BED: A new dataset for EEG-based biometrics

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

Abstract—Various recent research works have focused on the use of electroencephalography (EEG) signals in the field of biometrics. However, advances in this area have somehow been limited by the absence of a common testbed that would make it possible to easily compare the performance of different proposals. In this work, we present a dataset that has been specifically designed to allow researchers to attempt new biometric approaches that use EEG signals captured by using relatively inexpensive consumer-grade devices. The proposed dataset has been made publicly accessible and can be downloaded from https://doi.org/10.5281/zenodo.4309471. It contains EEG recordings and responses from 21 individuals, captured under 12 different stimuli across three sessions. The selected stimuli included traditional approaches, as well as stimuli that aim to elicit concrete affective states, in order to facilitate future studies related to the influence of emotions on the EEG signals in the context of biometrics. The captured data were checked for consistency and a performance study was also carried out in order to establish a baseline for the tasks of subject verification and identification.

Index Terms—Biometrics, consumer-grade device, dataset, EEG, session.

I. INTRODUCTION

MANY applications in the currently emerging digital world require person identification methods to secure access control. In this context, biometrics are turning out into an alternative to other more traditional access methods based on keys, ID cards or passwords. Traditional access control approaches require the individuals to remember or possess some information or item that must be presented to the access system. Presenting the correct information/item grants access to the individual. Biometrics are defined as the "*automated recognition of individuals based on their biological and behavioural characteristics*" [1]. Typically, biometrics systems consist of a device that captures a characteristic, e.g. a camera or microphone, a database that stores information about the persons registered in the biometrics system, algorithms for processing the acquired characteristics, e.g. signal processing algorithms, and finally a decision system that compares the stored and the captured characteristics and decides whether they belong to the same person [2]. Success in the comparison

of the biometrics trait grants access to a system or identifies an individual. Traditional biometrics approaches include fingerprint recognition, face recognition, iris recognition, voice recognition, and others.

Electroencephalography (EEG) signals, i.e., the recording of the electrical activity of the brain, present some major advantages when compared to other biometrics modalities: they are resilient to physical injuries, extremely hard to reproduce, and cannot be furtively captured at a distance [3]. These properties have motivated extensive research in the last few years, aimed at proposing reliable EEG-based solutions for biometrics (e.g. [4], [5], [6]) and included studies on using different types of stimuli [7], as well as on the influence on the EEG signal of the emotion they potentially elicit [8], [9]. However, the use of EEG signals for biometrics also faces important challenges, mainly related to practicality. EEG signals are contaminated by unwanted artefacts caused by, e.g., ocular, muscle, cardiac and respiratory activity, requiring a relatively intensive preprocessing of the signal. Hence, the time it takes to perform EEG-based user authentication is substantially higher than the time required by other competing biometric authentication schemes [10]. In addition, high precision devices are generally expensive, and require specific capturing protocols to ensure an adequate placement of the electrodes and the absence of electrical and electromagnetic interference.

EEG-based biometrics is still at its early stages. The absence of a public benchmark dataset has forced some researchers to evaluate their methods using datasets that were designed for a purpose other than biometrics; or proprietary databases of generally small size that have not been made public, thus preventing a fair comparison between different proposals. In addition, and despite the high performance of medical-grade devices for EEG-based biometrics, their practical suitability has been called into question not only because of their cost, but also due to the tedious preparation needed for acquiring the signals. This has motivated the development of consumer-grade EEG recording devices with a smaller number of electrodes as a more practical alternative. Although these devices offer a lower spatial resolution than medical-grade systems, they also simplify deployment in real-life scenarios.

Given this context, we believe that a database specifically designed for EEG-based biometrics with consumer-grade devices is a definite contribution to the future development of the state-of-the-art. First, it is a step forward to define a common ground that allows for a fair comparison between different proposals. Second, there are specific factors that should be considered when assessing any biometrics approach, and a careful design may help in establishing research routes that were partially disregarded in previous works. This includes a) using different types of stimuli to be able to properly test

both identification and verification scenarios; and b) using recordings from multiple sessions to assess the permanence property of the suggested mechanisms.

In this paper, we present BED (Biometric EEG Dataset), a dataset specifically designed to test EEG-based biometric approaches that use relatively inexpensive consumer-grade devices. This dataset, along with usage instructions, can be downloaded from https://doi.org/10.5281/zenodo.4309471 and includes EEG responses from 21 subjects to 12 different stimuli, across 3 different chronologically disjointed sessions. We have also considered stimuli aimed to elicit different affective states, so as to facilitate future research on the influence of emotions on EEG-based biometric tasks. In addition, we provide a baseline performance analysis to outline the potential of consumer-grade EEG devices for subject identification and verification. It must be noted that, in this work, EEG data were acquired in a controlled environment in order to reduce the variability in the acquired data stemming from external conditions.

The rest of this paper is organised in five sections. Section II describes the related literature and provides a general background covering consumer-grade EEG-based biometrics. Section III provides a detailed description of the dataset, including the stimuli used, the equipment employed and the acquisition setting and protocol. Section IV contains an analysis of the participants' responses that supports the subjects' engagement during the signal capturing process. Section V describes baseline experiments for cross-session subject identification and verification. Finally, section VI summarises the major conclusions that can be drawn from this work.

II. Background

Although EEG signals were initially used to assist the diagnosis of certain pathological conditions and disorders [11], [12], [13], [14], [15], their suitability and great potential to discern between individuals has attracted the interest of biometrics researchers. During the two first decades of this century, EEG signals have been extensively employed with the aim to properly identify individuals [16], [17], [18]. The vast research in this area has motivated a number of surveys, some covering the future perspectives and the theoretical aspects of EEG-based biometric identification systems [19], [20] and others more focused on the practicality and usability issues of EEG as a biometric signal [21].

However, and despite the large amount of research publications found in the area, there is a lack of a standard benchmark that allows for a fair comparison between methods. This has led many authors to create proprietary datasets, designed according to their particular experimental setting [17], [18], [22], [23], [24], [25], [26], [27], [28], [29], [30]. Other authors have used public EEG datasets that were originally designed for purposes other than biometrics [8], [31], [32], [33]. This includes the one by UCI (University of California, Irvine) [34] and VEP (Visual Evoked Potentials) [35], which were initially conceived for image speech and alcoholism detection, respectively; or DEAP [36], MAHNOB-HCI [37], DREAMER [38], SEED [39], and the Lakhan et al. [40] datasets, which were

constructed with emotion recognition in mind. In addition, little attention has been given to consumer-grade EEG devices, despite their importance in easing deployment in practical applications [21].

To achieve a sound experimental setting for practical biometrics, the dataset used needs to satisfy certain conditions related to data capturing and type of stimuli used. More specifically, it should facilitate practical applicability of the registration and verification/identification procedures, it should contain a variety of stimuli for evaluating their suitability, and it should allow the evaluation of the temporal stability of the extracted biometrics patterns by containing data from multiple acquisition sessions.

Regardless of the signal acquisition device, the person-specific patterns contained in brain signals may strongly depend on many factors, including the task type the subject is performing. Hence, biometric systems based on EEG signals should consider different stimuli in order to study their potential capabilities. In this line, different approaches are found in the literature, mainly focusing on three different types of pattern elicitation mechanisms: resting-state, cognitive tasks, and sensory stimuli [21]. Resting-state and sensory stimuli are the most common protocols used for identifying individuals, e.g., [41], [42], [43], [44], while cognitive tasks are more common for authentication purposes, e.g., [45], [46]. If one objective of the proposed dataset is to cover both the identification and authentication scenarios, it should not be limited to a single type of pattern elicitation and cover at least the three basic mechanisms mentioned above.

A common mistake in the evaluation of EEG-based biometric approaches relates to template ageing, i.e., "*the increase in error rates caused by time-related changes in the biometric pattern, its presentation and the sensor*" [47]. To properly evaluate a biometric system, data acquisition should happen over time, along several sessions. This practice is encouraged since it allows to ensure the temporal stability of the extracted patterns and the proposed solution. Surprisingly, only a few studies on EEG-based biometrics have considered the aspects of time and template ageing [25], [26], [48]. Most research described in a recent survey [21] and other remarkable works in the area use data acquired during a single session, e.g., [16], [18], [22], [27], [28], [49], or have used several sessions but constructed the training and validation sets mixing the samples from all sessions, disregarding the acquisition date [50]. Although they have generally claimed high accuracy results [22], such setting is generally biased towards high classification rate [23] and can be affected by a large number of session-specific factors that include, between many others, the exact positioning of the electrodes in the scalp, capacitative coupling of electrodes and cables with other devices, induction loops created between the employed equipment and the body, power supply artefacts, and others [24]. Therefore, although such studies provide proof that subject-related patterns can successfully be extracted from EEG signals for EEG-based biometrics applications, they do not examine the permanence of such patterns across time and thus their applicability for practical real-world EEG biometrics systems.

