | | AROUSAL | | POLARITY | |
	SAME	DIFF	SAME	DIFF	SAME	DIFF	SAME	DIFF	SAME	DIFF
RSVM	**0.9111**	0.8917	**0.8994**	0.8922	**0.7176**	0.6825	**0.7710**	0.6176	**0.7958**	0.5899
LSVM	**0.9897**	0.9863	**0.9813**	0.9718	**0.9589**†	0.9418	**0.9796**†	0.8661	**0.9475**†	0.8572
1-NN	**0.9915**†	0.9896	**0.9896**†	0.9852	**0.9196**	0.8688	**0.9309**	0.8603	**0.6916**	0.6397
3-NN	**0.9805**	0.9733	**0.9808**	0.9718	**0.8805**	0.8383	**0.8766**	0.8475	**0.5780**	0.5593
5-NN	**0.9639**	0.9567	**0.9697**	0.9591	**0.8613**	0.8191	**0.8419**	0.8284	**0.5525**	0.5454
AdaBOOST	**0.9773**	0.9711	**0.9715**	0.9671	**0.8911**	0.8331	**0.9010**	0.8110	**0.9200**	0.7887
MLP	**0.7704**	0.7560	**0.7382**	0.7340	0.6876	**0.7100**	**0.7156**	0.6588	**0.4158**	0.3520

† indicates the highest classification accuracy achieved for the respective dataset and emotional dimension

was plotted in a graphic form by encoding distance values using a colour scale from dark blue to dark red, with dark blue denoting a distance close to 0 and dark red the maximum distance. Results are shown in Figures 1a, 1b and 1c, for DEAP, MANHOB, and a representative session in SEED, respectively.

The visual structure of the result indicates that the lowest distances (in dark blue) appear in the diagonal, which correspond to samples that belong to the same subject.

Nevertheless, the results reported in Figures 1a and 1b, which refer to data taken during a relatively short period of time, only indicate that the uniqueness property holds in this case, and hence that the features extracted are potentially adequate for subject identification. To study the potential influence of the emotion in this context, we need to determine whether the fact that two samples from the same subject share the same emotional label has an impact on the distance value.

To this end, we have studied each subject on an individual basis, by computing the average distance between samples that share the same affective label, and comparing it to the average distance between data samples with a different label. Figure 2(a-d) shows the results for the two affective dimensions (valence and arousal) in the DEAP and MAHNOB datasets, while results for the SEED dataset are provided in Figure 2e. Despite some individual exceptions (subject 23 in DEAP-Arousal and subject 18 in MAHNOB-both dimensions), the bar plots presented in these figures reveal that distances between samples that were labeled with the same emotion are generally lower than the rest.

4.2 Results on biased scenario

The lower distances obtained in the same label case suggest an improvement of the uniqueness property if all samples share the same affective label, and highlight the potential benefits of knowing the subject's emotional state when recordings are taken. In order to analyze whether the distance reduction achieved is sufficient to yield a higher identification accuracy, we have carried out a second experiment. In order to evaluate the effect of the affective label on the subject identification accuracy, this time we used a classification setting, according to the protocol P1 described in section 3.3.1. The same experiment was repeated for valence and arousal in DEAP and MANHOB, and for the expected polarity of the emotion in SEED.

The results of these experiments are reported in Table 3. In all cases, the accuracy obtained in the *SAME* case was consistently better than in the *DIFF* case. In addition, the high performance in terms of the mean classification accuracy achieved with the Linear SVM classifier across the three repositories indicates that uniqueness is naturally present and supports the suitability of the feature set used in this case. This is also confirmed by the high performance of the 1-NN which suggests that data can be linearly separated to a reasonable level in the original space and there is no need to map it to a higher dimensional one. The better performance of the linear SVM over the Radial kernel version may be due to a less accurate computation of the optimal SVM parameters in the non-linear case, derived from the use of a relatively small training set in a high dimensional space of 230 features. This would also justify the poorer performance of all algorithms in MAHNOB, as less samples per subject were used for training. We shall also remark on the generally lower accuracy obtained in SEED. This is very likely because SEED is the only repository of the ones we have used that stores data captured over different sessions, and hence the uniqueness and permanence properties are simultaneously tested. It is also worth noting that the highest accuracy achieved for the DEAP dataset in all the examined scenarios (*SAME/DIFF*, Valence/Arousal), as shown in Table 3, is consistent with the subject identification accuracy recently reported by del Pozo-Panos et al. [20] for DEAP using PSD-based features (97.97%).

