

Online SSVEP-based BCI using Riemannian geometry

Emmanuel Kalunga, Sylvain Chevallier, Quentin Barthélemy, Karim Djouani, Eric Monacelli, Yskandar Hamam

▶ **To cite this version:**

Emmanuel Kalunga, Sylvain Chevallier, Quentin Barthélemy, Karim Djouani, Eric Monacelli, et al.. Online SSVEP-based BCI using Riemannian geometry. Neurocomputing, 2016, 191, pp.55-68. 10.1016/j.neucom.2016.01.007 . hal-01351623

HAL Id: hal-01351623
https://hal.science/hal-01351623

Submitted on 4 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Neurocomputing 191 (2016) 55–68

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Online SSVEP-based BCI using Riemannian geometry

Emmanuel K. Kalunga [a,b], Sylvain Chevallier [b,*], Quentin Barthélemy [c], Karim Djouani [a], Eric Monacelli [b], Yskandar Hamam [a]

[a] Department of Electrical Engineering and the French South African Institute of Technology, Tshwane University of Technology, Pretoria 0001, South Africa
[b] Laboratoire d'Ingénierie des Systèmes de Versailles, Université de Versailles Saint-Quentin, 78140 Velizy, France
[c] Mensia Technologies, S.A. ICM, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France

ARTICLE INFO

Article history:
Received 10 April 2015
Received in revised form
17 December 2015
Accepted 8 January 2016
Communicated by S. Fiori
Available online 17 February 2016

Keywords:
Riemannian geometry
Online
Asynchronous
Brain–Computer Interfaces
Steady State Visually Evoked Potentials

ABSTRACT

Challenges for the next generation of Brain Computer Interfaces (BCI) are to mitigate the common sources of variability (electronic, electrical, biological) and to develop online and adaptive systems following the evolution of the subject's brain waves. Studying electroencephalographic (EEG) signals from their associated covariance matrices allows the construction of a representation which is invariant to extrinsic perturbations. As covariance matrices should be estimated, this paper first presents a thorough study of all estimators conducted on real EEG recording. Working in Euclidean space with covariance matrices is known to be error-prone, one might take advantage of algorithmic advances in Riemannian geometry and matrix manifold to implement methods for Symmetric Positive-Definite (SPD) matrices. Nonetheless, existing classification algorithms in Riemannian spaces are designed for offline analysis. We propose a novel algorithm for online and asynchronous processing of brain signals, borrowing principles from semi-unsupervised approaches and following a dynamic stopping scheme to provide a prediction as soon as possible. The assessment is conducted on real EEG recording: this is the first study on Steady-State Visually Evoked Potential (SSVEP) experimentations to exploit online classification based on Riemannian geometry. The proposed online algorithm is evaluated and compared with state-of-the-art SSVEP methods, which are based on Canonical Correlation Analysis (CCA). It is shown to improve both the classification accuracy and the information transfer rate in the online and asynchronous setup.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Human–machine interactions without relying on muscular capabilities is possible with Brain–Computer Interfaces (BCI) [1] They are the focus of a large scientific interest [2–4], especially those based on electroencephalography (EEG) [5]. From a large literature based on the BCI competition datasets [6–8], one can identify the two most challenging BCI problems: on the one hand, the inter-individual variability plagues the models and leads to BCI-inefficiency effect [9–11], on the other hand, the intra-individual changes calls for the development of online algorithms and adaptive systems following the evolution of the subject's brain waves [12–14]. To alleviate these variations, several signal processing and machine learning techniques have been proposed, such as filtering, regularization or clustering [15,16] without the emergence of an obvious "best candidate" methodology.

A common vision is shared by all the most successful approaches to reduce signal variabilities: they are applied on covariance matrices instead of working in the input signal space. Common Spatial Pattern (CSP) [17–19], which is the most known preprocessing technique in 2-class BCI, try to maximize the covariance of one class while minimizing the covariance of the other. Similarly, Principal Components Analysis (PCA) [6,7], also applied for spatial filtering in BCI, is based on the estimation of covariance matrices. Canonical Correlation Analysis (CCA) is another example of a technique relying on covariance estimates successfully applied on EEG for spatial filtering [15,20]. Covariance matrices are also found in classifiers such as the Linear Discriminant Analysis (LDA), which is largely used in BCI. In all cases, they are handled as elements of an Euclidean space. However, being Symmetric and Positive-Definite (SPD), covariance matrices lie on a subset of the Euclidean space, with reduced dimensionality and specific properties, the *Riemannian manifold*. Considering covariance matrices in their original space would reduce the search area for an optimization problem [21,22]. As Riemannian manifolds inherently define a metric, the distance between SPD matrices takes into account the space where they lie on; approximating it to an Euclidean space introduces inaccuracies and results in ill-conditioned matrices.

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2016.01.007
0925-2312/© 2016 Elsevier B.V. All rights reserved.

Recently, studies have been done to consider covariance matrices obtained from multichannel brain signals in their original space [23–25]. Covariance matrices are the input features of the BCI system and the classifier algorithms rely on Riemannian metric for partitioning the feature space. The authors propose building specific covariance matrices in order to emphasize the spatial and frequential information of the multichannel brain signals [25]. The outcome of this approach is a simple processing tool chain, which achieves state-of-the-art classification performances.

This paper introduces an online version of the *minimum distance to Riemannian mean* (MDRM) algorithm [23], with an application to Steady-State Visually Evoked Potential (SSVEP) signals. In SSVEP, the subjects concentrate on stimuli blinking at fixed frequencies. Depending on the focus of their attention, brain waves will arise with the same phase and frequency as the stimulus chosen by the subject. The signals are recorded in an application of assistive robotics,[1] with a shared control scheme relying on an SSVEP-based BCI and a 3D touchless interface based on IR-sensors to operate an arm exoskeleton [26]. The long term objective is to equip a home environment with assistive technologies, including BCI, as proposed in [27,28]. In this context, it is important to design an online system, i.e. that adapt continuously to the user's brain signals, and asynchronous, i.e. that could be activated "on demand".

Our online implementation[2] is similar to the unsupervised or semi-unsupervised learning scheme proposed in [29,30]; that has the potential of shortening (or even removing) the calibration phase. We apply a similar approach to the dynamic stopping criterion used in [31] to increase the speed of the BCI system. This approach allows to dynamically determine the trial length and ensure robustness in classification results. Our MDRM approach outperforms state-of-the-art algorithms in the offline setup. Moreover, these state-of-the-art algorithms, that are based on CCA, are inherently limited as they could not handle resting state. They must rely on an external command to be turn on or off, and are thus only suitable to lab environment.

When working with covariance matrices, a crucial point is to correctly estimate the covariance when the number of samples is small or heavily corrupted by noise. Several approaches have been proposed to build the covariance matrices, relying on normalization or regularization of the sample covariances. To assess the quality of the covariance matrices obtained from EEG samples, a comparative study of these estimators is conducted.

Hence, the contributions of this works are:

- a comprehensive review of the literature on Riemannian geometry applied to EEG and time-series,
- a thorough analysis of the covariance estimators and their impact on tools derived from information geometry,
- first online application of a Riemannian classification algorithm on SSVEP-based BCI,
- introduction of a novel algorithm for online and asynchronous BCI, including a resting state class, yielding better performance than state-of-the-art SSVEP algorithms. No phase synchronization is required for the SSVEP.

The paper is divided as follows: Section 2 reviews the state of the art in SSVEP-based BCI and the applications of Riemannian geometry in machine learning for BCI. Section 3 presents concepts of Riemannian geometry relevant to this work and estimators of covariance. In Section 4, the proposed classification algorithm for online SSVEP is introduced and the experimental results are presented in Section 5 for offline and online setups as well as without and with a resting state class.

2. State of the art

2.1. Steady-state visually evoked potential

Sensory evoked potentials often oppose Event Related Potential (ERP) and Steady-State Response (SSR) [32]. This distinction originates from the idea that the SSR may be generated by neural oscillations elicited by the repeated stimulations [33] whereas the ERP is the transient response to an event occurring at sufficiently long time interval to allow the system to return to its initial state [34]. We will focus on the visual SSR, called SSVEP and its application to BCI.

The SSVEP-based BCI is often employed as a dependent BCI [35], that is, some residual muscular capabilities are required to move the eye toward the blinking stimulus as opposed to independent BCI, such as Motor Imagery (MI), where the communication does not rely on any motor capability. It has been shown that SSVEP could be used as an independent BCI [36,37] as the brain oscillations are strongly related to the focus of attention. Using covert attention, i.e. shifting the focus of attention without moving the eyes, subjects can generate different SSVEP responses.

BCI have highly variable subject-specific performances. 20–30% of the subjects cannot operate correctly brain interfaces. This phenomenon is referred to as BCI illiteracy [9–11]. It affects SSVEP-based BCI and it is correlated with age and gender, male subjects being more afflicted than female ones [38]. Offline BCI, that is approaches where the learning algorithms are trained on a large dataset of subject's EEG recording, are also afflicted which indicate that a source of variability at the subject level is not handled correctly by the existing approaches. BCI illiteracy is also afflicting online approaches, where the algorithms are adapted to the subject's EEG as the experiment goes by.

Visual stimulus plays a crucial role, affecting the BCI performance, and should be designed carefully. An in-depth review of the literature [39] shows that LED stimuli provide better results than those obtained on computer screen. A cognitive study [40] indicates that any stimulation between 2 and 50 Hz induces visible oscillations in the visual cortex. Another study shows that a peak in signal to noise ratio is visible at around 15 Hz [41]. Common values employed in SSVEP studies are between 12 and 25 Hz, as they induce oscillations with higher amplitudes [39]. One should note that safety of the subject should be taken into account as some frequency ranges of the stimulation train could trigger epileptic seizure [42].

The phase of the stimulation signal can also be modulated, enhancing the BCI performance by boosting the Information Transfer Rate (ITR) [43,44]. An important constraint in that case is that the experimental setup requires a synchronization between the display and the recording system, to ensure the correct estimation of the stimulus' phase. Better alternatives are available when considering systems with such constraints: code-modulated VEP (c-VEP) has yield the highest ITR in BCI [45,46]. In c-VEP, the sole difference is that the stimulus flickering is based on pseudorandom sequences instead of the fixed frequencies of SSVEP. All these successful approaches in SSVEP and c-VEP rely on CCA. Given two sets of signals, CCA aims at finding the projection space that maximizes their cross-covariance while jointly minimizing their covariance [20,15,44]. The common methodology is to find the canonical space between the multichannel EEG trial on the one hand and reference signals, usually sine and cosine of target frequencies and harmonics, on the other hand.

[1] This dataset is freely available from https://github.com/sylvchev/dataset-ssvep-exoskeleton.

[2] The open source code is available on https://github.com/emmanuelkalunga/Online-SSVEP.

This study is part of our efforts to conceive a smart and adapted environment for people with disabilities, with a standing point similar to [27,28]. In our case, the device generating flickering stimulus should not be connected to the EEG processing system, to allow, for example, the design of "smart switches" distributed in the home environment in further studies. Hence, we could not rely on phase-dependent setups, such as c-VEP or phase-modulated SSVEP.

Another requirement of our system is to be self-paced, a property also called asynchronous, to provide the user with the ability to use the system "on demand", i.e. when needed. As pointed out by [47], it is thus necessary to provide a "no-control" state to cope with situation where the user does not want to produce any command. Several methods could be considered, such as including a reject threshold in the system, as in [47], or directly provide the system with a reject class, see for example [48]. This "no-control" state or "resting" state is not always included in the existing studies, see for example [27,20,44]. The high ITR obtained with these systems are thus confined to the lab environment and could not be directly applied to realistic assistive scenarios.

2.2. Riemannian geometry in BCI

Information geometry provides useful tools for various machine learning and optimization problems. In machine learning, SPD matrices have been used in various applications where features and data are only considered in the Euclidean space. Indeed, covariance matrices lie in the space of SPD matrices which is a subset of the Euclidean space when considered with the scalar product. But the same space of SPD matrices, endowed with a differential structure, induces a Riemannian manifold.