TABLE I provides a summary of the results achieved in

TABLE I: Summary of relevant literature in EEG-based biometrics

Reference	# of Participants	# of Sessions	Cross-Session	Public Dataset	Recording Device	Stimuli	Metrics
[16]	50	1	No	No	Brain Vision	Sine gratings Low-frequency words Images	Acc. 100%
[17]	45 (1 session) 15 (2 Sessions) 9 (3 sessions)	3	Yes	No	Undisclosed	Acronyms inter-mixed with fillers from other lexical types	Acc. 93%
[18]	6	1	No	No	Emotiv EPOC[†]	Eyes open/ Eyes closed	Acc. 88%
[50]	10	5	No	No	gMobilab+	Eyes open/ Eyes closed	Acc. 97%
[22]	29	1	No	No	g.tec Brain Products	Images	10.7% FAR / 100% TAR
[23]	9	2	Yes	No	Custom-made in-ear [51]	No stimuli	HTER 17.2
[24]	45	6	Yes	No	Galileo BE Light Amplifier	Eyes open / Eyes closed Mathematical Computation Speech Imagery	EER 6.7
[25]	9	2	Yes	No	Undisclosed	Muscle Movement	Acc. 77.8%
[26]	15	3	Yes	No	Undisclosed	Acronyms	Acc. 93%
[27]	10	1	No	No	Undisclosed	VEP	Acc. 90%
[28]	120	1	No	Yes [35]	Undisclosed	VEP	Acc. 98.12%
[29]	12	12	Yes	No	Biosemi	Motor Imagery Speech Imagery	HTER 34.9%
[30]	15	2	No	No	Neurosky Mindset[†]	Breathing Motor Imagery Speech Imagery Audiovisual Stimuli Cognitive Tasks Pass-thought	HTER 1.1%
[31]	6 (UCI) 120 (VEP)	Undisclosed	No	Yes [34], [35]	Undisclosed	Imagined Speech VEP	Acc. 99.6%
[46]	12	1	No	No	Emotiv EPOC[†]	Thought	Acc. 66.67%
[49]	21	2	No	No	Emotiv EPOC[†]	Card Counting	Acc. 72%
[8]	32 (DEAP) 28 (MAHNOB) 15 (SEED)	1 (DEAP) 1 (MAHNOB) 3 (SEED)	Yes	Yes [36], [37], [39]	Biosemi Active II[‡] ESI NeuroScan[‡]	Audiovisual Stimuli	Acc. 79.34%
[32]	23	1	No	Yes [38]	Emotiv EPOC[†]	Audiovisual Stimuli	Acc. 94%
[33]	26	3	Yes	No	Emotiv EPOC+[†]	Emotion Images	Acc. 34.9%

Notes: [†] denotes consumer-grade devices and [‡] medical-grade devices. Acc: Accuracy, FAR: False Acceptance Rate, TAR: True Acceptance Rate, HTER: Half Total Error Rate, EER: Equal Error Rate.

some of the most relevant studies, along with the characteristics of the databases that were used in each of them. The performance metrics used are as reported in the original works, and include Accuracy (Acc), False Acceptance Rate (FAR), True Acceptance Rate (TAR), Half Total Error Rate (HTER) and Equal Error Rate (EER). Despite the relatively high accuracy achieved by some of the methods, none of these works has considered a cross-session study in a public dataset specifically designed for biometrics. Therefore, they do not ease comparison with the rest of the literature.

III. DATASET DESIGN

The aspects described above have all been taken into account to produce a new public dataset specifically designed to test biometric approaches. This repository contains three sessions separated in time, considers a wide set of stimuli and has been produced by using a low-cost off-the-shelf EEG recording device. In addition, and as a consequence of recently published studies that report further gains when the emotional state is taken into account [8], [9], [33], we have also considered image-based stimuli that were specifically selected to elicit a set of representative emotional responses. This section describes the specific characteristics of this dataset. Baseline experimental results obtained with it are later reported in Section V.

A. Stimuli used during EEG acquisition

We have considered the use of an extensive set of stimuli in order to provide support for a large variety of different settings. The presented stimuli contain the most common ones found in the literature [4], [16], [24], and others aimed at supporting future studies on the potential influence of affect. A total of 4 categories of stimuli were recorded for posterior analysis, which are described below:

1) Affective Stimulus (AS): These are a set of images that aimed to elicit typical emotional responses during EEG acquisition. They were obtained from two publicly available image datasets that are commonly used in affect detection studies [52], i.e., the Geneva Affective Picture Dataset (GAPED) [53] and the Open Affective Standardised Image Set (OASIS) [54]. GAPED contains 730 different images with a resolution of 640×479 pixels encoded in JPEG format and rated in terms of arousal and valence within the range $[0, 100]$. These images are organised along the following six categories: snakes, spiders, human concerns (depicting scenes violating human rights), animal mistreatments, neutral and positive. OASIS has 900 different images with a resolution of 500×400 pixels, also encoded in JPEG format. The images are rated in terms of arousal and valence within the range $[1, 7]$, and responses by gender are also reported. This dataset contains images that belong to one of the following four mutually exclusive groups: animals, objects, people, and scenes. It must be noted that image #I537 was discarded from the set due to
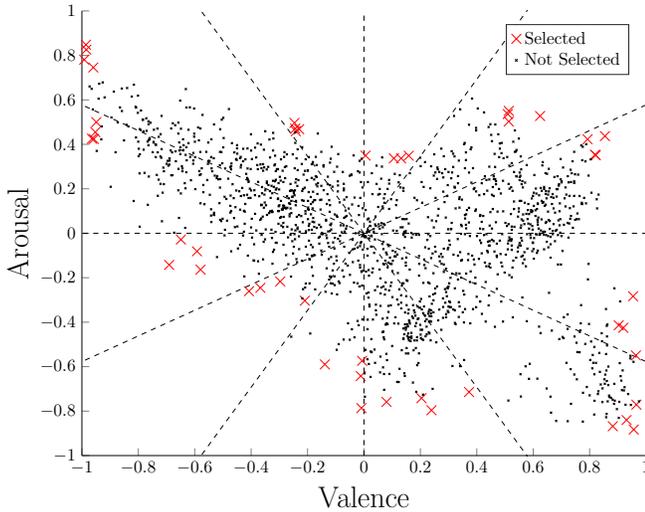
Fig. 1: Positioning of images from the GAPED and OASIS datasets in Valence-Arousal space. Images marked with a red "×" are the ones selected for this study.

Fig. 2: Electrodes' location on the human scalp.

Fig. 3: EEG signal (1 s) from participant #1 (session #1).

very explicit adult content.

These two datasets contain a total of $730 + 900 = 1\,630$ images, from which 48 were selected according to their valence and arousal labels, in order to obtain a representative set with intense emotional content. To this end, the valence and arousal labels in each dataset were first normalised to the range $[-1,\ 1]$. Then, the resulting valence/arousal space was divided into 12 equal regions, as shown in Fig. 1. Finally, we selected the 4 images from each region whose valence and arousal values were the farthest from the absolute neutral emotion $(0, 0)$. These 48 images were divided into 4 different sets of 12 images each, with each set containing exactly one image from each of the regions, thus each of the 4 images from each region was allocated to a different set. One of these sets was randomly selected to be shown to the participants in all three sessions, while the other three sets were assigned to one session each. In each session, each subject would first see the 12 common images shared between sessions, and then the 12 additional images assigned to the specific session.

