Transfer Learning for SSVEP-based BCI using Riemannian similarities between users

Emmanuel K. Kalunga
VASTech
Stellenbosch, South Africa
emmanuelkalunga.k@gmail.com

Sylvain Chevallier
LISV
Université de Versailles St Quentin
Velizy, France
sylvain.chevallier@uvsq.fr

Quentin Barthélemy
Mensia Technologies SA
Paris, France
qb@mensiatech.com

***Abstract*—Brain-Computer Interfaces (BCI) face a great challenge: how to harness the wide variability of brain signals from a user to another. The most visible problem is the lack of a sound framework to capture the specificity of a user brain waves. A first attempt to leverage this issue is to design user-specific spatial filters, carefully adjusted with a lengthy calibration phase. A second, more recent, opening is the systematic study of brain signals through their covariance, in an appropriate space from a geometric point of view. Riemannian geometry allows to efficiently characterize the variability of inter-subject EEG, even with noisy or scarce data. This contribution is the first attempt for SSVEP-based BCI to make the most of the available data from a user, relying on Riemannian geometry to estimate the similarity with a multi-user dataset. The proposed method is built in the framework of transfer learning and borrows the notion of composite mean to partition the space. This method is evaluated on 12 subjects performing an SSVEP task for the control of an exoskeleton arm and the results show the contribution of Riemannian geometry and of the user-specific composite mean, whereas there is only a few data available for a subject.**

***Index Terms*—Brain-computer interface, transfer learning, Riemannian geometry, SSVEP.**

I. Introduction

Brain-computer interfaces (BCI) endow a user with the ability to interact with a system, such as a physical interface or an application, based on the brain activity [1]. These BCI are of prime interests for users with physical disabilities or with difficulties to interact physically with a system. Brain activity is decoded in real-time to control or to provide insight on the user intentions and decisions. To ensure the portability and a reduced cost for BCI system, a common choice is to rely on electroencephalography (EEG) for recording brain activity [2], or in a lesser extend functional near-infrared spectra [1]. The EEG is most appropriate to detect fast temporal variations or transient events, but is subject to the volume conduction effect [3], mixing cerebral sources and smoothing spatial information.

The strongest limitation for a wide adoption of BCI, and the emergence of out-of-the-lab applications, is known as the BCI deficiency problem [4]. This effect is visible for circa 15 to 30% of the BCI users, who achieve a deceptive near-chance performance, and is not yet correlated with a specific psychological or neurological traits. Some tentative solutions may lie in a better protocol design, which could benefit from advances in human-machine interfaces [5]. Still, one of the main sources of this problem is the important variations between users, and for a given user, the variation from day to day, and in some case from hour to hour. To mitigate this issue, most of the existing approaches rely on a calibration phase. During this calibration, the various preprocessing and filtering steps could be tailored to the specific brain waves of a user. Nonetheless, this calibration is a source of fatigue, frustration and could induce loss of performance for the real task to come [6], [7].

The calibration phase should be performed before each session and is common to all neurobiological signals used for BCI. Existing BCI protocols rely mostly on Event-Related Potentials (ERP), Event-Related Desynchronization and Synchronization (ERD/S) and Steady-State Visually Evoked Potentials (SSVEP) [1]. The ERP is a transient activation induced by an exogenous event, resulting often from the combination of several cognitive components being activated concurrently. Using the oddball paradigm, it is possible to induce a visible ERP component 300 ms after the apparition of an awaited stimulus, called P300 and notoriously employed for BCI spelling tasks. Following another paradigm, the ERD/S has been mostly employed in motor imagery tasks for BCI, where the (de)synchronization results in a change of amplitude in a given frequency band and is related with the movement preparation of a body part. The SSVEP is a brain wave exhibited as a response from a repetitive stimulation with a fixed frequency. The most common practice being to generate visual patterns that are "blinking", which induce a cortical activity in visual areas synchronized with the stimulation. This paper focus on the visual SSVEP for the experimental part, even if the general approach could be adapted to any type of BCI.

