

# Spatial covariance improves BCI performance for late ERPs components with high temporal variability

Ruslan Aydarkhanov<sup>1</sup>, Marija Ušćumlić<sup>2</sup>, Ricardo Chavarriaga<sup>3,4</sup>, Lucian Gheorghe<sup>5</sup>, José del R Millán<sup>4,6,7</sup>

<sup>1</sup>Medical Image Processing Laboratory, Center for Neuroprosthetics, Interschool Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Campus Biotech H4, 1202 Geneva, Switzerland

<sup>2</sup>Nissan International SA, La Pièce 12, 1180 Rolle, Switzerland

<sup>3</sup>ZHAW Datalab, Zurich University of Applied Sciences, Winterthur, Switzerland

<sup>4</sup>École Polytechnique Fédérale de Lausanne (EPFL), Campus Biotech H4, 1202 Geneva, Switzerland

<sup>5</sup>Advanced Materials and Processing Laboratory, Nissan Research Center, Nissan Motors Co. LTD, 1, Natsushima, Yokosuka-shi, Kanagawa-ken, 237-8523, Japan

<sup>6</sup>Dept. of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA

<sup>7</sup>Dept. of Neurology, The University of Texas at Austin, Austin, TX 78712, USA

E-mail: [ruslan.aydarkhanov@epfl.ch](mailto:ruslan.aydarkhanov@epfl.ch)

**Abstract.** *Objective.* Event Related Potentials (ERPs) reflecting cognitive response to external stimuli, are widely used in Brain Computer Interfaces (BCI). ERP waveforms are characterized by a series of components of particular latency and amplitude. The classical ERP decoding methods exploit this waveform characteristic and thus achieve a high performance only if there is sufficient time- and phase-locking across trials. The required condition is not fulfilled if the experimental tasks are challenging or if it is needed to generalize across various experimental conditions. Features based on spatial covariances across channels can potentially overcome the latency jitter and delays since they aggregate the information across time. *Approach.* We compared the performance stability of waveform and covariance-based features as well as their combination in two simulated scenarios: 1) generalization across experiments on Error-related Potentials and 2) dealing with larger latency jitter across trials. *Main results.* The features based on spatial covariances provide a stable performance with a minor decline under jitter levels of up to  $\pm 300$  ms, whereas the decoding performance with waveform features quickly drops from 0.85 to 0.55 AUC. The generalization across ErrP experiments also resulted in a significantly more stable performance with covariance-based features. *Significance.* The results confirmed our hypothesis 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, Event Related Potentials, Covariance, Riemannian geometry

## 1. Introduction

Humans are constantly engaged in evaluating sensory percepts that are critical to achieve their goals. Such decision making processes give rise to brain responses that can be measured as electroencephalography (EEG) event-related potentials (ERP). Real-time detection of these ERPs provide tremendous opportunities for brain-computer interfacing (BCI). As an illustration, when the user observes an error committed by the system she is interacting with, an Error-Related Potential (ErrP) is elicited and its detection can be used to correct it [1]. Upon each occurrence of such an event, the perception and the evaluation processes take variable periods of time. We investigate the robustness of various ERP properties to these temporal variations for the purpose of improving their decoding performance for BCI applications.

ERP waveforms consist of multiple components characterized by a particular latency, polarity and spatial distribution [2]. The early components are triggered by a sensory stimulation whereas the late components reflect cognitive processing.

ERPs can be successfully detected directly from the waveform in the specific frequency band under controlled experimental conditions where time- and phase-locking is consistent across trials [3]. When facing real world conditions, our cognitive processes may be challenged by complex and diverse stimuli as well as the fact that the environment is dynamic. As a result, the later ERP components are more prone to latency variability across trials [4], which compromises the decoding performance.

An alternative way to detect ERP is to use features based on the spatial covariance matrices across EEG channels. Such features emphasize the pair-wise interaction between channels and the total variance of single channels. Since spatial covariance does not contain precise temporal information we hypothesize that the covariance-based features may be a suitable choice to address the challenge of decoding ERPs with high latency variability inherent to real world BCI applications. Typically, it is advisable to augment the covariance features estimated on a single trial with the covariances between channels in single trial and channels in the template ERP (i.e. grand average ERP) of one of the classes. The rationale for this, is that it allows to keep the temporal information of ERP dynamics which is considered to be a valuable source of discriminant information [5, 6]. In the case of ERP latency variability, however, augmented covariance matrices may have the same limitations as waveforms. The impact of latency shifts and jitter have previously been studied for augmented covariance matrices [6] and the performance dropped quickly for delays as short as 60 ms. An alternative is to consider simple channel-based covariance matrices. Covariance matrices can be used directly as features [7], however it has been hypothesized that treating them under the Riemannian geometry framework provides better performance [8, 9].