2) Cognitive Stimulus - Mathematical Computations (MC): Cognitive tasks in the form of two-digit additions were used in this study with a threefold purpose: (a) generate cognitive imagery-related patterns in the participant, (b) bring the participant back to a neutral emotion state after been exposed to an affective stimulus, and (c) check the participants engagement with the experimental process by checking the correctness of their answers. Following a similar approach to the *Affective Stimuli*, we randomly created 4 sets of 12 two-digit addition operations. These operations were designed so that the numbers to be added would be between 11 and 99, and the participant would have to carry for both digits. Out of the 4 sets, one was randomly selected to be used in all three sessions, while the remaining 3 sets were randomly assigned to one session each. At each session, participants would first be presented with the set of operations shared across the three sessions, followed by the set of operations for each respective session.
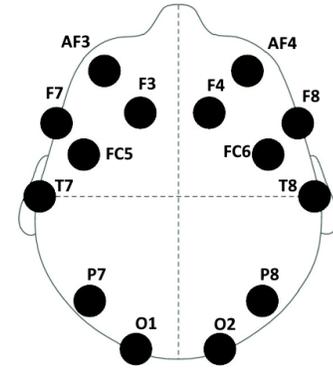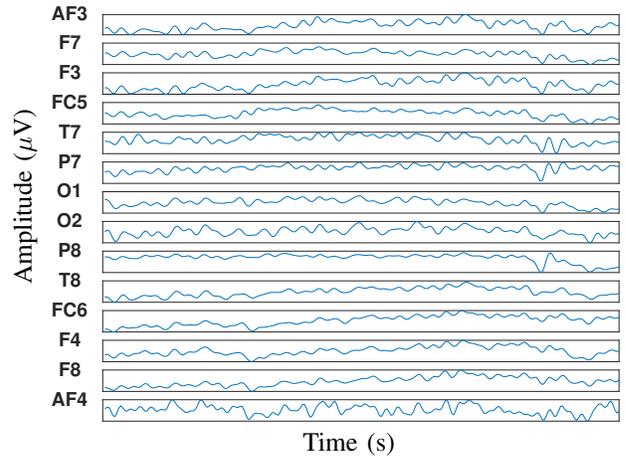
3) Visual Evoked Potentials (VCx and VFx): Visual Evoked Potentials (VEP) have traditionally been used for the diagnosis of different conditions [55], such as Alzheimer [56]. More recently, various researchers have proposed the use of VEP for the extraction of user-specific patterns [4], [31]. In this work, the subjects were presented with VEP at four different frequencies, i.e., 3, 5, 7, and 10 Hz, as commonly performed in the literature [57]. Shown patterns included the standard checker-board pattern with pattern reversal (VC3, VC5, VC7, VC10), as well as flashing VEP with a plain colour, set as black (VF3, VF5, VF7, VF10). The decision of using the second type of VEP was influenced by the indication in the VEP standard [57] that the flashing versions yield higher correlates with the individual.

4) Resting-state: The resting-state protocol has been largely adopted in the literature due to its simplicity [4], [24], [58], [59]. The protocol consists of relaxation with eyes closed (RC), or with the eyes open (RO). The screen in front of the participants remained switched-off during the resting-state protocol in order to avoid any potential effect on the participants.
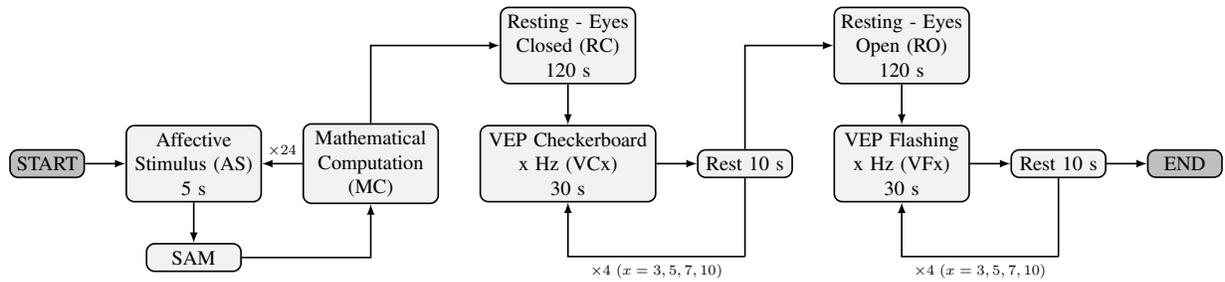
Fig. 4: Flow diagram illustrating the experimental protocol

B. Signal Acquisition

The experiments were performed in a laboratory environment, with controlled illumination and isolated from sources of noise or distraction. No specific protection against electrical and electromagnetic interference was used. 14-channel electroencephalography (EEG) signals were captured at a sampling frequency of 256 Hz using the Emotiv EPOC+ [60] wireless EEG headset. The Emotiv EPOC+ system is a commercially available low-cost EEG capturing device that is equipped with 16 contact sensors, fixed on flexible plastic arms that are placed against the scalp of the user. Fourteen of the contact sensors are placed in locations (Fig. 2) that closely align with the AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 locations of the Modified Combinatorial Nomenclature (MCN) [61], which extends the international 10-20 system and are used for signal recording, while two contact sensors located at the M1 and M2 mastoid locations are used as reference. Although the number of channels of the captured EEG signals (signal sample shown in Fig. 3) is lower than commonly used medical-grade devices due to the reduced number of electrodes, studies have shown that the Emotiv system is a viable alternative to expensive and non-portable medical-grade EEG devices [60], [62], as also evidenced by its successful use in various studies on affective computing [8], [38], [63], [64] and on EEG-based biometrics [18], [46], [49], [65], [66], [67]. Furthermore, the Emotiv EPOC+ headset has a built-in digital 5th order sinc filter and applies digital notch filters at 50Hz and 60Hz for reducing noise and artefacts in the acquired signal [68].

A computer equipped with an Intel Core i7-7700K @4.20 GHz CPU and 64 GB of DDR4 RAM memory, running MS Windows 10, was used for signal recording. Since the Emotiv EPOC+ headset connects to the computer via proprietary radio communication, participants and supervising researchers were instructed to switch off any electronic devices that may transmit signals, such as mobile phones, smart-watches, Bluetooth devices, etc., in order to avoid any potential interference with EEG signal transmission during the experimental sessions.

C. Acquisition Protocol

A total of 26 healthy participants (22 male and 4 female) were initially recruited for the creation of the presented dataset. Unfortunately, data from five of these participants had to be discarded due to erroneous signal acquisition, resulting in a dataset composed of data from 21 participants (18 male and 3 female), aged between 23 and 47 ($\mu_{\text{age}} = 30$) years old.

Prior to the experiment, potential participants were informed about its purpose and procedure and were also warned that images with some explicit adult content might be shown. Furthermore, potential participants were instructed not to participate if they suffered from any form of epilepsy, photosensitivity or related conditions, due to the viewing of VEP. After asking any questions that they may had and agreeing to participate, the subjects were asked to sign a consent form. Hereby, they explicitly stated that they did not suffer from any form of epilepsy, photo-sensitivity or related conditions; were informed in detail about the protocol, including the type of images that they would see; agreed to participate in the study; and granted permission to use or publish the acquired data in anonymised form for research purposes. Immediately after, participants were verbally provided with more detailed instructions about the experimental procedure and the use of the Self Assessment Manikins (SAM) to provide feedback about the emotion they felt, according to the directions included in [69]. The Emotiv EPOC+ EEG headset was then positioned on the head of the participants and the supervising researchers proceeded to check the quality of the signal acquisition. Once the quality of electrode contact was verified using the *Emotiv Xavier Control Panel* software and successful signal acquisition was confirmed via manual inspection of the plotted EEG signals, the supervising researchers instructed the participant to start the experiment whenever they felt ready and comfortable, and left the room.

The software for displaying the described stimuli was developed in Python using the pygame library (https://www.pygame.org). This library provides tools for drawing 2-D graphics and also controlling the frame rate of the displayed content, which was needed to generate the VEP. An overview of the experimental protocol followed in each session of the experiment is provided through the flowchart in Fig. 4.