The most common approach for all the BCI paradigms is to rely on spatial filters to enhance the signal of interest and to remove bad and noisy electrodes. These spatial filters allow to project the sensor data in a surrogate sensor space, where a proper combination of electrodes augment the signal to noise ratio. The XDAWN spatial filter [8] focuses on maximizing the signal-to-signal plus noise ratio by finding components best correlated with the ERP timings. With ERD, the common spatial patterns [9] is a popular and efficient approach to find

ISBN 978-90-827970-1-5 © EURASIP 2018

spatial filters that maximize the variance change during the ERD. For SSVEP, the canonical correlation analysis [10], [11] has proven to be a good candidate to compute spatial filters that maximize the canonical correlation between EEG data and reference signals of known frequencies. In all those cases, the spatial filters are also exploited as a dimensionality reduction technique, by selecting only an appropriate subset of filters. The correct estimation of those filters requires nonetheless a large amount of data, both in terms of number of electrodes and of length of calibration period. The estimation of spatial filters is especially sensitive to noise and label error, and a significant part of the literature is focusing on the possibilities to reduce the adverse effects of noise.

An orthogonal approach of the signal processing behind BCI is to consider subspaces that include all possible surrogate sensor space, that is linear transformation of the input signal. In that case, there is no need to estimate session- and user-specific spatial filters as they are encompassed in the geometry of the considered space. One such approach is known as the Riemannian BCI [12], [13], where spatial covariance matrices are estimated from the EEG signal and classified in the space of symmetric and positive-definite matrices. This Riemannian point of view of the BCI has demonstrated its interests in several occasions, and is now systematically found in the top tier submissions in BCI and EEG competitions. The main challenge is to reformulate the classification problems in the adequate geometrical space [14], [15]. Even without the need of estimating spatial filters, a calibration phase is still required to parametrize the classifier. One possibility to reduce the calibration phase is to rely on a smart initialization, that is using previously acquired data to produce a fast and precise parameterization of the classifier.

The contribution of this paper is to propose such smart initialization, using Riemannian geometry tools in a transfer learning framework. This is the first contribution attempt for SSVEP-based BCI. Section II introduces the geometric formalization of the EEG signal processing and explains how the classification is achieved. In Section III, the reduction of the calibration phase is formulated from a transfer learning point of view and novel approaches are introduced. Section IV demonstrates the interest of the proposed approaches on a real SSVEP dataset. Section V concludes this paper.

II. Riemannian BCI

In the following, we will consider a differentiable manifold $\mathcal{M}$ characterized by a Riemannian metric, that is a collection of inner products on the tangent space $T_\Sigma\mathcal{M}$ varying smoothly at each point Σ of the manifold. Endowed with this inner product, it is possible to compute the length of any curve. The shortest curve between any two points of the manifold is called a geodesic $\gamma(t)$. The length of the geodesic curve between Σ_1 and Σ_2 is the Riemannian distance δ:

$$\delta(\Sigma_1,\Sigma_2) = \left\| \log(\Sigma_1^{-\frac{1}{2}}\Sigma_2\Sigma_1^{-\frac{1}{2}}) \right\|_F \ . \tag{1}$$

It is known as the affine-invariant Riemannian (AIR) distance [16].

The notion of mean (or center of mass) of a set of points Σ_i can be extended in the context of Riemannian manifold, called Karcher (or Fréchet) mean. In that case, the mean $\bar{\Sigma}$ is the point minimizing the dispersion on the manifold, captured by the square of the distances between $\bar{\Sigma}$ and Σ_i. This optimization problem does not have a closed form and is written as:

$$\bar{\Sigma} = \mu(\{\Sigma_i\}) = \arg\min_\Sigma \sum_{i=1}^{N} \delta^2(\Sigma_i,\Sigma) \ . \tag{2}$$

There is an exact solution for $N = 2$, but for $N > 2$ it should be estimated iteratively [17].