Riemannian geometry can improve machine learning algorithms, taking explicitly into consideration the underlying structure of the considered space. Three kinds of approaches in the literature use the geometry of data in machine learning. The first one relies on the mapping of the Riemannian manifold onto an Euclidean vector space. One such mapping, called logarithmic mapping, exists between the manifold and its tangent space, which is an Euclidean space, and has been used in classification task for BCI [24]. Some kernels have been applied successfully to this end: Stein kernel, Log-Euclidean kernels as well as their normalized versions [49]. The main idea is to map the input data to a high dimensional feature space, providing a rich and hopefully linearly separable representation. The so-called kernel trick is to provide a kernel function, which computes an inner product in the feature space directly from points lying in the input space, defining a Reproducing Kernel Hilbert Space (RKHS). The family of kernels defined on the Riemannian manifold allows the implementation of extensions of all kernel-based methods, such as SVM, kernel-PCA or kernel k-means [50]. Apart from the kernel approaches, once the data are mapped onto a vector space, any machine learning algorithm working in Euclidean space, such as LDA, could be applied [23].

A second kind of machine learning approach exploits the underlying geometry of the data. Instead of mapping the data to an Euclidean space, either a tangent space or an RKHS, the algorithms are adapted to Riemannian space. For instance, sparse coding algorithm has been adapted to Riemannian manifold, using the geodesic distance to estimate the data point and its sparse estimate [51]. Similarly nonlinear dimensionality reduction techniques have been adapted to Riemannian manifold, such as Laplacian Eigenmaps (LE), Locally Linear Embedding (LLE), and Hessian LLE. This adaptation was used to cluster data using their probability density functions (pdf) [52] or covariance matrices [53] as features. Another example is the adaptation of interpolation and filtering of data to Riemannian space performed in [54], where an affine-invariant Riemannian metric is also proposed to offer a geodesically complete manifold i.e. a manifold with no edge and no singular point that can be reached in finite time.

In the last kind of approach, instead of adapting existing algorithm from Euclidean to Riemannian geometry, new algorithms are developed directly for Riemannian manifolds. The *minimum distance to Riemannian mean* (MDRM) relies on a Riemannian metric to implement a multi-class classifier and have been applied on EEG. New EEG trials are assigned to the class whose average covariance matrix is the closest to the trial covariance matrix [23]. The MDRM classification can be preceded by a filtering of covariance matrices, like in [55] where covariance matrices are filtered with LDA component in the tangent space, then brought back to the Riemannian space for classification with MDRM. Another example is the *Riemannian Potato* [56], an unsupervised and adaptive artifact detection method, providing an online adaptive EEG filtering (i.e. outliers removal). Incoming signals are rejected if their covariance matrix lies beyond a predefined z-score, computed from a sliding window. With the same objective of achieving robustness to noise that affects covariance matrices, Riemannian geometry is used to solve divergence functions of pdfs [57]. This allows to reformulate the CSP as the maximization of the divergence between the distributions of data from two different classes corresponding to two cognitive states [58,59]. Using the *beta divergence* the obtained CSP is robust to outliers in sample covariance matrices and this algorithm is successfully applied to EEG filtering for BCI. Riemannian metrics are also used for EEG channel selection [60] and the selection of the most discriminatory spatial filters in CSP [61].

In MI experiment, the subject is asked to imagine a movement (usually hand, feet or tongue), generating Event-Related Synchronization and Desynchronization (ERD/ERS) in pre-motor brain area. Riemannian BCI is well suited for MI experiment as the spatial information linked with synchronization is directly embedded in covariance matrices obtained from multichannel recordings. However, for BCI that rely on Evoked Potential such as SSVEP or Event Related Potential (ERP), as P300, both frequential and temporal information are needed; the spatial covariance matrix does not contain these information. To apply Riemannian geometry to SSVEP and ERP, the sample covariance matrices can be defined from a rearrangement of the recorded data. The rearrangement is done such that the temporal or frequency information are captured [25]. With similar motivations, [62,63] defined a new Riemannian distance between SPD matrices that would take into account a weighting factor on matrices. They use this new distance as a dissimilarity between weighted matrices of power spectral density to classify EEG into different sleep state by k-nearest neighbors.

3. Covariance matrices and their geometry

This section presents some formal definitions for the information geometry concepts used in this paper. The link with the covariance matrices is explicated in Section 3.2, along with the covariance estimators proposed in the literature.

3.1. Riemannian manifold

An m-dimensional *manifold* $\mathcal{M}$ is a Hausdorff space for which every point has a neighborhood that is homeomorphic to an open subset of $\mathbb{R}^m$ [64]. When a tangent space is defined at each point, $\mathcal{M}$ is called a differential manifold. A *geodesic* γ is the shortest smooth curve between two points, Σ_1 and Σ_2. The tangent space $T_{\Sigma}\mathcal{M}$ at point Σ is the vector space spanned by the tangent vectors of all geodesics on $\mathcal{M}$ passing through Σ. A *Riemannian* manifold is

a manifold endowed with an inner product defined on the tangent space, which varies smoothly from point to point.

For the rest of this paper, we will restrict to the analysis of the manifold $\mathcal{M}_C$ of the $C \times C$ symmetric positive definite matrices, defined as:

$$\mathcal{M}_C = \left\{ \Sigma \in \mathbb{R}^{C\times C} : \Sigma = \Sigma^\top \text{ and } x^\top \Sigma x > 0, \forall x \in \mathbb{R}^C \backslash 0 \right\}.$$

The tangent space $T_\Sigma \mathcal{M}_C$ is identified to the Euclidean space of symmetric matrices:

$$\mathcal{S}_C = \left\{ \Theta \in \mathbb{R}^{C\times C} : \Theta = \Theta^\top \right\}.$$

The dimension of the manifold $\mathcal{M}_C$, and its tangent space $T_\Sigma \mathcal{M}_C$, is $m = C(C+1)/2$.

The mapping from a point Θ_i of the tangent space to the manifold is called the exponential mapping $\mathrm{Exp}_\Sigma(\Theta_i)$: $T_\Sigma \mathcal{M}_C \rightarrow \mathcal{M}_C$ and is defined as:

$$\mathrm{Exp}_\Sigma(\Theta_i) = \Sigma^{\frac{1}{2}} \mathrm{Exp}(\Sigma^{-\frac{1}{2}} \Theta_i \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}}. \quad (1)$$

Its inverse mapping, from the manifold to the tangent space is the logarithmic mapping $\mathrm{Log}_\Sigma(\Sigma_i)$: $\mathcal{M}_C \rightarrow T_\Sigma \mathcal{M}_C$ and is defined as:

$$\mathrm{Log}_\Sigma(\Sigma_i) = \Sigma^{\frac{1}{2}} \mathrm{Log}(\Sigma^{-\frac{1}{2}} \Sigma_i \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}}. \quad (2)$$

$\mathrm{Exp}(\cdot)$ and $\mathrm{Log}(\cdot)$ are the matrix exponential and matrix logarithm, respectively. The computation of these operators is straightforward for SPD matrices of $\mathcal{M}_C$. They are obtained from their eigenvalue decomposition (EVD):

$$\Sigma = U \,\mathrm{diag}(\lambda_1, \ldots, \lambda_C) U^\top,$$
$$\mathrm{Exp}(\Sigma) = U \,\mathrm{diag}(\log(\lambda_1), \ldots, \log(\lambda_C)) U^\top,$$
$$\mathrm{Log}(\Sigma) = U \,\mathrm{diag}(\exp(\lambda_1), \ldots, \exp(\lambda_C)) U^\top,$$

where $\lambda_1, \ldots, \lambda_C$ are the eigenvalues and U the matrix of eigenvectors of Σ. As any SPD matrix can be diagonalized with strictly positive eigenvalues, $\mathrm{Log}(\cdot)$ is always defined. Similarly the square root $\Sigma^{\frac{1}{2}}$ is obtained as:

$$\Sigma^{\frac{1}{2}} = U \,\mathrm{diag}(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_C^{\frac{1}{2}}) U^\top,$$

and is unique. The same goes for $\Sigma^{-\frac{1}{2}}$.

The tangent vector of the geodesic $\gamma(t)$ between Σ_1 and Σ_2, where $\gamma(0) = \Sigma_1$ and $\gamma(1) = \Sigma_2$ is defined as $v = \overrightarrow{\Sigma_1 \Sigma_2} = \mathrm{Log}_{\Sigma_1}(\Sigma_2)$. A Riemannian distance between Σ_1 and Σ_2 can thus be defined as [65]:

$$\delta(\Sigma_1, \Sigma_2) = \| \mathrm{Log}(\Sigma_1^{-1} \Sigma_2) \|_F = \left[\sum_{c=1}^{C} \log^2 \lambda_c \right]^{1/2}, \quad (3)$$

where λ_c, $c = 1, \ldots, C$, are the eigenvalues of $\Sigma_1^{-1} \Sigma_2$. From Eq. (3), the geometric mean of I points Σ_i on the manifold, $i = 1, \ldots, I$, can be defined as the point that minimizes the sum of squared distances to all Σ_i:

$$\mu(\Sigma_1, \ldots, \Sigma_I) = \arg\min_{\Sigma \in \mathcal{M}_C} \sum_{i=1}^{I} \delta^2(\Sigma_i, \Sigma). \quad (4)$$

This mean has no closed form, and can be computed iteratively [66].

3.2. Covariance matrix estimation

Let $x_n \in \mathbb{R}^C$, $n = 1, \ldots, N$, denotes a sample of a multichannel EEG trial recorded on C electrodes. N is the trial length. Let $X \in \mathbb{R}^{C\times N}$ be the EEG trial such as $X = [x_1, \ldots, x_N]$. Under the hypothesis that all N samples x_n are randomly drawn from a distribution, it follows that $\boldsymbol{x}$ is a variable of random vectors and its expected vector is $\omega = E\{\boldsymbol{x}\}$ [67]. The covariance matrix of the random variable $\boldsymbol{x}$ is defined by $\Sigma = E\{(\boldsymbol{x} - \omega)(\boldsymbol{x} - \omega)^\top\}$ and is unknown, thus an estimate $\hat{\Sigma}$ should be computed. The choice of the appropriate estimator is crucial to verify that the obtained covariance matrices fulfill the following properties: they should be accurate, SPD, and well-conditioned. The last property requires that the ratio between the maximum and minimum singular value is not too large. Moreover, to ensure the computational stability of the algorithm, the estimator should provide full-rank matrices, and its inversion should not amplify estimation errors.

3.2.1. Sample covariance matrix estimator

The most usual estimator is the empirical *sample covariance matrix* (SCM), defined as:

$$\hat{\Sigma}_{\mathrm{scm}} = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \overline{x})(x_n - \overline{x})^\top = \frac{1}{N-1} X \left(\mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^\top \right) X^\top, \quad (5)$$

where $\overline{x} \in \mathbb{R}^C$ is the sample mean vector $\overline{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$. In the matrix notation, $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{1}_N$ is the vector $[1, \ldots, 1]$. The SCM is often normalized [67] as:

$$\hat{\Sigma}_{\mathrm{nscm}} = \frac{C}{N} \sum_{n=1}^{N} \frac{(x_n - \overline{x})(x_n - \overline{x})^\top}{\sigma_{x_n}^2}, \quad (6)$$

with the inter-channel variance at time $n$defined as $\sigma_{x_n}^2 = (x_n - \overline{x})^\top (x_n - \overline{x})$. Other normalization techniques could be used.

This estimation is fast and computationally simple. However when $C \approx N$, the SCM is not a good estimator of the true covariance. In the case $C > N$, the SCM is not even full rank.

3.2.2. Shrinkage covariance matrix estimators

To overcome the shortcomings of SCM, the shrinkage estimators have been developed as a weighted combination of the SCM and a target covariance matrix, which is often chosen to be close to the identity matrix, i.e. resulting from almost independent variables of unit variance.

$$\hat{\Sigma}_{\mathrm{shrink}} = \kappa \Gamma + (1 - \kappa) \hat{\Sigma}_{\mathrm{scm}}, \quad (7)$$

where $0 \leq \kappa < 1$. This estimator provides a regularized covariance that outperforms the empirical $\hat{\Sigma}_{\mathrm{scm}}$ for small sample size, that is $C \approx N$. The shrinkage estimator has the same eigenvectors as the SCM, but the extreme eigenvalues are modified i.e. the estimator is shrunk or elongated toward the average.