The first part of the experiment consisted of alternating between an Affective Stimulus and a Cognitive Stimulus. First, an image was shown for 5 s. Afterwards, the participant was asked to provide feedback about the emotion felt after being exposed to the image, by using the SAM that was displayed in the middle of the screen. This was done in order to (a) validate that the image stimuli were correctly selected, and (b) allow future researchers to create affect-based strategies, as
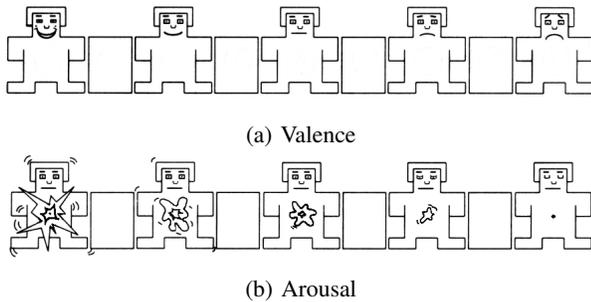
(a) Valence

(b) Arousal

Fig. 5: Self-Assessment Manikins (SAM) used.

suggested in [8]. Assessment was provided in terms of valence and arousal by double clicking on the respective manikins, or on the boxes between them, as shown in Fig. 5. The responses were later normalised to the range $[-1, +1]$ for further analysis. After providing the feedback for the image, the participant was shown a two-digit addition operation that they needed to solve using the keyboard to fill the answer field. Both the numeric pad and the typewriter keys could be used for typing the answer. Mistakes could be corrected using the *Backspace* or the *Delete* buttons and the answer was submitted by pressing the *Enter* button. No feedback regarding the correctness of the submitted answer was provided to the participant in order to avoid distracting them with feelings of success or failure that may affect the EEG signals. This procedure was repeated 24 times, until all the 24 images and the 24 two-digit addition operations selected for the session were shown.

After the Affective and the Cognitive stimuli, the participant was instructed to rest with the eyes closed for a total of 120 s. A sound emitted by the computer notified the participant that the 120 s had elapsed. Next, standard checker-board VEP were shown with increasing frequency ($\chi = 3, 5, 7, 10$) Hz for 30 s at each frequency. In between VEP, participants were given a sort rest period of 10 s. Then, the screen was switched-off and the participant was instructed to rest for 120 s keeping their eyes open. Afterwards, the subject was presented with the final stimulus, i.e., flashing VEP shown with increasing frequency ($\chi = 3, 5, 7, 10$) Hz for 30 s at each frequency. A sort rest period of 10 s between each frequency was also given, similar to the checker-board VEP procedure. Once all flashing VEP stimuli had been shown to the participant, the session was considered complete and the researcher returned to the room to remove the EEG sensor from the participant's head and check that the data had been recorded correctly. In the mean time, the participant was offered some complimentary sweets, coffee, water, or other non-alcoholic beverages.

The protocol described above was repeated three times for each participant in different days, a week apart from each other. The protocol followed for each session was the same, with the only difference being the sets of images and two-digit addition operations used at each session, as explained in Sections III-A1 and III-A2.

IV. ANALYSIS OF THE PARTICIPANTS RESPONSES

In order to ensure a consistent labelling and that subjects were engaged during the sessions, we performed an in-depth analysis of the participants' answers. First, we evaluated the subjects' answers to the SAM by computing their correlation with the expected responses according to the image labels in the corresponding image datasets. Second, we evaluated the consistency of the emotional labels provided by the participants across the three sessions. Finally, we analysed the success of the participants at solving the cognitive task (two-digit additions).

A. Correlation of emotional labels

In order to quantitatively evaluate the consistency of the participants' answers to the SAM, we computed the Pearson's correlation coefficient (ρ) between the average values provided by all participants and the answer expected according to the labels provided in the corresponding dataset.

For valence, this analysis led to a Pearson's $\rho = \{0.9710, 0.9824, 0.9752\}$, for the first, second, and third session, respectively. The strong linear correlation between the participants' and the expected responses is easily observable in Fig. 6a, 6c, and 6e. This trend can also be observed when data from all sessions are examined together. In this case, the Pearson's correlation coefficient was $\rho = 0.9788$ (Fig. 6g). For arousal, the computed correlation was $\rho = \{0.8431, 0.8568, 0.7538\}$ for the first, second, and third session, respectively. This high correlation can also be observed in Fig. 6b, 6d, and 6f. When all three sessions are jointly considered, the correlation coefficient was $\rho = 0.8223$ (Fig. 6h). The strong linear correlation between the average emotion ratings provided by the participants and the expected ratings, according to the available labelling, supports that a) our population voted similarly as the participants in OASIS and GAPED; b) the rating scales were correctly understood by the participants; and c) the selected images are suitable for the secondary purpose of this dataset, i.e., affect-enabled subject-recognition.

If the correlation with the expected answer is computed per subject, the average value is slightly smaller for valence, with $\rho = \{0.8231, 0.8506, 0.8507\}$ for the first, second, and third session, respectively. However, the arousal dimension presents significantly lower average correlation scores ($\rho = \{0.3701, 0.4183, 0.3299\}$). These figures show that the arousal scale is far more subjective than the valence one, and ratings may have a stronger dependency on the subjects' background, thus showing a higher variance. This finding is also consistent with the results reported in [54] for the OASIS dataset, stating that the relatively lower reliability of the arousal scale was in part due to a lack of internal consistency across gender groups, as well as with the findings of Warriner et al. [70] who showed that ratings of valence are relatively consistent across participants while arousal is much more variable.

B. Agreement between sessions

Additionally, the inter-session agreement of the participants' responses across the three sessions has been assessed by

(a) Session 1 - Valence

(b) Session 1 - Arousal

(c) Session 2 - Valence

(d) Session 2 - Arousal

(e) Session 3 - Valence

(f) Session 3 - Arousal

(g) All Sessions - Valence

(h) All Sessions - Arousal
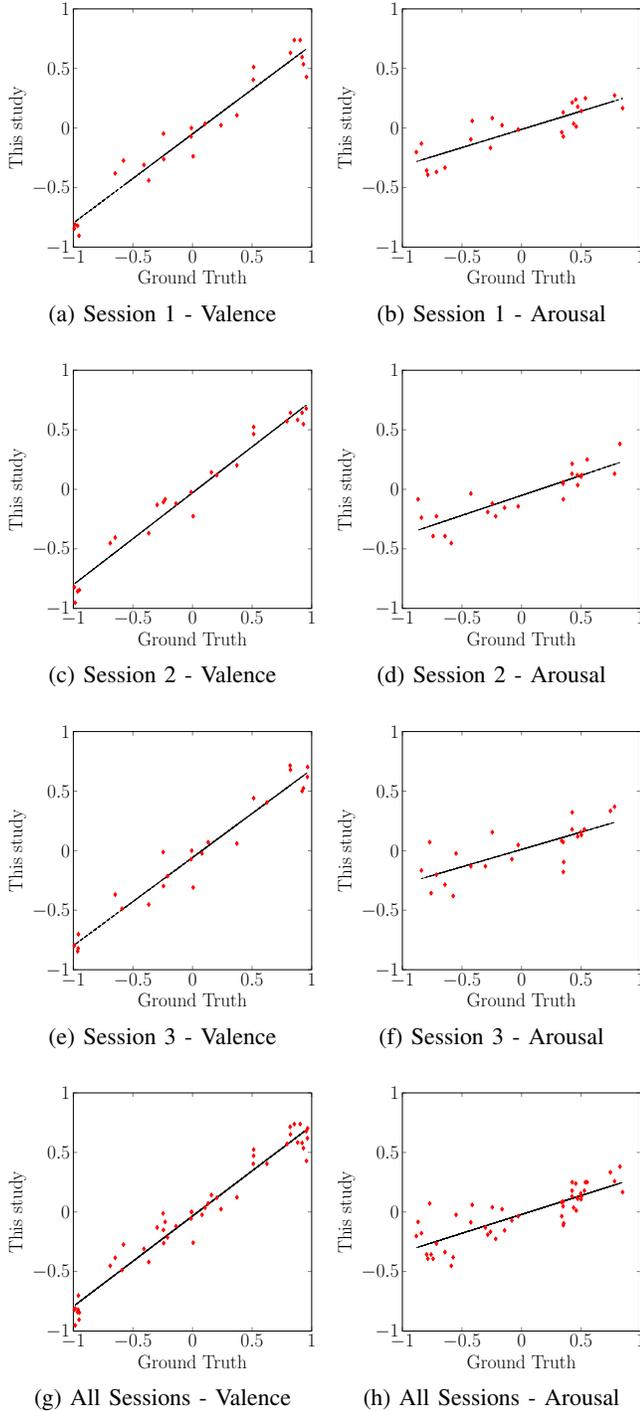
Fig. 6: Scatter plot, expected response vs average response, per session and emotional dimension

computing the Kendall's coefficient of concordance (Kendall's W) [71] for each of the participants separately, using the 12 images that were repeated in every session. Since our population and number of measurements are relatively small, we have followed the recommendation in [72] and replaced the traditional Friedman's χ^2 statistic by the F statistic in Eq. 1 in the computation of the p-values. The F distribution has two parameters that represent the degrees of freedom for the numerator and denominator. These are $\nu_1 = n - 1 - \left(\frac{2}{m}\right)$ and
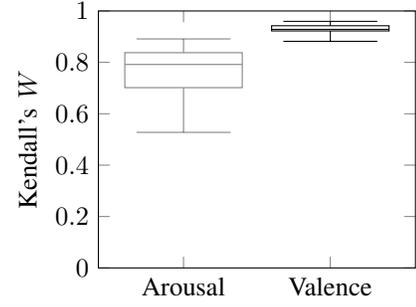
Fig. 7: Kendall's coefficient of concordance.

