Chapter 11

# EEG-based biometrics: Effects of template ageing

*Pablo Arnau-González[1], Stamos Katsigiannis[1], Miguel Arevalillo-Herraez[2], and Naeem Ramzan[1]*

This chapter discusses the effects of template ageing in EEG-based biometrics. The chapter also serves as an introduction to general biometrics and its main tasks: Identification and verification. To do so, we investigate different characterisations of EEG signals and examine the difference of performance in subject identification between single session and cross-session identification experiments. In order to do this, EEG signals are characterised with common state-of-the-art features, i.e. Mel Frequency Cepstral Coefficients (MFCC), Autoregression Coefficients, and Power Spectral Density-derived features. The samples were later classified using various classifiers, including Support Vector Machines and k-Nearest Neighbours with different parametrisations. Results show that performance tends to be worse for cross-session identification compared to single session identification. This finding suggests that temporal permanence of EEG signals is limited and thus more sophisticated methods are needed in order to characterise EEG signals for the task of subject identification.

## 11.1 Introduction

The ever increasing use of information systems and digital locking mechanisms that safeguard access to critical information and infrastructure requires the use of sophisticated security measures for restricting access to only authorised users. Techniques for authenticating users vary from the more traditional approaches of using usernames and passwords, to requiring specific hardware such as key passes or security tags, and to two-step authentication procedures and biometrics. The utilisation of biometric data has been broadly contemplated in the security domain, as an approach that meets the security requirements of such systems. Biometric modalities include fingerprints, iris, voice, face, or other physiological traits, such as, the electrocardiogram (ECG) [1, 2].

[1]School of Computing, Engineering and Physical Sciences, University of the West of Scotland, UK
[2]Departamento de Informatica, Universitat de Valencia, Spain

Due to the already established connection between an electroencephalography (EEG) signal and the person, EEG-based subject identification has also attracted attention in the subject identification area over the past 10 years [3, 4]. EEG signals encode information about the affective and mental state of a person and have therefore been widely studied in a multitude of applications, e.g. early diagnosis of Alzheimer's disease [5], detection of epilepsy episodes [6], assessment of user experience [7] and identification of individual emotional status and reactions [8, 9, 10].

Many studies in the field of EEG-based biometrics have been published lately; [11] and [12] provided a thorough overview of its opportunities and theoretical considerations. It is evident across the literature that the EEG signal procurement protocol is critical to the efficacy of an EEG-based biometrics system, as the identification content within the brain signal may be influenced by the task in which the subjects are involved. Subjects are generally asked to conduct particular tasks or are subjected to predefined stimuli (e.g. pictures or videos). The use of different tasks, typically resting state, audio-visual impulses (sensory activity), or cognitive tasks that are frequently suggested in the literature [13], has been studied by Ruiz-Blondet et al. [14] and Maiorana and Campisi [15]. Imagined speech [16] and custom tasks [17], as well as the use of event-related potentials (ERPs) [18] were also used and resulted in increased accuracy and stability over time (increased permanence).

Attempts to identify individuals by only recording one session of data are commonly found in the literature [14, 19, 20, 21, 22]. These approaches typically have disregarded the issue of the permanence in EEG signals. Some studies have argued that there is not a significant decrease in performance, therefore it is not necessary to record data in more than one session [22, 23]. However, recent studies in the field have shown that these approaches are erroneous [15, 24] and that there is a significant difference in performance when taking into account the degradation of the template quality over time. The effects of template ageing on EEG-based biometrics are examined in this chapter using a dataset created across various sessions spaced one week apart each.

The rest of this chapter is organised in five sections. General background in biometrics and electroencephalography is provided in Section 11.2. The data acquisition protocol, the data pre-processing, the feature extraction, and the classification methodology are described in Section 11.3, while Section 11.4 provides the experimental procedure and acquired results. Finally, conclusions are drawn in Section 11.5.

## 11.2 Background

### *11.2.1 Biometrics*

The ability to unequivocally identify individuals and associate personal traits with a subject has been of high relevance in history. Humans typically use corporal traits, such as face, voice or gait, together with other contextual information in order to recognise others. The set of attributes associated with a person constitutes their personal identity.

The handling of personal identity plays a critical role in a variety of applications. Examples of such applications are border crossing management, physical access restrictions to certain facilities, such as power plants or airports, access control to shared resources, performing financial transactions or others, The proliferation of web resources and the deployment of non-centralized services have highlighted the relevance of the risk of identity theft, now *vox populi*, becoming a topic of interest for users, increasing the demand for secure systems for the handling of personal identities.

A biometric system measures one or more physical or behavioural characteristics, including, but not limited to, fingerprint, face, iris, ear, voice, odour, or even the DNA of an individual, in order to determine their identity. This characteristics are referred to in different terms, for instance, traits, indicators, identifiers or modalities. Biometric systems are usually employed for two distinct tasks: a) verification/authentication and b) identification.

