Stability of spatial covariance in ERPs with high temporal variability

R Aydarkhanov[1], M Uscumlic[2], R Chavarriaga[3], J d R Millan[4]

[1]EPFL, Switzerland
[2]EPFL, Switzerland
[3]EPFL, Switzerland
[4]TU Austin, USA

E-mail: `ruslan.aydarkhanov@epfl.ch`

Abstract. Event Related Potentials (ERPs) are widely used in Brain Computer Interfaces (BCI) by providing a cognitive response to external stimuli. ERPs are characterized and typically decoded through a fixed set of components with particular amplitude and latency. However, the classical methods which rely on waveform features achieve a high decoding performance only with standardized and well aligned single trials. Since the amplitude and latency are sensitive to the experimental conditions, waveform features cannot be applied for challenging tasks and to generalize across various experimental protocols. Signal features based on spatial covariances can potentially overcome the latency jitter and delays since they aggregate the information across time. We compared the performance stability of waveform and covariance-based features as well as their combination in simulated scenarios such as classifier transfer and increased single-trial latency variability. Our findings suggest that covariance-based features can be used to: 1) classify more reliably ERPs with higher intrinsic variability in more challenging real-life applications and 2) generalize across related experimental protocols.

Keywords: Brain-computer interfaces, Electroencephalography, Riemannian geometry

1. Introduction

Diverse cognitive responses to external stimuli – error detection, object recognition etc. – reflected as distinguishable modulations of the Event-Related Potential (ERP) waveforms provide tremendous opportunities for brain computer interfacing (BCI). As an illustration, when an error committed by the user or machine is observed, an Error-Related Potential (ErrP) is elicited and can be used to correct it. Upon each occurrence of such an event the perception and the evaluation processes take variable period of time. We investigate the robustness of various ERP properties to these temporal variations for the purpose of improving decoding performance in BCI applications.

ERP waveforms consist of multiple components characterized by a particular latency, polarity and spatial distribution [1]. The early components are triggered by a sensory stimulation whereas the late components reflect cognitive processing. ERPs can be successfully detected in various paradigms under controlled conditions using preprocessed time signal of the EEG channel activity over the time-window of interest [2]. When facing real world conditions, our cognitive processing may be challenged by complex and diverse stimuli and dynamic environment. As a result, the later ERP components are more prone to latency variability [3] which compromises the decoding performance.

An alternative way to detect ERP is to use features based on the spatial covariance matrices across channels. Such features emphasize the pair-wise interaction between channels and the total variance of single channels. We hypothesize that these features may be a suitable choice to address the challenge of decoding ERPs with high latency variability inherent to real world BCI application. Typically, it is advisable to augment the matrix with the covariances between single trial and average ERP of one of the classes because the temporal ERP dynamics is considered to be a valuable source of discriminant information [4, 5]. Under ERP latency variability challenge, however, augmented covariance matrices may have the same issue as waveforms. The impact of latency shifts and jitter have previously been studied for augmented covariance matrices [5] and the performance dropped quickly for short delays of 60 ms. Thus, we need to consider simple channel-based covariance matrices.

Covariance matrices can be used directly [6] as features, although Euclidean geometry does not properly capture the relationship between them. It is shown that treating them under Riemannian geometry framework which will be briefly introduced in the section Materials and Methods provides better performance [7, 8].

In order to investigate this hypothesis, we specifically compare waveform *vs* covariance features, the benefits of their combination and the impact of Riemannian *vs* Euclidean geometry under two scenarios. This analysis is done on the data from error-monitoring experiments. Firstly, we simulate the high temporal variability by artificially introducing the latency jitter when extracting the ErrP trials. This corresponds to the expected ERP variability when facing diverse stimuli in natural environment. Secondly, we study the robustness of the mentioned features to the systematic latency shifts between the training and test data. Such scenario corresponds to the classifiers transfer between related ERP protocols. Latency-robust features may be used to reduce BCI calibration time for new protocols as requested by the end users.