As a further analysis of the results, the classification accuracies for the best performing classifier in the *SAME* and *DIFF* cases are plotted in Figures 3, 4, and 5 in the form of box plots. This plots show that the classification accuracy achieved in the *SAME* case also has a smaller variance compared to *DIFF*, regardless of the dataset and the emotional dimension (arousal or valence).

To test the statistical significance of the results presented above, we have run a hypothesis contrast test, under the null hypothesis H_0 that the distribution defined by $x - y$ (where x and y are the classifier-wise distributions for the *SAME* and *DIFF* cases, respectively) comes from a distribution whose mean (or median, in case of non-parametric tests) is 0, against the alternate hypothesis H_a that the distribution defined by $x - y$ actually has a mean (or median) with a value other than 0, meaning that there are significant differences between both distributions. Additionally, for testing the data
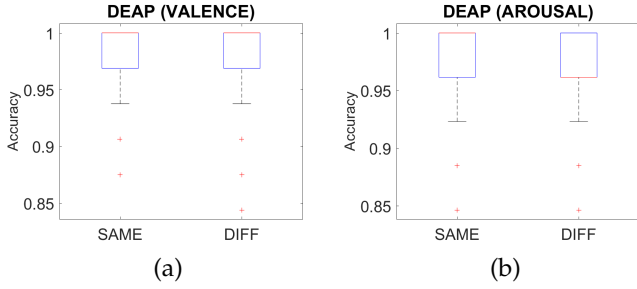
Figure 3: Box plots comparing differences between *SAME* and *DIFF* when using a Linear SVM in DEAP: (a) for arousal; and (b) for valence.
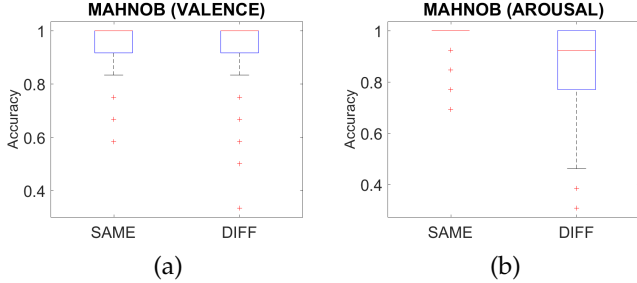

Figure 4: Box plots comparing differences between *SAME* and *DIFF* when using a Linear SVM in MAHNOB: (a) for arousal; and (b) for valence.
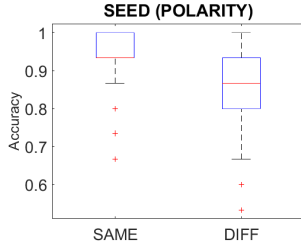

Figure 5: Box plot comparing differences between *SAME* and *DIFF* when using a Linear SVM in SEED

for normality, a Lilliefors test was done, rejecting the null hypothesis that the data comes from a distribution from the normal family ($p \ll 0.05$). Wilcoxon's signed rank test was chosen for testing these hypotheses due to being non-parametric, thus eliminating the assumption of normality inherent to Student's t-test. The results of the Wilcoxon's test show that there are significant differences between *SAME* and *DIFF* for all the five cases examined, independently of the classification algorithm used. All p-values were below 10^{-7}, providing strong evidence that there is a statistically significant gain in terms of identification performance when the training samples share the same emotional label.

4.3 Results on rigorous scenario

In order to complete our study on the influence of the affective state on the identification accuracy, we have exploited the multiple session structure of the SEED database to run a cross-session analysis, according to the experimental Protocol P2. In addition, we have computed other
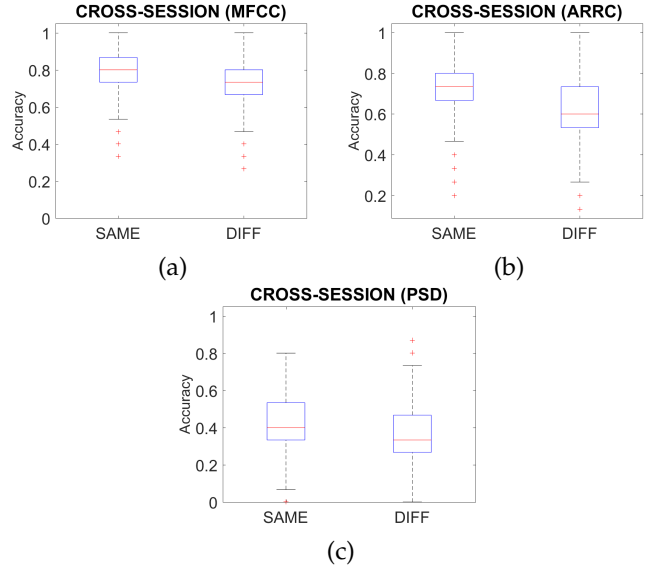

Figure 6: Box plots comparing differences between *SAME* and *DIFF* when using a Linear SVM in SEED: (a) for MFCC features; (b) for ARRC features; and (c) for PSD features.