$\nu_2 = \nu_1 \cdot (m - 1)$, with n being the number of measurements and m being the number of times a measurement has been taken. In our case, $m = 3$ is the number of sessions and $n = 12$ the number of images that were recurrently used across all sessions.

$$F = \frac{(m-1)W}{(1-W)} \qquad (1)$$

A median Kendall's $W = 0.9285$ ($p << 0.05$) was obtained across participants for valence, and a median $W = 0.7918$ for arousal ($p < 0.05$ in all but five subjects), as shown in Fig. 7. These values show a very high inter-session agreement for valence ratings and a moderate one for arousal ratings.

C. Validation of engagement

In order to validate that the participants were engaged in the labelling process and did not answer at random, we evaluated their performance on the addition operations that were presented to them. Fig. 8 shows the percentage of correct answers per user. On average, the participants correctly answered 88.89% of the calculations, which supports the participants' active engagement. Only participant #3 had a considerably lower performance (59.72%). Nevertheless, he was not discarded from the dataset since the Pearson's correlations between his emotion ratings and the expected ones ($\rho_{\text{Valence}} = 0.76$, $\rho_{\text{Arousal}} = 0.36$) was in line with the average values reported in Section IV-A.

V. BASELINE EXPERIMENTAL RESULTS

The aim of this study is to provide a benchmark dataset for EEG-based subject verification and identification. Verification refers to the task of deciding whether a user is whoever they claim to be. In this scenario, the query is only compared to the template of the requested identity, and the user is accepted or rejected depending on whether the result of the comparison is above or below a certain threshold. In contrast, identification refers to the task of deciding who the user is from a pool of possible profiles. In this context, the query is compared to all the available profiles and assigned to the identity that provides the best match.

To complement the dataset, we have carried out a set of supervised classification experiments in order to establish some baseline results in both the verification and identification scenarios. These experiments evaluated the performance when
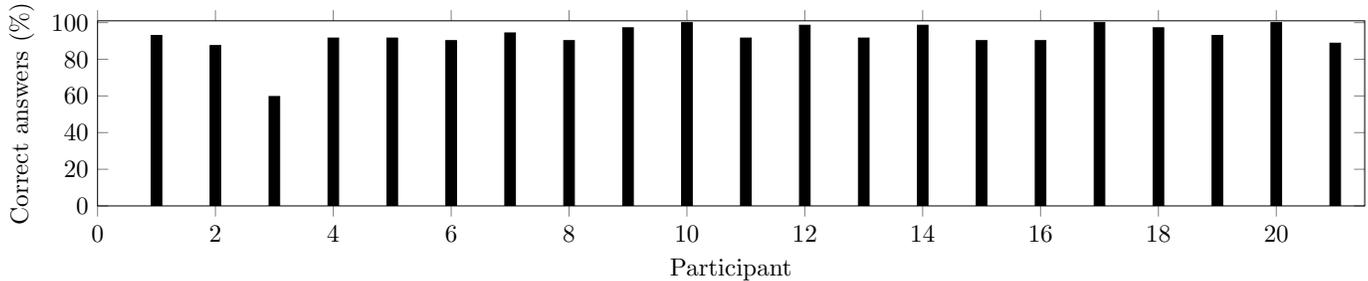
Fig. 8: Performance on the cognitive task per participant, in terms of the percentage of correctly solved mathematical operations.

using different combinations of features and stimuli. It must be noted that the intention of the conducted experiments was not to show the best feature extraction and/or classification methods, but rather to provide a better understanding of the data and their potential. Consequently, performance-related restrictions were not considered.

A. Data preparation and feature extraction

In order to remove artefacts, such as those originating from muscle movement, jaw clenching, or eye blinking, we used the EEGLAB toolbox [73] to pre-process the EEG signals by applying the PREP pipeline [74]. The PREP pipeline consists of the following steps: (1) removal of line-noise with filtering, (2) referencing of the signal relative to an estimate of the "true" average reference, and (3) detection and interpolation of bad channels, relative to the reference. After pre-processing, we extracted the following features from the EEG signals:

1) Mel Frequency Cepstral Coefficients (MFCC): MFCC features have been widely used in speech recognition applications [75], [76], and their use has recently been extended to EEG-based biometrics [24], [77]. The procedure for the extraction of MFCC from each EEG channel consists of applying the Fourier Transform, then a filterbank in the Mel scale, and finally the Discrete Cosine Transform. In this case, we applied a filterbank with 18 filters and obtained the first 12 coefficients after dropping the D_C coefficient, as proposed in [78]. The final feature vector was created as the concatenation of the 12 cepstral coefficients from each of the 14 EEG channels, leading to a total of $12 \times 14 = 168$ features.

2) Autoregression Reflection Coefficients (ARRC): ARRC features have been used for EEG-based verification and identification in various studies [24], [77]. In this work, reflection coefficients were extracted from each EEG channel from a 12-th order autoregressive model created by solving the Yule-Walker equations. This process led to 12 ARRC features per EEG channel, adding up to a total of $12 \times 14 = 168$ features.

3) Spectral Features (SPEC): Four spectral features were extracted from the commonly used [79] θ (4-8 Hz), β (12-30 Hz), γ (30+ Hz), and α (8-12 Hz) bands of each of the EEG signal's channels. The four computed spectral features were the spectral centroid, spectral bandwidth, spectral crest factor, and spectral flatness, and were computed as proposed in [80]. The final feature vector was created as the concatenation of the 4 spectral features from each of the 4 bands of the 14 EEG channels, leading to a total of $4 \times 4 \times 14 = 224$ features.

The extracted features are also provided in our dataset, along with the raw data, to ease future comparison of approaches against the baseline provided.

B. Baseline subject verification evaluation

Following previous work on the field of EEG-based biometrics [24] and in order to provide baseline subject verification results for the proposed dataset that adhere to the state-of-the-art, we used Hidden Markov Models (HMM) to model the EEG signal's temporal evolution for each user. We did this in the same way as in [24], except that we used regular HMM instead of *left-right* models because of their higher performance in our case. Furthermore, it must be noted that continuous variables with a multigaussian probability distribution were used for the HMM. The exact procedure used is explained below in greater detail. In this explanation, we refer to the signal associated to the c-th channel of an EEG recording r as $r^{(c)}$, with $c = 1, 2, ..., 14$.

The strategy followed for training the HMMs was to segment the recordings into consecutive overlapping epochs $e^{(c)}$. Each of these epochs had a length of 5 s with 50% overlapping. Due to possible disconnections of the recording device, captured signals with less than 3 s worth of data were discarded from further analysis and recordings with duration between 3 and 5 s were treated as a single epoch. The different epochs were split into H overlapping frames of 1 s and 50% overlapping, and represented as a sequence of observations $o^{(c)}$, so that $o^{(c)} = \left[f_0^{(c)}, f_1^{(c)}, ..., f_H^{(c)} \right]$ where $f_h^{(c)}$ is the h-th frame of the epoch $e^{(c)}$. Each frame was then used in order to extract the feature vectors $\hat{f}_h^c$, each containing Q features, with $Q_{\text{MFCC}} = 12$, $Q_{\text{ARRC}} = 12$, and $Q_{\text{SPEC}} = 16$. Finally, the sequence of observations in the feature space $\hat{o}^{(c)}$ was obtained as $\hat{o}^{(c)} = \left[\hat{f}_0^{(c)}, \hat{f}_1^{(c)}, ..., \hat{f}_H^{(c)} \right]$. The resulting observation sequences $\hat{o}^{(c)}$ are used to build a Markov Model $\lambda^{(c)}$ with $N = 4$ hidden states, by using the Baum-Welch algorithm [81].