In this paper, we will consider the weighted mean μ_w, where a coefficient w_i is associated with each point Σ_i of the considered set, with $\sum_i w_i = 1$ [17]:

$$\bar{\Sigma} = \mu_w(\{w_i\};\{\Sigma_i\}) = \arg\min_\Sigma \sum_{i=1}^{N} w_i \ \delta^2(\Sigma_i,\Sigma) \ . \tag{3}$$

The classifier *Minimum Distance to Mean* (MDM), introduced in [14], is presented for multi-class classification in the manifold. The covariance matrices of EEG trials are classified based on their distance to the $k = 1 \dots K$ centers of the classes $\bar{\Sigma}^{(k)}$. The predicted class $\hat{k}$ of the current matrix Σ is defined as:

$$\hat{k} = \arg\min_k \delta(\Sigma,\bar{\Sigma}^{(k)}) \ . \tag{4}$$

It is a simple Bayesian classifier, under the hypotheses that classes have identical dispersion and that it is operating on a manageable space.

III. Proposed approach for transfer learning

A. Related works

In BCI the need to transfer learning is important due to *inter-subject* variability and *inter-session* variability. Inter-subject variability is expressed by the difference of brain signals recorded from different subjects despite them being involved in the similar mental activities. This difference is mostly attributed to anatomical differences among users. Inter-session variability is visible between distinct recording sessions of a unique subject. This variability is attributed to changes in the mental states of the user, such as fatigue, and changes in experimental settings, e.g. electrodes placement, environment, stimulation.

Exposed to the same stimuli, BCI users do not produce similar EEG response. On top of user-specific brain waves, changes induced by different environmental conditions should also be taken into account. These cross-session changes have a lesser impact in Riemannian framework, as congruence invariance allows to reduce the sensitivity to spatial filtering of EEG [13]. In this work, we focus on cross-subject transfer to mitigate inter-subject variability [18], [19]. These efforts are meant to shorten the calibration phase, thus reducing the user fatigue and discomfort. In a cross-subject transfer perspective, the source domain is the dataset of all subjects previously

ISBN 978-90-827970-1-5 © EURASIP 2018

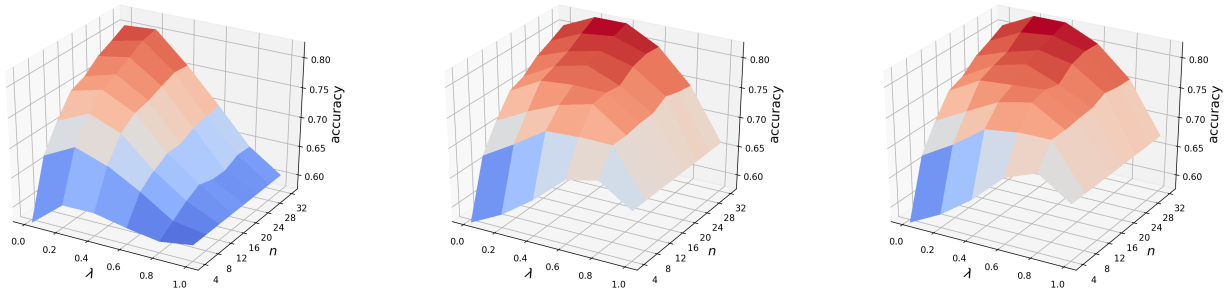

Fig. 1. Average accuracy value evaluated through grid search for λ and n, for minimum distance to E-MDPM 1(a), R-MDPM 1(b) and R-MDWM 1(c).

recorded, that is all the labeled covariance matrices for different users. The target domain is the covariance matrices that are acquired for a target subject. In the scope of this paper, we consider that only a few covariance matrices are available for the target subject.

Let $k = 1, \ldots, K$ be the number of classes and $j = 1, \ldots, J+1$ be the number of subjects. The resting-state class (no task) will be defined by $k = 0$. In order to estimate a model for a target user with few recordings, there are several possibilities at hand to make the most of source user models. In our case, the source models are described by the covariance matrices $\Sigma_j^{(k)}$ for a given user and a given class. Our first approach is to adapt composite mean [20], in the Euclidean space, to allow transfer from user-to-user. This approach could be adapted for the curved space of covariance matrices, relying on Riemannian geometry. The covariance matrices of different users could be averaged into a unique matrix, hypothesizing that each existing model contributes equally to estimate a user model. We called this approach Minimum Distance to Pooled Means (MDPM). We introduce a more advanced approach, called Minimum Distance to Weighted Means (MDWM), is to adjust the influence each model depending on the similarity between the EEG database and the newly observed EEG.