The different shrinkage estimators differ in their definition of the target covariance matrix Γ. Ledoit and Wolf [68] ($\hat{\Sigma}_{\mathrm{shrink_ledoit}}$ in Fig. 1) have proposed $\Gamma = \nu \mathbf{I}_C$, with $\nu = \mathrm{Tr}(\hat{\Sigma}_{\mathrm{scm}})$. Blankertz [69] ($\hat{\Sigma}_{\mathrm{shrink_blank}}$) defines Γalso as $\nu \mathbf{I}_C$ but with $\nu = \frac{\mathrm{Tr}(\hat{\Sigma}_{\mathrm{scm}})}{C}$. Schäfer ($\hat{\Sigma}_{\mathrm{shrink_schaf}}$) proposes several ways of defining Γdepending on the observed $\hat{\Sigma}_{\mathrm{scm}}$ [70].

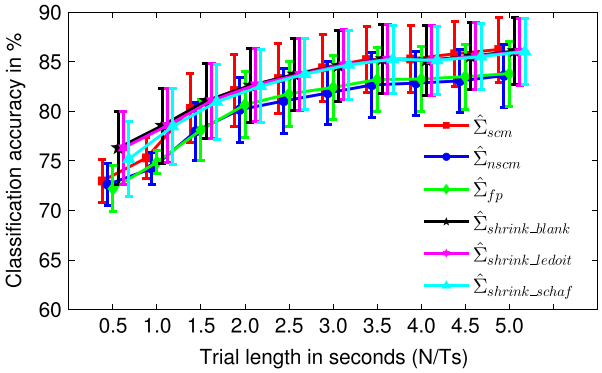


Fig. 1. Comparison of covariance estimators in terms of classification accuracy obtained with MDRM with increasing EEG trial length. For each trial length, the average accuracy across all subjects and across all replications is shown. Bars indicate the error of the mean, i.e. standard deviation divided by the square root of $n-1$, n = number of samples.

3.2.3. Fixed point covariance matrix estimator

The Fixed Point Covariance Matrix [71] is based on the maximum likelihood estimator $\hat{\mathscr{C}}$ which is a solution to the following equation:

$$\hat{\Sigma}_{\text{fp}} = \hat{\mathscr{C}} = \frac{C}{N}\sum_{n=1}^{N}\left(\frac{(x_n-\overline{x})(x_n-\overline{x})^{\top}}{(x_n-\overline{x})^{\top}\hat{\mathscr{C}}^{-1}(x_n-\overline{x})}\right). \tag{8}$$

As there is no closed form expression to Eq. (8), it can be written as a function of $\hat{\mathscr{C}}: g(\hat{\mathscr{C}}) = \hat{\Sigma}_{\text{fp}}$. g admits a single *fixed point* $\hat{\mathscr{C}}^*$, where $g(\hat{\mathscr{C}}^*) = \hat{\mathscr{C}}^*$, which is a solution to Eq. (8). Using $\hat{\mathscr{C}}_0 := \hat{\Sigma}_{\text{nscm}}$ as the initial value of $\hat{\mathscr{C}}$, it is solved recursively as $\hat{\mathscr{C}}_t \underset{t\to\infty}{\longrightarrow} \hat{\mathscr{C}}^*$.

4. Online adaptation of the Riemannian classifier

Concerning Riemannian classification of SSVEP, the offline methodology is explained in [25]. In this paper, we propose an online classifier for SSVEP, composed of an offline training phase and an online and asynchronous test phase. This analysis is performed for each subject independently.

4.1. Offline Riemannian classification

The proposed classifier relies on the Minimum Distance to Riemannian Mean (MDRM) introduced in [55] and extended in [25,72] for possible offline applications on SSVEP signals. Let us consider an experimental SSVEP setup with F stimulus blinking at F different frequencies. It is a multiclass classification with $K = F+1$ classes: one class per stimulus and one resting state class. The covariance matrices are estimated from a modified version of the input signal X:

$$X \in \mathbb{R}^{C\times N} \to \begin{bmatrix} X_{\text{freq}_1} \\ \vdots \\ X_{\text{freq}_F} \end{bmatrix} \in \mathbb{R}^{FC\times N}, \tag{9}$$

where X_{freq_f} is the input signal X band-pass filtered around frequency freq_f, $f = 1,\ldots,F$. Thus the resulting covariance matrix $\hat{\Sigma}$ belongs to $\mathcal{M}_{FC}$. Henceforth, all EEG signals will be considered as filtered and modified by Eq. (9).

From I labelled training trials $\{X_i\}_{i=1}^{I}$ recorded per subject, K centers of class $\Sigma_\mu^{(k)}$ are estimated using Algorithm 1. When an unlabelled test trial Y is given, it is classified as belonging to the class whose center $\Sigma_\mu^{(k)}$ is the closest to the trial's covariance matrix (Algorithm 2, step 2).

Algorithm 1. Offline estimation of Riemannian centers of classes.

Inputs: $X_i \in \mathbb{R}^{FC\times N}$, for $i = 1,\ldots,I$, a set of labelled trials.
Inputs: $\mathcal{I}(k)$, a set of indices of trials belonging to class k.
Output: $\Sigma_\mu^{(k)}$, $k = 1,\ldots,K$, centers of classes.

1: Compute covariance matrices $\hat{\Sigma}_i$ of X_i
2: **for** $k=1$ **to** K **do**
3: $\Sigma_\mu^{(k)} = \mu(\hat{\Sigma}_i : i \in \mathcal{I}(k))$, Eq. (4)
4: **end**
5: **return** $\Sigma_\mu^{(k)}$

Algorithm 2. Minimum distance to Riemannian mean.

Inputs: $\Sigma_\mu^{(k)}$, K centers of classes from Algorithm 1.
Input: $Y \in \mathbb{R}^{FC\times N}$, an unlabelled test trial.
Output: k^*, the predicted label of Y.

1: Compute covariance matrix $\hat{\Sigma}$ of Y
2: $k^* = \arg\min_k \delta(\hat{\Sigma}, \Sigma_\mu^{(k)})$
3: **return** k^*

4.2. Curve-based online classification

In offline synchronous BCI paradigm, cue onset are used as reference for the localization of a brain response, e.g. an evoked potential. Nonetheless most of the BCI applications are online and asynchronous; cue onsets are not known, thus designing online version of BCI algorithms is not a trivial task. The approach introduced here identifies a period (i.e. time interval) in the online EEG $\chi \in \mathbb{R}^{FC\times\mathcal{N}}$, where $\mathcal{N}$ is the number of recorded samples, associated with a high probability (above threshold) of observing an SSVEP at a specific frequency, as illustrated in Algorithm 3.

Algorithm 3. Curve-based online classification.

Inputs: hyper-parameters $w, \Delta n, D$, and ϑ.
Inputs: $\Sigma_\mu^{(k)}$, $k = 1,\ldots,K$, centers of classes from Algorithm 1 (offline training).
Inputs: Online EEG recording $\chi(n)$.
Output: $\tilde{k}(n)$, online predicted class.

1: $d = 1$
2: **for** $n = w$ **to** $\mathcal{N}$ **step** Δn
3: Epoch X_d, Eq. (10), and classify it with Algorithm 2
4: **if** $d \geq D$
5: Find the most recurrent class in $\mathcal{K} = k^*_{j\in\mathcal{J}(d)}$: $\overline{k} = \arg\max_k \rho(k)$, Eq. (11)
6: **if** $\rho(\overline{k}) > \vartheta$
7: Compute $\tilde{\delta}_{\overline{k}}$, Eq. (12)
8: **if** $\tilde{\delta}_{\overline{k}} < 0$
9: **return** $\tilde{k} = \overline{k}$
10: **end**
11: **end**
12: **end**
13: $d = d+1$
14: **end**

To locate this interval, we focus on the last D recorded EEG overlapping epochs $\{X_j \in \mathbb{R}^{FC\times w}\}_{j\in\mathcal{J}(d)}$, with the set of indices $\mathcal{J}(d) = d-D+1,\ldots,d-1,d$; where d is the index of the current epoch X_d in the online recording $\chi(n)$. Epochs have size w, and the interval between two consecutive epochs is Δn, with $w > \Delta n$:

$$X_d = \chi(n-w,\ldots,n). \tag{10}$$

To obtain the first D epochs $X_{j\in\mathcal{J}(d)}$, at least $w+(D-1)\Delta n$ samples of χ should be recorded (step 4).

The classification outputs $k^*_{j\in\mathcal{J}(d)}$ obtained in step 3 by applying Algorithm 2 on $X_{j\in\mathcal{J}(d)}$ are stored in a vector $\mathcal{K}$, which always contains the latest D classification outputs. The class that occurs the most in $\mathcal{K}$ (step 5), with an occurrence probability $\rho(k)$ above a defined threshold ϑ, is considered to be the class, denoted $\overline{k}$, of the ongoing EEG recording $\chi(n)$. The vector ρ is defined as:

$$\rho(k) = \frac{\#\{k^*_{j\in\mathcal{J}(d)} = k\}}{D}, \quad \text{for } k = 1,\ldots,K, \tag{11}$$

with $\overline{k} = \arg\max_k \rho(k)$; then $\rho(\overline{k})$ is compared to the threshold ϑ. If ϑ is not reached within the last D epochs, the classification output is held back, and the sliding process continues until ϑ is reached. In the last D epochs, once a class $\overline{k}$ has been identified, a curve direction criterion is introduced to enforce the robustness of the result. For class $\overline{k}$ to be validated, this criterion requires that the direction taken by the displacement of covariance matrices

$\hat{\Sigma}_{j\in\mathcal{J}(d)}$ be toward the center of class $\Sigma_\mu^{(\bar{k})}$. Hence $\tilde{\delta}_{\bar{k}}$, the sum of gradients (i.e. differentials) of the curve made by distances from $\hat{\Sigma}_{j\in\mathcal{J}(d)}$ to $\Sigma_\mu^{(\bar{k})}$ should be negative (step 8):

$$\tilde{\delta}_{\bar{k}} = \sum_{j\in\mathcal{J}(d)} \frac{\Delta\delta_{\bar{k}}(j)}{\Delta j} = \sum_{j=d-D+2}^{d} \delta_{\bar{k}}(j) - \delta_{\bar{k}}(j-1) < 0 \quad \text{with}$$

$$\delta_{\bar{k}}(j) = \frac{\delta(\hat{\Sigma}_j, \Sigma_\mu^{(\bar{k})})}{\sum_{k=1}^{K} \delta(\hat{\Sigma}_j, \Sigma_\mu^{(k)})}. \tag{12}$$

The occurrence criterion is inspired by the dynamic stopping of [31]; there is no fixed trial length for classification. The occurrence criterion ensures that the detected user intention is unaffected by any short time disturbances due to noise or subject's inattention, as presented in Algorithm 3. This approach offers a good compromise to obtain robust results within a short and flexible time.

The curve direction criterion solves both the problems of latency in the EEG synchronization and of the delays inserted by the EEG epochs processing. Indeed, some EEG epochs gather signals from different classes and might be wrongfully classified if the decision is solely based on the distance with the center of the class. This situation and the effect of the curve direction criterion are well shown in Section 5.4. Ensuring that the covariance matrices are displaced toward the center of the detected class provides a guarantee that it matches with the current EEG state. Inversely, if the direction of the curve is moving away from the center of the detected class, it might indicate that there has been a change in the EEG state that has not been detected.

