Longitudinal Evaluation of EEG-based Biometric Recognition

Emanuele Maiorana, *Senior Member, IEEE*, Patrizio Campisi, *Senior Member, IEEE*

***Abstract*—Brain signals have recently attracted the attention of the scientific community as potential biometric identifiers. In more detail, there is a growing interest in evaluating the feasibility of using electroencephalography (EEG) recordings to perform automatic people recognition. In this framework, the study of the longitudinal behavior of EEG signals, that is, their permanence across time, is of paramount importance. This paper is the first extensive attempt, in terms of employed elicitation protocols, number of involved subjects, number of acquisition sessions, and covered time span, to evaluate the influence of aging effects on the discriminative capabilities of EEG signals over long-term periods. Specifically, we here report and discuss the results obtained from experimental tests conducted on a database comprising 45 subjects, whose EEG signals have been collected during 5 to 6 distinct sessions spanning a total period of 3 years, using 4 different elicitation protocols. The longitudinal behavior of EEG discriminative traits is evaluated by means of a statistical- and performance-related analysis, using different EEG features and hidden Markov models as classifiers. A characterization of each considered EEG channel in terms of uniqueness and permanence properties is also performed, with the purpose of ranking their relevance for biometric purposes, thus giving hints to contain their number in practical applications. Moreover, we design some possible countermeasures to mitigate aging effects on recognition performance and evaluate their effectiveness, thus paving the road for the future deployment of real-life cognitive recognition systems relying on brain-based biometric traits.**

EDICS: BIO-MODA-OTH

***Index Terms*—Biometrics, Electroencephalography, Permanence, Aging Effects, Longitudinal Data Analysis.**

I. Introduction

Brain sensing has interested researchers since the beginning of the twentieth century, when the first devices able to detect brain activities have been designed. To this end, different methodologies, based on either the measurement of blood flow, using approaches such as functional magnetic resonance imaging (fMRI), near-infrared spectroscopy (NIRS), and positron emission tomography (PET), or the detection of neuronal electrical activity, like in magnetoencephalography (MEG) and electroencephalography (EEG), are nowadays available. These techniques have allowed getting significant insights for the diagnosis and treatment of brain disorders such as epilepsy, schizophrenia, Alzheimer's and Parkinson's diseases, to cite a few [1], and for the development of brain-computer interfaces (BCIs) with rehabilitative or entertainment applications [2].

E. Maiorana and P. Campisi are with the Section of Applied Electronics, Department of Engineering, Roma Tre University, Via V. Volterra 62, 00146 Roma, Italy, e-mail: {emanuele.maiorana@uniroma3.it, patrizio.campisi@uniroma3.it}, phone: +39.06.57337365, fax: +39.06.57337026

Kindly cite this work as: *E. Maiorana and P. Campisi, "Longitudinal Evaluation of EEG-based Biometric Recognition," in IEEE Transactions on Information Forensics and Security, vol. 13, no. 5, pp. 1123 - 1138, May 2018.* doi: 10.1109/TIFS.2017.2778010

Only recently, it has been postulated that cerebral activity can discriminate between human subjects, thus allowing its use as biometric identifier in automatic recognition systems [3]. Within this framework, most of the investigation carried out so far has focused on EEG analysis, because of the relatively inexpensiveness of the associated acquisition devices and their ease of use [4].

Brain signals have several peculiarities making them advantageous for biometric purposes. In fact, in addition to the obvious universality and intrinsic liveness properties, they are also highly robust against presentation attacks, being their acquisition at a distance impossible at the present stage of technology. On the other hand, being the research on EEG biometric modality in its infancy, several other issues still have to be properly addressed before deployment in practical applications will be possible. Among them, a comprehensive analysis of a key property such as performance across time, namely stability or permanence, is still missing in literature.

In this contribution we intend to tackle the permanence issue in EEG signals, presenting the results of a longitudinal analysis conducted on a dataset comprising recordings collected from 45 healthy subjects during 5 to 6 sessions spanning a 3-year time period. A preliminary investigation on EEG longitudinal behavior has been presented in [5], where a limited time span of 1 month between acquisitions has been considered. In this work we extend [5] both in terms of the employed elicitation protocols, the time span of the dataset, the EEG features employed, and the comparison procedure. Specifically, in addition to the use of a dataset spanning a large 36-month period, we here analyze EEG responses to 4 different elicitation protocols, including brain conditions in resting states with either closed- or open-eye conditions, the only ones analyzed in [5], as well as in active states while performing cognitive tasks involving mathematical computation and speech imagery. Three different EEG representations, expressing the acquired signals through time-, frequency-, and time-frequency-dependent parameters, are exploited as templates for the considered biometric trait. Hidden Markov models (HMMs) are here applied for the first time to EEG signals for biometric purposes, in order to generate the comparison scores. An evaluation of both the uniqueness and the permanence properties of the EEG signals versus the electrodes placement on the scalp is also provided here. Eventually, countermeasures to be adopted in practical biometric systems to deal with the EEG aging effects are also proposed.

The paper is organized as follows. A review of the most relevant longitudinal studies performed on different biometric modalities is given in Section II. Section III describes the state of the art on multi-session EEG-based biometric recognition. Section IV details the characteristics of the dataset exploited in the performed experimental tests. The employed biometric system is described in Section V. The results of the performed longitudinal evaluation are then outlined in Section VI. Specifically, in Section VI-A we present an analysis on the statistical behavior of genuine scores over time, and in Section VI-B we discuss the variability over time of the obtained recognition performance. Possible approaches to mitigate the effects of aging on EEG-based biometric recognition systems are proposed in Section VII. Conclusions are eventually drawn in Section VIII.

II. Longitudinal studies on biometric traits

Performing a proper analysis on the permanence of any biometric trait is a challenging task. While several studies evaluating the effects of aging on the achievable recognition accuracy have been presented in literature, most of them have focused on *cross-sectional* evaluations, comparing performance attainable for groups of individuals having differences in age [6]. Proper *longitudinal* studies would instead require the availability of data captured from the same subjects at multiple instances, over periods in the order of years, according to modalities minimizing the influence of non-aging-related factors, such as using the same equipment during the whole acquisition campaign [7].

The scientific community has started investigating the permanence issue in biometric recognition systems only very recently, mainly focusing on fingerprint, face, and iris. In [8] a detailed analysis of fingerprint characteristics has been carried out on a longitudinal database collected by the Michigan State Police, comprising data from 15.597 subjects with at least 5 acquisitions over a minimum 5-year time span, exploited to analyze genuine and impostor scores obtained through commercial off-the-shelf (COTS) fingerprint comparators. Although a significant decrease in genuine comparison score, together with a negligible variability in impostor score, has been noticed, the achievable recognition accuracy remains quite stable even though the time interval between a fingerprint pair being compared increases. A similar analysis on face recognition has been presented in [9] using mugshots from two different law enforcement agencies, comprising data collected from 5.633 and 18.007 subjects over at least a 5-year time span. A decreasing trend for genuine scores generated by COTS devices has been noticed, with achievable recognition rates starting to be affected when considering query images acquired 5 years after the enrolment. Temporal stability in iris recognition has been the object of a detailed NIST analysis [10] conducted on 7.876 subjects, whose iris has been acquired on 40 or more occasions over a minimum 4-year time span. Although no evidence of widespread iris aging effects have been there reported, other studies have challenged such result [11], as [12] where recognition rates evaluated over data captured from 322 subjects in a 3-year period have shown a performance worsening with increasing time between enrolment and probe images.

Longitudinal studies have been also performed on signature, speech, gait, keystroke, hand geometry and even electrocardiogram (ECG) biometric modalities. Datasets acquired in laboratory conditions have been used for these evaluations, therefore involving a limited number of subjects and a limited time span with respect to the analysis conducted in [8], [9], and [10], where the employed biometric data have been collected by government or law enforcement agencies. In more detail, data from 29 subjects enrolled in both BIOSECURE [13] and BiosecurID [14] datasets, covering an overall period of 15 months, have been analyzed in [15] to address aging effects on on-line signatures. The Trinity College Dublin speaker ageing (TCDSA) database, containing speech recordings from 18 public figures spanning ranges from 30 to 60 years, has been employed in [16] to highlight that speech-based recognition systems may become unreliable when comparing samples distant more than 5 years. Aging effects in gait recognition have been discussed in [17], where the gait of 10 subjects has been observed during 12 months. A longitudinal study on keystroke, conducted upon data collected from 8 subjects at a 2-year time lapse, has been presented in [18]. A significant performance drop has been noticed in [19] when comparing, in a geometry-based recognition system, left hands captured from 74 subjects during two sessions separated in time by 6 months, with respect to comparing data from the same session. Signals from 47 subjects, recorded during 6 months in sitting posture, have been used in [20] to evaluate performance variability of ECG-based recognition systems.

The aforementioned studies highlight that aging effects on the comparison scores of the same individual can be found in any biometric trait, with the sole exception of DNA [21].

III. EEG-based biometric recognition

EEG signals are the result of the electrical field generated by the synchronous firing of specific spatially-aligned neurons of the cortex, namely, pyramidal neurons. Such activity can be measured by sensing the electric potential difference between specific positions on the scalp surface. Wet electrodes currently represent the gold standard to sense brain activity with the lowest possible noise level, yet their use implies the adoption of electrolyte gel on the scalp, resulting in subject inconvenience and non-negligible time to setup the recording process. Although alternative solutions based on dry electrodes exist, they still need to be improved in order to achieve the desired performance level, in terms of both signal-to-noise ratio (SNR) and user comfort [22].

Most of the studies on EEG for biometric purposes have focused on single session datasets [23], often claiming the ability to reach perfect recognition performance with no error [24]. However, the reliability of such evaluations may be questionable, since it is hard to state whether the achieved recognition rates are only dependent on the discriminative brain characteristics of each subject, or if session-specific exogenous conditions, such as the capacitative coupling of electrodes and cables with other devices, induction loops created between the employed equipment and the body, power supply artifacts, and so on, may significantly differ among distinct acquisition sessions, thus affecting both the inter- and the intra-class variability of EEG recordings.

The collection of datasets to perform longitudinal studies is a very challenging task since it requires the availability of the same population involved in multiple data acquisition

TABLE I
STATE-OF-THE-ART LONGITUDINAL STUDIES EVALUATING EEG SIGNALS AS BIOMETRIC IDENTIFIERS.

Paper	Database					System		
	Sessions	**Covered Period**	**Users**	**Channels**	**Protocol**	**Performance**	**Features**	**Classifier**
Marcel *et al.* [25]	3	3 days	9	8	MI	HTER = 19.3%÷42.6%	PSD	GMM
Näpflin *et al.* [26]	2	15 months (median)	20	60	EC	IR = 88.0%	PSD	Linear regression
Brigham *et al.* [27]	4	n.a.	6	128	Imagined speech	IR = 78.6%÷99.8%	AR (2^{nd} order)	SVM
Kostílek *et al.* [28]	2	1 year	9	53	EC	IR = 87.1%	FZ-AR (7^{th} order)	Mahalanobis dist.
La Rocca *et al.* [29]	2	1÷3 weeks	9	5	EC	IR = 100.0%	AR (10^{th} order)	Linear classifier
				3	EO	IR = 90.5%		
Armstrong *et al.* [30]	2	1 week	15	1	ERP	IR = 89.0%	ERP signal	Correlation
		6 months	8	1	ERP	IR = 93.0%		
Wang *et al.* [31]	2	1 week	4	128	EC	IR = 92.58%	CWT	L2 dist.
Das *et al.* [32]	3	1 month	50	19	VEP	EER ≊ 13.0%	Evoked potential	Cosine dist.
Maiorana *et al.* [33]	2	1 month	30	19	EC	IR = 87.9%	EigenBrains	L1, L2, cosine dist.
					EO	IR = 75.4%		
Ruiz-Blondet *et al.* [34]	2	9 months (mean)	20	26	ERP	IR = 100.0%	ERP signal	Correlation
Maiorana *et al.* [5]	3	1 month	50	19	EC	IR = 90.8%	AR, PSD, COH	L1, L2, cosine dist.
					EO	IR = 85.6%		