Similarly to previous works [24], the similarity between each observation sequence of the verification session and the generated models has been calculated by computing the *a posteriori* log likelihood $l^{(c)} = P(\hat{o}^{(c)}|\lambda^{(c)})$, once the most probable path of hidden states is estimated using the Viterbi algorithm. The decision rule for accepting or rejecting an epoch according to the C models of any given subject is:

TABLE II: Verification results for Session 2 when the system is trained with data from Session 1

Metric	Feature	AS	MC	RC	RO	VC3	VC5	VC7	VC10	VF3	VF5	VF7	VF10	Avg. (St. Dev.)
	MFCC	**0.7699**†	0.6917	**0.7567**	0.6617	**0.6099**	0.6846	0.6476	0.6250	0.6602	0.6632	0.7017	0.7431	0.6846 (0.0486)
AUC	ARRC	0.7320	0.7034	0.7285	**0.6765**	0.5420	0.6611	0.6714	0.5725	0.6425	**0.6813**	**0.7478**	0.7479†	0.6756 (0.0626)
	SPEC	0.7465	**0.7035**	0.7192	0.6528	0.5934	**0.7567**	**0.6771**	**0.6424**	**0.6687**	0.6606	0.7380	**0.7571**†	0.6930 (0.0499)
	MFCC	**0.2633**†	**0.3085**	**0.2792**	0.3601	**0.4092**	0.3315	0.3643	0.3886	**0.3503**	0.3573	0.3183	0.2826	0.3344 (0.0434)
EER	ARRC	0.2911	0.3180	0.2961	**0.3355**	0.4594	0.3282	0.3409	0.4234	0.3604	**0.3314**	**0.2865**	0.2733†	0.3370 (0.0531)
	SPEC	0.2880	0.3135	0.3164	0.3660	0.4209	**0.2669**	**0.3392**	**0.3729**	0.3593	0.3667	0.2837	**0.2622**†	0.3296 (0.0471)

Notes: † denotes the best performance per EEG feature used for the respective metric. Results in **bold** denote the best performance per stimulus used for the respective metric. AUC: The higher the better. EER: The lower the better.

TABLE III: Verification results for Session 3 when the system is trained with data from Session 1

Metric	Feature	AS	MC	RC	RO	VC3	VC5	VC7	VC10	VF3	VF5	VF7	VF10	Avg. (St. Dev.)
	MFCC	**0.7109**	**0.6706**	0.7030	**0.7321**	**0.6505**	0.7070	0.6514	0.5539	0.6044	**0.7575**†	**0.6870**	0.6493	0.6731 (0.0538)
AUC	ARRC	0.6487	0.6668	**0.7643**†	0.6797	0.5932	**0.7251**	0.6069	**0.6059**	0.5905	0.7061	0.6514	**0.6636**	0.6585 (0.0524)
	SPEC	0.6385	0.6204	0.6922	0.7122†	0.6322	0.6958	**0.6579**	0.5569	**0.6356**	0.6756	0.6803	0.6255	0.6519 (0.0410)
	MFCC	**0.3113**	0.3389	0.3190	**0.2963**	0.3574	0.3172	**0.3616**	0.4386	0.4010	**0.2642**†	**0.3209**	0.3728	0.3416 (0.0457)
EER	ARRC	0.3587	**0.3348**	**0.2567**†	0.3354	0.3959	**0.2922**	0.3860	**0.3874**	0.4008	0.2997	0.3582	**0.3454**	0.3459 (0.0432)
	SPEC	0.3821	0.3847	0.3357	0.3152†	0.3776	0.3322	0.3637	0.4445	**0.3495**	0.3278	0.3337	0.3842	0.3609 (0.0346)

Notes: † denotes the best performance per EEG feature used for the respective metric. Results in **bold** denote the best performance per stimulus used for the respective metric. AUC: The higher the better. EER: The lower the better.

$$D_{oC} = \begin{cases} 1 & \text{if } \frac{1}{C}\sum_{c=1}^{C} d^{(c)} \geq \phi_C \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where ϕ_C is the minimum number of channels that have to be accepted in order to accept the epoch and $d^{(c)}$ is defined as:

$$d^{(c)} = \begin{cases} 1 & \text{if } l^{(c)} \geq \phi_t \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where ϕ_t is the threshold for deciding whether to accept or reject the c-th channel of the epoch.

The evaluation process was designed so as to resemble a real-life usage scenario of the proposed approach. This was done by using the data acquired during one session for enrolment, and data acquired at another later session to test the verification performance. To also evaluate performance degradation due to time ageing, we have trained the subject models with data from the first session and independently tested them by using data from the second and third sessions.

The verification performance was evaluated in terms of Equal Error Rate (EER) and Area Under the Curve (AUC). TABLES II and III show these results for each stimulus (columns) and EEG feature (rows). The values reported correspond to the best average results, obtained by varying the thresholds ϕ_C and ϕ_t as in [24]. The average of the performance metrics across all types of stimuli per EEG feature was also computed and reported in the last column of TABLES II and III, in order to provide an indication of the overall performance of the different types of features. Results presented are not conclusive with regard to the most convenient features, as their performance varies along the different types of stimuli. However, the type of stimuli seems to have a higher influence on verification performance. In particular, affective image stimuli (AS) and resting-state with the eyes closed (RC) seem to perform slightly better.

A clearer and more interesting effect can also be observed by comparing the last column of the two tables, if we take into

TABLE IV: Subject identification accuracy for different acquisition protocols

Stimulus	ARRC	SPEC	MFCC
AS	0.2320	0.3593	0.4025†
MC	0.2207	0.3392†	0.2641
RC	**0.3100**	**0.3790**	**0.4779**†
RO	0.1006	0.2306†	0.1583
VC3	0.1429	0.3117†	0.3117†
VC5	0.1223	0.2314	0.2926†
VC7	0.1637	0.0664	0.2434†
VC10	0.0873	0.1179	0.2140†
VF3	0.2254	0.2910	0.2992†
VF5	0.2061	0.2193	0.2719†
VF7	0.1397	0.1747†	0.1703
VF10	0.1921	0.1747	0.2838†

Notes: Values in **bold** refer to the best accuracy per feature. † denotes the best accuracy per stimulus.

account that the difference in the time elapsed between training and test in TABLE III is double than the one in TABLE II. The average performance along all different types of stimuli decreases consistently with time, and suggests the existence of an ageing effect that has a negative impact on performance as the time elapsed between enrolment and test measurements increases.

C. Baseline subject identification evaluation

In addition to the presented verification results, we have also used the proposed dataset for biometric identification. Subject identification was modelled as a multi-class classification problem with one class per participant, using the 5 s EEG epochs previously created as the input. Data from sessions 1 and 2 were used for training and data from session 3 were used to test the trained models. Such setting maximises the amount of training data without simultaneously using data from a same session for training and test, and also benefits from the *incremental learning* effect reported in [29] by having

training data from more than one session. For each of the stimuli and each of the features, one multi-class ensemble classifier was trained and fine-tuned using Matlab's (R2016b) built-in hyperparameter optimisation, which selects the best hyperparameters, including the ensemble aggregation method. Features in this experiment were extracted directly from each of the epochs, instead of dividing the epochs into frames.

Results for the subject identification experiments are displayed in TABLE IV. From this table, it can be observed that MFCC features provide the best performance for the majority of stimuli. This is true for all cases except MC, RO, and VF7, for which the SPEC features provided the highest accuracy. ARRC features did not provide the best results in any case.

The highest subject identification accuracy reached 47.79% for the MFCC features and the RC stimulus. The second highest accuracy was 40.25%, for the MFCC features and the AS stimulus. Furthermore, the second best identification performance for all stimuli other than VC7 and VF10 was achieved using the SPEC features, with the ARRC features providing the worst performance for all stimuli except for VC7 and VF10. A Wilcoxon's signed rank test between the results for the MFCC and the ARRC features resulted in a $p \ll 0.05$ but was not conclusive when comparing the results for MFCC and SPEC ($p = 0.1016$). On the contrary, a comparison between the results for ARRC and SPEC resulted in a statistically significant difference ($p < 0.05$).