Algorithm 3 has 4 hyperparameters: $w, \Delta n, D$, and ϑ. The values of w, D, and ϑ are set through cross validation and are given in Section 5.4. Although a large window size w is expected to increase the classification accuracy, it increases the response time, thus reducing the time resolution, and extends the overlap between different EEG states. The step size Δn should be set to a minimum value to allow a maximum number of overlapping epochs (D) within a short time. However, it should be large enough to avoid too many calculations within a time interval with small or inexistent changes in EEG states. If the number of epoch D is too small, the classification will be sensitive to non-intentional and abrupt changes in the EEG. A too large D will increase the momentum and reinforce the influence of the past EEG signals. It should also be mentioned that both the occurrence and the curve direction criteria cannot have a significant impact if the value of D is too small. The probability threshold parameter ϑ acts like a rejection parameter: high ϑ values correspond to a high rejection rate.

4.3. Outliers removal with Riemannian potato

Outliers in the training data might affect the Riemannian mean of classes in the MDRM classification scheme. To alleviate this effect, an approach called the Riemannian potato, introduced in [56], is exploited. In this approach, all trials are represented by their covariance matrices Σ_i. A reference covariance matrix is estimated, e.g. Riemannian mean of all trials Σ_μ. The Riemannian distances δ_i between each Σ_i and Σ_μ are computed. Any trial that lies too far, i.e. beyond a certain threshold, from the reference matrix Σ_μ in terms of Riemannian distance is rejected. In [56], the distance z-score thresholding is defined as:

$$z(\delta_i) = \frac{\delta_i - \mu}{\sigma} > z_{th} \tag{13}$$

where μ and σ are respectively the mean and standard deviation of distances $\{\delta_i\}_{i=1}^{I}$. In other words, any trial Σ_i whose z-score $z(\delta_i)$ is larger than the threshold $z_{th} = 2.5$ is rejected.

In this work, we propose a slightly different application of the Riemannian potato where the outliers are removed per class. Hence for K class, K Riemannian potatoes are defined $\left\{\Sigma_\mu^k, \mu^k, \sigma^k\right\}_{k=1}^{K}$. Since Riemannian distances to geometric mean do not have a Gaussian distribution, we make use of the geometric mean for μ, the geometric standard deviation for σ and the geometric z-score. They are defined as follows [72]:

$$\mu^k = \exp\left(\frac{1}{I}\sum_i \ln(\delta_i^k)\right)$$

$$\sigma^k = \exp\left(\sqrt{\frac{1}{I}\sum_i \left(\ln\left(\delta_i^k/\mu^k\right)\right)^2}\right)$$

$$z(\delta_i^k) = \frac{\ln\left(\delta_i^k/\mu^k\right)}{\ln(\sigma^k)}. \tag{14}$$

Through cross-validation, the z-score threshold is set to $z_{th} = 2.2$.

Moreover, outliers are removed iteratively. Each time outliers are rejected, a new center of class is computed and used as reference for the next iteration. Iterations continue until convergence, i.e. no more outlier found.

5. Experimental validation

Covariance matrix estimators, Algorithms 2 and 3 are applied to SSVEP signals for offline and online analysis. This section presents the analysis and results obtained.

5.1. Data description

The signals are recorded from 12 subjects during an SSVEP experiment. EEG are measured on $C = 8$ channels: O_Z, O_1, O_2, PO_Z, PO_3, PO_4, PO_7, and PO_8. The ground and the reference electrodes were placed respectively on F_Z and the right hear mastoid respectively. The acquisition rate is $T_s = 256$ Hz on a gTec MobiLab Amp (gTec, Graz, Austria). The subjects are presented with $F = 3$ visual target stimuli blinking respectively at freq = 13 Hz, 17 Hz and 21 Hz. It is a $K = 4$ classes BCI setup made of the $F = 3$ stimulus classes and one resting class (no-SSVEP). In a session, which lasts 5 min, 32 trials are recorded: 8 for each visual stimulus and 8 for the resting class. The number of sessions recorded per subject varies from 2 to 5. Thus the longest EEG recorded for a single subject is 25 min or 160 trials. The trial length is 6 s, that is $N = 6 \times T_s = 1536$ samples. Since data are rearranged as detailed in (9), trials $X \in \mathbb{R}^{FC\times N}$, where $FC = 24$ corresponding to 8 channel times 3 stimulus frequencies. For each subject, a test set is made of 32 trials whereas the remaining trials (which might vary from 32 to 128) make up for the training set.

5.2. Covariance estimators comparison

In this section, the effectiveness of covariance matrix estimators is evaluated for SSVEP signals. The evaluation is done in terms of classification accuracy and integrated discrimination improvement (IDI), obtained by each estimator (see Section 3.2) with respect to SCM estimator while using the offline MDRM classifier. The different conditioning of covariance matrices are also investigated.

A bootstrapping with 1000 replications is performed to assess the performances of each estimator. Estimators are compared on 10 trial lengths $t \in \{0.5, 1.0, \ldots, 5.0\}$ s, as these are known to affect the estimators performance. Here $N \in \{128, 256, \ldots, 1280\}$ is computed as $N = t \times T_s$.

Fig. 1 shows the classification accuracies of each estimator computed across all subjects. Even if the error bars show an

important inter-subject variability, the increase in the accuracy can be attributed to the fact that the relevant patterns in EEG accumulate with the trial length, producing better estimation of the covariance matrices. This is known to be particularly true for the SCM estimator and it could be seen in Fig. 1. It appears that shrinkage estimators (especially Ledoit and Schäfer) are less affected by the reduction of epoch sizes than the other estimators. This is a direct consequence of the regularization between the sample covariance matrices and the targeted (expected) covariance matrix of independent variables.

For computational purposes, it is important to look at the matrix conditioning. Fig. 2 shows the ratio $\mathcal{C}$ between the largest and smallest eigenvalues: in well-conditioned matrices, $\mathcal{C}$ is small. Shrinkage estimators offer better conditioned matrices whereas the SCM, NSCM, and Fixed Point matrices are ill-conditioned below 2 s of trial length, and may result in singular matrices.

In Fig. 2b, the Integrated Discrimination Improvement (IDI), as defined in [73], is computed for the different estimators and trial lengths. The SCM is used as a reference for improvement, as this is the most popular estimator in the literature. Negative IDI means a deterioration in the method discrimination ability. It is clear that shrinkage estimators increase the discrimination power of the classifier. However, despite being more complex than the SCM, the NSCM and the Fixed Point estimators decrease the discrimination ability of classifiers. From Figs. 1 and 2b, it is apparent that the difference in performance between the SCM and shrinkage estimators reduces as the trial length increases. The simplicity of the SCM plays a favorable role: it is an attractive method for longer trials. The p-values under the hypothesis that there is no improvement (i.e. IDI=0) from one estimator to another are all inferior to 10^{-47}, ($p<10^{-3}$ indicating a statistically significant discriminatory improvement); hence the improvement is significant. It should be noted that the estimation of covariance matrices is a trade-off between the quality of the estimate and the computation time required; this should be considered for real-time processing.

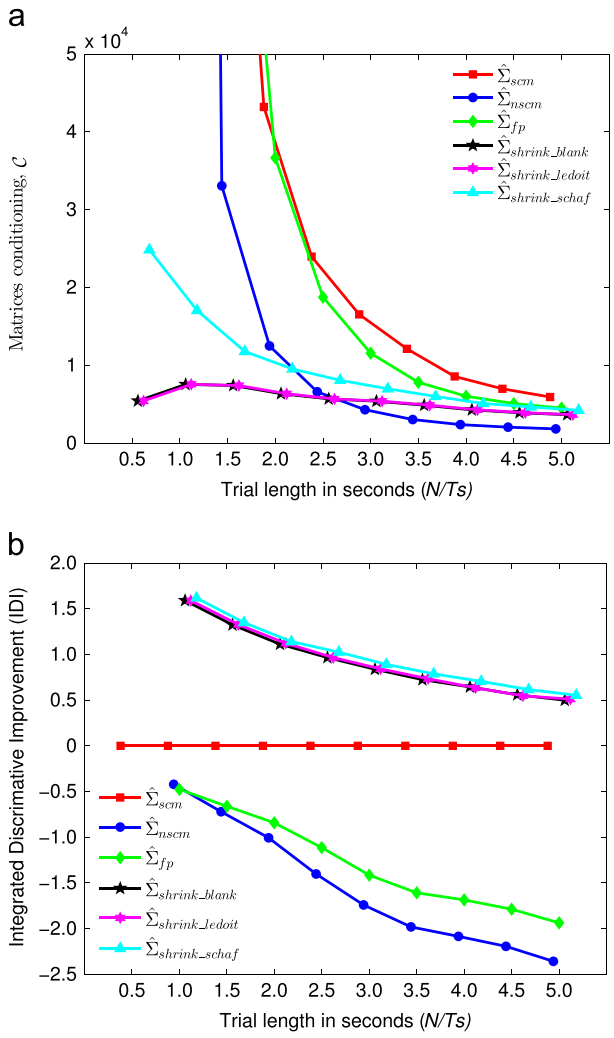


Fig. 2. (a) Covariance matrices condition expressed as the ratio $\mathcal{C}$ between largest and smallest eigenvalues for the different covariance estimators. The comparison is done for increasing EEG trial length. (b) Integrated discrimination improvement brought to the classification task by various estimators along varying trail length. The indicated IDI values are multiplied by 10^2. $\hat{\Sigma}_{\mathrm{scm}}$ is used as a baseline.

5.3. Effect of outliers on center estimations

Outliers can affect the offline training of the K centers of class $\Sigma_{\mu}^{(k)}$ by Algorithm 1, which is crucial for the evaluation phase and online application. Fig. 3 shows representations of training covariance matrices Σ_i in the tangent space (Θ_i), projected at the mean of all training trials, for the subjects with the lowest (Fig. 3a and b) and the highest (Fig. 3c and d) BCI performance. To obtain this visualization, the first two principal components of a PCA applied on $\{\Theta_i\}_{i=1}^{I}$ are selected. In Fig. 3b and d, the Riemannian potato presented in Section 4.3 is applied; outliers in each class are removed. The interest of using a Riemannian potato is well seen in Fig. 3a and b. In Fig. 3a, the outliers are so distant from the rest of the class matrices that the center of class is stretched away. Applying a Riemannian potato removes the outliers, and the center of class is better estimated (Fig. 3b).

When training trials are not noisy, their covariance matrices are compact around their Riemannian mean. In this case the removal of outliers by the Riemannian potato does not influence, at least not significantly, the Riemannian mean. This is the case in Fig. 3c and d. Thus, applying the Riemannian potato is crucial for noisy data and will have a limited effect on clean data. The impact of the Riemannian potato on the classification accuracy is discussed in Section 5.4.

5.4. Classification results and analysis

In this section, the performance of the proposed method is presented. First, the performance of the MDRM approach in an offline setup is analyzed, then the results of the online algorithm are presented. In the offline analysis, the relevance of identifying the latency between cue onset and SSVEP response is shown. The results of the MDRM approach are compared to two state-of-the-art methods [20,44]. The online evaluation is divided into two parts: in the first one the algorithm discriminates between $K=F=3$ SSVEP classes (i.e. 13, 17 and 21 Hz) and in the second one is applied on $K=4$ classes, i.e. the $F=3$ SSVEP class and the resting class.