Table 4: Cross-session classification accuracy for SEED

	MFCC		ARRC		PSD	
	SAME	DIFF	SAME	DIFF	SAME	DIFF
RSVM	**0.7309**	0.6292	**0.6467**	0.5745	**0.1786**	0.1487
LSVM	**0.7934**[†]	0.7108	**0.7079**[†]	0.6210	**0.4224**	0.3614
1-NN	**0.4925**	0.433	**0.3337**	0.2765	**0.2338**	0.2070
3-NN	**0.4953**	0.4281	**0.3513**	0.2905	**0.2271**	0.1946
5-NN	**0.4961**	0.4292	**0.3578**	0.2964	**0.2343**	0.2054
AdaBoost	**0.5112**	0.4321	**0.5017**	0.3861	**0.4601**[†]	0.3924
MLP	**0.6294**	0.5348	**0.5656**	0.4704	**0.2338**	0.1835

[†] indicates the highest classification accuracy for each of the features used.

alternative features (see section 3.3.2) and repeated the experimentation, to further support the consistency of our results and also ensure that the differences found are not due to particularities of the PSD features.

Table 4 reports the accuracy results obtained when the training and validation datasets do not share data from the same session. Results reported in Tables 3 and 4 reveal that accuracy results are considerably lower when using the rigorous scenario, as also highlighted in the box plots of Figure 6. This finding is consistent with the bias reported in [32] that leads to unrealistically high results when using protocol P1, and also with the phenomenon of *Template Aging* which effects when using recordings of cognitive tasks performed with long time intervals between them, as discussed in [25], [52], and [53]. However, results from both cases support this work's argument that the emotional state of the individual affects identification accuracy. The identification performance is clearly and consistently better in the *SAME* case, regardless of the features, classifier or concrete accuracy achieved. The best results are obtained when using MFCC features. With these features, the SVM behaves the best, reaching accuracies of 0.79 and 0.73 when using a linear and a radial kernel, respectively. In the case of the ARRC features, results are generally lower, but the SVMs also behave as the best classification algorithms, with

also reasonable accuracies of 0.71 for a linear kernel and 0.65 for a radial one. The PSD features behave the worst, but still achieve a 0.46 identification accuracy when using AdaBoost. Accuracy values under the *SAME* setting reported in Table 4 are from 11.6% to 29.9% higher than in the *DIFF* case.

As in the biased scenario, a paired Wilcoxon's signed rank test between the results for *SAME* and *DIFF* was also conducted in order to examine the statistical significance of the acquired results in the cross-session study for all the features and classification algorithms used, leading to $p \ll 0.001$ in all cases and supporting the statistical significance of the results.

The observation of the same trend for both the single-session and cross-session studies, as well as for different types of EEG features, indicates that the effect of emotion in the performance of EEG-based subject identification transcends the nature of the features used and highlights the importance of considering the emotional state of the subjects in EEG-based biometric applications.

5 CONCLUSIONS

In this work, we have studied how the difference in the emotional state affects the classification performance of EEG-based subject identification systems. For this purpose, three publicly available datasets containing EEG recordings, acquired under stimuli selected to elicit specific emotional responses, have been examined for the task of subject identification. Initial observational results suggested that there was a dependency linking the distance values between samples and the emotional labels associated with them. In particular, feature vectors from a subject that shared the same label were generally closer to each other. This finding led to a more thorough study of this phenomenon, by designing further experiments to study potential implications in a subject identification setting, with regard to the accuracy achieved.

The new experiments showed a significant difference in performance when the emotional labels in the training and test sets match, compared to when they differ. A further cross-session study performed over the SEED dataset revealed that these differences in performance hold when data is captured across different sessions, and also when different sets of features are used. These results lead to the conclusion that the user's affective state plays a determinant role that influences the performance of EEG-based identification systems, and indicate that the integration of affective data would help improve the performance and reliability of such systems.