2. Materials and Methods

2.1. Data

The dataset that we chose was specifically recorded to address the issue of systematic latency shift in transfer learning [9]. This dataset contains recordings from multiple

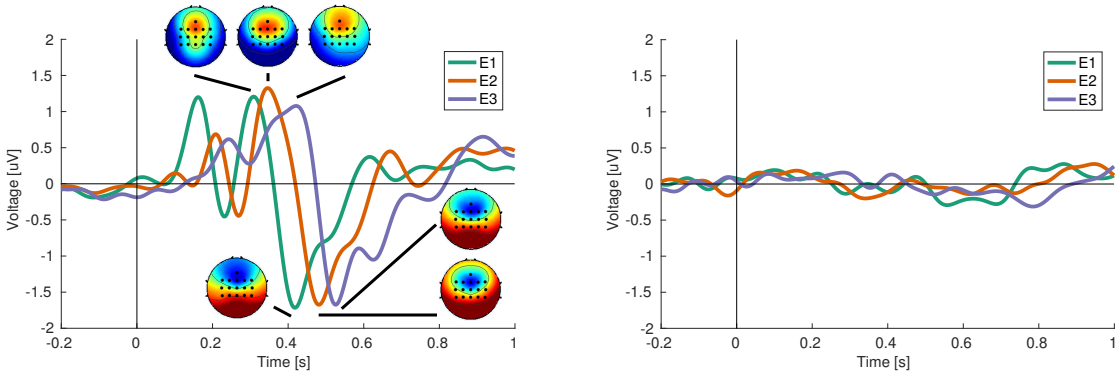

Figure 1. Grand averages for all three experiments. Left: The waveforms show the difference between Error and Correct trials at FCz channel. Topographies show the amplitudes for P3 and N4 peaks. Right: Waveform differences between Error and Correct trials at FCz channel after adding jitter with uniform noise in range [-300, 300] ms.

protocols on Error-related Potentials (ErrP). It will allow us to test the transfer learning in real application in addition to theoretical investigations and simulations. In all the protocols participants observed a discrete movement of an object on the visual display or in physical space. They evaluated the movement as correct if the object moved towards a marked target or erroneous otherwise. The evaluation was made mentally without an overt behavioral response.

In Experiment 1 (E_1) a one-dimensional space with 9 positions was shown on a visual display and a blue square was moving towards a red square (target). In Experiment 3 (E_3) participant were seated in front of a robotic arm, 2 meters away with a transparent panel in between. The panel contained squared marks in a regular grit forming a two-dimensional space. The robotic arm moved behind the marks up, down, left or right towards one of the corners (targets). Experiment 2 (E_2) was a virtual version of Experiment 3 in which marks and the robotic arm were rendered on a visual display. The probability of the erroneous movement was approximately 30%. The time between consecutive movements was randomly sampled from the range [1.7, 4.0] s. 6 subjects participated in three ErrP experiments. Further details about the protocols can be found in [9].

The three experiments yielded a similar grand average ERP waveform while having systematic differences in latency (Figure 1). The latency differences are previously estimated to be: 60.42 $\pm$ 25.24, 108.85 $\pm$ 22.86 and 41.02 $\pm$ 12.95 ms for the E_1E_2, E_1E_3, and E_2E_3 pair of experiments. A linear classifier based on waveform features cannot cope with it when trained and applied on different datasets, the performance drops to random level for considerable latency shifts. However, in an offline analysis when the latency shift can be estimated a simple correction of latency allows to recover the performance up to the level obtained within the single protocol dataset [9].

2.2. Covariance-based features

Different sources of neural activity project at the scalp with a specific distribution. Spatial covariances between EEG channels capture the shape of this distribution and the strength of the activity averaged across the time-window of interest. Covariance matrices can be built in different ways. The simple estimation is done as follows: $\mathbf{C} = \frac{1}{s}\mathbf{X}^T\mathbf{X}$ where $\mathbf{X} \in \mathbb{R}^{s\times n}$ is a multichannel EEG epoch with n channels and s time points. In case of ERPs, this estimation looses the precise temporal dynamics of the waveform. In order to compensate for this, a modified version of covariance matrices was suggested [4]. The epoch $\mathbf{X}$ is augmented with a template $\mathbf{T} \in \mathbb{R}^{s\times k}$ along the channel dimension to build a super epoch $\mathbf{Z} = [\mathbf{X} \quad \mathbf{T}] \in \mathbb{R}^{s\times(n+k)}$. The covariance matrix estimated on the augmented signal becomes:

$$\mathbf{C_Z} = \frac{1}{s}\mathbf{Z}^T\mathbf{Z} = \frac{1}{s}\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{T} \\ \mathbf{T}^T\mathbf{X} & \mathbf{T}^T\mathbf{T} \end{pmatrix} \tag{1}$$

The covariance between $\mathbf{X}$ and $\mathbf{T}$ allows to capture the temporal dynamics specifically with the relation to the template $\mathbf{T}$. In the context of ERP experiments, the template is typically an average ERP of one or several classes. In this case it will reflect how similar a particular epoch is to the average ERP. Thus the covariance matrix $\mathbf{C_T}$ contains the combination of two sources of information: covariances between channels and the similarity to the template waveform.