In order to investigate this hypothesis, we specifically compare waveform *vs*

covariance features, the benefits of their combination and the impact of Riemannian geometry under two scenarios. This analysis is done on the data from error-monitoring experiments [10]. Firstly, we simulate the high temporal variability by artificially introducing a latency jitter on ErrP trials with a low temporal variability. This corresponds to the expected ERP variability when facing events that require high cognitive processing because either their complexity or limited attentional resources due to multi-tasking. 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 generalization (i.e. classifier transfer) across related ERP protocols. Latency-robust features may be used to reduce BCI calibration time by direct transfer of the available classifier trained on one protocol to other related protocols.

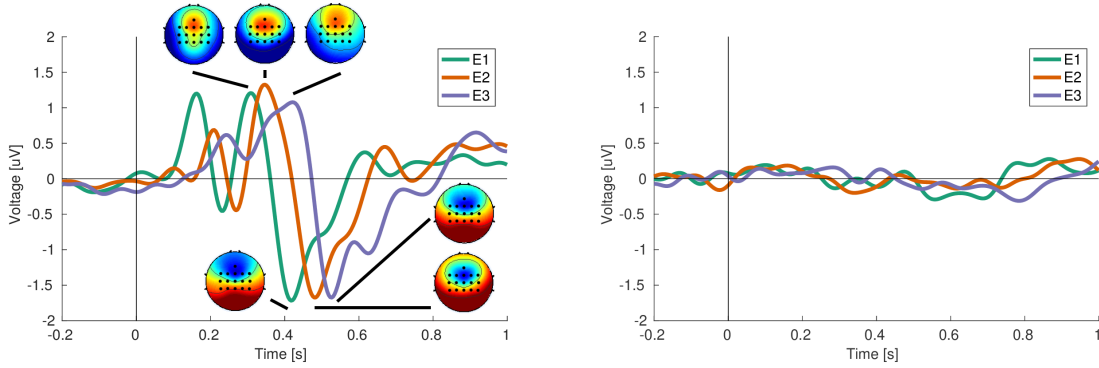
## 2. Materials and Methods

### 2.1. Data

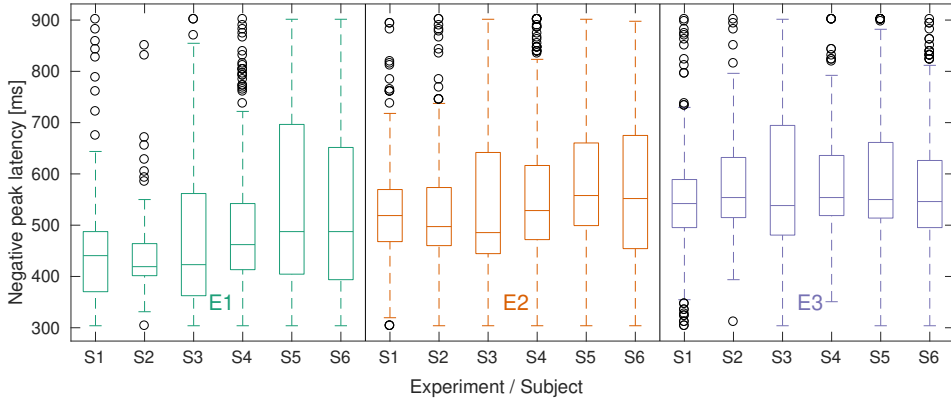
The dataset that we chose was specifically recorded to address the issue of systematic latency shift in transfer learning across protocols [10]. This dataset contains recordings from multiple protocols on Error-related Potentials (ErrP). In all the protocols participants observed a discrete movement of an object on the computer screen 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 ( $E_3$ ), participants 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 grid 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 the three ErrP experiments. Further details about the protocols can be found in [10].