Possible ways to make explicit use of the research presented in this paper in a real setting would be a) by using stimuli to elicit a particular emotion both while training the biometric system and when identifying a user; or b) by asking the user to report his/her emotional state during measurements. Furthermore, subject identification accuracy could easily be improved by training multiple systems, one per emotion, and fusing the biometric responses. Nevertheless, there are still intrinsic difficulties to all these approaches that require further investigation. These include both technical and usability-related issues, such as the implementation of a reliable and personalized procedure to consistently elicit the same emotion during training and test; or the selection of an appropriate self-reporting mechanism to seamlessly capture the emotion in a non-intrusive way.

We plan to extend the current research to study under which conditions the results reported can be generalised to other feature sets in typical classification settings. In addition, we shall study the potential further benefits obtained by combining the classification results computed for several emotional labels, e.g. arousal and valence. This may lead to the use of data fusion methods that yield more consistent, accurate, and useful results than those reported in Table 3 when using a single label.

REFERENCES

[1] M. del Pozo-Banos, J. B. Alonso, J. R. Ticay-Rivas, and C. M. Travieso, "Electroencephalogram subject identification: A review," *Expert Systems with Applications*, vol. 41, no. 15, pp. 6537–6554, 2014.

[2] K. Revett, "Cognitive biometrics: A novel approach to person authentication," *International Journal of Cognitive Biometrics*, vol. 1, no. 1, pp. 1–9, 2012.

[3] P. Campisi and D. La Rocca, "Brain waves for automatic biometric-based user recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 782–800, 2014.

[4] J. J. Gross and O. P. John, "Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior," *Journal of Personality and Social Psychology*, vol. 72, no. 2, pp. 435–448, 1997.

[5] J. Berkhout and D. O. Walter, "Temporal stability and individual differences in the human EEG: An analysis of variance of spectral values," *IEEE Transactions on Biomedical Engineering*, vol. BME-15, no. 3, pp. 165–168, 1968.

[6] A. Keil, M. Stolarova, S. Heim, T. Gruber, and M. M. Müller, "Temporal stability of high-frequency brain oscillations in the human EEG," *Brain topography*, vol. 16, no. 2, pp. 101–110, 2003.

[7] R. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles, "The electroencephalogram as a biometric," in *Electrical and Computer Engineering, 2001. Canadian Conference on*, vol. 2. IEEE, 2001, pp. 1363–1366.

[8] J.-S. Kang, G. Jang, and M. Lee, "Stress status classification based on EEG signals," *The Journal of The Institute of Internet, Broadcasting and Communication*, vol. 16, no. 3, pp. 103–108, 2016.

[9] S. Patidar, R. B. Pachori, A. Upadhyay, and U. R. Acharya, "An integrated alcoholic index using tunable-Q wavelet transform based features extracted from EEG signals for diagnosis of alcoholism," *Applied Soft Computing*, vol. 50, pp. 71–78, 2017.

[10] A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi, "Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 4, pp. 888–896, 2017.

[11] M. Light, T. Casimire, C. Chua, V. Koushyk, O. Burschtin, I. A. Ayappa, and D. M. Rapoport, "Addition of frontal EEG to home sleep apnea testing to diagnose adult obstructive sleep apnea: Does a more accurate determination of sleep time make a difference?" in *A80-C. NOVEL DIAGNOSTIC APPROACHES TO SDB*. Am Thoracic Soc, 2017, pp. A2607–A2607.

[12] A. Yadollahpour, M. Mirzaiyan, and S. Rashidi, "Quantitative EEG for early and differential diagnosis of bipolar disorders: a comprehensive review of the literature," *International Journal of Mental Health and Addiction*, vol. 15, no. 2, pp. 387–393, 2017.

[13] E. Grossi, C. Olivieri, and M. Buscema, "Diagnosis of autism through EEG processed by advanced computational algorithms: A pilot study," *Computer Methods and Programs in Biomedicine*, vol. 142, pp. 73–79, 2017.

[14] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, "A review on the computational methods for emotional state estimation from the human EEG," *Comp. Math. Methods in Medicine*, vol. 2013, pp. 573 734:1–573 734:13, 2013.

[15] X. Hu, J. Yu, M. Song, C. Yu, F. Wang, P. Sun, D. Wang, and D. Zhang, "EEG correlates of ten positive emotions," *Frontiers in Human Neuroscience*, vol. 11, 2017.