#### 11.2.1.1 Verification/Authentication

In verification, the user takes an identity and the system verifies if the user is truly whoever they say they are. In this scenario, the query is compared only against the corresponding template of the requested identity. The identity is usually stored under a Personal Identification Number (PIN), a username, or a token. If the user input matches or in some cases has a high enough similarity with the template of the requested identity, then the request and therefore the user, is considered legitimate and the user is verified/authenticated. Otherwise, if the samples are not similar enough, the request is rejected and the user is considered as an impostor. Verification is very frequent in applications whose goal is to stop unauthorised users from using a service or accessing a place.

Formally, verification can be posed as the following binary problem: Let $I$ be a requested identity and $R$ be the user input during a request for the identity $I$. The decision that the system needs to take is that $R$ is similar enough to $I$ for the user to be considered either legitimate or an impostor. The decision rule will then be:

$$D_{(I,R)} = \begin{cases} 1 & \text{if } s_{I,R} \leq \eta \\ 0 & \text{if } s_{I,R} > \eta \end{cases} \tag{11.1}$$

Where $s_{I,R}$ is the similarity score and $\eta$ the threshold for acceptance/rejection of the input.

#### 11.2.1.2 Identification

Identification can be divided into positive and negative identification. In positive identification, the user tries to identify himself in the system without explicitly assuming an identity. A positive identification system answers the question ”Are you someone known by the system?”, obtaining the answer from a set of stored profiles. On the other hand, in negative identification, the user is assigned with an identity, and the system assumes it to be correct. The system then tries to determine whether the user is not the person whose identity has been assigned. The purpose of this kind of identification is to stop users from assuming multiple identities. A clear example of this application would be a system that decides if a certain person is presenting a false passport and the profile matches that stored in a watch-list. Independently of the type of identification (positive or negative) the user input is compared with all the profiles stored in the database and the system assigns an identity based on the similarity between the input and the template.

Formally, the problem of identification can be described as follows: Let $R$ be the user input during a request for identification. The system has to assign the identity $I$ to the user, where $I \in \{I_1, I_2, ..., I_N, I_{N+1}\}$, with $I_1$ to $I_N$ being the $N$ identities known to the system and $I_{N+1}$ corresponding to the unknown identity. The decision rule is:

$$R \in \begin{cases} I_{n_0} & \text{if } n_0 = argmax\ s_n \text{ and } s_{n_0} > \eta \\ I_{N+1} & \text{otherwise} \end{cases} \tag{11.2}$$

#### 11.2.1.3 Characteristics of biometric traits

Different biometric traits have been used in a number of applications, each of them having its own advantages and disadvantages. Therefore, the selection of a trait depends on the particular applications and their requirements, apart from performance or accuracy. In general, there are seven different facts that influence the selection of a biometric trait:

1. **Universality**: All the users must posses the trait.
2. **Uniqueness**: The trait should be sufficiently different across different subjects, being able to identify them unequivocally.
3. **Permanence**: The trait should have stability over a period of time. A trait that changes drastically over time, is not useful for biometrics.
4. **Measurability**: It must be possible to acquire the biometric trait, using available hardware, without causing too much hustle to the users.
5. **Performance**: Besides recognition accuracy, the computational cost required for the matching algorithm and the throughput should meet the application restrictions.
6. **Acceptability**: The users of the application should be willing to present the trait.
7. **Circumvention**: The easiness with which a trait can be replicated or obfuscated by an attacker.

In any case, it is not expected for a single trait to match all seven factors for all possible applications. In other words, there is no ideal trait but many are generally admissible. The relevance of a given trait to a specific application will depend upon how well the trait complies with the application requirements.

### *11.2.2 Electroencephalography (EEG)*

Electroencephalography (or EEG) is the recording of the electrical activity from the cerebral cortex, measured in microvolts ($\mu V$) [25]. The measured electrical potentials present in the scalp originate from the superposition of all the electrical fields generated by the dendrites during the synapses [26]. The recorded signals have been used in a number of applications, although mostly for medical purposes, such as seizure detection [27]. For the acquisition of conventional scalp EEG, electrodes are placed on the scalp of the person on locations specified by the *International 10-20*

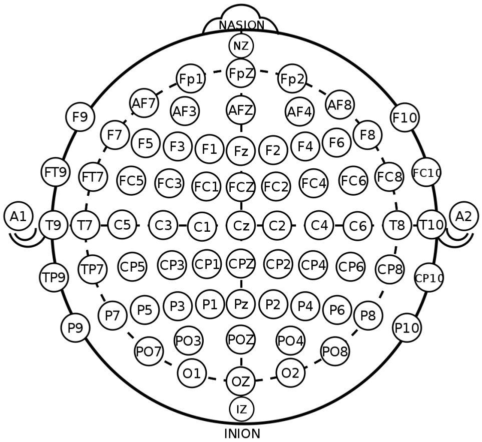


Figure 11.1: International 10-20 system for EEG electrode placement. *Source: wikimedia.org. Published under the Creative Commons CC0 1.0 Universal Public Domain Dedication licence.*