B. Composite Euclidean Mean

Composite Euclidean Mean is an *instance transfer* technique, *i.e.* re-weighting existing labeled data to be applied on newly recorded data, inspired from the composite common spatial patterns method proposed by [20]. It could be seen as a regularization of the newly observed data from user i, where the observed covariance matrices are modified to include information from existing models of users j:

$$\bar{\Sigma}_i^{(k)} = (1-\lambda)\Sigma_i^{(k)} + \lambda \sum_{j \neq i} \frac{1}{J} \Sigma_j^{(k)} \ . \tag{5}$$

The obtained $\bar{\Sigma}_i^{(k)}$ are then diagonalized in order to obtain transferred CSP filters [20], [21]. The hyperparameter $\lambda \in [0, 1]$ allows to choose if the centers of class rely more on the existing data, that is the source domain, or more on newly acquired data, that is the target domain. With $\lambda = 0$, there is no transfer, only the data acquired from a subject are used. With $\lambda = 1$, this is a calibration-free BCI system as no data are required from the subject.

C. Minimum Distance to Pooled Means (MDPM)

The former formulation does not take into account the specific geometry of covariance matrices. The rightmost term in Eq. (5) could be replaced with a Riemannian mean. The whole right-hand side could also be written as Riemannian mean. This geometric formulation could be written as:

$$\bar{\Sigma}_i^{(k)} = \mu_w \left(\{1-\lambda, \lambda\} ; \left\{ \Sigma_i^{(k)}, \bar{\Sigma}_{w,j}^{(k)} \right\} \right) \ , \tag{6}$$

with $\bar{\Sigma}_{w,j}^{(k)} = \mu \left(\left\{ \Sigma_j^{(k)} \right\}_{j \neq i} \right)$.

Even if Eq. (5) has never been used in a MDM based BCI, a MDM applied after it will be called E-MDPM (Euclidean-MDPM), and will be considered as the state-of-the-art. MDM applied with mean from Eq. (6) will be referred to as R-MDPM (Riemannnian-MDPM)

D. Minimum Distance to Weighted Means (MDWM)

The similarity between two subjects i and j is defined as the inverse of the AIR distance δ between their covariance matrices of the resting-state class $k = 0$:

$$s_{i,j} = \frac{1}{\bar{s}_i} \times \frac{1}{\delta(\Sigma_i^{(0)}, \Sigma_j^{(0)})} \ , \tag{7}$$

where $\bar{s}_i$ is a normalization factor integrating all distances to the subjects of the database:

$$\bar{s}_i = \sum_{j \neq i} \frac{1}{\delta(\Sigma_i^{(0)}, \Sigma_j^{(0)})} \ . \tag{8}$$

These weights are obtained in an unsupervised way; no labels are required to estimate this similarity, only a clean recording of 2min of resting-state.

Taking into account these similarities, it is possible to rewrite Eq. (6) as a weighted mean:

$$\bar{\Sigma}_i^{(k)} = \mu_w \left(\{(1-\lambda), \lambda\} ; \left\{ \Sigma_i^{(k)}, \bar{\Sigma}_{w,j}^{(k)} \right\} \right) \ , \tag{9}$$

with $\bar{\Sigma}_{w,j}^{(k)} = \mu_w \left(\left\{ s_{i,j}^{(k)} \right\}_{j \neq i} ; \left\{ \Sigma_j^{(k)} \right\}_{j \neq i} \right)$.