[16] P. Arnau-González, M. Arevalillo-Herráez, and N. Ramzan, "Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals," *Neurocomputing*, vol. 244, pp. 81 – 89, 2017.

[17] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, Jan 2018.

[18] P. Arnau-González, N. Ramzan, and M. Arevalillo-Herráez, "A method to identify affect levels from EEG signals using two dimensional emotional models," in *The 30th European Simulation and Modelling Conference-ESM'2016*, 2016.

[19] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, July 2014.

[20] M. del Pozo-Banos, C. M. Travieso, J. B. Alonso, and A. John, "Evidence of a task-independent neural signature in the spectral shape of the electroencephalogram," *International Journal of Neural Systems*, vol. 28, no. 01, p. 1750035, 2018.

[21] M. V. Ruiz-Blondet, Z. Jin, and S. Laszlo, "CEREBRE: A novel method for very high accuracy event-related potential biometric identification," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 7, pp. 1618–1629, 2016.

[22] B. C. Armstrong, M. V. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, Z. Jin, and S. Laszlo, "Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics," *Neurocomputing*, vol. 166, pp. 59–67, 2015.

[23] K. P. Thomas, A. P. Vinod, and N. Robinson, "Online biometric authentication using subject-specific band power features of EEG," in *Proceedings of the 2017 International Conference on Cryptography, Security and Privacy*. ACM, 2017, pp. 136–141.

[24] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, "State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals," *IET Biometrics*, vol. 4, no. 3, pp. 179–190, 2015.

[25] S. Yang and F. Deravi, "On the usability of electroencephalographic signals for biometric recognition: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 958 – 969, 2017.

[26] S. J. Luck, *An introduction to the event-related potential technique*. MIT press, 2014.

[27] R. Palaniappan and P. Raveendran, "Individual identification technique using visual evoked potential signals," *Electronics Letters*, vol. 38, no. 25, pp. 1634–1635, 2002.

[28] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: A machine learning approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 738–742, 2007.

[29] S. Marcel and J. D. R. Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 743–752, April 2007.

[30] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I think, therefore I am: Usability and security of authentication using brainwaves," in *International Conference on Financial Cryptography and Data Security*. Springer, 2013, pp. 1–16.

[31] K. Brigham and B. V. K. V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sept 2010, pp. 1–8.

[32] T. Nakamura, V. Goverdovsky, and D. P. Mandic, "In-ear EEG biometrics for feasible and readily collectable real-world person authentication," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 648–661, 2018.

[33] B. C. Armstrong, M. V. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, Z. Jin, and S. Laszlo, "Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics," *Neurocomputing*, vol. 166, pp. 59 – 67, 2015.

[34] E. Maiorana, D. L. Rocca, and P. Campisi, "On the permanence of EEG signals for biometric recognition," *IEEE Transactions on*

[35] E. Maiorana and P. Campisi, "Longitudinal evaluation of EEG-based biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1123–1138, 2018.

[36] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *Affective Computing, IEEE Transactions on*, vol. 3, no. 1, pp. 18–31, Jan 2012.

[37] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, Jan 2012.

[38] W. L. Zheng and B. L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, Sept 2015.

[39] J. A. Russell, "Affective space is bipolar," *Journal of Personality and Social Psychology*, vol. 37, no. 3, pp. 345–356, 1979.

[40] J. D. Morris, "Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response," *Journal of Advertising Research*, vol. 35, no. 6, pp. 63–68, 1995.

[41] S. Liu, D. Zhang, M. Xu, H. Qi, F. He, X. Zhao, P. Zhou, L. Zhang, and D. Ming, "Randomly dividing homologous samples leads to overinflated accuracies for emotion recognition," *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 96, no. 1, p. 29—37, April 2015.

[42] P. Petrantonakis and L. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, March 2010.

[43] R. Das, E. Maiorana, and P. Campisi, "EEG biometrics using visual stimuli: a longitudinal study," *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 341–345, 2016.

[44] M. Fraschini, A. Hillebrand, M. Demuru, L. Didaci, and G. L. Marcialis, "An EEG-based biometric system using eigenvector centrality in resting state brain networks," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 666–670, 2015.

[45] E. Piciucco, E. Maiorana, O. Falzon, K. P. Camilleri, and P. Campisi, "Steady-state visual evoked potentials for EEG-based biometric identification," *BIOSIG 2017*, 2017.