sessions. This task is even more challenging than usual when considering EEG signals, due to the possible users' displeasure in the acquisition process. Only few works, summarized in Table I, have therefore studied EEG-based biometric recognition using multi-session datasets. Specifically, data collected from 9 subjects performing motor imagery (MI) tasks during 3 consecutive days, with 4 sessions each day, have been used in [25]. A half-total error rate (HTER) of 19.3%, using power spectral density (PSD) as EEG features and Gaussian mixture models (GMM) for incremental learning across multiple sessions, has been there achieved. Signals from 20 people, recorded during 2 sessions at a median distance of 15 months, and represented through PSD characteristics, have been used in [26] to estimate a rank-1 identification rate (IR) at about 88%. A biometric system based on an imagination task performed by 6 subjects, whose EEG signals have been recorded in 4 different days, has been analyzed in [27], where IRs ranging from 78.6% to 99.8% have been reported. Details on the time distances between two acquisitions are not given, being therefore impossible to derive proper information on the stability of the obtained performance. A database collected from 9 subjects during 2 one-year-apart sessions has been considered in [28], applying a frequency-zooming auto-regressive (FZ-AR) modeling to 53 channels to achieve IR = 87.1%. Signals from 9 subjects have been recorded during 2 sessions spanning up to 3 weeks in [29], and exploited to achieve perfect IR for EEG data acquired in eyes-closed (EC) conditions, and IR = 90.53% for the eyes-open (EO) scenario. Event-related potentials (ERPs), obtained as responses to visual stimuli, have been exploited in [30] to achieve IRs at 89.0% and 93.0%, respectively for a database collected from 15 people during 2 sessions spanning 1 week, and a dataset comprising signals from 8 persons acquired during 2 six-month-separated sessions. Since these results have been evaluated over two distinct databases, it is impossible to argue on performance permanence for varying time distances between enrolment and test data. Signals recorded from 4 subjects in EC conditions during 2 one-week-apart sessions have been processed through continuous wavelet transform (CWT) in [31] guaranteeing IR = 92.58%. Visual-evoked potentials (VEPs) to both target and non-target stimuli have been evaluated in [32] to provide equal error rates (EERs) respectively at about 18% and 13%, over a database comprising signals acquired from 50 users during 3 sessions taken during a period of 1 month. Parsimonious representations in the frequency domain have been proposed in [33], where IR = 87.9% and IR = 75.4% have been respectively achieved in EC and EO conditions, using EEG signals taken from 30 subjects during 2 recording sessions spanning one month. Perfect accuracy has been achieved in [34] applying the system proposed in [23] to 20 subjects whose EEG signals have been recorded during 2 sessions at an average distance of 9 months. The most detailed analysis on permanence so far performed for EEG-based biometric recognition systems has been presented in [5], where the performance behavior achievable when comparing data captured from 50 subjects during 3 different sessions spanning a 1-month period, and represented through auto-regressive (AR), PSD and spectral coherence (COH) features, has been discussed. IR at 90.8% comparing signals captured in EC conditions, and IR = 85.6% for the EO scenario, have been reported almost regardless of the sessions being compared out of the 3 available ones.

Despite their higher reliability with respect to single-session studies, it is worth observing that all the mentioned evaluations have considered either short time distances between the available EEG acquisitions, in the order of days or months, or a small population, or both. Therefore it is hard to speculate on EEG permanence on the basis of the aforementioned works. Within this framework, this paper represents a significant improvement with respect to the state of the art on longitudinal analysis of EEG biometric trait, since the analysis here reported is the most extensive one in the literature of EEG-based biometric recognition systems, involving multiple acquisition sessions, spanning a wide time frame, employing a large number of subjects, using several elicitation protocols, and considering different EEG representations.

IV. LONGITUDINAL EEG DATABASE

In this paper we carry out the permanence analysis of EEG discriminative capabilities on a longitudinal database collected using a 19-channel GALILEO BE Light amplifier, recording EEG signals at an original sampling rate of 256 Hz. Specifically, the dataset comprises 5 different sessions where 45 healthy subjects have donated their EEG signals elicited using 4 different protocols. Out of these 45 subjects, 30 have

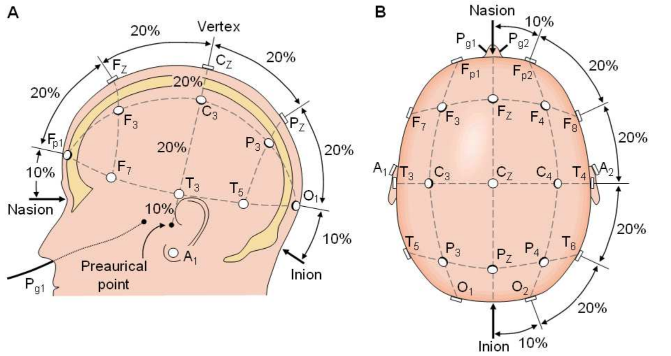


Fig. 1. The 10-20 International system seen from left (A) and above the head (B). The letters F, T, C, P, and O stand for frontal, temporal, central, parietal, and occipital lobes. (Jaakko Malmivuo and Robert Plonsey, Bioelectromagnetism, Oxford University Press, 1995, WEB version).

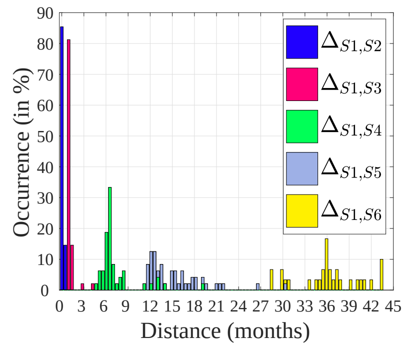


Fig. 2. Acquisitions' sessions temporal distribution histograms.

donated their EEG traits also during a sixth acquisition session. The subjects' age at the time of the first acquisition ranges from 21 to 34 years, with an average of 25 years. During each recording session, subjects have been comfortably seated on a chair in a dimly lit room, with the electrodes placed on their scalp according to the *de-facto* standard $10-20$ montage depicted in Figure 1. Conductive gel has been used to reduce the scalp impedance under 10 kΩ. Neither dietary nor activity restrictions have been suggested to the subjects, neither between consecutive EEG acquisitions nor during the days of the recordings. This lack of restrictions goes in the direction of using acquisition conditions close to real life.

A. Dataset time span

Six distinct recording sessions, indicated in the following as $S_1, S_2, \ldots, S_6$, have been carried out to collect the multi-session database we exploit in our analysis. Specifically, the average distances between the first and the other five acquisition sessions are: $\bar{\Delta}_{S_1,S_2} = 1$ week, $\bar{\Delta}_{S_1,S_3} = 1$ month, $\bar{\Delta}_{S_1,S_4} = 7$ months, $\bar{\Delta}_{S_1,S_5} = 16$ months, $\bar{\Delta}_{S_1,S_6} = 36$ months. More in detail, Figure 2 shows the distributions of the time distances Δ_{S_1,S_n}, $n = 2, \ldots, 6$, elapsed between the first and the n-th EEG recording for each considered subject. The performed activities have covered an overall period of more than 3 years.

B. Elicitation protocols

The following elicitation protocols have been adopted during each subject's acquisition session:

- **resting state with eyes closed (EC):** EEG signals have been collected for 4 minutes in EC conditions. Resting state with EC is one of the most commonly used acquisition modality, widely investigated for both medical and biometric applications;
- **resting state with eyes open (EO):** EEG signals have been acquired for 4 minutes in EO conditions, with subjects asked to fix a light point on the screen;
- **mathematical computation (MC):** a cognitive protocol consisting in asking the subject to perform sums and differences of integers has been considered. During each performed session, 28 operations are shown on the screen, each for 5s interleaved by 2s from the previous one, for a total duration of 3min and 14s;
- **speech imagery (SI):** subjects have been asked to mentally reproduce the sound of a vowel observed on the screen. Each of the 5 vowels is shown 20 times in random order for 3s, with a 2s separation time between any presentation, for an overall period of 8min and 18s.

It is worth remarking that, as also evident from Table I, the collected database is the largest EEG multi-session dataset ever employed in literature, both in terms of enrolled subjects, employed elicitation protocols, and time span employed to test the feasibility of EEG signals as biometric identifiers.

V. EEG-based biometric recognition system

The proposed biometric recognition system employed to evaluate the permanence of EEG discriminative characteristics across long-time periods is depicted in Figure 3. Specifically, the acquired EEG signals are first preprocessed as described in Section V-A. The epochs obtained by segmenting the original signals are then processed through HMMs as outlined in Section V-B, modeling each channel of a given epoch as a sequence of hidden states generating observations given by the features described in Section V-C. The adopted verification strategy is then detailed in Section V-D.

A. Preprocessing

A spatial common average referencing (CAR) filter [35] is first applied to the acquired data in order to improve their signal-to-noise ratio (SNR), by subtracting from each raw EEG signal $\mathbf{r}^{(c)}$, with $c = 1, \ldots, C$ being C the number of considered channels, the mean voltage sensed over the entire scalp. The obtained signals are then band-pass filtered to extract the EEG subband of interest. Specifically, in the following we always refer to signals within the $\alpha - \beta = [8, 30]$Hz subband. In fact, we have verified in our experimental tests that the information there contained always guarantees the best achievable recognition performance for all the evaluated acquisition protocols. Given the considered subband, the filtered data are then downsampled at 64Hz to reduce the computational complexity of the subsequent processing, without negatively affecting the accuracy achievable through the employed EEG representations, described in Section V-C. The signals $\mathbf{g}^{(c)}$ are so obtained as output of the preprocessing step.

A segmentation process is then carried out, generating for each acquisition the consecutive epochs $\mathbf{g}_i^{(c)}$, with $i =$

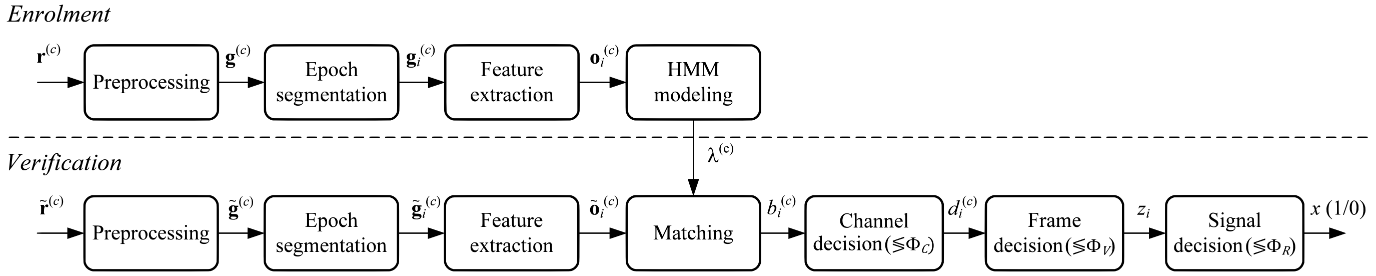


Fig. 3. Framework of the considered EEG-based biometric recognition system.

$1, \dots, E$ during enrolment and $i = 1, \dots, V$ during verification. Epoch segmentation is performed differently depending on the considered acquisition protocol. Specifically:

- EC: the available data are divided into overlapping epochs lasting 5s, with a 40% overlap factor between consecutive epochs;
- EO: as for EC, data are divided into overlapping epochs lasting 5s with a 40% overlap factor. It is worth specifying that the epoch length and the overlap factor employed for segmenting EC and EO EEG recordings have been selected experimentally in order to generate, from the acquired signals, a number of epochs guaranteeing low recognition rates, while also keeping the required computational cost and processing time at acceptable levels;
- MC: an epoch is selected correspondingly to the EEG signals recorded when a required operation is shown, relying on a synchronization signal between the shown operation and the acquired EEG data;
- SI: an epoch is selected as the period a vowel is shown on the screen, relying on a synchronization signal between the display inputs and the acquired EEG data.

B. HMM modeling of EEG signals

A novel approach based on HMMs is here employed to model EEG signals for biometric recognition purposes. It is worth remarking that, although HMMs are here exploited for the first time in an EEG-based biometric recognition system, they have been already applied to EEG signals for medical applications [36] or for the design of BCI systems [37]. In fact, they can properly model the brain's non-stationary and non-localized sources of information [38], [39], and represent the dynamic behavior of the spatio-temporal EEG patterns with the associated changes of brain states over time [40].