*system*. The 10-20 system is the name given to the standardisation of the names of the electrodes and their placement on the scalp [28], as shown in Figure 11.1. The system gets its name from the separation of the adjacent electrodes that are separated by a 10% or 20% of the total front-back or right-left distance of the skull. The 10-20 system names the electrodes according to the positioning in the scalp, with the following naming system, *XXN*, where *XX* refers to the lobe, being Fp, F, C, O, T, P, referring to Pre-Frontal, Frontal, Central, Occipital, Temporal, and Parietal respectively, and *N* referring to the positioning, where odd numbers refer to the left hemisphere and even numbers to the right hemisphere. It is also possible to find Z electrodes. In this case, the electrodes are placed on the central part of the scalp of their corresponding lobe. Typical reference electrodes are placed in Cz or Tz. The system has been extended for higher resolution strategies, including the following regions, AF, FC, FT, CP, TP, PO, being those regions placed in the middle. For example, according to that convention, AF would be placed between the Anterior and Frontal lobes.

### *11.2.3 EEG in biometrics*

EEG signals have been traditionally restricted to the field of medicine. However in recent years and with the use of machine learning technologies, EEG signals have been used for an ever-increasing pool of applications, mainly focused on Brain-Computer Interfaces (BCI). More recently, researchers showed interest in the uniqueness of EEG for each individual and attempted to create biometric systems based on EEG signals [19, 28, 20]. Available works have typically disregarded the influence of the so called *template ageing*, an effect that reflects how a given biometric trait changes over a period of time. However, as some studies suggest [15], EEG signals not only change with the passage of time, but the characterisations of said signals are also time-dependent. In this chapter, we will compare both approaches and study how the systems behave in the case of single-session approaches versus multi-session approaches, in order to examine the effects of template ageing in EEG-based biometrics.

## 11.3 Data acquisition and experimental protocol

A dataset with EEG recordings belonging to different people over a time period of three weeks was captured in order to test the temporal permanence of EEG signals. In order to do this, different classification experiments were designed with the aim to compare the accuracy of signals in a single-session acquisition versus a plausible scenario in which data from one or more sessions is used for enrolling users in the system and a recording from any other given day is used for validating these users.

### *11.3.1 Stimuli*

Images with powerful emotional content were used to generate emotional responses to users, while recording EEG signals, according to recommendations of previous work [21]. The stimuli were acquired from two openly accessible picture datasets, i.e. the Geneva Affective Picture Dataset (GAPED) [29] and the Open Affective Standardized Image Set (OASIS) [30]. The images of both datasets are annotated in terms of the emotional response they elicit to human viewers using Russel's *Circumplex Model of Affect* [31] that considers emotion as being distributed in the two-dimensional Valence/Arousal space. Valence is a measure of the positiveness of an emotion, varying from negative to positive, while arousal is a measure of the excitement associated with an emotion, varying from low to high.

GAPED contains 730 different JPEG images. The pictures are annotated within the range [0, 100] in terms of arousal and valence. The pictures of the dataset belong to the following classifications: *snakes, spiders, natural problems (representing scenes that violate human rights), animal mistreatment (representing animal mistreatment scenes), Neutral, and Positive*. OASIS includes 900 distinct JPEG pictures. The pictures are annotated within the range [1, 7] in terms of arousal and valence. It also includes the median answers for each picture separated by gender. The OASIS dataset includes pictures from one of the following four mutually exclusive classifications: *Animals, Objects, People, and Scenes*. It should be noted that some of the pictures contain explicit sexual material, leading to the removal of image #I537 from this research.

The two datasets contain a total of 1,630 images from which 48 were selected according to their associated valence/arousal ratings in order to obtain a representative set of images with an intense emotional content. To this end, each dataset's valence/arousal values were first standardised to the range $[-1, 1]$. The resulting valence/arousal space was then uniformly split into 12 regions of $\frac{\pi}{6}$ radians, as shown in Figure 11.2. Finally, 4 pictures were selected from each region, whose (Valence,