The three experiments yielded a similar grand average ERP waveform while having systematic differences in latency (Figure 1). The latency differences were 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, e.g. by using a few labelled trials of the new experiment, a simple correction of latency allows to recover the performance up to the



**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 the range [-300, 300] ms.



**Figure 2.** The box plots of latency variability estimation per experiment and per subject. The latencies are estimated from a negative peak of Error trials at channel FCz in the window [0.3, 0.9] s.

level obtained within the single protocol dataset [10]. We further estimated latency variability of Error trials per subject and per experiment based on the position of the characteristic negative peak at FCz within the time window [0.3, 0.9] s (Figure 2). Standard deviations of latencies across trials per experiment and subject range from 90 to 170 ms with mean standard deviation of 130 ms.

## 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 loses the precise temporal dynamics of the waveform. In

order to compensate for this, a modified version of covariance matrices was suggested [5]. 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  $\mathbf{Z} = [\mathbf{X} \ \mathbf{T}] \in \mathbb{R}^{s \times (n+k)}$ , where  $k$  is the number of template channels. The covariance matrix estimated on the augmented signal becomes:

$$\mathbf{C}_{\mathbf{Z}} = \frac{1}{s} \mathbf{Z}^T \mathbf{Z} = \frac{1}{s} \begin{pmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{T} \\ \mathbf{T}^T \mathbf{X} & \mathbf{T}^T \mathbf{T} \end{pmatrix} \quad (1)$$

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  $\mathbf{C}_{\mathbf{T}}$  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 [11].

*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 [12]. Covariance matrices are always Symmetric and Positive semi-Definite (SPD), however, Euclidean geometry is not well suited for SPD matrices due to various drawbacks [13], 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 the 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 to project them on a tangent space [14]. 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:

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

where  $\mathbf{C}_i$  denotes a covariance matrix, and  $\delta$  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  $\text{Logm}(\cdot)$ :

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

where  $\logm$  denotes the logarithm of a matrix [15].

### 2.3. Data processing and feature extraction

EEG signals were filtered with a 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 was spatially filtered with common-average-reference (CAR). Further feature extraction was 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.

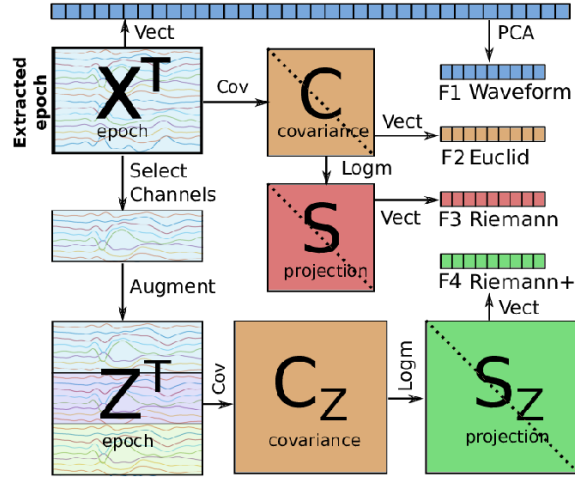
For the classification we compute the following features (Figure 3):

- F1 **Waveform**. Waveform of the ERP epoch  $\mathbf{X}$ , i.e. the amplitude at all channels and time points in the selected window. Due to high dimensionality (1648) we applied 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, yielding 136 features.
- F3 **Riemann**. Projections of shrinkaged covariance matrices on a tangent space  $\mathbf{S}$ . The projection reference point was the geometric mean estimated from the training data. The projection does not affect the dimensionality, yielding 136 features (see the equation 3).
- F4 **Riemann+**. Shrinkaged covariance matrices  $\mathbf{C}_{\mathbf{Z}}$  estimated on super epochs augmented with the averaged ERPs of each class  $\mathbf{Z}$ . ERPs were estimated from the training data. Such an augmentation produced a high number of features (1176), so we decided to preselect 8 channels which drastically reduced the number of features to 300. We automatically chose the channels according to the mean Fisher score across all time points [16]. The covariance matrices  $\mathbf{C}_{\mathbf{Z}}$  were projected on a corresponding tangent space  $\mathbf{S}_{\mathbf{Z}}$ .