In high dimensional settings a simple Maximum Likelihood Estimator of covariance matrix is not stable so a regularization is often applied. In this paper we use a shrinkage as described in [10].

2.2.1. Riemannian geometry on covariance matrices. The natural choice to treat covariance matrices is to vectorize them and use as a feature vector for further processing and classification. But most vector-based algorithms assume a Euclidean space, such as PCA [11]. Euclidean geometry is not well suited for SPD matrices due to various drawbacks [12], so they can be characterized better with Riemannian geometry while improving the performance.

Riemannian geometry is the branch of mathematics that studies smooth spaces locally behaving like a Euclidean space. The main feature of Riemannian space consists in the way the distances are defined. While Euclidean space has a constant distance metric in all points, the Riemannian metric smoothly changes along the space. As a simple visual example, imagine a curvy surface of a sphere. Although the surface is a 2-dimensional space, we cannot directly compute the distance between 2 remote points. However, around each point distances can be locally approximated with $\mathbb{R}^2$ Euclidean space.

One of the simple ways to introduce the benefits from Riemannian geometry into classification of covariance matrices is project them on a tangent space [13]. The tangent space is a Euclidean space which allows to leverage the standard vector-based algorithms. The approach requires two steps:

(i) To find a geometric mean $\bar{\mathbf{C}}$ of the sample of covariance matrices:

$$\bar{\mathbf{C}} = \text{argmin}_{\mathbf{C}} \sum_{i=1}^{N} \delta^2(\mathbf{C}_i, \mathbf{C}) \tag{2}$$

where $\mathbf{C}_i$ denotes a covariance matrix, and δ is a Riemannian distance between covariance matrices.

(ii) The relationship between covariance matrices can be approximated with Euclidean geometry around the geometric mean $\bar{\mathbf{C}}$ by projecting them on a tangent space $\mathbf{S}_{\bar{\mathbf{C}}}$ which is called logarithmic mapping Logm$(\cdot)$:

$$\mathbf{S}_{\bar{\mathbf{C}}} = \text{Logm}(\mathbf{C}) = \bar{\mathbf{C}}^{1/2}\text{logm}(\bar{\mathbf{C}}^{-1/2}\mathbf{C}\bar{\mathbf{C}}^{-1/2})\bar{\mathbf{C}}^{1/2} \tag{3}$$

where logm denotes the logarithm of a matrix [14].

2.3. Data processing and feature extraction

All the EEG was filtered with Butterworth band-pass filter of order 4 within the band [1, 10] Hz forward and backward and downsampled from 512 Hz to 128 Hz. Then the signal is spatially filtered with common-average-reference (CAR). Further feature extraction is done in the time window of [200, 1000] ms after the event (correct or erroneous movement). We explore and compare different sets of features, which include ERP waveform and covariance matrices.

For the classification we prepare the following features (Figure 2):

F1 **Waveform**. Waveform of the ERP epoch $\mathbf{X}$, i.e. the amplitude at all channels and time points. Due to high dimensionality (1648) we apply PCA and keep components with highest eigenvalue which explain 90% of the variance. It resulted into 60 +/- 8 features.

F2 **Euclid**. Covariance matrices $\mathbf{C}$ computed with shrinkage to improve stability bringing 136 features.

F3 **Riemann**. Projections of shrinkaged covariance matrices on a tangent space $\mathbf{S}$. The projection reference point is the geometric mean estimated from the training data. The projection does not affect the dimensionality yielding 136 features (see the equation3).