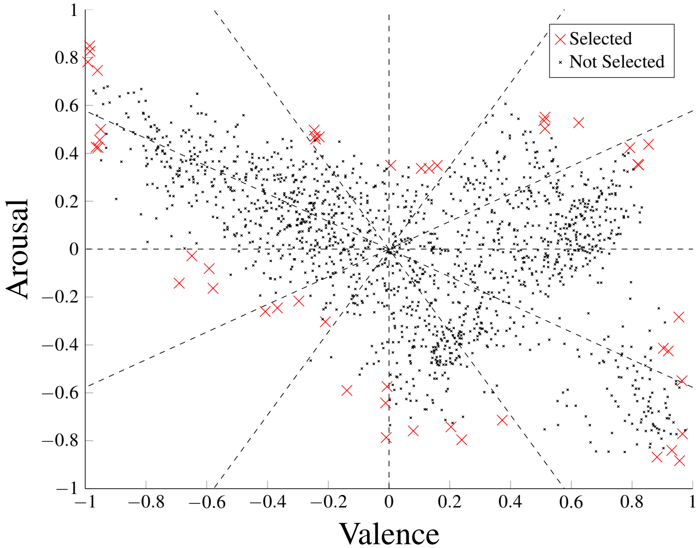


Figure 11.2: Images of the GAPED and OASIS datasets in Valence/Arousal space. Marked images were selected for this study.

Arousal) locations were farthest from the neutral emotion $(0,0)$, as they contained the most intense emotional content in that region.

## *11.3.2 Experimental protocol*

26 healthy subjects, aged between 23 and 55 years old ($\mu_{age} = 31.9$), were recruited as volunteers for data acquisition. The experiment took place inside a quiet room with ambient light and no physical supervision, in order to not alter the response of the participants and not introduce any artefacts related to stress or distractions. Before starting the experiment, the experimental procedure and the used valence-arousal scale were explained, and participants were then asked to fill a consent form indicating that they agree to participate in the study and to the viewing of images that may depict strong emotional content.

Three sessions were recorded for each participant, each of them spaced 7 days apart. Out of the four images selected per region, one was randomly selected to be displayed in all the sessions as baseline, and each of the other three was assigned to one session, leading to a total of three sets of 24 images (12 repeated and 12 unique images per set). The selected images were presented to the participants for 5 s and immediately after seeing each image, the participants were asked to report the felt emotion using Self Assessment Manikins (SAM) [32] on a 9-point Likert scale. After the self assessment, participants were asked to perform a simple mathematical operation and report the result, in order to reduce any effects of the emotional stimulus to their emotional state. The EEG activity was recorded during the whole session, using the 14-channel Emotiv EPOC+® wireless EEG device with a sampling rate of

256 Hz. Furthermore, timestamps with millisecond precision were used to synchronise the acquired EEG signals with the image stimuli viewing.

## *11.3.3 Data preparation and feature extraction*

The recorded timestamps were used in order to segment the acquired EEG recordings to segments referring to specific images. Furthermore, the EEG segments were annotated with the ID of each respective participant, as well as with the valence and arousal values reported in the study by each respective participant. Then, the EEGLAB toolbox [33] was used to pre-process the EEG signals by applying the PREP pipeline for EEG data pre-processing as described in [34], in order to remove artefacts such as the ones stemming from muscle movement, jaw clenching, or eye blinking. Then, to create a machine learning model for subject identification from EEG signals, various features were extracted from the pre-processed EEG recordings, namely the Mel Frequency Cepstral Coefficients (MFCC), Power Spectral Density (PSD), and Autoregression Reflection Coefficients (ARRC).

### 11.3.3.1 Mel Frequency Cepstral Coefficients (MFCCs)

Mel Frequency Cepstral Coefficients (MFCCs) are a parametric representation of the Fourier Spectrum and have been commonly used in voice recognition [35, 36] and more lately implemented to EEG-based person recognition [37, 38]. In this work, MFCCs were computed using HTK-like filterbanks and the Discrete Cosine Transform. MFCC features were computed using 18 filterbanks, as described in [37], generating a total of 12 cepstral coefficients per channel, after discarding the $D_C$ coefficient. The feature vector was generated by concatenating each channel's cepstral coefficients, resulting in a total of 168 features (12 coefficients $\times$ 14 channels).

### 11.3.3.2 Power Spectral Density (PSD)

Power Spectral Density (PSD)-based features have been frequently used to identify emotional states from EEG signals [39, 40]. In this work, PSD features were calculated as described in [41]: First, on each channel, the PSD is calculated using Welch's algorithm. For each channel, the PSD is calculated using a 2 s Hamming window (512 samples) with a 75% overlap (384 samples) and the FFT is generated over each of these windows and averaged across time. Finally, the feature vector was created by concatenating the resulting PSD values of the $[1-40]$ Hz frequency band. This process resulted into a total of 38 features per channel, leading to a total of 532 features (38 features $\times$ 14 channels).