ISBN 978-90-827970-1-5 © EURASIP 2018

Eq. (5) can also be rewritten taking with similarity measures as:

$$\bar{\Sigma}_i^{(k)} = (1-\lambda)\Sigma_i^{(k)} + \lambda \sum_{j \neq i} s_{i,j}^{(k)} \Sigma_j^{(k)} \ . \tag{10}$$

Thus, an MDWM applied with means from Eq. (9) will be called R-MDWM (Riemannian-MDWM), and E-MDWM (Euclidean-MDWM) when applied with Eq. (10).

IV. Results

A. SSVEP experimental setup and data

The experimental study is conducted on multichannel EEG signals recorded during a SSVEP-based BCI experiment. There are 3 groups of 4 LEDs blinking at different frequencies: $F = 3$ visual target stimuli blinking respectively at 13, 21 and 17 Hz. A sequence of trials is proposed to the user. When he do not intend to activate any SSVEP command, this constitutes the reject class, *i.e.* reference state or no-SSVEP state. This, plus the 3 groups of LEDs make a 4-class BCI ($K = 4$). The EEG was recorded at a sampling rate of 256 Hz with $C = 8$ electrodes/channels (PO7, PO3, POz, PO4, PO8, O1, Oz, and O2).

In a session, 32 trials were recorded: 8 for each visual stimulus and 8 for the resting-state class. A trial is 4 seconds long. There were 12 subjects and the number of sessions recorded per subject varied from 2 to 5. The full description of the experiment and dataset can be found in [11].

The covariance matrices are estimated from a modified version of the input signal $X \in \mathbb{R}^{C \times T}$:

$$X \in \mathbb{R}^{C \times T} \rightarrow \begin{bmatrix} X_{\text{freq}_1} \\ \vdots \\ X_{\text{freq}_F} \end{bmatrix} \in \mathbb{R}^{FC \times T} \ , \tag{11}$$

where X_{freq_f} is the input signal X band pass filtered around frequency freq_f, $f = 1, \ldots, F$. The sample covariance matrix estimator $\tilde{\Sigma}_{\text{scm}} = \frac{1}{C} X X^\intercal$ is a possible choice, but a shrinkage estimator [22] could produced a better conditioned estimator $\tilde{\Sigma}_{\text{skrinkage}} = \kappa \text{tr}(\hat{\Sigma}_{\text{scm}}) \mathbf{I}_C + (1-\kappa)\hat{\Sigma}_{\text{scm}}$. In the rest of the document, we will write Σ to denote matrices estimated with the shrinkage estimator.

This dataset is accessible at https://github.com/sylvchev/dataset-ssvep-exoskeleton.

B. Experimental comparison

Four transfer learning approaches are compared: E-MDPM (Eq. 5) considered as the state-of-the-art, R-MDPM (Eq. 6), E-MDWM (Eq. 10), and R-MDWM (Eq. 9).

The four approaches are evaluated at different values of $\lambda \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The objective of our transfer learning approaches being to eliminate or reduce the number of training samples required for the target domain subjects by using data from source domain subjects, the approaches are also evaluated with various numbers of available training samples from the target subject $n \in \{4, 8, 12, 16, 20, 24, 28, 32\}$.

The target domain is made of a test subject, while the remaining data from other subjects are used as the source domain. The number of labeled samples in the target domain corresponds to n. The classifier is evaluated on unlabeled samples of the target domain (varying from 32 to 128). For statistical significance, bootstrapping is used. The classifier is implemented on 10 bootstraps of the target domain training data by re-sampling with replacement.

C. Results and discussion

1) E-MDPM *and* R-MDPM*:* The results obtained with the classifier trained as described in Eq. (5) and Eq. (6) over various λ and n are shown in Fig. 1. It can be seen that the accuracy is always better when the test subject or target domain has more label samples (*i.e.* higher values of n). Taking class means from the source domain ($\Sigma_j^{(k)}$) improves the results as λ grows from zero. Although this trend is observed on all values of n, it is significantly bigger for small values of n, indicating that, when a subject has very few training samples available, the composite Riemannian mean transfer learning approach is a good BCI initialization.