F4 **Riemann+**. Shrinkaged covariance matrices $\mathbf{C_Z}$ estimated on super epochs augmented with the averaged ERPs of each class $\mathbf{Z}$. ERPs were estimated from the training data. Such an augmentation produces a high number of features (1176) so we decided to preselect 8 channels which drastically reduces the number of features to 300. We automatically choose the channels according to the mean Fisher score across all time points [15]. The covariance matrices $\mathbf{C_Z}$ are projected on a corresponding tangent space $\mathbf{S_Z}$.

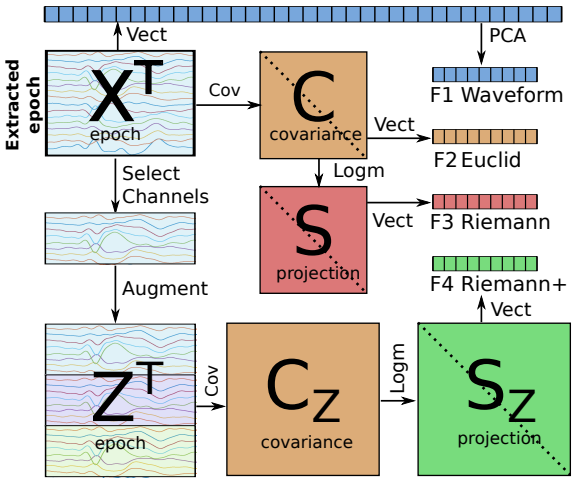

Figure 2. Steps of feature construction for each EEG epoch. The following operations are used: Vect - vectorization of unique values (only upper triangle for SPD matrices), Cov - estimation of covariance matrix with shrinkage, Select Channels - selection of 8 most discriminant channels, Augment - augmentation of the epoch with grand average of erroneous and correct classes, Logm - logarithmic mapping on the tangent space.

2.4. Classification performance evaluation with respect to the latency

To classify erroneous vs correct trials we applied penalized logistic regression (PLR), robust to overfiting and outliers [16]. The regularization parameter is chosen by 4-fold cross-validation in the training dataset. The data are split to preserve the temporal relationships (non-randomized trials).

Additionally, we standardize all features by z-score with mean and standard deviation obtained from the training data because we use a regularized classifier [17]. Since the classes are not balanced in ErrP protocols, we measure the performance by area under the ROC curve (AUC) [18].

Baseline performance. As a baseline we estimate the performance of all features in a leave-one-run-out cross validation. Together with the hyperparameter optimization the overall procedure is structured as a nested cross validation.

Latency jitter. The authors of the used dataset reported high decoding performance of AUC > 80 with a linear classifier. This suggests that ERP waveforms are consistent across single trials. In order to investigate the robustness of different features to the temporal variability we introduce a random jitter to the time window for all data which included both training and test epochs. The amount of jitter is sampled from a uniform distribution. Three gradual jitter levels are [-50, 50], [-100, 100], [-200, 200] and [-300, 300] ms. This procedure is repeated on all three experiments.

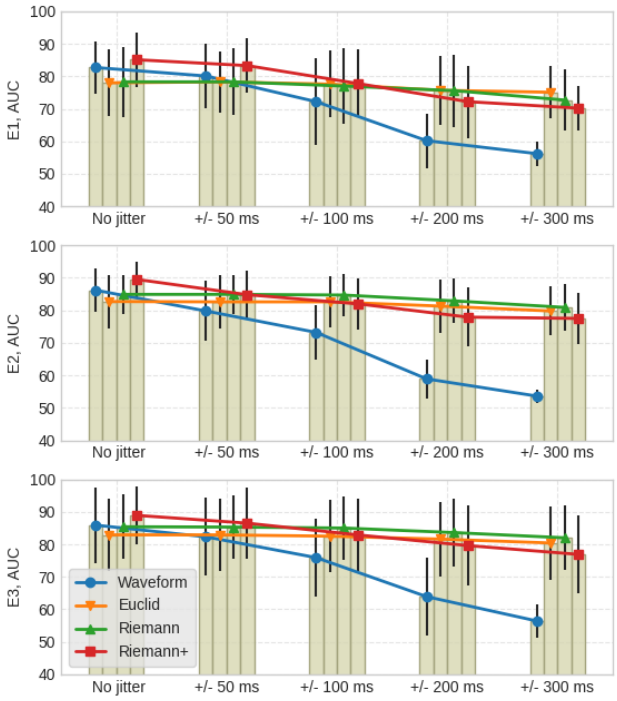

Figure 3. Robustness of the features to simulated temporal jitter. Each figure represents one of the experiments. The data shows AUC as mean and standard deviation across all subjects.