### 11.3.3.3 Autoregression Reflection Coefficients (ARRCs)

Autoregression reflection Coefficients (ARRCs) have been used extensively for EEG signal analysis [42]. Recently, various research works examined their effectiveness for EEG-based biometrics [24, 37, 43]. An EEG signal can be characterised as an output of a causal stable linear time-invariant stationary AR($P$-th order) system based on the EEG spectrum's autoregressive (AR) or all-pole model, with the AR parameters being estimated using the Yule-Walker equations [44]. Individual ARRCs were obtained by estimating an AR model of order 10 for each channel. Then, the feature vector was created by combining the 10 reflection coefficients of each channel, resulting to a total of 140 features (10 coefficients $\times$ 14 channels).

### *11.3.4 Classification*

The acquired feature vectors were then used in order to create various supervised classification models for EEG-based subject identification, i.e. the prediction of the subject ID associated with a feature vector. The subject identification problem was thus modelled as a multi-class classification problem where each class referred to a specific subject's ID. The examined classification algorithms included the $k$-Nearest Neighbour ($k$NN) for $k = 1,3,5,7$, Linear Support Vector Machines (LSVM), SVM with the Radial Basis Function kernel (RSVM), SVM with a second order polynomial kernel (QSVM), and SVM with third order polynomial kernel (CSVM).

## 11.4 Experimental results

In order to show the difference in performance that occurs when performing single session or multiple session identification tasks, two different experiments were considered, both using the same features and the same classifiers. The main difference between the two experiments is in the training and testing schemes. For the first experiment, samples from the same session were used for training and testing, following a *leave-one-out* cross validation strategy. In the second experiment, a cross-session subject identification task was examined by adopting a validation strategy that identifies the samples against samples collected in the past, e.g. training with samples from the first session and testing with samples from the second or third sessions. In this case, a single or more sessions were used for training the classifiers, and data from the remaining sessions were used for testing.

### *11.4.1 Single-session subject identification*

Eight different classifiers were trained in order to evaluate the performance of the extracted features in a hypothetical identification system. The system in this case is trained and tested with samples extracted from the same recording session. In Table 11.1, the accuracy of the different classifiers is displayed for the different Sessions recorded in this dataset and the different features computed. The highest accuracy of 0.885 was achieved using the LSVM classifier with the MFCC features. For all sessions, the combination of MFCC features and the LSVM classifier provided the highest classification accuracy. This shows that MFCC features provided the best characterisation of the individuals out of the three features examined. This fact is also supported by MFCC features achieving in general higher performance than the other features regardless of the classifier ($p < 0.05$), with the exception of the CSVM, where ARRC features performed considerably better than both MFCC and PSD features. The performance of the different classifiers seems more dependent on the features than on the training session. Although, as can be seen in Table 11.1,

*Table 11.1 Classification accuracy for single-session subject identification*

| Session | Features | LSVM | RSVM | QSVM | CSVM | 1-NN | 3-NN | 5-NN | 7-NN |
|---|---|---|---|---|---|---|---|---|---|
| 1 | MFCC | 0.885 | 0.848 | 0.868 | 0.052 | 0.855 | 0.835 | 0.798 | 0.756 |
| | PSD | 0.351 | 0.032 | 0.050 | 0.036 | 0.457 | 0.426 | 0.415 | 0.404 |
| | ARRC | 0.378 | 0.239 | 0.409 | 0.399 | 0.216 | 0.242 | 0.220 | 0.205 |
| 2 | MFCC | 0.849 | 0.817 | 0.778 | 0.047 | 0.833 | 0.806 | 0.768 | 0.754 |
| | PSD | 0.363 | 0.032 | 0.041 | 0.036 | 0.457 | 0.442 | 0.410 | 0.371 |
| | ARRC | 0.420 | 0.308 | 0.473 | 0.429 | 0.284 | 0.293 | 0.310 | 0.318 |
| 3 | MFCC | 0.750 | 0.742 | 0.689 | 0.040 | 0.717 | 0.683 | 0.661 | 0.645 |
| | PSD | 0.383 | 0.033 | 0.098 | 0.040 | 0.441 | 0.398 | 0.392 | 0.364 |
| | ARRC | 0.408 | 0.283 | 0.433 | 0.415 | 0.255 | 0.288 | 0.256 | 0.250 |