5.4.1. Offline analysis

A close inspection of the filtered signals shows that almost all signals are synchronized with the trial frequency 2 s after cue onset $\tau_0=0$, as shown in Fig. 4. This delay is mainly due to protocol design and user specific cognitive processes. The protocol is aimed to provide an asynchronous setup close to real application. The user are not required to look at a fixation point or to directly gaze toward the target, as in [74,44], during inter-trial periods. This is a tentative explanation for the higher delay observed in our study and it is consistent with literature observations [75,76]. In fact, before τ_0+2 s, for some users the signal could still be synchronized with the previous trial frequencies. An important

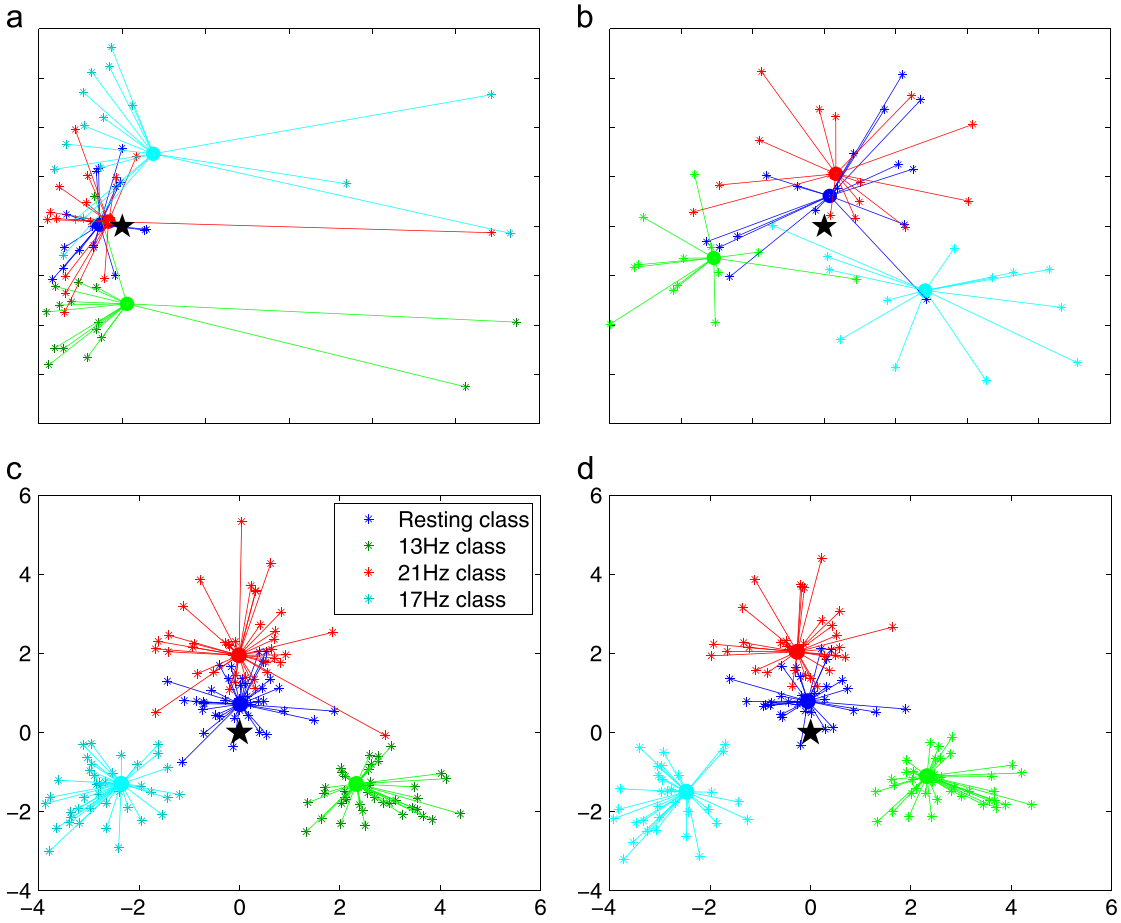


Fig. 3. Scatter plot of covariance matrices for all trials mapped on the tangent space. The distance between each trial covariance matrix Σ_i and its Riemannian mean class $\Sigma_\mu^{(k)}$ is shown as connection line. The black/thick star represents the Riemannian mean of all trials. Matrices of resting class, 13Hz class, 21Hz class, and 17Hz class are represented with starts in black, grey, dark grey, and light grey respectively. Subject with lowest BCI performance, (3a) before and (3b) after Riemannian potato filtering. Subject with highest BCI performance, (3c) before and (3d) after Riemannian potato filtering.

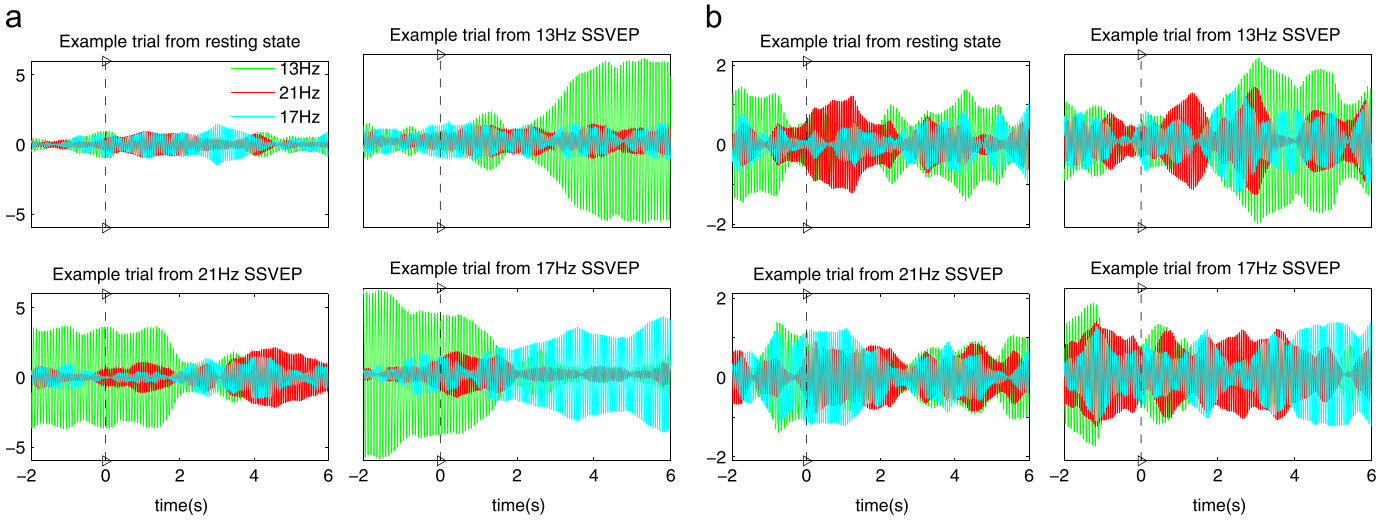


Fig. 4. Signal amplitude at each stimulus frequency, showing synchronization of EEG with respect to time (seconds). The raw signal of the trial measured on Oz is band filtered using a Butterworth of order 8 at each stimulus frequency and the resulting signals are shown in blue (light grey), green (grey), and red (dark grey) for the same signal filtered respectively at 13, 17, and 21 Hz. The cue onset τ_0 at time 0 on the x-axis is shown with a vertical discontinued line. 4 trials are shown, one for each class. Signals are from the subjects with the highest (4a) and with the lowest BCI performance (4b). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

increase in average classification accuracy (almost 10%) could be obtained by taking the trial from 2 s after cue onset. It is therefore crucial to consider the latency between the cue onset of trial and the actual synchronization of SSVEP at stimulus frequency. Thus in the offline synchronous processing, the confident window for classification is set 2 s after the cue onset (τ_0+2).

Table 1 shows the offline classification accuracies for each subject obtained by the application of the *MDRM* as described in

Table 1
Offline performance in terms of accuracy and ITR. Five methods are compared: (1) CCA approach introduced by Lin et al. [20], (2) CCA approach introduced by Nakanishi et al. [44], (3) MDRM described in Section 4.1 (Algorithm 2), (4) MDRM where processed epochs are taken 2 s from the beginning of the trial, and (5) MDRM-Potato, where outliers are removed using the Riemannian potato approach described in Section 4.3.

	Offline algorithms									
	Lin et al. [20]		Nakanishi et al. [44]		MDRM (τ_0)		MDRM		MDRM-Potato	
	acc (%)	itr (bpm)	acc (%)	itr (bpm)	acc (%)	itr (bpm)	acc (%)	itr (bpm)	acc (%)	itr (bpm)
S1	91.7	16.3	84.7	12.2	67.6	3.5	84.7	12.2	84.5	12.1
S2	45.8	0.7	47.9	1.0	66.0	3.2	79.4	9.7	79.3	9.6
S3	100.0	23.8	93.0	17.2	90.2	10.3	99.3	22.7	99.3	22.7
S4	97.9	21.3	96.6	20.0	78.3	6.1	89.7	15.0	89.7	15.0
S5	83.3	11.5	82.2	11.0	76.0	5.5	89.5	14.9	89.4	14.9
S6	77.1	8.7	76.2	8.3	72.2	4.5	87.2	13.6	87.2	13.6
S7	98.6	22.0	96.7	20.1	90.0	10.2	99.8	23.5	99.8	23.4
S8	97.9	21.3	65.5	4.7	90.4	10.3	99.7	23.2	99.7	23.2
S9	91.7	16.3	77.9	9.0	64.0	2.8	85.8	12.8	85.7	12.7
S10	80.2	10.0	76.9	8.6	79.2	6.4	93.1	17.3	93.0	17.2
S11	89.6	15.0	82.7	11.2	54.8	1.4	78.2	9.2	78.2	9.1
S12	95.8	19.4	93.8	17.8	82.3	7.4	98.6	22.0	98.6	22.0
Mean	**87.5 ± 15.1**	**15.5 ± 6.8**	**81.2 ± 14.1**	**11.8 ± 6.0**	**75.9 ± 11.4**	**6.0 ± 3.1**	**90.4 ± 7.8**	**16.3 ± 5.3**	**90.4 ± 7.8**	**16.3 ± 5.3**

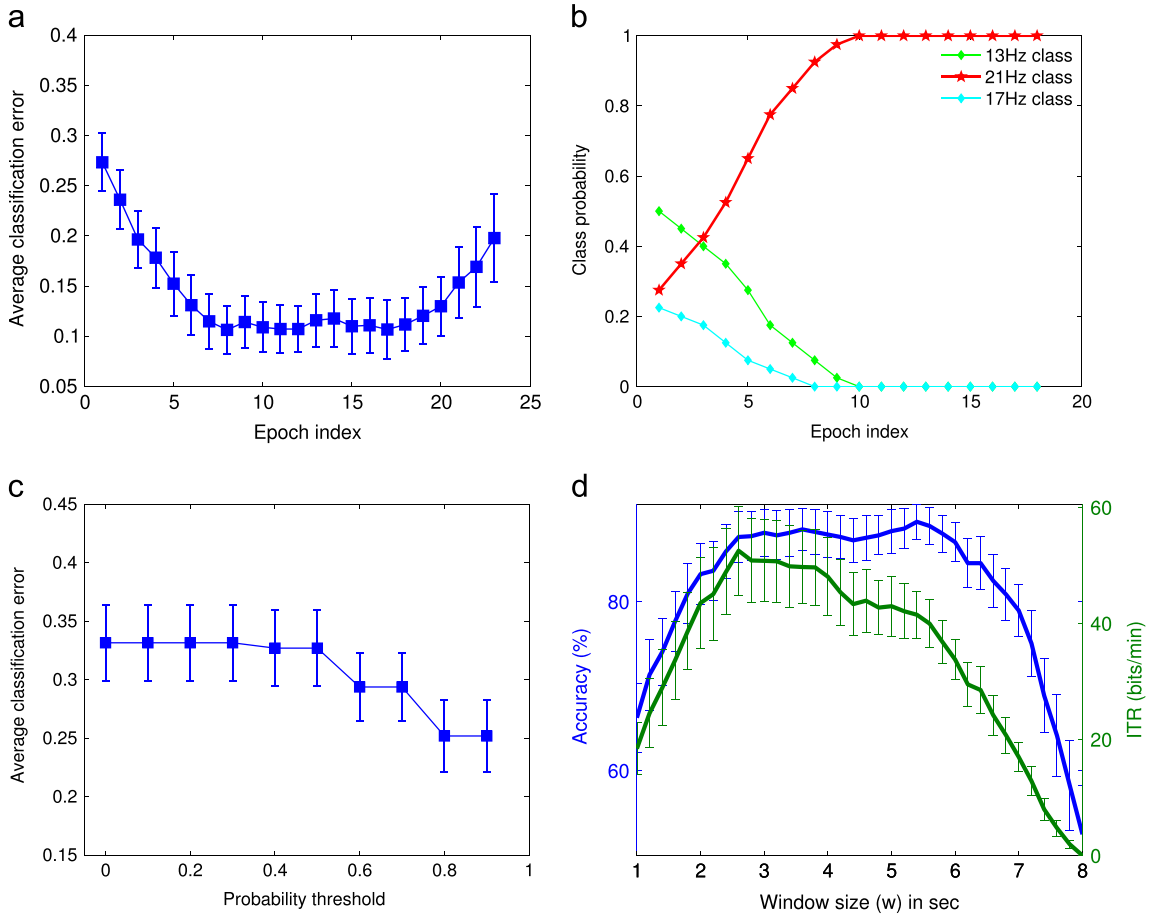


Fig. 5. Evaluation of the online algorithm parameters. 5a shows the decrease of the average classification error over all subjects during the successive epochs after the beginning of the trial. 5b is an example taken from the subject with the best performance showing how the probability of the actual class varies with epoch position from beginning of trial. The groundtruth class probability is represented with a thick-and-star line, while other classes probability lines are thin-and-diamond. 5c shows the variation of the average classification error for different probability threshold ($0 \leq \vartheta < 1$) and its influence on the classifier output (Algorithm 3 step 6). 5d shows how the average online performance varies with respect to the epoch size (w). It shows both the classification accuracy (left y-axis, black curve) and the ITR (right y-axis, grey curve). In 5a, 5c, and 5d, the bars represent the error of the mean i.e. standard deviation divided by the square root of $n-1$, $n=$ number of samples.