[46] Y. Chen, A. D. Atnafu, I. Schlattner, W. T. Weldtsadik, M. C. Roh, H. J. Kim, S. W. Lee, B. Blankertz, and S. Fazli, "A high-security EEG-based login system with rsvp stimuli and dry electrodes," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2635–2647, Dec 2016.

[47] M. V. Ruiz-Blondet, Z. Jin, and S. Laszlo, "Cerebre: A novel method for very high accuracy event-related potential biometric identification," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 7, pp. 1618–1629, July 2016.

[48] M. K. Abdullah, K. S. Subari, J. L. C. Loong, and N. N. Ahmad, "Analysis of effective channel placement for an EEG-based biometric system," in *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2010*. IEEE, 2010, pp. 303–306.

[49] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[50] P. Dohnálek, P. Gajdoš, T. Peterek, and M. Penhaker, *Pattern Recognition in EEG Cognitive Signals Accelerated by GPU*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 477–485.

[51] K. Xielifuguli, A. Fujisawa, Y. Kusumoto, K. Matsumoto, and K. Kita, "Pleasant/unpleasant filtering for affective image retrieval based on cross-correlation of EEG features," *Applied Computational Intelligence and Soft Computing*, vol. 2014, article ID 415187, pp. 1–10, 2014.

[52] M. Kostílek and J. Štástný, "EEG biometric identification: Repeatability and influence of movement-related EEG," in *2012 International Conference on Applied Electronics*, Sept 2012, pp. 147–150.

[53] M. V. R. Blondet, S. Laszlo, and Z. Jin, "Assessment of permanence of non-volitional EEG brainwaves as a biometric," in *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*, March 2015, pp. 1–6.

Information Forensics and Security, vol. 11, no. 1, pp. 163–175, Jan 2016.

Pablo Arnau-González obtained his degree in Computer Engineering by the University of Valencia in 2015 and is currently working towards his PhD at the School of Engineering and Computing of the University of the West of Scotland (UWS) under the supervision of Prof. Naeem Ramzan. He is currently a member of the Artificial Intelligence Visual Communication and Networks research group in UWS. His research interests include affective computing and applied machine learning.

Miguel Arevalillo-Herráez received the first degree in computing from the Technical University of Valencia, Spain, in 1993; the B.Sc. (Hons.) degree in Computing from Liverpool John Moores University, UK, in 1994; the PgCert degree in Teaching and Learning in Higher Education; and the Ph.D. degree in 1997, both from Liverpool John Moores University, UK. He worked as a post-doctoral research fellow and as a senior lecturer at Liverpool John Moores University, until 1999. Then, he left to work for private industry for a one-year period, and came back to academia in 2000. He was the program leader for the computing and business degrees at the Mediterranean University of Science and Technology until 2006. Since 2006, he has worked as an associate professor at the University of Valencia (Spain), and has gained accreditation as full-professor by the Spanish National Agency for Assessment and Accreditation (ANECA).

Stamos Katsigiannis received the B.Sc. (Hons.) degree in Informatics and Telecommunications in 2009 from the National and Kapodistrian University of Athens, Greece, the M.Sc. degree in Computer Science in 2011 from the Athens University of Economics and Business, Greece, and the Ph.D. in Computer Science (Biomedical Image and General Purpose Video Processing) in 2016 from the National and Kapodistrian University of Athens, Greece. He is currently a Postdoctoral Research Fellow at the University of the West of Scotland, School of Engineering and Computing. He has participated in four national and international research projects and has authored and co-authored over 20 research publications, including peer reviewed journals, book chapters, and conference proceedings. His research interests include affective computing, image analysis, machine learning, video coding, image and video quality, and GPU computing..

Naeem Ramzan received the M.Sc. degree in telecommunication from University of Brest, France, in 2004 and the Ph.D. degree in electronics engineering from Queen Mary University of London, London, U.K, in 2008. Currently he is a full Professor at the School of Engineering and Computing, University of the West of Scotland. He has authored or co-authored over 110 research publications, including journals, book chapters, and standardization contributions. He co-edited a book entitled Social Media Retrieval (Springer, 2013). He is a fellow of the Higher Education Academy and a senior member of IEEE. He served as a Guest Editor for a number of special issues in technical journals. He has organized and co-chaired three ACM Multimedia Workshops, and served as the Session Chair/Co-Chair for a number of conferences. He is the Co-Chair of the Ultra HD Group of the Video Quality Experts Group (VQEG) and the Co-Editor-in-Chief of VQEG E-Letter. He has participated in more than 20 projects funded by European and U.K. research councils.