*Table 11.2 Classification accuracy for cross-session subject identification*

| Train Session | Test Session | Features | LSVM | RSVM | QSVM | CSVM | 1-NN | 3-NN | 5-NN | 7-NN |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | MFCC | 0.112 | 0.130 | 0.110 | 0.037 | 0.130 | 0.144 | 0.151 | 0.146 |
| | | PSD | 0.118 | 0.039 | 0.041 | 0.036 | 0.075 | 0.058 | 0.071 | 0.062 |
| | | ARRC | 0.070 | 0.050 | 0.083 | 0.083 | 0.050 | 0.052 | 0.068 | 0.062 |
| | 3 | MFCC | 0.195 | 0.198 | 0.213 | 0.040 | 0.253 | 0.261 | 0.266 | 0.261 |
| | | PSD | 0.132 | 0.040 | 0.063 | 0.040 | 0.116 | 0.144 | 0.134 | 0.126 |
| | | ARRC | 0.102 | 0.078 | 0.119 | 0.111 | 0.060 | 0.074 | 0.119 | 0.122 |
| 2 | 3 | MFCC | 0.198 | 0.187 | 0.215 | 0.030 | 0.213 | 0.208 | 0.238 | 0.235 |
| | | PSD | 0.132 | 0.031 | 0.046 | 0.040 | 0.107 | 0.126 | 0.137 | 0.131 |
| | | ARRC | 0.091 | 0.066 | 0.091 | 0.091 | 0.078 | 0.078 | 0.079 | 0.081 |
| 1-2 | 3 | MFCC | 0.337 | 0.311 | 0.334 | 0.040 | 0.276 | 0.279 | 0.261 | 0.264 |
| | | PSD | 0.071 | 0.040 | 0.015 | 0.040 | 0.147 | 0.164 | 0.174 | 0.167 |
| | | ARRC | 0.136 | 0.111 | 0.154 | 0.149 | 0.081 | 0.089 | 0.096 | 0.106 |

CSVM was not capable to properly model the individuals with the exception of when the ARRC features were used.

## *11.4.2 Cross-session subject identification*

Following a similar approach as in the previous experiment, the different classification algorithms were trained using the features extracted from one or more sessions. The difference in this case is that the trained models were tested against the samples from future acquisition sessions, i.e. sessions recorded later. This process led to a total of 4 possible experiments. Table 11.2 shows the identification accuracy achieved for the different classification experiments. The highest accuracy of 0.3370 was achieved using the MFCC features and the LSVM classifier when training with sessions 1 and 2 and testing with session 3. It is not clear why the accuracy increases so much in this case compared to when training with the first session and testing with the second or third sessions, or when training with the second session and testing with the third. It may be attributed to the different algorithms being able to generalise more knowledge, or simply because there are more data, therefore the border between the different subjects is more defined. In general, MFCC features provided better accuracy than the other tested features regardless the time elapsed between the training and test session ($p < 0.05$). Regarding the performance of the different clas-

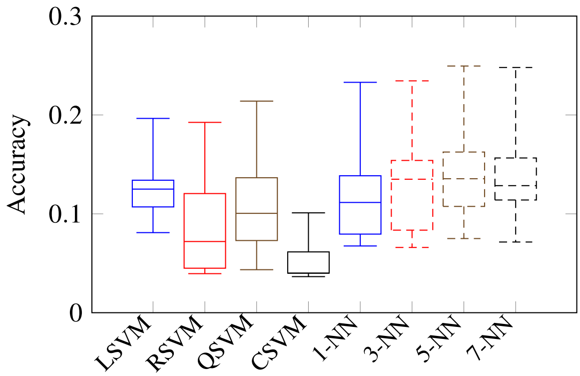


Figure 11.3: Distribution of classification accuracy for the different classifiers in the cross-session approach.

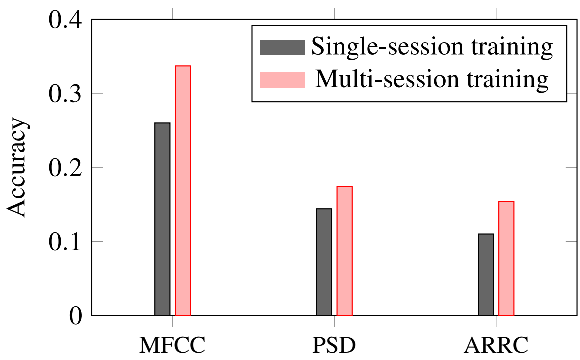


Figure 11.4: Maximum classification accuracy obtained for each of the studied features when using one session for training vs using multiple sessions for training.

sifiers, as shown in Figure 11.3, the Nearest Neighbour classifiers provide generally better results compared to the performance of the SVM-based classifiers.

As previously noted in [45], the performance of the identification task is significantly increased when training with more than one session, an effect called *incremental learning*. The effects of incremental learning are displayed in Figure 11.4 where the maximum accuracy of the single-session training classification experiments is compared to the accuracy of the multiple-session classification experiments. As shown in that figure, the effects of incremental learning are evident regardless of the features employed, reinforcing the conclusions drawn in [45]. Moreover, in Figure 11.5, the distribution of results for the different classifiers in the case of the single-session learning is displayed against the case of incremental learning. That figure further demonstrates the potential benefits of incremental learning.