Similarly to what commonly performed on speech recognition [41], each EEG signal $\mathbf{g}_i^{(c)}$, associated to the c-th channel of the i-th obtained epoch, $c = 1, \dots, C$ and $i = 1, \dots, E$, is first split into H overlapping frames, and thus represented as a sequence $\mathbf{o}_i^{(c)}[h]$ of H observations, $h = 1, \dots, H$. Each observation consists of a set of Q parametric features extracted from the corresponding frame, as detailed in Section V-C. Frames lasting 1s, with a 50% overlap between consecutive frames, are employed for each considered protocol, in order to generate observation sequences with reasonable lengths, while allowing the extracted features having the required resolution and accuracy.

It is then assumed that each sequence of observations $\mathbf{o}_i^{(c)}$ can be modeled as generated by a process that, at a given h-th frame, is in one of the N admissible hidden states, generates a measurable observation characterized by a specific distribution, and then moves to another admissible state at the next frame. This process can be learnt through an iterative procedure [42] as a statistical left-right HMM model $\lambda^{(c)} = \{A^{(c)}, B^{(c)}, \pi^{(c)}\}$, on the basis of the E observation sequences considered for enrolment purposes, being:

- $A^{(c)}$ the state transition matrix, describing the probabilities of moving from one of the N considered hidden states to another, for consecutive frames;
- $B^{(c)}$ the observation probability distributions in the N states, each modeled as a mixtures of M multivariate Gaussian distributions;
- $\pi^{(c)}$ the initial state distributions for the N states.

As outlined in Section V-D, the models $\lambda^{(c)}$ are employed, for each c-th channel, to estimate the similarity scores between the enrolment data and each probe sequence $\tilde{\mathbf{o}}_i^{(c)}$ associated to the i-th epoch of a verification EEG sample.

C. Feature extraction

In both enrolment and verification, each EEG signal corresponding to the c-th channel of the h-th frame, extracted from the i-th epoch taken from the acquired recording, is individually processed in order to derive a feature-based representation $\mathbf{o}_i^{(c)}[h]$, employed as h-th observation of the i-th sequence modeled through HMMs for channel c. In order to provide a comprehensive analysis of EEG signals, three different kinds of parametric features are here exploited: AR reflection coefficients, mel-frequency cepstrum coefficients (MFCCs), and bump representation characterize EEG signals in the time domain, the frequency domain, and the time-frequency domain, respectively.

1) AR modeling: AR modeling is among the most commonly employed approaches for EEG signals analysis and it has been often adopted for biometric recognition purposes [4]. In this work, we use as features the AR reflection coefficients, which are expressed in terms of the AR filter parameters and of the variance of the white noise feeding the filter estimated by means of the Yule-Walker equations. As in [5], we estimate the reflection coefficients of an AR model of order $Q = 12$ by using the Burg method [43].

2) MFCC modeling: MFCC modeling has been widely used in speech recognition, while it has been only recently exploited for biometric analysis of brain data [44]. In our implementation, a filter-bank made of 18 mel-scaled triangular band-pass filters is first applied to the EEG spectrum. The natural logarithm of the resulting cepstral bins is then evaluated to perform an homomorphic filtering operation, separating the underlying neural activation signals from the effects of their propagation through the skull [45]. A discrete cosine transform (DCT) is eventually performed on the resulting values. The desired MFCC representation is obtained by selecting the first $Q = 12$ DCT coefficients, with the exclusion of the DC component.

3) Bump modeling: Bump modeling has been first proposed to process invasive EEG potentials [46], and later to investigate brain oscillatory dynamics in the medical field, especially regarding EEG data from patients with early stage of epilepsy and Alzheimer's disease [47]. It has been applied for biometric purposes in [48] to model EEG signals acquired in EC and EO resting states, although only single-session data are considered there. Bump modeling is obtained through discrete wavelet transform (DWT) decomposition, using complex Morlet wavelets of Gaussian shape in time to accurately represent EEG oscillations in both time and frequency domains. A sparse representation is then derived from the DWT domain by extracting the most prominent bursts within a normalized time-frequency map, and modeling them into a sum of parametric functions, indicated as bumps. The $Q = 13$ parameters suggested in [48] to characterize the observed EEG behavior are here employed to represent the considered signals.

D. EEG signals comparison

The comparison between two EEG acquisitions is performed by first evaluating, for each considered channel c, the similarity between individual verification epochs and the enrolment signal, represented through the HMM $\lambda^{(c)}$. Being $\tilde{\mathbf{o}}_i^{(c)}$ the sequence of observations associated with the i-th epoch of the verification probe, a similarity score $b_i^{(c)}$ is computed as the a posteriori log likelihood $b_i^{(c)} = \log P(\tilde{\mathbf{o}}_i^{(c)}|\lambda^{(c)})$, once the path of HMM hidden states which the observed sequence has followed with maximum probability has been estimated through the Viterbi algorithm [41].

For each i-th verification epoch, the scores obtained from each channel are compared against a threshold Φ_C, obtaining:

$$d_i^{(c)} = \begin{cases} 1 \text{ if } b_i^{(c)} \geq \Phi_C \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

A fusion strategy is then implemented to combine the information derived from the C available channels as:

$$z_i = \begin{cases} 1 \text{ if } \dfrac{1}{C}\displaystyle\sum_{c=1}^{C} d_i^{(c)} \geq \Phi_V \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

Eventually, the decision regarding the identity of the presented user is taken fusing the information extracted from the available epochs, as:

$$x = \begin{cases} 1 \text{ (user verified)} & \text{if } \dfrac{1}{V}\displaystyle\sum_{i=1}^{V} z_i \geq \Phi_R \\ 0 \text{ (user not verified)} & \text{otherwise.} \end{cases} \tag{3}$$

It is worth pointing out that the $b_i^{(c)}$ scores computed for each c-th channel and i-th verification epoch could be fused also according to a score-level strategy, instead of following a decision-level approach as in the aforementioned description. Nevertheless, experimental tests conducted on the employed database show that the approach here proposed provides better recognition rates rather than score-level fusion strategies.

VI. Longitudinal analysis of EEG biometric trait

The longitudinal analysis performed on the multi-session EEG database described in Section IV is outlined in the following. With regard to the collected dataset, it is worth pointing out that the distributions of the distances Δ_{S_1,S_n}, $n = 2,\ldots,6$, shown in Figure 2, are characterized by non-negligible variance and overlap, especially for sessions beyond the fourth one. Since this issue could potentially affect the reliability of the performed tests, the available data are rearranged as detailed hereafter. Specifically, we refer in the following to comparisons of EEG recordings taken at distances $\Delta_1,\ldots,\Delta_5$, where such time intervals are characterized by well-separated distributions, each with relatively small standard deviation, obtained as:

- time distances Δ_1: comparisons of EEG data collected at enrolment and verification stages having temporal distances within the range $\Delta_1 = [23;43]$ days. Considering the available EEG database and comparing data in sessions S_1 and S_3 (S_1 *vs* S_3), as well as samples in sessions S_2 and S_3 (S_2 *vs* S_3), the instances falling into the desired time interval are 67, with data taken from 45 subjects and characterized by an overall average time distance between acquisition sessions of $\bar{\Delta}_1$ = 1 month;
- time distances Δ_2: lapses between enrolment and verification within the range $\Delta_2 = [150;271]$ days. Given the available EEG database, and comparing S_1 *vs* S_4, S_2 *vs* S_4, S_3 *vs* S_4 and S_4 *vs* S_5, the instances falling into the desired time interval are 148 with data taken from 45 subjects, with an overall average time distance between recording sessions of $\bar{\Delta}_2$ = 7 months;
- time distances Δ_3: lapses between enrolment and verification within the range $\Delta_3 = [400;552]$ days. Comparing S_1 *vs* S_5, S_2 *vs* S_5 and S_3 *vs* S_5, the instances falling into the desired time interval are 58 with data taken from 45 subjects, with an overall average time distance between recording sessions of $\bar{\Delta}_3$ = 16 months;
- time distances Δ_4: lapses between enrolment and verification within the range $\Delta_4 = [631;861]$ days. Comparing S_4 *vs* S_6 and S_5 *vs* S_6, the instances falling into the desired time interval are 30, one for each of the 30 available subjects, with an overall average time distance between recording sessions of $\bar{\Delta}_4$ = 26 months;
- time distances Δ_5: lapses between enrolment and verification within the range $\Delta_5 = [982;1300]$ days. Comparing S_1 *vs* S_6, S_2 *vs* S_6 and S_3 *vs* S_6, the instances falling into the desired time interval are 70 with data taken from 30 subjects, with an overall average time distance between recording sessions of $\bar{\Delta}_5$ = 36 months.

Figure 4 shows the distributions of the time distances $\Delta_1,\ldots,\Delta_5$. The performed longitudinal analysis thus evaluates the aging effects on EEG signals over a period going

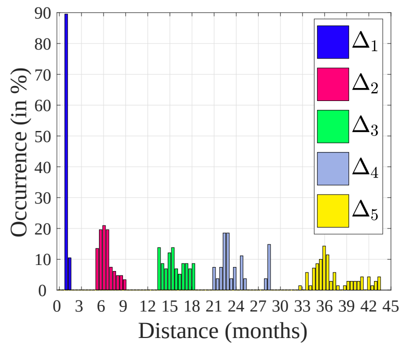


Fig. 4. Considered comparisons' temporal distribution histograms.

from 1 month to 3 years. Time intervals below 1 month are not here covered, having been already analyzed in our previous work [5].

In the following, we present two longitudinal analyses. First, in Section VI-A, a statistical evaluation regarding the genuine score distributions obtained comparing EEG signals captured at different times is given. Then, in Section VI-B, a detailed analysis on the verification performance achievable exploiting EEG signals as biometric identifiers is provided. It is worth specifying that a 10-run cross-validation is performed to estimate the score distributions and recognition rates. Specifically, at each iteration, EEG signals lasting 3min overall are randomly selected from enrolment data to generate the employed templates. For each enrolment selection, 10 different verification probes lasting 45s are then taken from the verification session to evaluate the desired scores and performance metrics, for all the considered protocols.

A. Statistical analysis

In this section we investigate the behavior of the distance separating the genuine and the impostor score distributions evaluated in the proposed system, as well as the mutual divergence among the obtained genuine score distributions, for increasing time lapses between enrolment and verification. EEG signals from the 30 subjects whose characteristics have been acquired during all the 6 scheduled sessions are here employed. An HMM with $N = 4$ hidden states, each with $M = 4$ Gaussian distributions modeling the available observations, is employed for the following analysis.

Let us indicate with $\psi_{\Delta_t}(b^{(c)})$ the distribution of genuine scores generated when comparing the c-th channel of an EEG epoch collected at time distance Δ_t from the corresponding enrolment set, with $t = 1, \ldots, 5$. Let $\phi(b^{(c)})$ be the overall impostor score distribution obtained comparing the c-th channel of an EEG epoch with the enrolment set of a different user, selected at any time distance from the test probe. A single impostor score distribution, regardless of the time distance between enrolment and test data, is considered in the performed tests, thus obtaining a highly-robust estimation through the exploitation of a large number of similarity scores. The ideal behavior of the aforementioned distributions would consist in a constant and wide distance between the genuine score distributions, generated comparing EEG data with different time lapses between two acquisitions of the same user and the impostor one. Limited variations in the obtained genuine score distributions would be also desired. The performed longitudinal statistical analysis takes into account these two aspects, evaluating:

- the Bhattacharyya distances $D_{Bh}(\psi_{\Delta_t}(b^{(c)}), \phi(b^{(c)}))$ between the genuine score distributions computed with enrolment and verification at time distances Δ_t, $t = 1, \ldots, 5$, and the impostor score distribution, for each c-th channel. Besides allowing to detect the presence of aging effects, this measure also provides information about discriminative capabilities of EEG signals: the higher the measured values, the better such characteristics are. Results obtained for each considered channel using AR features for EEG representation are reported in Figure 5;
- the Kullback-Leibler divergences $D_{KL}(\psi_{\Delta_1}(b^{(c)}), \psi_{\Delta_t}(b^{(c)}))$ between the score probability distribution evaluated at the minimum considered time distance Δ_1, and the one evaluated at time lapses Δ_t, $t = 2, \ldots, 5$, between enrolment and verification. Large values of this measure imply significant variability of the observed characteristics. With respect to the aforementioned Bhattacharyya distance, such measure provides further details on the variability of the genuine score distributions over time, even in case of a constant distance from the impostor score distribution. Figure 6 shows the results obtained when modeling each EEG channel with AR features.