Classifier transfer. The three experiments proved to evoke similar ErrP waveforms on grand averages with a systematic shift in latency. Without any fine-tuning and additional transformation we transfer classifiers within the same subject between experiments. Classifier trained on all data of one experiment is tested on all data of the other two, providing a single AUC value per pair. This procedure is repeated for all pairs of the experiments and all feature sets.

3. Results

3.1. Baseline performance

First of all, we can see that all features perform well on the original datasets without jitter (Figure 3). The average AUC across subjects is between 75 and 90. There is a statistically significant difference across feature sets obtained with two-way repeated measures ANOVA, the factors being the feature and the experiments (p-value < 0.001 (0.000186)). By breaking down the difference into pair-wise comparison with post-hoc two-way repeated measures ANOVA, we can see that the difference is mainly driven by "Riemann+" features (p-value < 0.01 for all pair-wise comparisons with other features). So the combination of discriminant information by means of covariance matrices on super epochs gives the best results. Processing of covariance matrices with Riemannian geometry significantly improves the performance as opposed to Euclidean geometry (p-

value = 0.03).

3.2. Robustness to latency jitter

We introduce different levels of latency jitter to compare the temporal robustness of different feature sets. The performance change for different levels of jitter is shown on Figure 3. The overall trend is a reduction in performance with the increase in temporal jitter for all features. The significance of the effect is tested on each experiment independently with 2-way repeated measures ANOVA for the jitter level and the feature as factors. Both main effects are significant (p-values < 0.001). Regardless the experiment, the strongest effect is obtained for the interaction of factors (p-values < 0.0001) for all the experiments.

The performance decreases more rapidly with jitter for Waveform features, which drops to nearly chance level for jitter level above 200 ms. The post-hoc comparison of "Euclid" or "Riemann" against "Riemann+" features gives non-significant difference on the features, however, the interaction is significant (p-values < 0.0001). It allows us to conclude that simple covariance matrices are most stable against the jitter regardless of whether they are treated with Euclidean or Riemannian geometry.

3.3. Classifier transfer between protocol

We trained separate classifiers on the datasets of each protocol and assessed their performance on the two remaining protocols (Figure 4). As it was the case for the latency variations, waveform features demonstrate the biggest drop in performance. The transfer from E_1 to either E_2 or E_3 leads to a random classification performance. The respective grand average latency differences are 60 and 100 ms.

In contrast to the within-dataset performance, the "Riemann+" approach which combines both types of information does not provide best results when transferring between protocols. In the meantime covariance matrices are more robust especially when the latency shift is bigger. There is a statistically significant difference across features for all training/test pairs, with p-values < 0.0001 (one-way ANOVA). Interestingly, there is no difference in performance between simple covariance matrices and the ones transformed with Riemannian geometry.

4. Discussion

Various studies show high performance ERP-based BCI in controlled experimental conditions. Stimuli diversity or any other change in experimental conditions, however, can affect ERP waveforms and degrade classification performance. On one side, as we have seen through the set of experiments on error monitoring a small modification such as a transition from 1D to 2D scenario or virtual to physical can influence the ERP latency. On the other side, more complex stimuli could lead to a higher temporal variability of ERP within the same protocol [3].

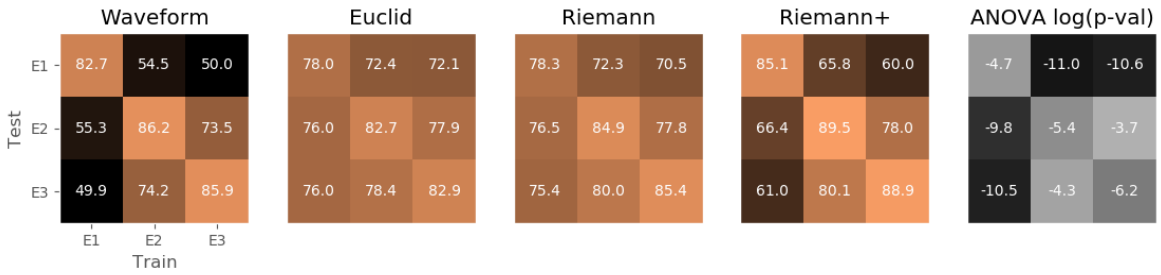

Figure 4. Cross-protocol classifier transfer. First three matrices show AUC estimated on test dataset (y-axis) with a classifier trained on a training dataset (x-axis). The right-most matrix provides p-values in $\log_{10}(\cdot)$ obtained with ANOVA when comparing features for each train/test pair.