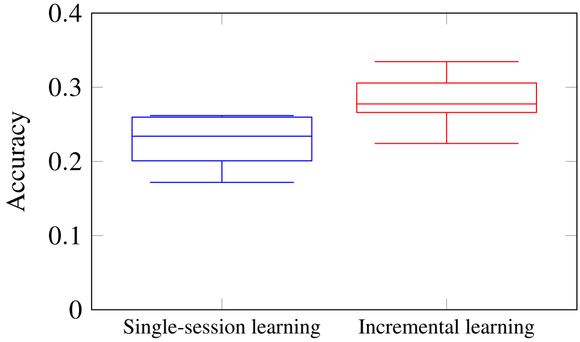


Figure 11.5: Mean classification accuracy using single-session learning vs incremental learning (multi-session learning).

## 11.5 Conclusions

The correct identification of individuals from EEG signals is still a challenge that many researchers are working to solve. A good solution would provide a very convenient alternative for identifying individuals. Since EEG signals cannot be captured at a distance, it is extremely challenging to capture furtively the EEG of a given individual with malicious intentions. This trait is highly desired in critical applications involving very sensitive information. However, there is still much research to be done before this biometric modality can be exploited in practical applications. Typically, researchers have focused on identification in only one session. However, from the presented results it is clear that EEG signals are heavily affected by session-specific noise and phenomena. A number of artefacts are present in the EEG recordings, such as the precise electrode positioning (EEG signal caps are not always fitted in the exact same position), line or other background-specific noises. The results provide evidence on why the developed methods should be tested in a multi-session scenario, since it is not possible to generalise the results obtained from a single-session classification approach. Good performance in that scenario cannot guarantee good performance later in time. Moreover, the presented results show that the accuracy of the identification task will be worse in any given case in a multi-session environment. Furthermore, results also show that from the examined representations, MFCC modelling of EEG signals performs consistently better in both scenarios regardless of the experimental setup, a finding that is consistent with similar studies [21, 15], although the reported accuracy in those studies is considerably higher. This may be due to the different EEG recording devices, since a consumer non-medical-grade device was used in this study, while the cited studies used medical-grade devices. The difference in performance between this and other works suggests that consumer-grade EEG devices have poorer quality compared to medical-grade devices, hence robust features are necessary for lower quality devices. Nevertheless, the phenomenon of *incremental learning*, present in medical-grade recording devices, also appears in the proposed low-cost setup, showing that the development of more robust feature extraction approaches can potentially allow the use of consumer-grade EEG devices in practical EEG-based biometrics systems.

## References


[1] Bowyer KW, Hollingsworth K, Flynn PJ. Image understanding for iris biometrics: A survey. Computer vision and image understanding. 2008;110(2):281–307.

[2] Louis W, Komeili M, Hatzinakos D. Continuous Authentication Using One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP) in ECG Biometrics. IEEE Transactions on Information Forensics and Security. 2016;11(12):2818–2832.

[3] Ruiz-Blondet MV, Jin Z, Laszlo S. CEREBRE: A novel method for very high accuracy event-related potential biometric identification. IEEE Transactions on Information Forensics and Security. 2016;11(7):1618–1629.

[4] Palaniappan R, Mandic DP. Biometrics from brain electrical activity: A machine learning approach. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007;29(4):738–742.

[5] Cichocki A, Shishkin SL, Musha T, et al. EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease. Clinical Neurophysiology. 2005;116(3):729–737.

[6] Kannathal N, Choo ML, Acharya UR, et al. Entropies for detection of epilepsy in EEG. Computer methods and programs in biomedicine. 2005;80(3):187–194.

[7] Arnau-Gonzalez P, Althobaiti T, Katsigiannis S, et al. Perceptual video quality evaluation by means of physiological signals. In: Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on. IEEE; 2017. p. 1–6.

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

[9] Katsigiannis S, Ramzan N. 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. 2018 Jan;22(1):98–107.

[10] Koelstra S, Muhl C, Soleymani M, et al. Deap: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing. 2012;3(1):18–31.

[11] Campisi P, La Rocca D. Brain waves for automatic biometric-based user recognition. IEEE Transactions on Information Forensics and Security. 2014;9(5):782–800.

[12] Abo-Zahhad M, Ahmed SM, Abbas SN. State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals. IET Biometrics. 2015;4(3):179–190.

[13] Yang S, Deravi F. On the Usability of Electroencephalographic Signals for Biometric Recognition: A Survey. IEEE Transactions on Human-Machine Systems. 2017;PP(99):1–12.

[14] Ruiz-Blondet MV, Jin Z, Laszlo S. CEREBRE: A novel method for very high accuracy event-related potential biometric identification. IEEE Transactions on Information Forensics and Security. 2016;11(7):1618–1629.

[15] Maiorana E, Campisi P. Longitudinal Evaluation of EEG-Based Biometric Recognition. IEEE Transactions on Information Forensics and Security. 2018;13(5):1123–1138.

[16] Brigham K, Kumar BVKV. Subject identification from electroencephalogram (EEG) signals during imagined speech. In: 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS); 2010. p. 1–8.