It is worth pointing out that the highest identification accuracy for each type of feature (MFCC, ARRC, SPEC) was always achieved for the resting-state with eyes closed (RC) stimulus, leading to an accuracy considerably higher than under the second best performing stimulus for each type of features (+7.54%, +1.97%, +7.8% respectively). Furthermore, the second best identification accuracy for each type of features was achieved when using the images as the stimulus (AS), demonstrating its superiority compared to other more commonly used stimuli.

The subject identification procedure was then repeated separately for each EEG channel, using only the features computed from each channel for training and testing the machine learning models. Unfortunately, no consistent conclusions could be extracted regarding each channel's performance using the three examined features. Consistent to our findings shown in TABLE IV, the highest identification accuracies were achieved using the resting-state with eyes closed (RC) stimulus for the vast majority of channels and features, reaching a highest accuracy of 0.4027 for channel 1 (AF3) using the MFCC features, 0.2544 for channel 7 (O1) using the ARRC features, and 0.3790 for channel 9 (P8) and the SPEC features. It is evident that the highest single-channel accuracies for each feature type are lower than the highest accuracies achieved using all the EEG channels, as shown in TABLE IV.

VI. Conclusions

In this work, we have introduced BED, a new dataset for EEG-based biometrics that takes into consideration the specific characteristics of the biometric context. This dataset contains EEG recordings from 21 different individuals when using 12 different stimuli, captured along three different recording sessions, each separated across time by a week. Both the raw signals and the features used in this paper are provided as part of the dataset. Preliminary results have been evaluated in two typical biometric scenarios, namely verification and identification. In the first of these tasks, the best results were obtained when using resting-state with eyes closed and images as the stimuli, with an AUC above 0.7 in most cases. For identification, best results were achieved when using the resting-state with eyes closed stimulus, independently of the type of features; and MFCC features provided in general better performance, showing the best accuracy for 9 out of 12 stimuli. In general, these results suggest a better performance of the resting-state with eyes closed stimulus in both verification and identification tasks.

The presented dataset and baseline results provide researchers with a valuable tool to evaluate their proposals and develop other characterisations for EEG signals that help in identifying patterns that are closer related to the individual. In addition, the dataset also supports attempts to consider the influence of emotions in the EEG signal, by using image-based stimuli and self-reported emotional labels related to the emotion that the participants experienced. Furthermore, the BED dataset will allow researchers to study multiple aspects of low-cost EEG biometrics, such as the effects of template ageing in relation to different stimuli for signal acquisition, the effects of different stimuli and/or the emotional state of the individual on identification/verification performance, the effects of different stimuli and emotional state on EEG signals, and others.

Nevertheless, EEG-based biometrics is at its early stages and there are still many open questions that will have to be answered before it becomes a practical proposition. For example, whether EEG biometrics is sufficiently reliable to recognise a person after months or years from when the system was trained is an issue which is not currently supported by any existing dataset. Our dataset provides some support to evaluate template ageing, but this is limited to a relatively short period of time (two weeks). Some previous research works also support that template ageing does occur, and report a larger decrease in system performance for longer periods of time, e.g., [26], [82]. However, further research is required to quantify the potential impact of ageing on the applicability of EEG biometrics when a large time span between training and recognition is expected.

References

[1] ISO Central Secretary, "Information technology — Vocabulary — Part 37: Biometrics," International Organization for Standardization,

Geneva, CH, Standard ISO/IEC 2382-37:2017, 2017. [Online]. Available: https://www.iso.org/standard/66693.html

[2] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, and C. Busch, "Demographic bias in biometrics: A survey on an emerging challenge," *IEEE Trans. Technol. Soc.*, vol. 1, no. 2, pp. 89–103, May 2020.

[3] K. Revett, "Cognitive biometrics: A novel approach to person authentication," *Int. J. Cognitive Biometrics*, vol. 1, no. 1, pp. 1–9, 2012.

[4] E. Maiorana, D. La Rocca, and P. Campisi, "Eigenbrains and eigentensorbrains: Parsimonious bases for EEG biometrics," *Neurocomputing*, vol. 171, pp. 638–648, Jan. 2016.

[5] R. Das, E. Maiorana, and P. Campisi, "EEG biometrics using visual stimuli: A longitudinal study," *IEEE Signal Process. Lett.*, vol. 23, no. 3, pp. 341–345, Mar. 2016.

[6] G. Hine, E. Maiorana, and P. Campisi, "Resting-state EEG: A study on its non-stationarity for biometric applications," in *BIOSIG*, Darmstadt, Germany, 2017.

[7] S. Yang and F. Deravi, "On the effectiveness of EEG signals as a source of biometric information," in *IEEE EST*, Sep. 2012, pp. 49–52.

[8] P. Arnau-González, M. Arevalillo-Herráez, S. Katsigiannis, and N. Ramzan, "On the influence of affect in EEG-based subject identification," *IEEE Trans. Affect. Comput.*, 2018, Early Access. DOI: 10.1109/TAFFC.2018.2877986.

[9] M. Arevalillo-Herráez, M. Cobos, S. Roger, and M. Garcí a Pineda, "Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from EEG signals," *Sensors*, vol. 19, no. 13, 2019.

[10] F. Gondesen, M. Marx, and D. Gollmann, *EEG-Based Biometrics*. Cham: Springer International Publishing, 2019, pp. 287–318.

[11] 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.

[12] 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 J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 888–896, 2017.

[13] 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?" *Sleep Breath*, vol. 22, p. 1179–1188, Oct 2018.

[14] A. Yadollahpour, M. Mirzaiyan, and S. Rashidi, "Quantitative EEG for early and differential diagnosis of bipolar disorders: a comprehensive review of the literature," *Int. J. Ment. Health Ad.*, vol. 15, no. 2, pp. 387–393, 2017.

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

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

[17] B. Armstrong, M. Ruiz-Blondet, N. Khalifian, K. 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.

[18] K. P. Thomas, A. P. Vinod, and N. Robinson, "Online biometric authentication using subject-specific band power features of EEG," in *ICCSP*. New York, NY, USA: ACM, 2017, pp. 136–141.

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

[20] 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.

[21] S. Yang and F. Deravi, "On the usability of electroencephalographic signals for biometric recognition: A survey," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 6, pp. 958 – 969, 2017.

[22] 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 Trans. Inf. Forensics Security*, vol. 11, no. 12, pp. 2635–2647, Dec 2016.

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

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

[25] M. Kostílek and J. Štástný, "EEG biometric identification: Repeatability and influence of movement-related EEG," in *AE*. IEEE, Sept 2012, pp. 147–150.

[26] M. V. Ruiz-Blondet, S. Laszlo, and Z. Jin, "Assessment of permanence of non-volitional EEG brainwaves as a biometric," in *IEEE ISBA*, Mar 2015, pp. 1–6.

[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 Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 738–742, Feb 2007.

[29] S. Marcel and J. D. R. Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 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 *FC*. Okinawa, Japan: Springer, 2013, pp. 1–16.

[31] K. Brigham and B. V. K. V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *IEEE BTAS*, Washington, DC, USA, 2010, pp. 1–8.

[32] P. Arnau-González, S. Katsigiannis, N. Ramzan, D. Tolson, and M. Arevalillo-Herráez, "ES1D: A deep network for EEG-based subject identification," in *IEEE BIBE*, 2017, pp. 81–85.

[33] P. Arnau-González, S. Katsigiannis, M. Arevalillo-Herráez, and N. Ramzan, "Image-evoked affect and its impact on eeg-based biometrics," in *IEEE ICIP*, 2019.

[34] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, "Toward EEG sensing of imagined speech," in *Human-Computer Interaction. New Trends*, J. A. Jacko, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 40–48.

[35] H. Begleiter, "Eeg dataset," Oct 1999. [Online]. Available: http://kdd.ics.uci.edu/databases/eeg/eeg.data.html

[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," *IEEE Trans. Affect. Comput.*, 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 Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan 2012.

[38] 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 J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 98–107, Jan 2018.