### 2.4. Classification performance evaluation with respect to the latency

To classify erroneous vs correct trials we applied penalized logistic regression (PLR), robust to overfitting and outliers [17]. The logistic regression model is a generalized linear model that describes probability of data sample  $\mathbf{x}$  to belong to class  $c = t$ , it is parametrized by the projection vector  $\mathbf{w}$ :

$$p(c = t|\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}} \quad (4)$$



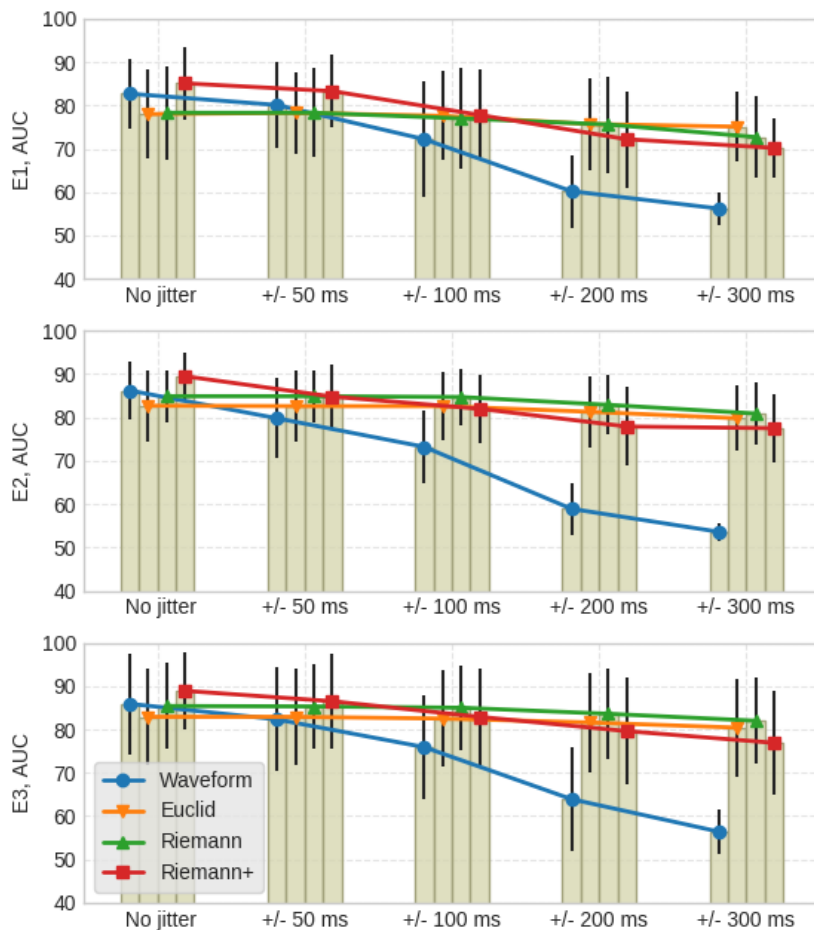
**Figure 3.** Steps of feature construction for each EEG epoch. The following operations were used: Vect - vectorization of unique values (only upper triangle for SPD matrices), Cov - estimation of covariance matrix with shrinkage, Select Channels - selection of 8 most discriminant channels, Augment - augmentation of the epoch with grand average of erroneous and correct classes, Logm - logarithmic mapping on the tangent space.

The optimal projection vector  $\mathbf{w}$  is obtained as a maximum likelihood estimator. However, in penalized logistic regression the estimator is regularized by L1 and/or L2-norm of vector  $\mathbf{w}$ , we use only L2-norm. The regularization parameter was chosen by 4-fold cross-validation in the training dataset. The data was split to preserve the temporal relationships (non-randomized trials).

Additionally, we standardized all features by z-score with mean and standard deviation obtained from the training data because we used a regularized classifier [18]. Since the classes were not balanced in ErrP protocols, we measured the performance by area under the ROC curve (AUC) [19].

*Baseline performance within protocol.* As a baseline we estimated the performance of all features in a leave-one-run-out cross validation. Together with the hyperparameter optimization the overall procedure was 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 when trained and applied within the same protocol. This suggests that ERP waveforms are consistent across single trials and classifier can tolerate latency variability of up to 130 ms of standard deviation when used with waveform features. In order to investigate the robustness of different features to the temporal variability we introduced a random jitter to the time window for all data which included both training and test epochs. The amount of jitter was sampled from a uniform distribution. Three gradual jitter levels were  $[-50, 50]$ ,  $[-100, 100]$ ,  $[-200, 200]$  and  $[-300, 300]$  ms. This procedure was repeated on all three experiments.