[17] Chuang J, Nguyen H, Wang C, et al. I think, therefore I am: Usability and security of authentication using brainwaves. In: International Conference on Financial Cryptography and Data Security. Springer; 2013. p. 1–16.

[18] Armstrong BC, Ruiz-Blondet MV, Khalifian N, et al. Brainprint: Assessing the uniqueness, collectability, and permanence of a novel method for ERP biometrics. Neurocomputing. 2015;166:59–67.

[19] Arnau-Gonzalez P, Katsigiannis S, Ramzan N, et al. ES1D: A deep network for EEG-based subject identification. In: 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE; 2017. p. 81–85.

[20] Thomas KP, Vinod AP, Robinson N. 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. p. 136–141.

[21] Arnau-González P, Arevalillo-Herráez M, Katsigiannis S, et al. On the influence of affect in EEG-based subject identification. IEEE Transactions on Affective Computing. 2018;In press. DOI: 10.1109/TAFFC.2018.2877986.

[22] Chen Y, Atnafu AD, Schlattner I, et al. A High-Security EEG-Based Login System with RSVP Stimuli and Dry Electrodes. IEEE Transactions on Information Forensics and Security. 2016 Dec;11(12):2635–2647.

[23] Palaniappan R, Mandic DP. Biometrics from brain electrical activity: A machine learning approach. IEEE transactions on pattern analysis and machine intelligence. 2007;29(4):738–742.

[24] Maiorana E, Rocca DL, Campisi P. On the Permanence of EEG Signals for Biometric Recognition. IEEE Transactions on Information Forensics and Security. 2016 Jan;11(1):163–175.

[25] Marcuse LV, Fields MC, Yoo JJ. Rowan's Primer of EEG. Elsevier Health Sciences; 2015.

[26] Sanei S, Chambers JA. EEG signal processing. Wiley Online Library; 2007.

[27] Tzallas AT, Tsipouras MG, Fotiadis DI. Epileptic seizure detection in EEGs using time–frequency analysis. IEEE transactions on information technology in biomedicine. 2009;13(5):703–710.

[28] Homan RW, Herman J, Purdy P. Cerebral location of international 10–20 system electrode placement. Electroencephalography and clinical neurophysiology. 1987;66(4):376–382.

[29] Dan-Glauser ES, Scherer KR. The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. Behav Res Methods. 2011;43(2):468–477.

[30] Kurdi B, Lozano S, Banaji MR. Introducing the open affective standardized image set (OASIS). Behav Res Methods. 2017;49(2):457–470.

[31] Russell J. A Circumplex Model of Affect. Journal of Personality and Social Psychology. 1980 12;39:1161–1178.

[32] Morris JD. Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response. Journal of advertising research. 1995;35(6):63–68.

[33] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004;134(1):9–21.

[34] Bigdely-Shamlo N, Mullen T, Kothe C, et al. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Front Neuroinform. 2015;9:16.

[35] Ittichaichareon C, Suksri S, Yingthawornsuk T. Speech recognition using MFCC. In: Proc. ICGSM. Pattaya, Thailand; 2012. p. 28–29.

[36] Juvela L, Bollepalli B, Wang X, et al. Speech Waveform Synthesis from MFCC Sequences with Generative Adversarial Networks. In: Proc. ICASSP. Alberta, Canada; 2018. p. 5679–5683.

[37] Piciucco E, Maiorana E, Falzon O, et al. Steady-State Visual Evoked Potentials for EEG-Based Biometric Identification. BIOSIG 2017. 2017;.

[38] Nguyen P, Tran D, Huang X, et al. A proposed feature extraction method for EEG-based person identification. In: Proc. ICAI. Las Vegas, NV, USA; 2012. p. 826–831.

[39] Arnau-Gonzlez P, Arevalillo-Herrez M, Ramzan N. Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals. Neurocomputing. 2017;244:81 – 89.

[40] Katsigiannis S, Ramzan N. 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. 2017;.

[41] del Pozo-Banos M, Travieso CM, Alonso JB, et al. Evidence of a Task-Independent Neural Signature in the Spectral Shape of the Electroencephalogram. International journal of neural systems. 2018;28(01):1750035.

[42] Rahman MM, Chowdhury MA, Fattah SA. An efficient scheme for mental task classification utilizing reflection coefficients obtained from autocorrelation function of EEG signal. Brain Informatics. 2018;5(1):1.

[43] Hine GE, Maiorana E, Campisi P. Resting-state EEG: A Study on its non-Stationarity for Biometric Applications. In: Proc. BIOSIG. Darmstadt, Germany; 2017. .

[44] Kay SM. Modern spectral estimation. Pearson Education India; 1988.

[45] Marcel S, R Millan JD. Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007 April;29(4):743–752.