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

[40] P. Lakhan, N. Banluesombatkul, V. Changniam, R. Dhithijaiyratn, P. Leelaarporn, E. Boonchieng, S. Hompoonsup, and T. Wilaiprasitporn, "Consumer grade brain sensing for emotion recognition," *IEEE Sens. J.*, vol. 19, no. 21, pp. 9896–9907, 2019.

[41] L. A. Moctezuma and M. Molinas, "Subject identification from low-density EEG-recordings of resting-states: A study of feature extraction and classification," in *FICC*. Springer, 2019, pp. 830–846.

[42] L. Ma, J. W. Minett, T. Blu, and W. S. Wang, "Resting state EEG-based biometrics for individual identification using convolutional neural networks," in *IEEE EMBC*, 2015, pp. 2848–2851.

[43] Y. Di, X. An, F. He, S. Liu, Y. Ke, and D. Ming, "Robustness analysis of identification using resting-state EEG signals," *IEEE Access*, vol. 7, pp. 42 113–42 122, 2019.

[44] D. Kim and K. Kim, "Resting state EEG-based biometric system using concatenation of quadrantal functional networks," *IEEE Access*, vol. 7, pp. 65 745–65 756, 2019.

[45] Z. A. A. Alyasseri, A. T. Khader, M. A. Al-Betar, J. P. Papa, and O. ahmad Alomari, "EEG-based person authentication using multi-objective flower pollination algorithm," in *IEEE CEC*, 2018.

[46] I. Jayarathne, M. Cohen, and S. Amarakeerthi, "BrainID: Development of an EEG-based biometric authentication system," in *IEEE IEMCON*, Oct 2016, pp. 1–6.

[47] A. Mansfield, "Information technology–biometric performance testing and reporting–part 1: Principles and framework," *ISO/IEC*, pp. 19 795–1, 2006.

[48] E. Maiorana, D. L. Rocca, and P. Campisi, "On the permanence of EEG signals for biometric recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 163–175, Jan 2016.

[49] T. Koike-Akino, R. Mahajan, T. K. Marks, Y. Wang, S. Watanabe, O. Tuzel, and P. Orlik, "High-accuracy user identification using EEG biometrics," in *IEEE EMBC*. IEEE, 2016, pp. 854–858.

[50] 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 IECBES*. IEEE, 2010, pp. 303–306.

[51] V. Goverdovsky, D. Looney, P. Kidmose, and D. P. Mandic, "In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring," *IEEE Sens. J.*, vol. 16, no. 1, pp. 271–277, 2015.

[52] P. Sarma and S. Barma, "Review on stimuli presentation for affect analysis based on EEG," *IEEE Access*, vol. 8, pp. 51 991–52 009, 2020.

[53] E. Dan-Glauser and K. Scherer, "The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance," *Behav. Res. Methods*, vol. 43, no. 2, pp. 468–477, 2011.

[54] B. Kurdi, S. Lozano, and M. Banaji, "Introducing the open affective standardized image set (OASIS)," *Behav. Res. Methods*, vol. 49, no. 2, pp. 457–470, 2017.

[55] K. H. Chiappa, *Evoked potentials in clinical medicine*. Lippincott Williams & Wilkins, 1997.

[56] I. Bodis-Wollner and M. D. Yahr, "Measurements of visual evoked potentials in Parkinson's disease," *Brain*, vol. 101, no. 4, pp. 661–671, Dec 1978.

[57] J. V. Odom, M. Bach, C. Barber, M. Brigell, M. F. Marmor, A. P. Tormene, and G. E. Holder, "Visual evoked potentials standard (2004)," *Doc. Ophthalmol.*, vol. 108, no. 2, pp. 115–123, 2004.

[58] R. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles, "The electroencephalogram as a biometric," in *IEEE CCECE*, vol. 2, 2001, pp. 1363–1366.

[59] D. La Rocca, P. Campisi, B. Vegso, P. Cserti, G. Kozmann, F. Babiloni, and F. D. V. Fallani, "Human brain distinctiveness based on EEG spectral coherence connectivity," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 9, pp. 2406–2412, 2014.

[60] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur, "Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs," *PeerJ*, vol. 1, p. e38, Feb. 2013.

[61] J. N. Acharya, A. Hani, J. Cheek, P. Thirumala, and T. N. Tsuchida, "American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature," *J. Clin. Neurophysiol.*, vol. 33, no. 4, pp. 308–311, Aug 2016.

[62] H. Ekanayake. (2015, January) P300 and Emotiv EPOC: Does Emotiv EPOC capture real EEG? (Accessed 13 Dec. 2019). [Online]. Available: http://neurofeedback.visaduma.info/EmotivResearch.pdf

[63] T. Althobaiti, S. Katsigiannis, D. West, and N. Ramzan, "Examining human-horse interaction by means of affect recognition via physiological signals," *IEEE Access*, vol. 7, pp. 77 857–77 867, 2019.

[64] J. A. Miranda Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE Trans. Affect. Comput.*, pp. 1–1, 2018, (Early Access).

[65] J. Klonovs, C. K. Petersen, H. Olesen, and A. Hammershoj, "ID proof on the go: Development of a mobile EEG-based biometric authentication system," *IEEE Veh. Technol. Mag.*, vol. 8, no. 1, pp. 81–89, Mar 2013.

[66] K. Reshmi, P. I. Muhammed, V. Priya, and V. Akhila, "A novel approach to brain biometric user recognition," *Proc. Tech.*, vol. 25, pp. 240–247, 2016.

[67] R. S. Harshit, K. P. Thomas, K. G. Smitha, and A. P. Vinod, "Online electroencephalogram (EEG) based biometric authentication using visual and audio stimuli," in *IEEE IECBES*, Dec 2016, pp. 454–459.

[68] Emotiv, "Emotiv EPOC technical specifications," [Online] https://emotiv.gitbook.io/epoc-user-manual/introduction-1/technical_specifications, 2020.

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

[70] A. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behav. Res. Methods*, vol. 45, p. 1191–1207, Dec 2013.

[71] M. G. Kendall and B. B. Smith, "The problem of m rankings," *Ann. Math. Stat.*, vol. 10, pp. 275–287, 1939.

[72] P. Legendre, "Coefficient of concordance," in *Encyclopedia of Research Design*, N. J. Salkind, Ed. SAGE Publications, Inc., 2010, p. 164–169.

[73] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[74] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K. Su, and K. Robbins, "The PREP pipeline: standardized preprocessing for large-scale EEG analysis," *Front. Neuroinform.*, vol. 9, p. 16, 2015.

[75] C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk, "Speech recognition using MFCC," in *ICGSM*, Pattaya, Thailand, 2012, pp. 28–29.

[76] L. Juvela, B. Bollepalli, X. Wang, H. Kameoka, M. Airaksinen, J. Yamagishi, and P. Alku, "Speech waveform synthesis from mfcc sequences with generative adversarial networks," in *ICASSP*, Alberta, Canada, April 2018, pp. 5679–5683.

[77] P. Nguyen, D. Tran, X. Huang, and D. Sharma, "A proposed feature extraction method for EEG-based person identification," in *ICAI*, Las Vegas, NV, USA, 2012, pp. 826–831.

[78] E. Piciucco, E. Maiorana, O. Falzon, K. Camilleri, and P. Campisi, "Steady-state visual evoked potentials for EEG-based biometric identification," in *BIOSIG*. Darmstadt, Germany: Gesellschaft für Informatik, Bonn, 2017.

[79] S. Jawed, H. U. Amin, A. S. Malik, and I. Faye, "Classification of visual and non-visual learners using electroencephalographic alpha and gamma activities," *Front. Behav. Neurosci.*, vol. 13, p. 86, 2019.

[80] J. Monge-Alvarez, C. Hoyos-Barcelo, L. M. San Jose-Revuelta, and P. Casaseca-de-la Higuera, "A machine hearing system for robust cough detection based on a high-level representation of band-specific audio features," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2319–2330, Aug. 2019.

[81] K. P. Murphy, *Machine learning: A probabilistic perspective*. MIT press, 2012.

[82] B. Kaur, P. Kumar, P. P. Roy, and D. Singh, "Impact of ageing on EEG based biometric systems," in *ACPR*. Springer, 2017, pp. 459–464.