Although the extents of aging effects are different for the four considered protocols, a consistent trend of decreasing D_{Bh} values and increasing D_{KL} values, for all the employed channels, can be seen when increasing the time distance between enrolment and verification. A non-significant variability in D_{Bh} is often observed for time distances beyond Δ_3. This implies that recognition performance should not notably vary, when there are more than 16 months between two EEG acquisitions. On the contrary, the D_{KL} divergences, evaluated for the largest time intervals, indicate that the distributions of genuine scores not cease to deviate from the one evaluated at Δ_1, highlighting a certain variability also over long time periods. The tests conducted using MFCC and bump EEG representations lead to analogous observations.

The longitudinal analysis on the genuine score distributions is also exploited to assess the discriminative capabilities of the employed EEG channels. Specifically, for each considered acquisition protocol, the Bhattacharyya distances and Kullback-Leibler divergences are used to evaluate the cumulative values

$$\begin{aligned} \Theta_{Bh}^{(c)} &= \sum_{t=1}^{5} D_{Bh}\left(\psi_{\Delta_t}(b^{(c)}), \phi(b^{(c)})\right), \\ \Theta_{KL}^{(c)} &= \sum_{t=2}^{5} D_{KL}\left(\psi_{\Delta_1}(b^{(c)}), \psi_{\Delta_t}(b^{(c)})\right), \end{aligned} \tag{4}$$

providing respectively, for each c-th channel, measures regarding its discrimination capability and stability over time. Figures 7 and 8 depict the obtained results in terms of topographic maps for EEG representations expressed through AR features. From these plots it can be observed that the EC protocol provides the best results in terms of discriminative

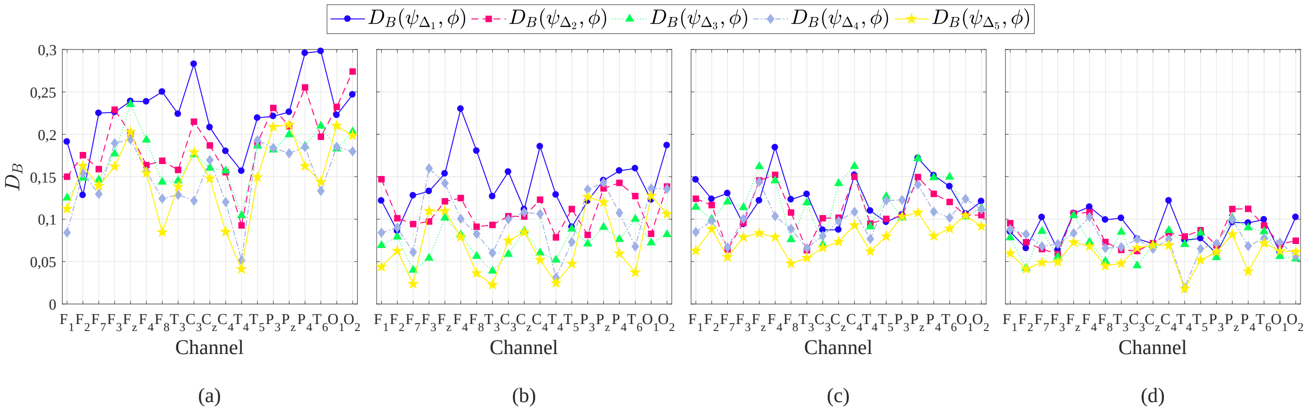


Fig. 5. Statistical analysis performed on EEG signals modeled with AR features using Bhattacharyya distances. (a): EC; (b): EO; (c): MC; (d): SI.

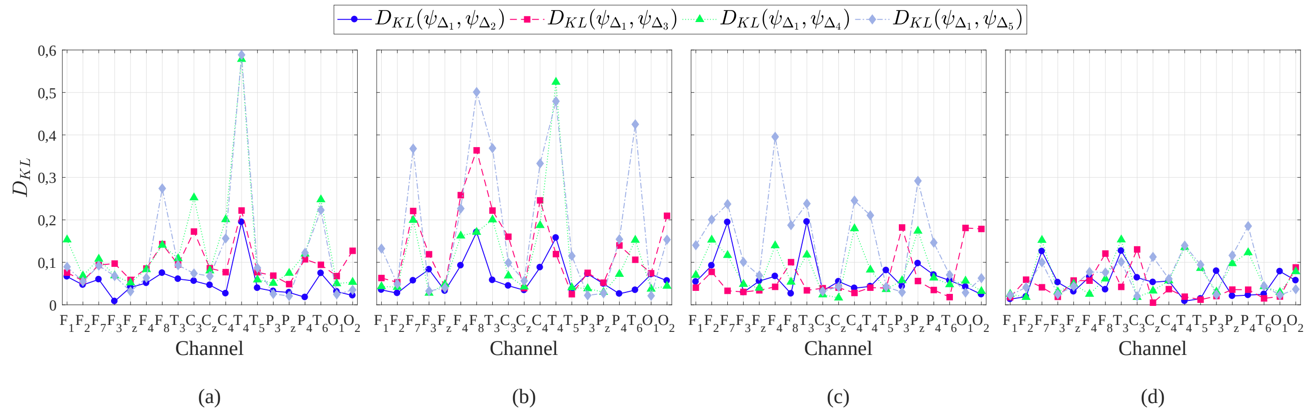


Fig. 6. Statistical analysis performed on EEG signals modeled with AR features using Kullback-Leibler divergences. (a): EC; (b): EO; (c): MC; (d): SI.

capability and that the parieto-occipital region is the most relevant head area for all the considered protocols.

Eventually, in order to derive for each channel c a single measure taking into account all the aforementioned aspects, we first evaluate the mean values $\bar{\Theta}_{Bh}^{(c)}$ and $\bar{\Theta}_{KL}^{(c)}$ of the considered distances over all the exploited protocols. After having normalized such measures to the [0;1] range with a *min-max* approach for AR, MFCC, and bump features, with no claim of optimality, we employ the value:

$$\begin{aligned}\Theta_{Bh,KL}^{(c)} = {}^{(AR)}\bar{\Theta}_{Bh}^{(c)} + {}^{(MFCC)}\bar{\Theta}_{Bh}^{(c)} + {}^{(Bump)}\bar{\Theta}_{Bh}^{(c)} \\ - {}^{(AR)}\bar{\Theta}_{KL}^{(c)} - {}^{(MFCC)}\bar{\Theta}_{KL}^{(c)} - {}^{(Bump)}\bar{\Theta}_{KL}^{(c)}\end{aligned} \quad (5)$$

as a measure of both uniqueness and permanence properties for each channel. Figure 9 confirms that the parieto-occipital region is the most informative area to be exploited when using EEG signals for biometric recognition purposes. In more detail, the computed values of $\Theta_{Bh,KL}$ are shown in Figure 10, where the considered channels are ranked according to the proposed measure. From the reported results, it can be clearly seen that the 4 worst performing channels show a behavior substantially worse from the other ones. Other notable differences can be found from the second- to the third-most-relevant channels (P_z and O_2), from the sixth- to the seventh-most-relevant channels (P_4 and F_4), and from the ninth- to the tenth-most-relevant channels (C_z and T_5). According to such observations, reasonable selections for the number of electrodes to be included in an EEG montage for biometric recognition purposes would consist in using either the first 2, 6, 9, or 15 channels from the ranking provided in Figure 10. It can be also observed that all these selections correspond to symmetrical distributions of the electrodes over the scalp, with the most discriminative and permanent information coming from the scalp midline and the parieto-occipital region.

To provide an illustrative example of the observed behaviors, Figure 11 reports the genuine and impostor score distributions evaluated when using AR modeling for EEG signals acquired according to the EC protocols through the best (F_z) and worst (T_4) channels. As can be seen, the impostor distribution is much more overlapped with the genuine distributions for the T_4 channel than for the F_z channel. Moreover, the genuine distributions for the T_4 channel show a larger variability, as an effect of EEG aging over a 3-year period, with respect to the F_z-related distributions, especially along the tails. Figure 11 makes it also clear that it would be hard achieving good recognition performance in an EEG-based biometric system when employing a single protocol, using a single channel and modeling EEG data with a single representation, due to the large overlap between impostor and genuine distributions. Similar behaviors can be found also when dealing with the other considered acquisition protocols and EEG representations.

B. Performance analysis

As mentioned in the previous section, the proposed HMM modeling with $N = M = 4$ is employed in the performed

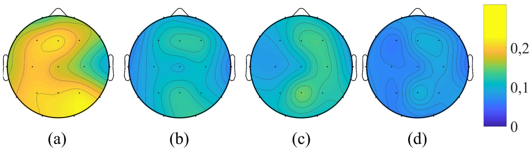


Fig. 7. Θ_{Bh} maps for AR modeling. (a): EC; (b): EO; (c): MC; (d): SI.

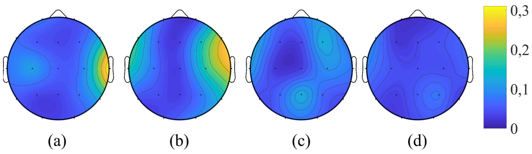


Fig. 8. Θ_{KL} maps for AR modeling. (a): EC; (b): EO; (c): MC; (d): SI.

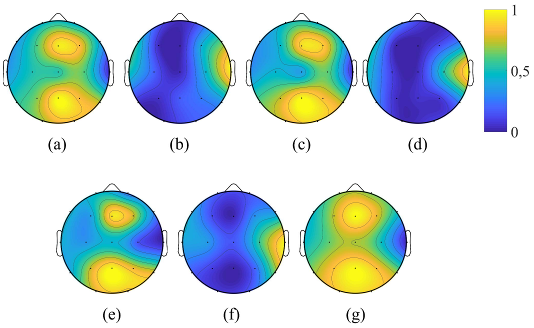


Fig. 9. Topographic maps obtained through statistical analysis. (a): ${}^{(AR)}\bar{\Theta}_{Bh}$; (b): ${}^{(MFCC)}\bar{\Theta}_{Bh}$; (c): ${}^{(Bump)}\bar{\Theta}_{Bh}$; (d): ${}^{(AR)}\bar{\Theta}_{KL}$; (e): ${}^{(MFCC)}\bar{\Theta}_{KL}$; (f): ${}^{(Bump)}\bar{\Theta}_{KL}$; (g): $\Theta_{Bh,KL}$

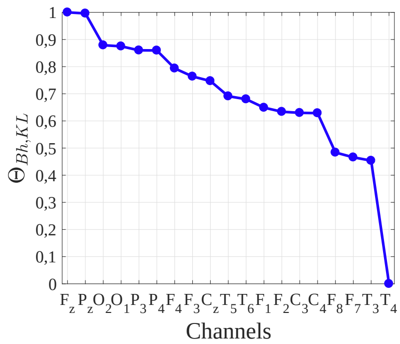


Fig. 10. Ranking of the considered EEG channels with respect to $\Theta_{Bh,KL}$ values, normalized to [0;1].

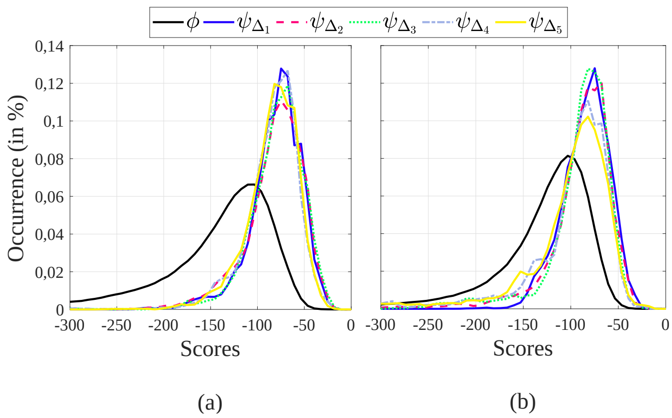


Fig. 11. Genuine ($\psi_{\Delta_t}, t = 1, \ldots, 5$) and impostor ($\phi$) score distributions obtained with AR features for EC EEG recordings. (a): F_z; (b): T_4.

statistical analysis to generate similarity scores between compared EEG signals. In order to show the effectiveness of HMMs as learning method for EEG discriminative characteristics, we report in Table II a comparison between different configurations of the proposed approach and other methods exploited in the literature of EEG-based biometric recognition systems, including distance-based comparators using L1, L2, cosine, and Mahalanobis distances. For these latter approaches, two EEG recordings are compared by first evaluating the distances between the representations extracted from a verification epoch and from each of the E epochs available in the enrolment dataset. The minimum among the computed distances is therefore selected as representative dissimilarity score $b_i^{(c)}$ for the i-th verification epoch. Furthermore, Gaussian mixture models (GMMs) are employed to represent features extracted from enrolment epochs through distributions made of M Gaussians. In this case, a similarity score $b_i^{(c)}$ is obtained as the probability that features taken from the examined i-th verification epoch belong to the estimated distribution.