Algorithm 2, with the epochs taken at τ_0+2. Column MDRM (τ_0) shows the results obtained when the epochs are taken from cue onset. The Riemannian potato technique presented in Section 4.3 was applied for outliers removal (*MDRM-Potato*). The performance of the MDRM approach is compared to two CCA-based state-of-the-art methods proposed by Lin et al. [20] and Nakanishi et al.

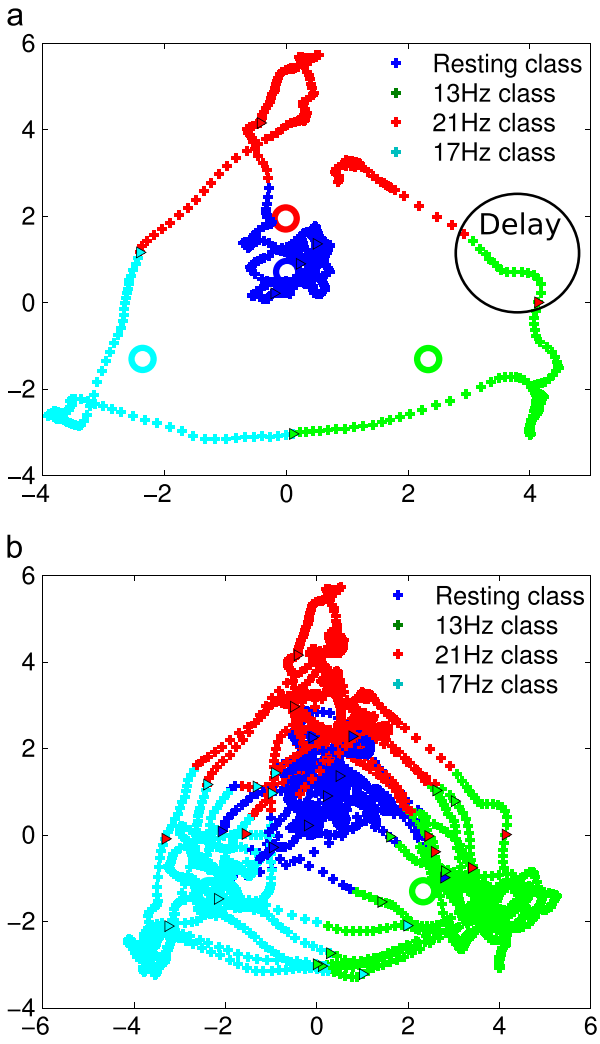


Fig. 6. Covariance matrices trajectory during a 4-class SSVEP online recording. The circles represent class centers. The triangles mark the beginning in the experiment of a new trial whose class is indicated by the triangle's color. Matrices of resting class, 13Hz class, 21Hz class, and 17Hz class are represented in black, grey, dark grey, and light grey respectively. 6a shows the first 7 trials. The first 3 trials are from the resting class, the remaining are respectively class 13 Hz, 17 Hz, and 21 Hz. 6b shows the entire recording. Data are taken from the subject with the highest BCI performance.

[44] respectively. In the implementation of these methods, the epochs are also taken from τ_0+2.

The MDRM approach outperforms both CCA-based method with an average classification accuracy of $90.4 \pm 7.8\%$ and ITR of 16.3 ± 5.3 bits/min. Lin et al. rank second with $87.5 \pm 15.1\%$ and 15.5 ± 6.8 bits/min. The method proposed by Nakanishi et al., which could be expected to achieve better results as reported in [44], only ranks third. This is mainly due to the fact that this method requires information on the phase of the stimuli. In fact, Nakanishi et al. use the average of all training trials belonging to a unique class as reference signal in the CCA. When SSVEP trials belonging to a unique trial are not in-phase, which is the case in the current work, averaging them will cancel the signal.

Within the MDRM approach, it is shown that taking into account the latency between the cue onset and the SSVEP response significantly increases the classification performances: accuracy and ITR rises from $75.9 \pm 11.4\%$ and 6.0 ± 3.1 bits/min to $90.4 \pm 7.8\%$ and 16.3 ± 5.3 bits/min. In turn removing outliers with the Riemannian potato does not bring significant change. This could be attributed to the fact that the recording have been conducted in controlled environment, with small or little external noise.

5.4.2. *Online analysis without resting class*

In an online asynchronous experiment, there is no cue onset, and the delay before SSVEP synchronization might differ from one trial to another and from one subject to another. To locate the trust EEG region for the classification, D and ϑ are set respectively to 5 and 0.7 through cross-validation. The performance of this online setup are analyzed and Fig. 5 shows the results. From the analysis shown in Fig. 5d, the epoch size is set to $w = 2.6$ s. The step size is set to $\Delta n = 0.2$ s, that is a new epoch is classified every 0.2 s.

In Fig. 5a, the classification error is plotted against the epoch index. It shows that the error decreases as epochs move from the beginning of the trial. The error increases in the last epochs of the trial, corresponding to the end of the SSVEP task. Fig. 5b details the evolution of the probability for each class as epochs index increases. It appears clearly that the class of the EEG trial (thick-and-star line) has the largest probability only a few epochs after the beginning of the trial. Moreover, one can see that this is an increasing trend over the whole trial. Thus by setting an appropriate probability threshold ϑ, the actual class can be identified with enough confidence. Fig. 5c shows the influence of the probability threshold ϑ on the classification error. The error is reduced when the probability threshold ϑ is increased. Fig. 5d shows how

Table 2
Classification performances (accuracy in %, delay before valid and confident classification in seconds, and ITR in bits/min) achieved using the online algorithm. The first column indicates the subjects. The following three columns show the results obtained without the curve direction criterion (Algorithm 3 up to 6): by stopping at step 6, $\overline{k}$ is taken to be the valid class. The next three columns contain the results of the complete online algorithm. The last three columns report the results obtain when outliers are removed in the training phase using the Riemannian potato technique described in Section 4.3.

	Online ($\rho(\overline{k}) > \vartheta$)			Online (full Algorithm 3)			Online-Potato		
	acc (%)	delay (s)	itr (bpm)	acc (%)	delay (s)	itr (bpm)	acc (%)	delay (s)	itr (bpm)
S1	68.8	0.8	26.3	77.1	1.1	27.9	77.1	1.1	27.9
S2	64.6	0.7	21.6	77.1	1.2	26.8	77.1	1.2	26.8
S3	81.2	0.7	54.3	95.8	1.0	73.0	95.8	1.0	73.0
S4	83.3	0.8	53.2	91.7	1.0	58.6	95.8	1.0	69.2
S5	72.9	0.7	37.1	83.3	1.0	42.5	83.3	1.0	42.5
S6	66.7	0.7	24.5	72.9	1.1	24.3	72.9	1.1	24.3
S7	93.1	0.7	89.6	98.6	0.9	87.0	98.6	0.9	86.8
S8	87.5	0.6	76.2	100.0	0.9	95.9	100.0	0.9	95.9
S9	60.4	0.7	15.7	77.1	1.2	27.6	77.1	1.2	27.6
S10	64.6	0.7	21.5	87.5	1.1	45.3	87.5	1.1	45.3
S11	54.2	0.7	9.9	87.5	1.3	38.9	87.5	1.3	38.9
S12	52.5	0.7	8.0	99.2	1.2	71.7	99.2	1.2	71.8
Mean	**70.8 ± 13**	**0.7 ± 0.0**	**36.5 ± 26.3**	**87.3 ± 9.8**	**1.1 ± 0.1**	**51.6 ± 25.1**	**87.7 ± 10**	**1.1 ± 0.1**	**52.5 ± 25.5**

Table 3
This table summarizes the performance achieved with the online algorithm with resting class identification, as in Table 2.

	Online ($\rho(\overline{k}) > \vartheta$)			Online (full Algorithm3)			Online-Potato		
	acc (%)	delay (s)	itr (bpm)	acc (%)	delay (s)	itr (bpm)	acc (%)	delay (s)	itr (bpm)
S1	67.2	0.7	37.6	71.4	1.1	32.4	71.4	1.1	32.4
S2	78.1	0.7	59.0	75.0	1.0	39.2	75.0	1.0	39.2
S3	89.1	0.8	85.2	89.1	1.0	67.6	89.1	1.0	67.6
S4	75.0	0.7	52.2	75.0	0.9	42.9	75.0	0.9	43.4
S5	71.9	0.7	46.7	70.3	1.1	31.0	70.3	1.1	31.0
S6	87.5	0.8	80.2	87.3	1.1	58.7	87.3	1.1	58.7
S7	84.4	0.7	76.3	85.4	1.0	62.5	88.5	1.0	69.1
S8	85.9	0.8	76.4	89.1	1.0	68.1	89.1	1.0	68.1
S9	67.2	0.7	37.2	75.0	1.0	39.6	76.6	1.1	40.3
S10	62.5	0.7	30.3	69.5	1.0	32.0	69.5	1.0	32.0
S11	59.4	0.8	23.5	68.8	1.1	29.1	68.8	1.1	29.1
S12	69.4	0.7	44.8	93.8	1.0	79.4	93.8	1.0	79.9
Mean	**74.8 ± 10.2**	**0.7 ± 0.0**	**54.1 ± 21.0**	**79.1 ± 9.1**	**1.0 ± 0.1**	**48.6 ± 17.6**	**79.5 ± 9.3**	**1.0 ± 0.1**	**49.2 ± 18.2**

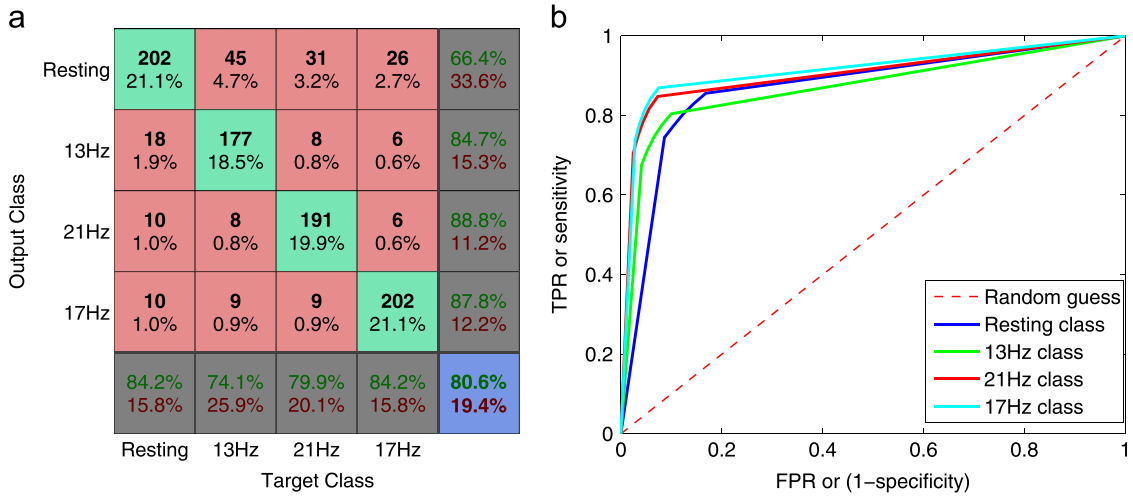


Fig. 7. (a) Confusion matrix for $K=4$ classes with Online-Potato. (b): ROC curve indicating the influence of the ϑ parameter.

the average online performance varies with respect to the epoch size (w). Both the classification accuracy and the ITR are shown. With short w values, the epoch size does not capture enough feature for a correct classification, and with long w, the epoch loses temporal resolution. The ITR increases with the classification rate but drops sensibly after a peak value.