**Figure 4.** 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 tested classifier transfer 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 BCI performance within protocol

First, we can see that all features perform well on the original datasets without jitter (Figure 4). The average AUC across subjects is between 75 and 90. There is a statistically significant difference in performance 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 introduced 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 4. 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.

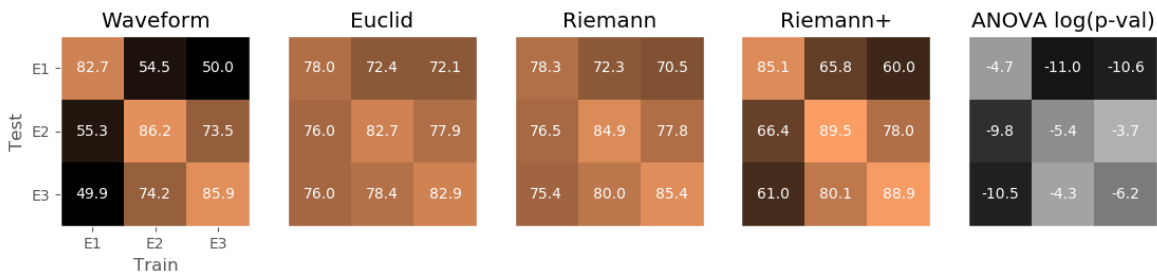
### 3.3. Classifier transfer between protocols

We trained separate classifiers on the datasets of each protocol and assessed their performance on the two remaining protocols (Figure 5). 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. Interestingly, 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). Furthermore, there is no difference in performance between simple covariance matrices and the ones transformed with Riemannian geometry.

## 4. Discussion and Conclusion

Various studies have shown high performance of ERP-based BCIs in controlled experimental conditions. Stimuli diversity or any other change in experimental



**Figure 5.** 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.

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 [4].

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 an ErrP latency shift is observed. Although our study is based on the later ERP’s components reflecting cognitive process of Error detection, we expect that these results can be generalized to any other type of ERPs prone to large latency variability across trials/conditions; yet further validation is needed.

We showed that both types of features – spatial covariances and the ERP waveform – carry sufficient information for high decoding performance ( $AUC > 75$ ) in 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 contained in the spatio-temporal relations between channels. When facing ERP latency shifts – as in the between-protocol generalization scenario – spatial covariance features are more robust in contrast to waveform features. Furthermore, the results of our simulated study on increased latency jitter (variability) suggest that spatial covariance can successfully cope also with this challenge for certain range of variability. Namely, the 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 do not provide the best performance, it is intermediate between the performances of covariance and waveform feature sets alone. This is expected as it is constrained by limitation of the waveform features. Confirmation of the potential of the considered method will require investigation of protocols that elicit ERPs characterized by large latency variance.

Covariance matrices can be used directly to classify EEG signals, however, they

are also widely applied in the estimation of spatial filters by means of Common Spatial Patterns (CSP) method [20] mainly in non-ERP paradigms [21]. The typical features extracted after CSP filter estimation are log variances of spatially projected signal. This feature space approximates a decomposition of Riemannian distances between covariance matrices [22]. Moreover, covariance matrices can be used to estimate spatio-temporal filters as in Regularized Spatio-Temporal Filtering and Classification (RSTFC) [23]. These filtering and denoising methods can take advantage of the spatial covariance robustness against latency variability. Nonetheless, we warn the readers that some methods of spatio-temporal filtering may result in none or negative effect on the performance with high latency variability datasets. This includes methods that model the synchronous ERP in temporal domain based on waveform features and assume constant latency, for example, Spatial-Temporal Discriminant Analysis (STDA) [24], ERP spatial and spectral patterns (ESSP) [25] or Signal-to-noise ratio Maximizer for event-related potentials (SIM) [26].