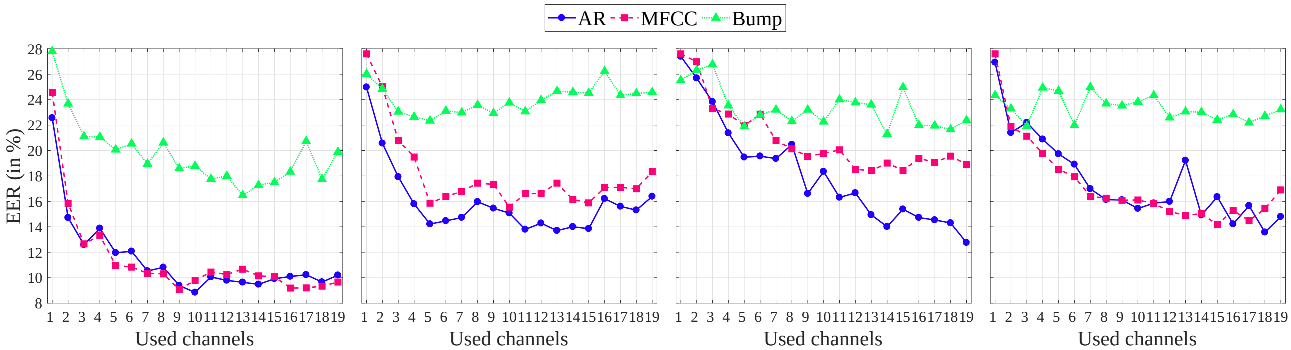
The results in Table II are expressed as EERs computed when comparing EEG signals captured with all the 19 considered channels during different sessions, from the 30 users whose traits are collected for a 3-year period, averaged over all the possible time distances Δ_t, $t = 1, \ldots, 5$ between enrolment and verification. The best average results obtained when varying the thresholds Φ_V and Φ_R, for 45s-long verification probes, are reported for each considered protocol, EEG representation, and comparison method. The reported rates show that HMMs outperforms the other approaches for all the employed EEG representations and all the exploited acquisition protocols. In more detail, the best results are commonly obtained when adopting $N = M = 4$. It is worth specifying that such configuration does not always guarantee the best possible results, regardless of the temporal distance between the signals to be compared. Yet, it guarantees *on average* the best expected outcomes, considering all the possible time lapses between enrolment and verification. Therefore, it is selected for both the performed statistical analysis and for the experiments described in the following.

Exploiting the channel ordering shown in Figure 10 obtained through the performed statistical analysis, the recognition performance for each considered acquisition protocol are estimated for an increasing number of employed electrodes in Figure 12. Specifically, we provide the mean EERs obtained over 10 performed runs for each scenario when comparing EEG signals captured at the longest considered time distance Δ_5, using verification probes lasting 45s, and representing EEG data with AR, MFCC, and bump features. As can be seen, the achievable recognition rates improve when increasing the number of employed channels, reaching a plateau when using the best 15 channels which will be therefore employed for the tests described in the following.

The results related to the longitudinal analysis are given in Figures 13, 14, and 15, where the recognition performance expressed in terms of 95% confidence intervals of the achievable EERs are depicted for the considered AR, MFCC, and

TABLE II
EER (IN %) AVERAGED OVER THE CONSIDERED TIME DISTANCES Δ_t, $t = 1, \ldots, 5$, FOR THE CONSIDERED LEARNING METHODS.

Protocol	EEG feature	Comparison method																	
		L1	L2	Cos.	Mahal.	GMM			HMM										
									N=2			N=4			N=8				
						M=2	M=4	M=8	M=2	M=4	M=8	M=2	M=4	M=8	M=2	M=4	M=8		
EC	AR	11,1	10,5	14,1	13,6	12,0	10,7	10,8	7,3	6,9	7,2	7,0	6,6	6,8	7,2	6,9	6,9		
	MFCC	12,7	12,4	22,1	15,5	14,2	13,2	13,9	7,0	6,7	7,0	6,8	6,5	6,5	6,9	6,7	6,8		
	Bump	20,2	19,4	17,3	21,3	19,8	18,9	18,9	17,1	16,8	17,2	16,4	15,9	16,2	16,9	16,6	16,7		
EO	AR	16,8	17,0	16,9	18,1	17,9	16,8	17,1	11,9	11,0	11,4	11,4	10,6	11,1	11,7	10,8	10,8		
	MFCC	17,6	17,5	30,3	17,3	17,7	17,4	17,1	12,3	11,4	11,8	11,7	10,8	11,3	12,1	11,1	11,0		
	Bump	24,1	24,2	24,7	32,4	25,7	24,2	28,2	23,4	21,1	21,6	22,9	20,4	22,7	23,2	20,9	20,8		
MC	AR	16,3	15,3	13,6	18,1	14,9	14,7	17,4	12,3	11,2	11,8	11,8	10,7	11,4	12,1	11,0	11,3		
	MFCC	16,7	16,7	26,1	17,2	17,1	16,6	18,1	12,9	12,3	12,8	12,3	11,6	11,8	12,7	11,9	12,1		
	Bump	27,7	27,9	19,2	32,4	29,2	30,9	35,6	21,5	19,9	20,2	20,3	18,8	20,1	21,0	19,5	19,7		
SI	AR	14,3	14,4	15,6	17,6	16,3	14,2	15,9	10,8	9,9	10,3	10,1	9,0	9,4	10,5	9,7	9,9		
	MFCC	18,9	19,0	26,2	18,4	19,1	17,9	19,1	11,6	10,5	11,2	11,2	10,1	10,5	11,5	10,4	10,6		
	Bump	36,2	35,4	19,3	34,3	38,3	36,0	40,0	19,1	17,4	18,1	18,3	16,9	17,4	18,9	17,1	17,4		



(a) (b) (c) (d)

Fig. 12. EER *vs* no. of used channels, for EEG signals compared at time distance Δ_5 and 45s-long verification probes. (a): EC; (b): EO; (c): MC; (d): SI.

bump EEG representations, respectively. In more detail, we report the performance obtained when considering both the EEG signals taken from the 30 subjects acquired during all the 6 scheduled sessions, as well as those from the 45 subjects acquired in the first 5 sessions. The reported recognition rates show a performance degradation over the considered 3-year time period. Such behavior has not been observed in [5], where 3 sessions spanning a 1-month period have been considered. In order to counteract the degraded recognition performance due to the observed aging effect, we propose some possible mitigation strategies in the next section.

VII. AGING EFFECTS COUNTERMEASURES

The longitudinal analysis reported in the previous sections highlights that aging effects in EEG signals cannot be neglected, when considering a wide time span between enrolment and verification. In the following we propose different strategies to counteract these undesirable effects.

A. Template update

Template update is one of the strategies most commonly used to contrast the effects of aging in biometric recognition systems [49]. In our tests, given the verification session S_j, the enrolment dataset is built by collecting EEG signals acquired during all sessions S_t, with $t < j$, at a distance of at least Δ_k, with $k = 1, \cdots, 5$ (see Fig. 4), from session S_j. The EEG data associated to each considered scenario are specified in Table III. For the sake of a fair comparison, we consider

TABLE III
DATA EMPLOYED FOR TESTS WITH TEMPLATE UPDATE.

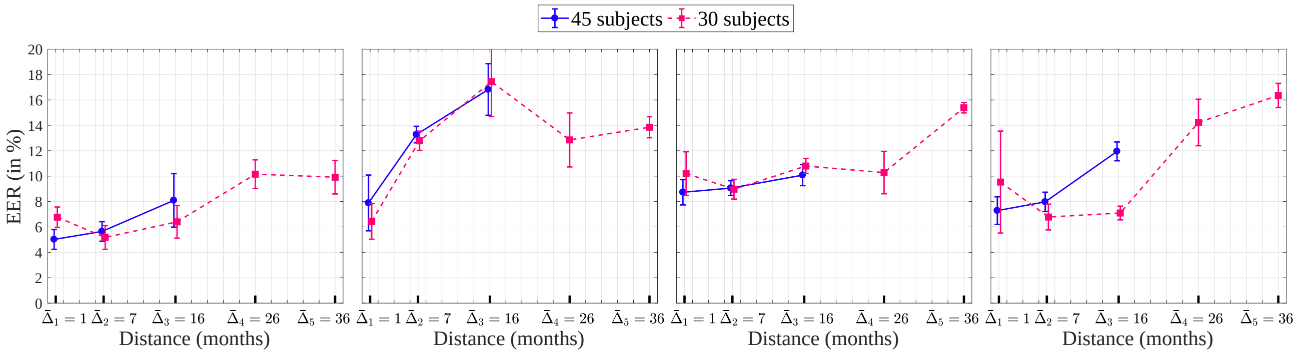
Distance	Enrolment	Verification
Δ_1	S_1, S_2	S_3
Δ_2	S_1, S_2, S_3	S_4
	S_1, S_2, S_3, S_4	S_5
Δ_3	S_1, S_2, S_3	S_5
Δ_4	S_1, S_2, S_3, S_4, S_5	S_6
Δ_5	S_1, S_2, S_3	S_6

an overall number of enrolment epochs E equal to that used when analyzing single-session enrolment scenarios.

Figures 16, 17, and 18 show the recognition rates respectively obtained with AR-, MFCC-, and bump-based EEG representations, when exploiting multi-session enrolment. Comparing these results with those in Figures 13-15, the performance improvement due to the template update approach is clearly noticeable, with the bump representation being the one that most benefits from template-update strategies. In more detail, it is possible to notice that, although an improvement in terms of absolute accuracy can be achieved when resorting to multi-session enrolment, as already evidenced in [5], such approach does not significantly reduce the performance variability observed across time. The strategies described in the following are proposed to further deal with this issue.

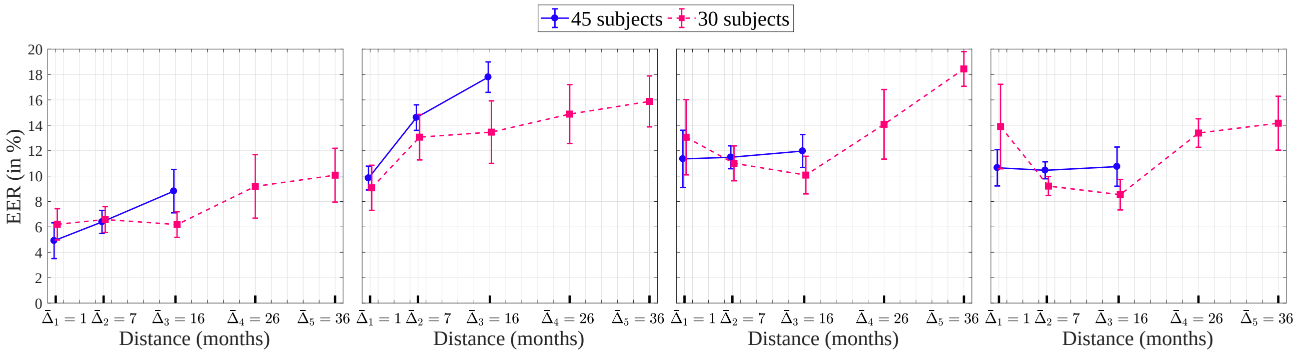
B. EEG representation

Multiple EEG representations can be jointly exploited to improve the achievable recognition rates and reduce their



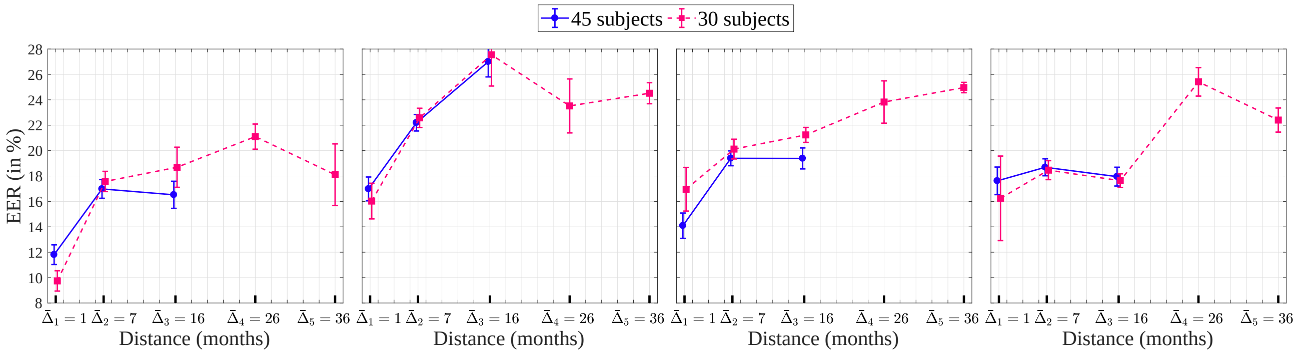
(a) (b) (c) (d)

Fig. 13. EER at time distances Δ_t, $t = 1, \dots, 5$, 45s-long verification probes, $C = 15$ channels, AR features. (a): EC; (b): EO; (c): MC; (d): SI.