Comparing Fig. 1(a) to Fig. 1(b), it is visible that *R-MDPM*, which is consistent with the Riemannian approach, outperforms *E-MDPM*. A one way ANOVA analysis shows significance in this improvement with f-value of over 40 and corresponding p-value in the order of 10^{-10}. It is therefore important to consider the fully Riemannian approach of the composite Riemannian mean introduced in Eq. (6) rather than the one of Eq. (5).

2) R-MDPM *versus* R-MDWM*:* The composite Riemannian mean of Eq. (9) introduces a weight based of a similarity measure between subjects.

Table I compares the results of *R-MDPM* and *R-MDWM*. For illustration purposes, the value of n is fixed to 12 and on the value of λ that yields the highest classification accuracy is considered. To test the statistical significance of the results, a paired Student t-test is run on the results (across all λ and all n) of *R-MDPM* on one side, and *R-MDWM* on the other side. The p-values show that these two methods yield significantly different results, and *R-MDWM* improves the overall results.

3) Euclidean *versus* Riemannian*, and* pooled *versus* weighted*:* A two-way ANOVA is performed on the four methods, and the interaction between Euclidean/Riemannian mean and pooled/weighted mean is shown in Fig. 2. It shows that using remannian mean in transfer learning of class mean (*i.e* Eq. (6) and Eq. (9)) significantly improves the classification performance, with F-value of 103 and equivalent p-value in the orders of 10^{-23}. On a smaller scale, weighting the data in the source domain based on their similarity to the target domain shows a trend in the improvement the performance (F-value close to 1, and p-value in the order of 0.1).

V. Conclusion

This paper proposes a Riemannian transfer learning approach for SSVEP-based BCI, inspired from composite mean for instance transfer where a user's model is combined with with those of similar subjects. The Riemannian geometry offers a very robust framework for inter-subject transfer learning

ISBN 978-90-827970-1-5 © EURASIP 2018

	Sub. 1	Sub. 2	Sub. 3	Sub. 4	Sub. 5	Sub. 6	Sub. 7	Sub. 8	Sub. 9	Sub. 10	Sub. 11	Sub. 12
	Accuracy											
E-MDPM	0.619	0.875	0.925	0.756	0.756	0.844	0.808	0.869	**0.938**	0.641	0.619	**0.845**
R-MDPM	**0.625**	0.856	0.944	0.862	**0.794**	0.844	0.821	**0.875**	0.812	0.656	**0.788**	0.842
E-MDWM	0.619	**0.881**	0.925	0.756	0.762	0.838	0.808	0.869	0.875	0.647	0.625	**0.845**
R-MDWM	**0.625**	0.85	**0.950**	**0.881**	**0.794**	**0.856**	**0.829**	**0.875**	0.812	**0.659**	**0.788**	0.842
	Best λ											
E-MDPM	0.0	0.4	0.0	0.0	0.2	0.4	0.0	0.0	1.0	0.2	0.2	0.4
R-MDPM	0.2	0.2	0.8	0.8	0.6	0.6	0.2	0.4	0.6	0.4	0.8	0.2
E-MDWM	0.0	0.4	0.0	0.0	0.2	0.4	0.0	0.0	1.0	0.4	0.2	0.4
R-MDWM	0.2	0.2	0.6	0.6	0.6	0.4	0.2	0.4	1.0	0.4	0.8	0.2

TABLE I
SNAPSHOT OF *E-MDPM*, *R-MDPM*, *E-MDWM*, *R-MDWM* PERFORMANCES AT $n = 12$.

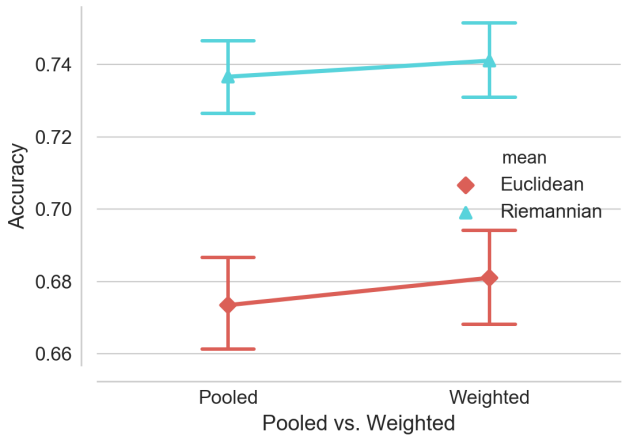

Fig. 2. Two-way ANOVA: interactions between Euclidean/Riemannian mean and pooled/weighted mean.