The observation of Fig. 6 provides a visualization of the principle guiding the online implementation of Eq. (12). This figure shows the trajectory on the tangent space taken by covariance matrices during a 4-class SSVEP experiment, and how they are classified epoch by epoch. It can be seen (encircled in Fig. 6a) that a change in the SSVEP stimulus might not be detected instantaneously by the classifier. The trials are erroneously attributed with confidence to the previous class. The proposed online algorithm, described in Algorithm 3, mitigates this issue and increases the classification accuracy as shown in Table 2. The "Online ($\rho(\overline{k}) > \vartheta$)" column shows the results of the online algorithm without the curve direction criterion (i.e., without steps 6–11), and "Online (full Algorithm 3)" shows the improvement brought by this criterion. The performances are in terms of average classification accuracy (acc (%)), average time taken into the trial before classification (delay (s)), and the ITR (itr (bits/min)).

The curve direction criterion increases the rejection of epochs that could be wrongly classified, it thus significantly increases the classification accuracy of the online algorithm ($70.8 \pm 13\%$ to $87.3 \pm 9.8\%$), while increasing the delay (0.7–1.1 s) before classification. When compared to the state-of-the-art offline MDRM, the online curve-based classification yields better results in terms of ITR as the delay before classification is much shorter in the latter than the trial length used in the former; classification outputs are reached faster with the online algorithm. Moreover, the online algorithm can be applied in both synchronous and asynchronous paradigms, whereas the offline algorithms are limited to synchronous paradigms which provide strongly limited user interaction.

Last, the impact of the Riemannian potato is analyzed. A bootstrapping with 50 replications was performed on the offline data to assess the effect of applying the Riemannian potato. The results show that for most subjects the results are unchanged when the Riemannian potato is applied: due to the fact that data are recorded in a controlled environment, most of them are thus clean. It does however improve the results of few subjects. It was then applied in the training phase of the online application, and a similar observation is made. We can conclude that the Riemannian potato can be used as a safety guard to ensure that the Riemannian mean used in the MDRM classification scheme is not affected by outliers, especially for BCI used in less controlled environment.

5.4.3. *Online analysis with resting class*

Using the MDRM approach it is possible to identify the resting class. In fact, covariance matrices of signal recorded during resting periods can be characterized with their own Riemannian mean. As such, they can be identified as any other class using the MDRM

approach. The state-of-the-art methods, Lin et al. [20] and Nakanishi et al. [44], are both based on CCA where a reference signal is needed. These methods do not handle resting class, since there is no reference signal for them. In this section, the performance of the proposed approach including the identification of the resting class is presented. Table 3 summarizes the classifier performance in the same format as Table 2, in terms of classification accuracy, delay before valid classification and ITR. Like in Table 2, the best performance is achieved by the complete online algorithm preceded with outliers removal with the Riemannian potatoes (i.e. *Online-Potato*). The identification of the resting class induces a drop of the overall classification accuracy by 8.2%, and a drop of ITR from 52.5 ± 25.5 to 49.2 ± 18.2.

The effect of the resting class is seen with more details in Fig. 7. Fig. 7a shows the classification confusion matrix. There are few misclassifications between SSVEP classes compared to the misclassifications between the resting class and any SSVEP class: the largest percentages are located in the first row and the first column, apart from the diagonal block. Fig. 7b displays a ROC curve showing how the classifier performs in discriminating each class versus the others depending on the value of the ϑ parameter. On this ROC curve, the performance of the *Online−Potato* algorithm are indicated in terms of False Positive Rate (FPR) and True Positive Rate (TPR).

Confirming the observation from the confusion matrix, the ROC curve indicates that the resting is the most prone to false positive. Despite the drop in performance, the identification of resting class is crucial for online BCI setup, allowing the subject to use the system at his own pace.

6. Conclusion

This work investigated the efficiency of Riemannian geometry when dealing with covariance matrices as classification features. A novel algorithm based on MDRM, enhanced by class probability and the curve direction in the space of covariance EEG signals, was introduced and applied on an SSVEP classification task for a 4-class brain computer interface. Existing covariance matrix estimators were investigated and their robustness was assessed on multichannel SSVEP signals to ensure that the obtained matrices are accurate estimates of data covariance, are well conditioned, and verify the positive-definiteness property. The Schäfer shrinkage estimator was found to be the best as it yielded the highest classification accuracy with the MDRM algorithm.

The MDRM approach is first analyzed in a offline classification setup. To prevent the effect of noisy signals on the MDRM approach, outliers in the training set of are removed using a modified version of the *Riemannian potato*. This approach is compared to two CCA-based state-of-the-art methods. The results show that offline MDRM achieves better classification performances than any of the CCA-based methods.

In the online setup, the proposed online algorithm enhances the stability of the BCI system, balancing between classification speed and prediction accuracy. The evaluation of the classification confidence over several epochs mitigates the short term perturbations in the experimental conditions and the attentional variations of the subject. The curve direction overcomes the misclassification of EEG trials that are still synchronized with past stimuli frequencies at classification time.

Unlike the CCA-based state-of-the-art methods considered in this work, the proposed online algorithm is capable of identifying the resting periods during an online EEG recording. These resting periods are considered as an additional class in the classification task.

All these contributions help to pave the way towards BCI used in non-controlled, assistive environment.

Acknowledgement

The authors would like to thank Louis Mayaud from Mensia Technologies for his contribution in discussions that led to the completion of this work.

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (6) (2002) 767–791.
[2] J.J. Vidal, Toward direct brain–computer communication, Annu. Rev. Biophys. Bioeng. 2 (1) (1973) 157–180.
[3] J.D. Bayliss, D.H. Ballard, Single trial P3 epoch recognition in a virtual environment, Neurocomputing 32–33 (2000) 637–642.
[4] W. Tu, S. Sun, A subject transfer framework for EEG classification, Neurocomputing 82 (2012) 109–116.
[5] E. Niedermeyer, F. Lopes da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 5th Ed., Lippincott Williams & Wilkins, 2004.
[6] B. Blankertz, K.-R.R. Müller, G. Curio, T.M. Vaughan, G. Schalk, J.R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, N. Birbaumer, The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1044–1051.
[7] B. Blankertz, K.R. Muller, D.J. Krusienski, G. Schalk, J.R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. Millan, M. Schroder, N. Birbaumer, The BCI competition III: validating alternative approaches to actual BCI problems, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2) (2006) 153–159.
[8] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K.J. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, B. Blankertz, Review of the BCI Competition IV, Front. Neurosci. 6 (55). http://dx.doi.org/10.3389/fnins.2012.00055.
[9] T. Dickhaus, C. Sannelli, K.-R. Müller, G. Curio, B. Blankertz, Predicting BCI performance to study BCI illiteracy, BMC Neurosci. 10 (Suppl 1) (2009) 1–2.
[10] B.Z. Allison, C. Neuper, Could anyone use a BCI? in: D.S. Nijholt (Eds.), Brain–Computer Interfaces, Human–Computer Interaction Series, Springer, London, 2010, pp. 35–54 (Chapter 3).
[11] C. Vidaurre, B. Blankertz, Towards a cure for BCI illiteracy, Brain Topogr. 23 (2) (2010) 194–198.
[12] B. Obermaier, C. Guger, C. Neuper, G. Pfurtscheller, Hidden Markov models for online classification of single trial EEG data, Pattern Recognit. Lett. 22 (12) (2001) 1299–1309.
[13] A. Lenhardt, M. Kaper, H. Ritter, An adaptive P300-based online brain computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (2) (2008) 121–130.
[14] L.F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, R. Hornero, Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain–computer interfaces, Neurocomputing 2016. http://dx.doi.org/10.1016/j.neucom.2015.02.005.
[15] E. Kalunga, K. Djouani, Y. Hamam, S. Chevallier, E. Monacelli, SSVEP enhancement based on Canonical Correlation Analysis to improve bci performances, in: IEEE, Africon 2013, pp. 1–5.
[16] H. Lu, H.-L. Eng, C. Guan, K. Plataniotis, A. Venetsanopoulos, Regularized common spatial pattern with aggregation for EEG classification in small-sample setting, IEEE Trans. Biomed. Eng. 57 (12) (2010) 2936–2946.
[17] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.R. Muller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Process. Mag. 25 (1) (2008) 41–56.
[18] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, IEEE Trans. Biomed. Eng. 58 (2) (2011) 355–362.
[19] Y. Yang, S. Chevallier, J. Wiart, I. Bloch, Automatic selection of the number of spatial filters for motor-imagery BCI, in: European Symposium on Artificial Neural Networks (ESANN), 2012, pp. 109–114.
[20] Z. Lin, C. Zhang, W. Wu, X. Gao, Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2610–2614.
[21] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, 2009.
[22] R. Bhatia, Positive Definite Matrices, Princeton University Press, Princeton, 2009.
[23] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain–computer interface classification by Riemannian geometry, IEEE Trans. Biomed. Eng. 59 (4) (2012) 920–928.
[24] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Classification of covariance matrices using a Riemannian-based kernel for BCI applications, Neurocomputing 112 (2013) 172–178.
[25] M. Congedo, A. Barachant, A. Andreev, A new generation of brain–computer interface based on Riemannian geometry, arXiv preprint arXiv:1310.8115.
[26] E.K. Kalunga, S. Chevallier, O. Rabreau, E. Monacelli, Hybrid interface: integrating BCI in multimodal human–machine interfaces, in: IEEE/ASME