(a) (b) (c) (d)

Fig. 14. EER at time distances Δ_t, $t = 1, \dots, 5$, with 45s-long verification probes, $C = 15$ channels, MFCC features. (a): EC; (b): EO; (c): MC; (d): SI.



(a) (b) (c) (d)

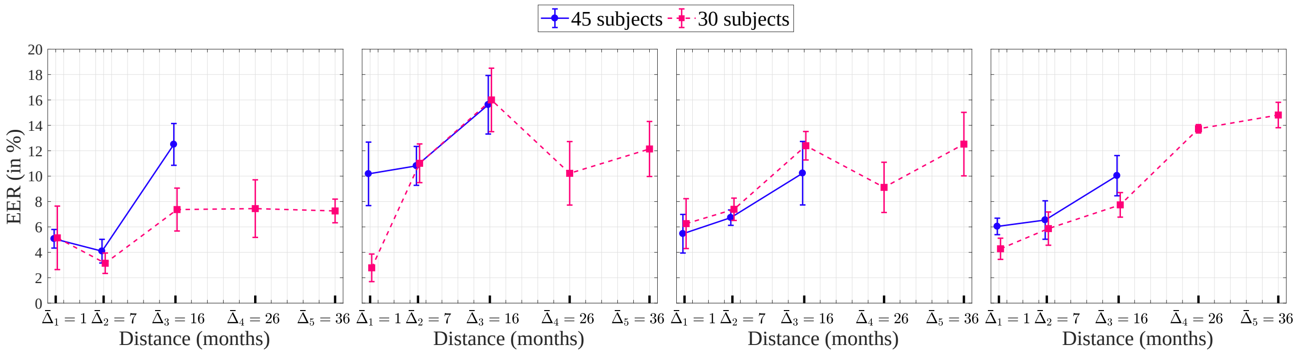
Fig. 15. EER at time distances Δ_t, $t = 1, \dots, 5$, with 45s-long verification probes, $C = 15$ channels, bump features. (a): EC; (b): EO; (c): MC; (d): SI.

variability over time. In our tests we have implemented fusion strategies at the feature-, score-, and decision-level, based on the employed AR, MFCC, and bump representations. Among the considered approaches, decision-level fusion provides the best recognition performance. Specifically, a positive decision on the i-th verification epoch is taken according to the following rule:

$$z_i = \begin{cases} 1 \text{ if } {}^{(AR)}z_i + {}^{(MFCC)}z_i + {}^{(Bump)}z_i \geq 2 \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

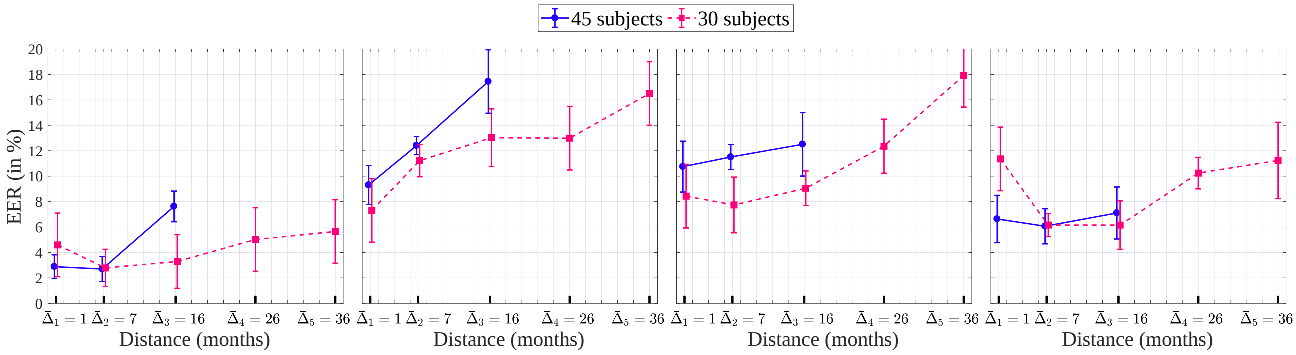
The obtained recognition rates are reported in terms of 95% confidence intervals in Figure 19. The improved recognition rates, and the reduced performance variability for all the considered acquisition protocols, testifies the effectiveness of the proposed fusion scheme, which exploits the heterogeneity of aging effects on different biometric representations to limit their impact on the achievable verification accuracy. High-level permanence is obtained especially for the EC scenario, where EERs below 4% can be guaranteed even when comparing EEG signals captured at a time distance of 3 years.

This system configuration is further analyzed in order to evaluate the effects of the number of employed channels and of the length of the verification probe on the achievable recognition performance. Specifically, Figure 20 shows the mean EERs obtained, at time distances Δ_3 and Δ_5, when varying the number of electrodes used for EEG acquisition, considering the channel ranking provided in Figure 10. It can



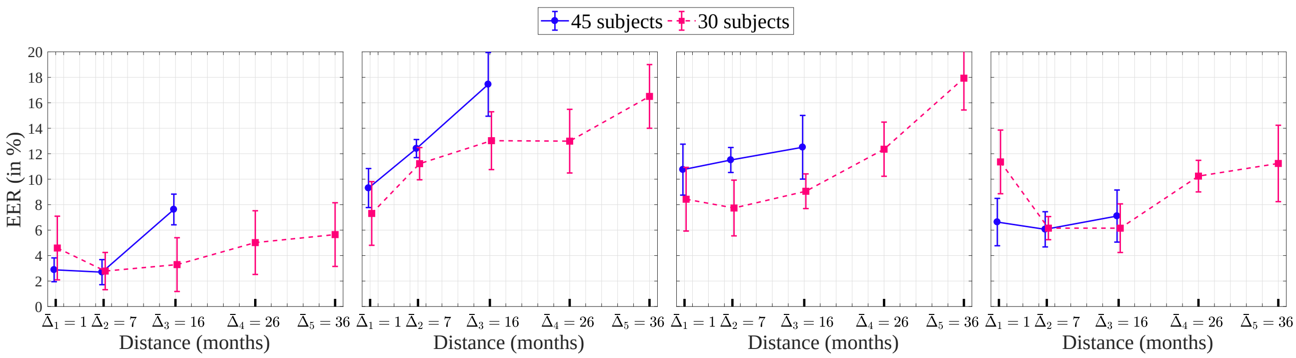
(a) (b) (c) (d)

Fig. 16. EER at time distances Δ_t, $t = 1, \ldots, 5$, with 45s-long verification probes, $C = 15$ channels, AR features, multi-session enrolment. (a): EC; (b): EO; (c): MC; (d): SI.



(a) (b) (c) (d)

Fig. 17. EER at time distances Δ_t, $t = 1, \ldots, 5$, with 45s-long verification probes, $C = 15$ channels, MFCC features, multi-session enrolment. (a): EC; (b): EO; (c): MC; (d): SI.

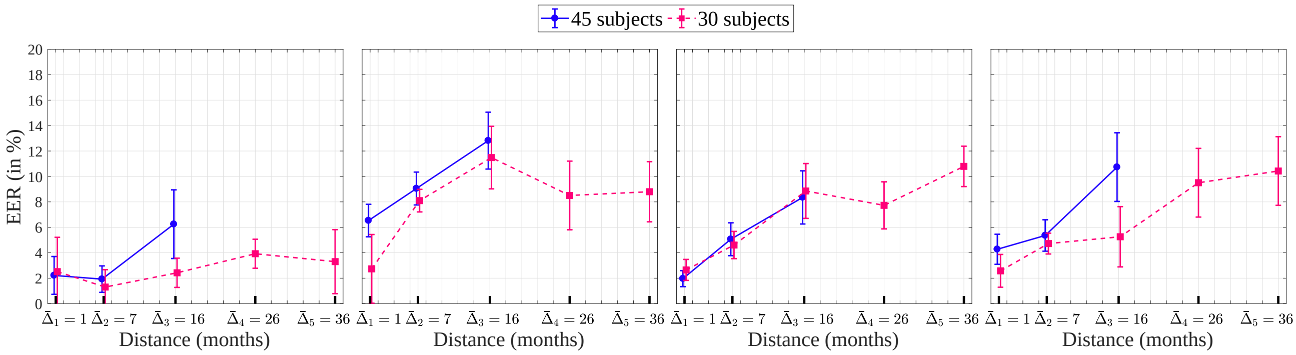


(a) (b) (c) (d)

Fig. 18. EER at time distances Δ_t, $t = 1, \ldots, 5$, with 45s-long verification probes, $C = 15$ channels, bump features, multi-session enrolment. (a): EC; (b): EO; (c): MC; (d): SI.

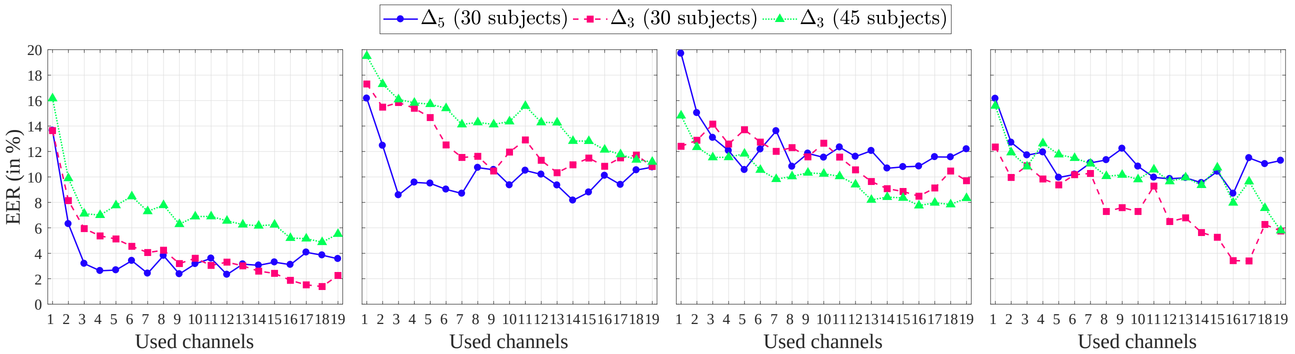
be seen that most of EEG discriminative capability can be exploited using only 9 electrodes, as outlined in Section VI-A.

Actually, an even lower number of channels could be considered to further improve the usability of EEG-based biometric recognition systems in practical applications, but the resulting performance would show a larger variability over time. In order to illustrate this effect, three different scenarios are evaluated in the following while taking into account the observations reported in Section VI-A, using $C = 4$, $C = 6$ and $C = 9$ channels in the adopted EEG montage. Figures 21-23 show the mean EERs obtained for each acquisition protocol when varying the number of epochs involved in the verification process, for the three considered scenarios. The reported results show that considerable recognition accuracies could be already achieved when exploiting only $C = 4$ channels, although a not negligible variation over time is noticeable in this case. Improved stability can be obtained when two additional electrodes are considered, while the use of $C = 9$ channels further improves the achievable recognition rates, attesting the reliability of the proposed statistical analysis and the resulting channel ranking in terms of both uniqueness and permanence properties.



(a) (b) (c) (d)

Fig. 19. EER at time distances Δ_t, $t = 1, \ldots, 5$, with 45s-long verification probes, $C = 15$ channels, multi-session enrolment, decision-level fusion of AR, MFCC and bump features. (a): EC; (b): EO; (c): MC; (d): SI.



(a) (b) (c) (d)

Fig. 20. EER *vs* no. of employed channels, for EEG signals compared at either Δ_3 and Δ_5 time distances, with 45s-long verification probes, multi-session enrolment, and decision-level fusion of AR, MFCC and bump features. (a): EC; (b): EO; (c): MC; (d): SI.