and we have shown that it outperforms the Euclidean formulation. We introduce two Riemannian approaches, the minimum distance to pooled means (R-MDPM) and the minimum distance to weighted means (R-MDWM). The results demonstrate that taking into account the similarities between subjects yields significantly better results for almost all considered subjects. This transfer learning has been applied to SSVEP, but could be easily used for MI or P300 paradigms.

REFERENCES

[1] C.S. Nam, A. Nijholt, and F. Lotte, *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*, CRC Press, 2018.

[2] E. Niedermeyer and F. H. Lopes da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Lippincott Williams & Wilkins, 2005.

[3] P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch, "EEG coherency: I: Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales," *Electroencephalogr Clin Neurophysiol*, vol. 103, no. 5, pp. 499–515, 1997.

[4] C. Vidaurre and B. Blankertz, "Towards a cure for BCI illiteracy.," *Brain Topography*, vol. 23, no. 2, pp. 194–198, 2010.

[5] C Jeunet, E Jahanpour, and F Lotte, "Why standard brain-computer interface (BCI) trainingprotocols should be changed: An experimental study," *J Neural Eng*, vol. 13, 2016.

[6] J. Faller, C. Vidaurre, T. Solis-Escalante, C. Neuper, and R. Scherer, "Autocalibration and recurrent adaptation: Towards a plug and play online ERD-BCI," *IEEE Trans Neural Syst Rehabil Eng*, vol. 20, no. 3, pp. 313–319, 2012.

[7] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proc IEEE*, vol. 103, no. 6, pp. 871–890, 2015.

[8] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface," *IEEE Trans Biomed Eng*, vol. 56, no. 8, pp. 2035–2043, 2009.

[9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing Spatial filters for Robust EEG Single-Trial Analysis," *IEEE Signal Process Mag*, vol. 25, no. 1, pp. 41–56, 2008.

[10] G. Bin, X. Gao, Z. Yan, B. Hong, and S. Gao, "An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method," *J Neural Eng*, vol. 6, no. 4, 2009.

[11] E. K. Kalunga, S. Chevallier, O. Rabreau, and E. Monacelli, "Hybrid interface: Integrating BCI in multimodal human-machine interfaces," in *Int Conf on Adv Int Mech (AIM)*, 2014, pp. 530–535.

[12] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: a review," *IEEE Trans Neural Syst Rehabil Eng*, vol. 25, no. 10, pp. 1753–1762, 2017.

[13] M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, pp. 1–20, 2017.

[14] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Trans Biomed Eng*, vol. 59, no. 4, pp. 920–928, 2012.

[15] E. K. Kalunga, S. Chevallier, Q. Barthélemy, K. Djouani, E. Monacelli, and Y. Hamam, "Online SSVEP-based BCI using Riemannian geometry," *Neurocomputing*, vol. 191, pp. 55–68, 2016.

[16] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 3, pp. 735–747, 2005.

[17] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Trans Med Imaging*, vol. 23, no. 8, pp. 995–1005, 2004.

[18] H He and D Wu, "Transfer learning enhanced common spatial pattern filtering for brain computer interfaces (BCIs): Overview and a new approach," in *NIPS*, 2017, pp. 811–821.

[19] N R Waytowich, V J Lawhern, A W Bohannon, K R Ball, and B J Lance, "Spectral transfer learning using information geometry for a user-independent brain-computer interface," *Frontiers in Neuroscience*, vol. 10, pp. 430, 2016.

[20] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process Lett*, vol. 16, no. 8, pp. 683–686, 2009.

[21] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans Biomed Eng*, vol. 58, pp. 355–362, 2011.

[22] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

ISBN 978-90-827970-1-5 © EURASIP 2018