The robustness of covariance-based features to latency shifts allows to optimize the classifier transfer between different ERP protocols. The common approaches to transfer learning in BCI try to match the distributions of various datasets by keeping the classifier trained on one of the datasets. For example, in the original study the authors suggested to estimate the ERP latency between different protocols and adjust it for single ERPs in new protocols [10]. It can also be done in a geometry-aware manner by treating covariance matrices with Riemannian geometry. For example, the two distributions of covariance matrices can be matched by alignment of geometric means (see equation 2) [9] or by translation, stretching and rotation of the whole distribution [27]. In all cases, achieving efficient cross-experiment, and even cross-subject, transfer requires substantial modeling of distributions, which is only possible after collecting a certain amount of new data. Here we show that in the case of latency shifts, our approach does not need to collect and model new data by relying on spatial covariances as features. It allows to reduce the calibration time to zero with only a mild drop in performance. Nevertheless, we point out that our approach may not be applicable to classifier transfer in case of substantial ERP differences such as in cross-subject transfer.

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 [28]. Moreover, there is a requirement on efficient calibration and a quick switch to a new setup, which could be fulfilled with the approach presented here.

## Acknowledgments

The authors thank Nissan Motor Co. Ltd for supporting this work.

## References

- [1] Ricardo Chavarriaga, Aleksander Sobolewski, and José Del R. Millán. Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Front Neurosci*, 8:208, 2014.
- [2] Connie C. Duncan, Robert J. Barry, John F. Connolly, Catherine Fischer, Patricia T. Michie, Risto Näätänen, 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.
- [3] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Müller. Single-trial analysis and classification of ERP components — A tutorial. *NeuroImage*, 56(2):814–825, May 2011.
- [4] Pietro Aricò, 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.
- [5] Marco Congedo, Alexandre Barachant, and Anton Andreev. A new generation of brain-computer interface based on Riemannian geometry. *arXiv:1310.8115 [cs, math]*, October 2013. arXiv: 1310.8115.
- [6] Alexandre Barachant and Marco Congedo. A Plug&Play P300 BCI Using Information Geometry. *arXiv:1409.0107 [cs, stat]*, August 2014. arXiv: 1409.0107.
- [7] 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: Corvallis, Oregon, USA.
- [8] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, 112:172–178, July 2013.
- [9] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu. Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, May 2018.
- [10] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. d. R. Millán. Latency correction of event-related potentials between different experimental protocols. *Journal of Neural Engineering*, 11(3):036005, April 2014.
- [11] 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.
- [12] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. Wiley-Interscience, New York, 1 edition edition, May 2001.
- [13] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, January 2007.
- [14] Alexandre Barachant, Stéphane 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.
- [15] Marcel Berger. *A Panoramic View of Riemannian Geometry*. Springer-Verlag, Berlin Heidelberg, 2003.
- [16] Richard O Duda, Peter E Hart, and David G Stork. Pattern classification. 2nd. *Edition*. New York, 2001.
- [17] 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.
- [18] 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.

- [19] 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.
- [20] Zoltan J. Koles, Michael S. Lazar, and Steven Z. Zhou. Spatial patterns underlying population differences in the background EEG. *Brain Topography*, 2(4):275–284, June 1990.
- [21] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446, December 2000. Conference Name: IEEE Transactions on Rehabilitation Engineering.
- [22] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Common Spatial Pattern revisited by Riemannian geometry. In *2010 IEEE International Workshop on Multimedia Signal Processing*, pages 472–476, October 2010.
- [23] Feifei Qi, Yuanqing Li, and Wei Wu. RSTFC: A Novel Algorithm for Spatio-Temporal Filtering and Classification of Single-Trial EEG. *IEEE Transactions on Neural Networks and Learning Systems*, 26(12):3070–3082, December 2015.
- [24] Yu Zhang, Guoxu Zhou, Qibin Zhao, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Spatial-Temporal Discriminant Analysis for ERP-Based Brain-Computer Interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(2):233–243, March 2013.
- [25] Chaohua Wu, Ke Lin, Wei Wu, and Xiaorong Gao. A Novel Algorithm for Learning Sparse Spatio-Spectral Patterns for Event-Related Potentials. *IEEE Transactions on Neural Networks and Learning Systems*, 28(4):862–872, April 2017.
- [26] Wei Wu and Shangkai Gao. Learning event-related potentials (ERPs) from multichannel EEG recordings: A spatio-temporal modeling framework with a fast estimation algorithm. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6959–6962, August 2011. ISSN: 1558-4615.
- [27] J. Adam Wilson, Jürgen Mellinger, Gerwin Schalk, and Justin Williams. A procedure for measuring latencies in brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 57(7):1785–1797, July 2010.
- [28] 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.