C. Elicitation protocol fusion

The results obtained with the previous analysis are exploited when evaluating the third proposed countermeasure against EEG aging effects, exploiting fusion at the protocol level. Specifically, Figures 21-23 show that limited improvements are typically gained when using, for verification purposes, EEG signals lasting more than 26s in EC and EO conditions, 40s for MC conditions, and 38s for SI. EEG signals with such lengths are therefore employed as verification probes for the considered elicitation protocols, whose final decisions x are fused according to the OR rule, which guarantees better recognition accuracy and stability with respect to other fusion rules, according to the performed experimental tests. Figure 24, 25, and 26 show the performance obtained when exploiting different combinations of the considered elicitation protocols, and adopting EEG montages with $C = 4$, $C = 6$ and $C = 9$ channels, respectively. The reported results show that although remarkable results can be achieved even exploiting only $C = 4$ EEG electrodes in the adopted EEG montage, the inclusion of more channels and more protocols not only improves the achievable recognition rates, but also their stability over time. Actually, an EERs below 2% for all the considered time distances, when employing $C = 9$ electrodes and all the employed elicitation protocols, can be achieved.

VIII. Conclusions

A detailed longitudinal analysis on the discriminative characteristics of EEG signals captured in both resting and active states, performed on a database comprising signals captured from 45 users during 5 to 6 sessions, covering an overall period of about 3 years, has been presented. HMMs have been employed to model and compare EEG representations expressed through AR, MFCC, and bump features. Both the performed statistical and performance analysis, respectively investigating the behavior of the genuine score distributions and of the achievable recognition rates for different time distances between enrolment and verification phases, have evidenced that aging actually affects EEG biometric traits. Besides providing a ranking of the EEG channels employed in the adopted montage, taking into account both uniqueness and permanence capabilities, several strategies for mitigating EEG aging effects have been also proposed, showing that EERs below 2% can be achieved also when comparing samples taken at temporal distances in the order of years.

References

[1] N. Al-Qazzaz, S. Ali, S. Ahmad, K. Chellappan, S. Islam, and J. Escudero, “Role of EEG as biomarker in the early detection and classification of dementia,” *The Scientific World Journal*, vol. 2014, pp. 1–16, 2014.

[2] M. Fouad, K. Amin, N. El-Bendary, and A. Hassanien, “Brain computer interface: A review,” in *Brain-Computer Interfaces: Current Trends and Applications*, A. Hassanien and A. Azar, Eds. Springer International Publishing, 2015, pp. 3–30.

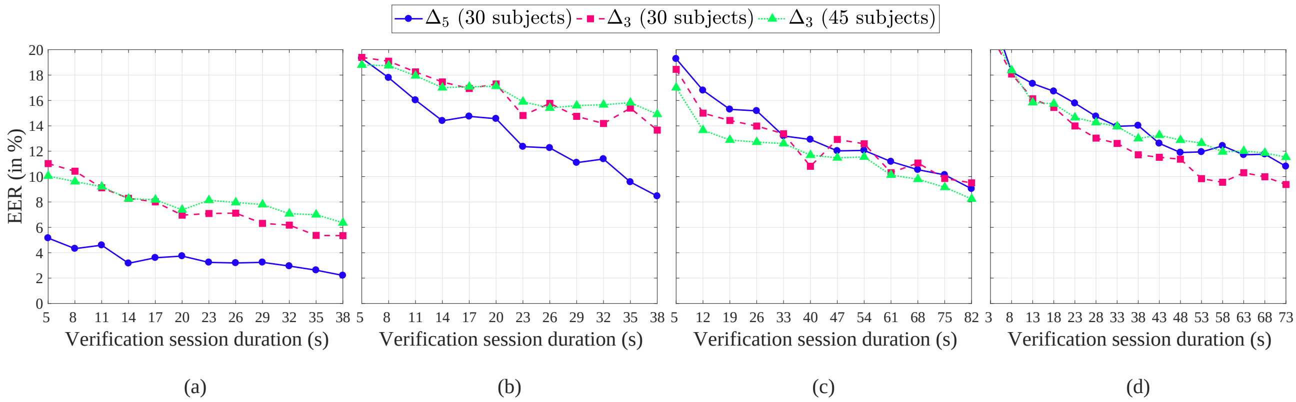


Fig. 21. EER *vs* duration of verification sessions, for EEG signals compared at either Δ_3 and Δ_5 time distances, using $C = 4$ acquisition channels, multi-session enrolment, and decision-level fusion of AR, MFCC and bump features. (a): EC; (b): EO; (c): MC; (d): SI.

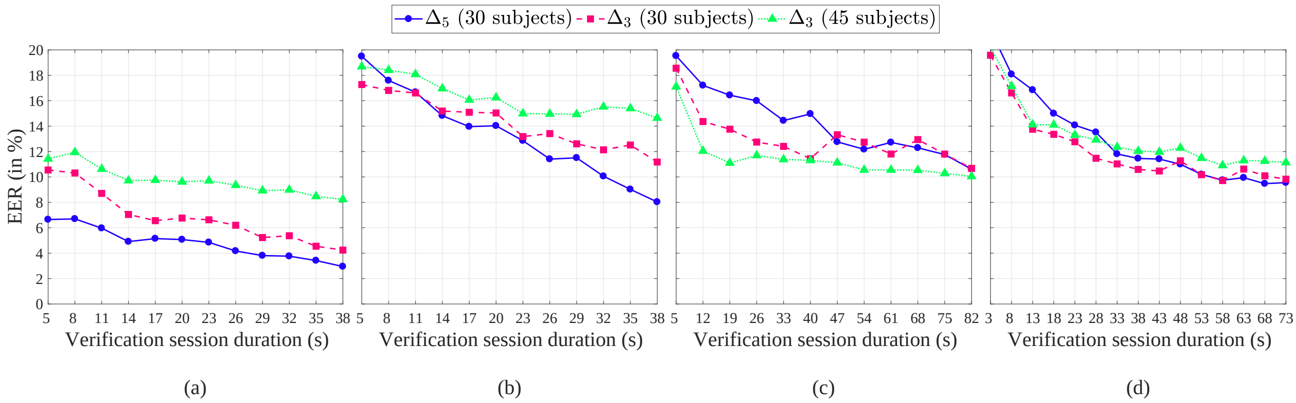


Fig. 22. EER *vs* duration of verification sessions, for EEG signals compared at either Δ_3 and Δ_5 time distances, using $C = 6$ acquisition channels, multi-session enrolment, and decision-level fusion of AR, MFCC and bump features. (a): EC; (b): EO; (c): MC; (d): SI.

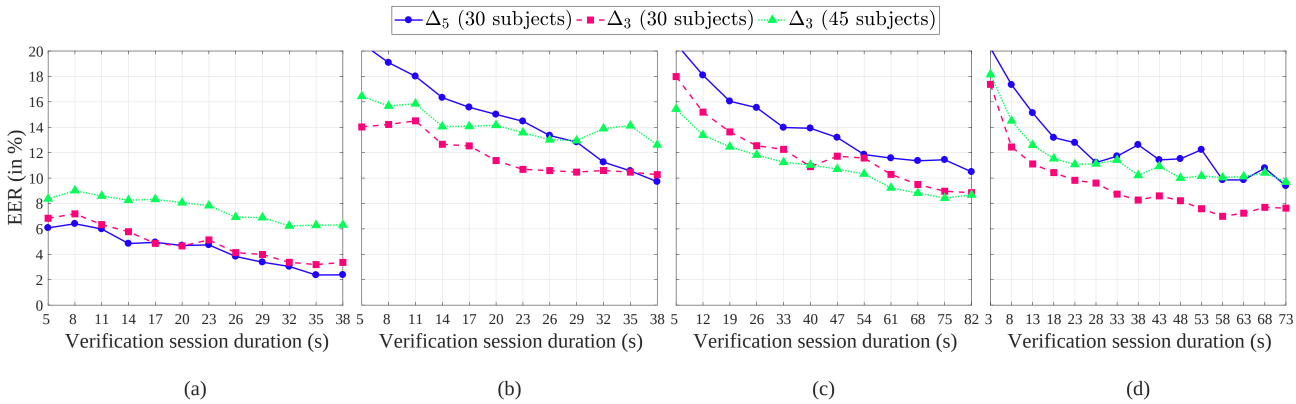


Fig. 23. EER *vs* duration of verification sessions, for EEG signals compared at either Δ_3 and Δ_5 time distances, using $C = 9$ acquisition channels, multi-session enrolment, and decision-level fusion of AR, MFCC and bump features. (a): EC; (b): EO; (c): MC; (d): SI.

[3] H. H. Stassen, "Computerized recognition of persons by EEG spectral patterns," *Electroen Clin Neuro*, vol. 49, pp. 190–194, 1980.

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

[5] E. Maiorana, D. La Rocca, and P. Campisi, "On the permanence of EEG signals for biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 163–175, 2016.

[6] A. Jeffreys, V. Wilson, and S. Thein, "Quantitative evaluation of the effects of aging on biometric templates," *IET Computer Vision*, vol. 5, no. 6, pp. 338–347, 2011.

[7] M. Fairhurst (Ed.), *Age Factors in Biometric Processing*. IET, 2013.

[8] S. Yoon and A. Jain, "Longitudinal study of fingerprint recognition," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 112, no. 28, pp. 8555–8560, 2015.

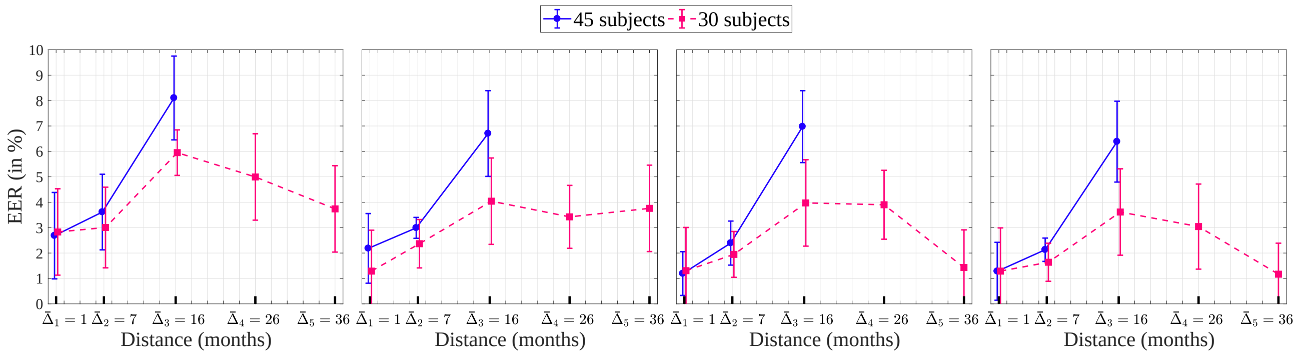
[9] L. Best-Rowden and A. Jain, "Longitudinal study of automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, 2017.

[10] P. Grother, J. Matey, E. Tabassi, G. Quinn, and M. Chumakov, "IREX VI: temporal stability of iris recognition accuracy," NIST Interagency Report 7948, https://www.nist.gov/node/578086, 2015, accessed: 2017-08-07.

[11] K. W. Bowyer and E. Ortiz, "Critical examination of the IREX VI results," *IET Biometrics*, vol. 4, no. 4, pp. 192–199, 2015.

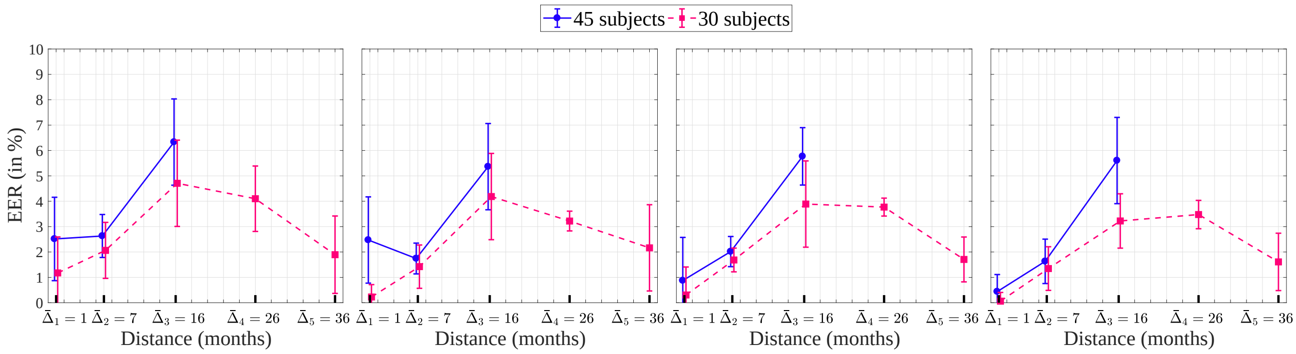
[12] S. Fenker, E. Ortiz, and K. Bowyer, "Template aging phenomenon in iris recognition," *IEEE Access*, vol. 4, no. 4, pp. 266–274, 2013.