In this study we investigated the robustness of different types of features to the variability and systematic shifts in latency using real ErrP data. We conducted two analyses: 1) simulated ErrP temporal variability (latency jitter) and 2) classifier transfer between protocols where ErrP latency shift is observed.

We showed that both types of features – spatial covariances and the ERP waveform – carry sufficient information for high decoding performance (AUC > 75) for all three experiments. Waveform features outperform covariance features yet their combination by the means of the augmenting covariance matrices results in the highest performance. This can be explained by complementary information seen in the spatio-temporal relations between channels. When facing ERP latency shifts or jitter, however, spatial covariance features are more robust in contrast to waveform features. Our results indicate that classification based on waveform features approach a chance level for jitter levels above 200 ms or latency differences between experiments that exceed 100 ms. The features combining both information sources lead to an intermediate performance as compared to simpler feature sets. This is expected as it is constrained by limitation of the waveform features.

We recognize a potential of the covariance-based features for real world BCI applications where ERP variability is inevitable. One cannot control all the aspects of the natural environment – diverse stimuli may require different levels of cognitive effort reflected in the ERP latency [19]. Moreover, there is a requirement on efficient calibration and a quick switch to a new setup.

to be extended...

References

[1] Connie C. Duncan, Robert J. Barry, John F. Connolly, Catherine Fischer, Patricia T. Michie, Risto Ntnen, John Polich, Ivar Reinvang, and Cyma Van Petten. Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, 120(11):1883–1908, November 2009.

[2] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Mller.

Single-trial analysis and classification of ERP components A tutorial. *NeuroImage*, 56(2):814–825, May 2011.

[3] Pietro Aric, Fabio Aloise, Francesca Schettini, Serenella Salinari, Donatella Mattia, and Febo Cincotti. Evaluation of the Latency Jitter of P300 Evoked Potentials during (C)overt Attention BCI. In *Proceedings of TOBI Workshop IV*, 2013.

[4] Marco Congedo, Alexandre Barachant, and Anton Andreev. A New Generation of Brain-Computer Interface Based on Riemannian Geometry. page 33, 2013.

[5] Alexandre Barachant and Marco Congedo. A Plug&Play P300 BCI Using Information Geometry. *arXiv:1409.0107 [cs, stat]*, August 2014. arXiv: 1409.0107.

[6] Ryota Tomioka and Kazuyuki Aihara. Classifying Matrices with a Spectral Regularization. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 895–902, New York, NY, USA, 2007. ACM. event-place: Corvalis, Oregon, USA.

[7] Alexandre Barachant, Stphane Bonnet, Marco Congedo, and Christian Jutten. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, 112:172–178, July 2013.

[8] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu. Transfer Learning: A Riemannian Geometry Framework With Applications to BrainComputer Interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, May 2018.

[9] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and JdR Milln. Latency correction of event-related potentials between different experimental protocols. *Journal of Neural Engineering*, 11(3):036005, April 2014.

[10] Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero. Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, October 2010.

[11] Aapo Hyvrinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. Wiley-Interscience, New York, 1 edition edition, May 2001.

[12] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric Means in a Novel Vector Space Structure on Symmetric PositiveDefinite Matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, January 2007.

[13] Alexandre Barachant, Stphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass Brain-Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, March 2012.

[14] Marcel Berger. *A Panoramic View of Riemannian Geometry*. Springer-Verlag, Berlin Heidelberg, 2003.

[15] Richard O Duda, Peter E Hart, and David G Stork. Pattern classification. 2nd. *Edition. New York*, 2001.

[16] Lucas C. Parra, Clay D. Spence, Adam D. Gerson, and Paul Sajda. Recipes for the linear analysis of EEG. *NeuroImage*, 28(2):326–341, November 2005.

[17] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer-Verlag, New York, 2 edition, 2009.

[18] Andrew P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, July 1997.

[19] M. Kutas, G. McCarthy, and E. Donchin. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, 197(4305):792–795, August 1977.