International Conference on Advanced Intelligent Mechatronics (AIM), 2014, pp. 530–535.
[27] X. Gao, D. Xu, M. Cheng, S. Gao, A BCI-based environmental controller for the motion-disabled, IEEE Trans. Neural Syst. Rehabil. Eng. 11 (2) (2003) 137–140.
[28] G. Edlinger, C. Holzner, C. Guger, A hybrid Brain–Computer interface for smart home control, in: J.A. Jacko (Ed.), Human–Computer Interaction. Interaction Techniques and Environments, Lecture Notes in Computer Science, vol. 6762, Springer, Berlin, Heidelberg, 2011, pp. 417–426.
[29] R.C. Panicker, S. Puthusserypady, Y. Sun, Adaptation in P300 brain–computer interfaces: a two-classifier cotraining approach, IEEE Trans. Biomed. Eng. 57 (12) (2010) 2927–2935.
[30] F. Schettini, F. Aloise, P. Aricò, S. Salinari, D. Mattia, F. Cincotti, Self-calibration algorithm in an asynchronous P300-based brain–computer interface, J. Neural Eng. 11 (3) (2014) 035004.
[31] H. Verschore, P.-J. Kindermans, D. Verstraeten, B. Schrauwen, Dynamic stopping improves the speed and accuracy of a P300 speller, in: Artificial Neural Networks and Machine Learning–ICANN 2012, Toronto, Canada, Springer, 2012, pp. 661–668.
[32] D. Regan, Comparison of transient and steady-state methods, Ann. N.Y. Acad. Sci. 388 (1) (1982) 45–71.
[33] T. Takahashi, K.H. Chiappa, Activation methods, Electroencephalography. in: Basic Principles, Clinical Applications, and Related Fields, 5th ed., Lippincott Williams & Wilkins, 2004, pp. 241–262.
[34] E. Niedermeyer, F.L. da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 5th ed., 2004.
[35] J. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (6) (2002) 767–791.
[36] S.T. Morgan, J.C. Hansen, S.A. Hillyard, Selective attention to stimulus location modulates the steady-state visual evoked potential, Proc. Natl. Acad. Sci. USA 93 (10) (1996) 4770–4774.
[37] M.M. Müller, S. Andersen, N.J. Trujillo, P. Valdés-Sosa, P. Malinowski, S.A. Hillyard, Feature-selective attention enhances color signals in early visual areas of the human brain, Proc. Natl. Acad. Sci. USA 103 (38) (2006) 14250–14254.
[38] B. Allison, T. Lüth, D. Valbuena, A. Teymourian, I. Volosyak, A. Gräser, BCI demographics: how many (and what kinds of) people can use an SSVEP BCI?, IEEE Trans. Neural Syst. Rehabil. Eng. 18 (2) (2010) 107–116.
[39] D. Zhu, J. Bieger, G.G. Molina, R.M. Aarts, A survey of stimulation methods used in SSVEP-based BCIs, Intell. Neurosci. (2010), http://dx.doi.org/10.1155/2010/702357.
[40] C.S. Herrmann, Human EEG responses to 1100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena, Exp. Brain Res. 137 (2001) 346–353.
[41] M.A. Pastor, J. Artieda, J. Arbizu, M. Valencia, J.C. Masdeu, Human cerebral activation during steady-state visual-evoked responses, J. Neurosci. 23 (37) (2003) 11621–11627.
[42] R.S. Fisher, G. Harding, G. Erba, G.L. Barkley, A. Wilkins, Photic- and pattern-induced seizures: a review for the epilepsy foundation of america working group, Epilepsia 46 (9) (2005) 1426–1441.
[43] J. Pan, X. Gao, F. Duan, Z. Yan, S. Gao, Enhancing the classification accuracy of steady-state visual evoked potential-based brain–computer interfaces using phase constrained canonical correlation analysis, J. Neural Eng. 8 (3) (2011) 036027. 〈http://iopscience.iop.org/article/10.1088/1741-2560/8/3/036027/meta〉.
[44] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, T.-P. Jung, A high-speed brain speller using steady-state visual evoked potentials, Int. J. Neural Syst. 24 (06) (2014) 1450019.
[45] M. Spüler, W. Rosenstiel, M. Bogdan, Online adaptation of a c-VEP brain–computer interface (BCI) based on error-related potentials and unsupervised learning, PLoS ONE 7 (12) (2012) e51077. 〈http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051077〉.
[46] G. Bin, X. Gao, Y. Wang, Y. Li, B. Hong, S. Gao, A high-speed BCI based on code modulation VEP, J. Neural Eng. 8 (2) (2011) 025015. 〈http://iopscience.iop.org/article/10.1088/1741-2560/8/2/025015/meta〉.
[47] H. Cecotti, A self-paced and calibration-less SSVEP-based brain–computer interface speller, IEEE Trans. Neural Syst. Rehabil. Eng. 18 (2) (2010) 127–133.
[48] S. Parini, L. Maggi, A.C. Turconi, G. Andreoni, A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication, Intell. Neurosci. (2009), http://dx.doi.org/10.1155/2009/864564.
[49] F. Yger, A review of kernels on covariance matrices for BCI applications, in: IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2013, pp. 1–6.
[50] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, M. Harandi, Kernel methods on the Riemannian manifold of symmetric positive definite matrices, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 73–80.
[51] Y. Xie, J. Ho, B. Vemuri, On a nonlinear generalization of sparse coding and dictionary learning, in: International Conference on Machine Learning (ICML), 2013, p. 1480.
[52] A. Goh, R. Vidal, Unsupervised Riemannian clustering of probability density functions, in: Machine Learning and Knowledge Discovery in Databases, Springer, 2008, pp. 377–392.
[53] A. Goh, R. Vidal, Clustering and dimensionality reduction on Riemannian manifolds, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–7.
[54] X. Pennec, P. Fillard, N. Ayache, A Riemannian framework for tensor computing, Int. J. Comput. Vis. 66 (1) (2006) 41–66.
[55] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Riemannian geometry applied to BCI classification, in: Latent Variable Analysis and Signal Separation, Springer, 2010, pp. 629–636.
[56] A. Barachant, A. Andreev, M. Congedo, et al., The Riemannian potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry, in: Proceedings of TOBI Workshop IV, 2013, pp. 19–20.
[57] S.-I. Amari, α-divergence is unique, belonging to both f-divergence and Bregman divergence classes, IEEE Trans. Inf. Theory 55 (11) (2009) 4925–4931.
[58] W. Samek, D. Blythe, K.-R. Müller, M. Kawanabe, Robust spatial filtering with beta divergence, in: Advances in NeuralInformation Processing Systems (NIPS), 2013, pp. 1007–1015.
[59] W. Samek, K.-R. Muller, Information geometry meets BCI spatial filtering using divergences, in: International Winter Workshop on Brain–Computer Interface, Seoul, Korea, IEEE, 2014, pp. 1–4.
[60] A. Barachant, S. Bonnet, Channel selection procedure using Riemannian distance for BCI applications, in: International IEEE/EMBS Conference on Neural Engineering (NER), 2011, pp. 348–351.
[61] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Common spatial pattern revisited by Riemannian geometry, in: IEEE International Workshop on Multimedia Signal Processing (MMSP), 2010, Saint Malo, France, pp. 472–476.
[62] Y. Li, K.M. Wong, H. De Bruin, EEG signal classification based on a Riemannian distance measure, in: Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference, IEEE, 2009, Toronto, Canada pp. 268–273.
[63] Y. Li, K. Wong, H. De Bruin, Electroencephalogram signals classification for sleepstate decision: a Riemannian geometry approach, IET Signal Process. 6 (4) (2012) 288–299.
[64] J. Jost, Riemannian Geometry and Geometric Analysis, vol. 62011, Springer, New-York, NY, USA, 2011.
[65] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 26 (3) (2005) 735–747.
[66] P.T. Fletcher, C. Lu, S.M. Pizer, S. Joshi, Principal geodesic analysis for the study of nonlinear statistics of shape, IEEE Trans. Med. Imag. 23 (8) (2004) 995–1005.
[67] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, San Diego, CA, USA, 1990.
[68] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivar. Anal. 88 (2) (2004) 365–411.
[69] B. Blankertz, S. Lemm, M. Treder, S. Haufe, K.-R. Müller, Single-trial analysis and classification of ERP components: a tutorial, NeuroImage 56 (2) (2011) 814–825.
[70] J. Schäfer, K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, Stat. Appl. Genet. Mol. Biol. 4 (1). http://dx.doi.org/10.2202/1544-6115.1175.
[71] F. Pascal, P. Forster, J.P. Ovarlez, P. Arzabal, Theoretical analysis of an improved covariance matrix estimator in non-gaussian noise, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, 2005.
[72] M. Congedo, EEG source analysis, Habilitation á diriger des recherches, Université de Grenoble, October 2013.
[73] M.J. Pencina, R.B. D'Agostino, R.S. Vasan, Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond, Stat. Med. 27 (2) (2008) 157–172.
[74] Y. Kimura, T. Tanaka, H. Higashi, N. Morikawa, SSVEP-based brain–computer interfaces using FSK-modulated visual stimuli, IEEE Trans. Biomed. Eng. 60 (10) (2013) 2831–2838.
[75] F.-B. Vialatte, M. Maurice, J. Dauwels, A. Cichocki, Steady-state visually evoked potentials: focus on essential paradigms and future perspectives, Prog. Neurobiol. 90 (4) (2010) 418–438.
[76] H. Bakardjian, T. Tanaka, A. Cichocki, Optimization of SSVEP brain responses with application to eight-command brain–computer interface, Neurosci. Lett. 469 (1) (2010) 34–38.



Emmanuel K. Kalunga is a Ph.D. candidate, in his second year, in the joint doctorate program between the Tshwane University of Technology (South Africa) and Université de Versailles Saint-Quentin, France. On completion of his Baccalaureus Technologiae (Cum Laude) in 2010 at the Tshwane University, Emmanuel was awarded the Mandela Rhodes Scholarship and pursued a double Masters degree (Mtech/Msc) in Control, Image and Signal processing at the same university on the topic "Development of brain computer interface (BCI) based intention detection approach for persons with limited neuro-muscular control". He graduated Cum Laude in September 2013 after spending a year of internship at the Laboratoire d'ingénierie des systémes de Versailles (France). Since January 2014, Emmanuel is enrolled for his PhD on the topic of "Pattern recognition techniques for implicit brain computer interfaces", under the supervision of Karim Djouani, Eric Monacelli, Sylvain Chevallier, and Yskandar Hamam.



Sylvain Chevallier obtained his Ph.D. in computer science at the LIMSICNRS, with a speciality in cognitive science as he was interested in visual attention process for robotics. He spent 2 years in the neurocybernetic team of ETIS-CNRS where he worked on biologically inspired decision processes and visual perception. He joined afterwards INRIA Saclay to study deep neural network and brain computer interface. He pursued this post-doctoral research theme within Telecom ParisTech laboratory. In 2011, he was hired as an assistant professor in Université de Versailles to work on rehabilitation robotics. He is interested in neural communication and brain-like bioinspired methods for assistive technologies. He is focusing on brain–computer interfaces using shared control and passive approaches.



Quentin Barthélemy received the Engineering degree and the M.Res. in signal and images analysis and processing (with distinction) from Grenoble Institut National Polytechnique (Grenoble INP), France, both in 2009; and the Ph.D. degree in signal processing from Grenoble University and CEA-LIST (Alternative Energies and Atomic Energy Commission), France, in 2013. His Ph.D. dissertation deals with sparse representations for multivariate signals, including invariances as shift, rotation and affine transformation. He joined Mensia Technologies at the Institut du Cerveau et de la Moelle Epiniére, Paris, France, in 2013, to develop signal processing and machine learning methods for real-time EEG analysis. His research interests are sparse representation, time-frequency analysis, source separation and Riemannian geometry.



Karim Djouani is a professor, scientist and technical group supervisor of soft computing, telecommunication, networking systems and Robotics. Since January 2011 he is Full professor at University Paris Est-Creteil (UPEC), France and Tshwane University of Technology, Pretoria, South Africa. From July 2008 to December 2010, he was seconded by the French Ministry of Higher Education to the French South African Institute of Technology (F'SATI) at Tshwane University of Technology (TUT), Pretoria, South Africa. He was also national and European projects manager at the LISSI Lab.His current works focus on the development of novel and highly efficient algorithms for reasoning systems with uncertainty as well as optimization, for distributed systems, networked control systems, wireless ad-hoc network, wireless and mobile communication, and wireless sensors networks as well as Robotics. He has authored/co-authored over 150 articles in archival journals and conference proceedings as well as five chapters in edited books.



Eric Monacelli is an HDR Associate Professor at Versailles University (UVSQ), working in the LISV laboratory. He is working on the development of analysis methods and experimental devices adapted to the assistance of specific end users. His research projects incorporate issues of man–machine interface, robotics, assistive technologies and ambient intelligence systems.



Yskandar Hamam graduated as a Bachelor of the American University of Beirut (AUB) in 1966. He obtained his M.Sc. in 1970 and Ph.D. in 1972 from the University of Manchester Institute of Science and Technology. He also obtained his “Diplôme d'Habilitation á Diriger des Recherches” (equivalent to D.Sc.) from the “Université des Sciences et Technologies de Lille” in 1998. He conducted research activities and lectured in England, Brazil, Lebanon, Belgium and France. He was the head of the Control department and dean of faculty at ESIEE, France. He was an active member in modelling and simulation societies and was the president of EUROSIM. He was the Scientific Director of the French South African Institute of Technology (F'SATI) at TUT in South Africa from 2007 to 2012. He is currently professor at the Department of Electrical Engineering of TUT. He has authored/co-authored about 300 papers in archival journals and conference proceedings as well as book contributions. He is a senior member of the IEEE.