[13] J. Ortega-Garcia *et al.*, "The multiscenario multi-environment biosecure multimodal database (BMDB)," *IEEE Transactions on Pattern Analysis*



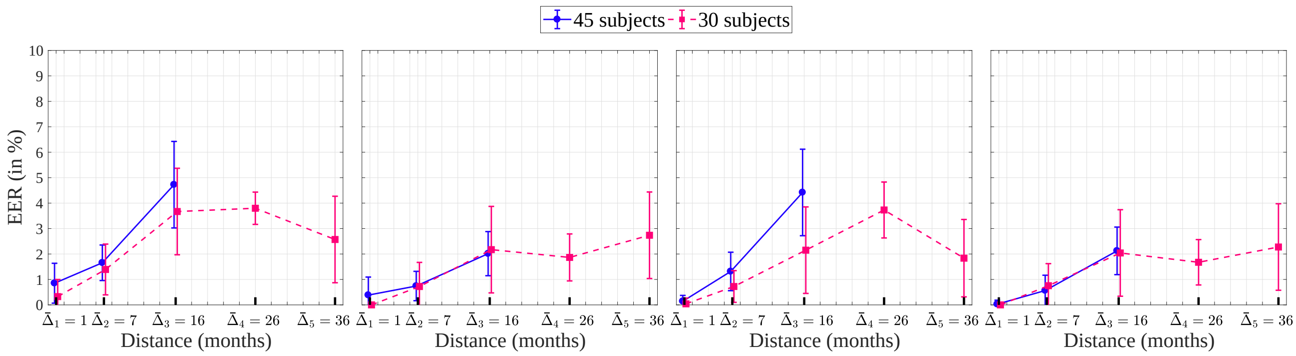
(a) (b) (c) (d)

Fig. 24. EER at different time distance Δ_t, $t = 1, \ldots, 5$, $C = 4$ channels, multi-session enrolment, decision-level fusion of AR, MFCC and bump features. (a): EC+EO (52s verification); (b): EC+EO+MC (92s verification); (c): EC+EO+SI (90s verification); (d): EC+EO+MC+SI (130s verification).



(a) (b) (c) (d)

Fig. 25. EER at different time distance Δ_t, $t = 1, \ldots, 5$, $C = 6$ channels, multi-session enrolment, decision-level fusion of AR, MFCC and bump features. (a): EC+EO (52s verification); (b): EC+EO+MC (92s verification); (c): EC+EO+SI (90s verification); (d): EC+EO+MC+SI (130s verification).



(a) (b) (c) (d)

Fig. 26. EER at different time distance Δ_t, $t = 1, \ldots, 5$, $C = 9$ channels, multi-session enrolment, decision-level fusion of AR, MFCC and bump features. (a): EC+EO (52s verification); (b): EC+EO+MC (92s verification); (c): EC+EO+SI (90s verification); (d): EC+EO+MC+SI (130s verification).

and Machine Intelligence, vol. 32, no. 6, pp. 1097–1111, 2010.

[14] J. Fierrez-Aguilar, J. Ortega-Garcia, D. Torre-Toledano, and J. Gonzalez-Rodriguez, “Biosec baseline corpus: a multimodal biometric database,” *Pattern Recognition*, vol. 40, no. 4, pp. 1389–1392, 2007.

[15] J. Galbally, M. Martinez-Diaz, and J. Fierrez, “Aging in biometrics: An experimental analysis on on-line signature,” *PlosOne*, vol. 8, no. 7, pp. 1–16, 2013.

[16] F. Kelly and N. Harte, “The impact of ageing on speech-based biometric systems,” in *Age Factors in Biometric Processing*, M. Fairhurst, Ed. IET, 2013, ch. 9, pp. 171–184.

[17] G. Veres, M. Nixon, and J. Carter, “Model-based approaches for predicting gait changes over time,” in *IEEE Intelligent Sensors, Sensor Networks and Information Processing Conference (ISSNIP)*, 2005.

[18] C. Tappert, M. Villani, and S.-H. Cha, “Keystroke biometric identification and authentication on long-text input,” in *Behavioral Biometrics for Human Identification: Intelligent Applications*, L. Wang and X. Geng, Eds. IGI Global, 2010, ch. 16, pp. 342–367.

[19] H. Dutağaci, G. Fouquier, E. Yörük, B. Sankur, L. Likforman-Sulem, and J. Darbon, “Hand recognition,” in *Guide to Biometric Reference Systems and Performance Evaluation*, D. Petrovska-Delacrétaz, G. Chollet, and B. Dorizzi, Eds. Springer-Verlag London, 2009.

[20] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, “On evaluating ECG biometric systems: Session-dependence and body posture,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp.

2002–2013, 2014.

[21] A. Jeffreys, V. Wilson, and S. Thein, "Individual-specific "fingerprints" of human DNA," *Nature*, vol. 316, no. 6023, pp. 76–794, 1985.

[22] A. Pinegger, S. Wriessnegger, J. Faller, and G. Muller-Putz, "Evaluation of different EEG acquisition systems concerning their suitability for building a braincomputer interface: Case studies," *Frontiers in Neuroscience*, vol. in print, 2016.

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

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

[25] S. Marcel and J. D. R. Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 743–748, 2006.

[26] M. Näpflin, M. Wildi, and J. Sarnthein, "Test-retest reliability of resting EEG spectra validates a statistical signature of persons," *Clinical Neurophysiology*, vol. 118, no. 11, pp. 2519–2524, 2007.

[27] K. Brigham and B. V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *Proceedings of the IEEE Fourth International Conference on Biometrics: Theory, Applications and Systems (BTAS'10)*, 2010.

[28] M. Kostílek and J. Šťástný, "EEG biometric identification: Repeatability and influence of movement-related EEG," in *IEEE International Conference on Applied Electronics*, 2012.

[29] D. La Rocca, P. Campisi, and G. Scarano, "On the repeatability of EEG features in a biometric recognition framework using a resting state protocol," in *BIOSIGNALS 2013*, 2013, pp. 419–428.

[30] 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.

[31] Y. Wang and L. Najafizadeh, "On the invariance of EEG-based signatures of individuality with application in biometric identification," in *IEEE International Conference of the Engineering in Medicine and Biology Society (EMBC)*, 2016.

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

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

[34] M. Ruiz-Blondet, Z. Jin, and S. Laszlo, "Permanence of the CEREBRE brain biometric protocol," *Pattern Recognition Letters*, vol. 95, pp. 37–43, 2017.

[35] D. McFarland, L. McCane, S. David, and J. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 386–394, Sep. 1997.

[36] B. Torrésani and E. Villaron, "Harmonic hidden Markov models for the study of EEG signals," in *European Signal Processing Conference (EUSIPCO)*, 2010.

[37] S. Chiappa and S. Bengio, "HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems," in *European Symposium on Artificial Neural Networks (ESANN)*, 2004.

[38] W. Penny and S. Roberts, "Gaussian observation hidden Markov models for EEG analysis," Technical report TR-98-12, Imperial College of Science, Technology and Medicine, London, 1998.

[39] S.-T. Pan, C.-E. Kuo, J.-H. Zeng, and S.-F. Liang, "A transition-constrained discrete hidden Markov model for automatic sleep staging," *BioMedical Engineering OnLine*, vol. 11, no. 52, pp. 1–19, 2012.

[40] C. N. B. Obermaier, C. Guger and G. Pfurtscheller, "Hidden Markov models for online classification of single trial EEG data," *Pattern Recognition Letters*, vol. 22, no. 12, pp. 1299–1309, 2012.

[41] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech processing," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[42] K. Murphy, *Machine Learning: a Probabilistic Perspective*. The MIT Press, 2012.

[43] S. Kay, *Modern Spectral Estimation. Theory and Applications*. Prentice-Hall, 1988.

[44] P. Nguyen, D. Tran, X. Huang, and D. Sharma, "A proposed feature extraction method for EEG-based person identification," in *International Conference on Artificial Intelligence (ICAI)*, 2012.

[45] P. Davis, C. Creusere, and J. Kroger, "Subject identification based on EEG responses to video stimuli," in *IEEE International Conference on Image Processing (ICIP)*, 2015.

[46] F. Vialatte, C. Martin, N. Ravel, B. Quenet, G. Dreyfus, and R. Gervais, "Oscillatory activity, behaviour and memory, new approaches for LFP signal analysis," in *European Brain and Behaviour Neuroscience Society (EBBS)*, 2003.

[47] F. Vialatte, C. Martin, R. Dubois, J. Haddad, B. Quenet, R. Gervais, and G. Dreyfus, "A machine learning approach to the analysis of time-frequency maps, and its application to neural dynamics," *Neural networks*, vol. 20, no. 2, pp. 194–209, 2007.

[48] D. La Rocca, P. Campisi, and J. Solé-Casals, "EEG-based user recognition using bump modelling," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2013.

[49] L. Didaci, G. L. Marcialis, and F. Roli, "Analysis of unsupervised template update in biometric recognition systems," *Pattern Recognition Letters*, vol. 37, pp. 151–160, 2014.

Emanuele Maiorana (SM'17) received the Ph.D. degree in biomedical, electromagnetism and telecommunication engineering with European Doctorate Label from Roma Tre University, Rome, Italy, in 2009. He was a visiting researcher at the Universidad Autonoma de Madrid, Madrid, Spain, in 2007 and 2008, at the Universidad de Vigo, Vigo, Spain, in 2010, at the University of Warwick, Coventry, UK, in 2012, at the Ecole Polytechnique de Nantes, Nantes, France, in 2013, at the University of Twente, Twente, The Netherlands, in 2013, and at the University of Salzburg, Salzburg, Austria, in 2015. He is currently a research engineer at the Section of Applied Electronics, Department of Engineering, of Roma Tre University, Rome, Italy. He is the recipient of the Lockheed Martin Best Paper Award for the Poster Track at the IEEE Biometric Symposium 2007, and of the Honeywell Student Best Paper Award at the IEEE Biometrics: Theory, Applications and Systems conference 2008. His research interests are in the area of digital signal and image processing with applications to multimedia communications and security of telecommunication systems. Specifically, he worked on biometric recognition and protection of biometric templates, high dynamic range images imaging and watermarking, synthesis of video textures, and stereo image analysis and enhancement.

Patrizio Campisi (SM'08) received the Ph.D. degree in electrical engineering from Roma Tre University, Rome, Italy, where he is currently a Full Professor with the Section of Applied Electronics, Department of Engineering. His current research interests are in the area of secure multimedia communications and biometrics. He is co-recipient of the IEEE ICIP06 and the IEEE BTAS 2008 Best Student Paper Award and the IEEE Biometric Symposium 2007 Best Paper Award. He is the General Chair of the 26th European Signal Processing Conference EUSIPCO 2018, September 2018, Rome, Italy. He has been the General Chair of the seventh IEEE Workshop on Information Forensics and Security (WIFS) 2015, Rome, Italy, and of the 12th ACM Workshop on Multimedia and Security, Rome, Italy in 2010. He has been Technical Co-Chair of the First ACM Workshop on Information Hiding and Multimedia Security, France, in 2013, and of the Fourth IEEE WIFS, Spain, in 2012. He is the Editor of the book Security and Privacy in Biometrics (Springer, 2013). He is a Coeditor of the book Blind Image Deconvolution: Theory and Applications (CRC press, 2007). He has been an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He is currently a Senior Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. He is the IEEE SPS Director Student Services and the Chair of the IEEE Technical Committee on Information Forensics and Security (IFS-TC). He is a member of the IEEE Technical Committee on Information Assurance and Intelligent Multimedia-Mobile Communications, System, Man, and Cybernetics Society and was a member of the IEEE Certified Biometric Program Learning System Committee.