

# Spatial filters yield stable features for error-related potentials across conditions

Fumiaki Iwane<sup>†,‡</sup>, Ricardo Chavarriaga<sup>†</sup>, Iñaki Iturrate<sup>†</sup>, José del R. Millán<sup>†</sup>

<sup>†</sup> Defitech Chair in Brain-Machine Interface (CNBI), Center for Neuroprosthetics,

Ecole Polytechnique Fédérale de Lausanne, Switzerland

<sup>‡</sup> ATR Computational Neuroscience Laboratories, Department of Brain Robot Interface, ATR, Kyoto, Japan

iwane@atr.jp, {ricardo.chavarriaga, inaki.iturrate, jose.millan}@epfl.ch

**Abstract**—Error-related potentials (ErrP) have been increasingly studied in psychophysical experiments as well as for brain-machine interfacing. In the latter case, the generalisation capabilities of ErrP decoders is a crucial element to avoid frequent recalibration processes, thus increasing their usability. Previous studies have suggested that ErrP signals are rather stable across recording sessions. Also, studies using protocols of serial stimuli presentation show that these potentials do not change significantly with the presentation rate. Here we complement these studies by analysing the decoding generalisation capabilities. Using data from monitoring experiments, we evaluate how much the performance degrades when tested in a condition different than the one the decoder was trained with. Moreover, we compare different spatial filtering techniques to see which preprocessing steps yield less-sensitive features for ErrP decoding.

## I. INTRODUCTION

The existence of brain activity correlated to error monitoring has been widely studied using different imaging methods [1]. Several works have uncovered electroencephalography correlates of errors of different nature, including errors committed by oneself [2], error related to the perceived feedback [3] as well as errors in the interaction with external devices [4]. These *Error-related potentials*, *ErrPs* are particularly interesting in the field of brain-machine interfaces [5]. Specifically, their decoding has been proposed as a mean of improving reliability by correcting erroneous commands [6], [7] or as a mean to adapt the BCI system [8].

The feasibility of performing single-trial recognition of ErrPs has been demonstrated in different paradigms and setups, including P300 spellers [7], Motor-imagery BCI [6], human-robot interaction [9], [8], and car driving [10]. Noticeably these signals, linked to cognitive monitoring processes, have also been reported to be rather stable across different recording days [11], and feedback characteristics [12], [9]. As a matter of fact, they are not strongly modulated by the stimulus presentation rate [13], although they may vary depending on factors such as the predictability of the stimulus [14]

These studies typically compare the signal across different conditions without assessing the classification performance across different experimental conditions (i.e., generalisation across days or feedback presentation speeds). In these studies, both the pre-processing steps and the classifier parameters are specifically suited for a given experimental condition.

In this work we evaluate whether spatial filtering techniques can be used to extract ErrP features that are stable across experimental conditions. In particular, we assess ErrPs decoding performance in two experiments where the subject monitors the behaviour of a computer cursor. In the first experiment, we assess the signal and decoding changes across two different recording dates [11]. The second experiment, evaluates these changes with respect to changes in the cursor speed (i.e., the pace at which the stimuli is provided to the subject) [13].

## II. METHODS

### A. Experimental protocols

Both experimental protocols are focused on ErrPs elicited when the subject monitors the movements of a cursor in the screen. During the experiment, the subject should evaluate if the cursor moves towards a target position marked in red. The subject has no control whatsoever on the cursor movements. Instead, cursor movements are programmed so that it usually moves towards the target, except for a proportion of trials where it moves in a different direction (i.e. error trials).

In the *first experiment* the cursor moves horizontally every 2000 ms across a 1D space composed of 20 positions as shown in Figure 1(a) [11]. When the cursor reaches the target position, a new target location is chosen. Six subjects (1 female, mean age  $27.8 \pm 2.2$ ) took part in the experiment which was composed of two recording sessions. The spanning days between these sessions is listed in Table I. The cursor moved correctly towards the target 80% of the trials (i.e. error rate was 0.2). For this experiment we considered each different recording day as a different condition to test the classifier generalisation<sup>1</sup>.

The *second experiment* follows the same approach where the user monitors the cursor movements. In this case the cursor can move in four directions within a 5x5 2D grid shown in Figure 1(b) [13]. Besides cursor movements, there was a fifth possible cursor action to denote that the system considered that it had reached the target, see Figure 1(c). In this case, the cursor shrank to 40% of its size during 400 ms. The probability for the cursor to execute the correct action in a given position was of 70%. Eight male subjects

<sup>1</sup>The data used in this study is publicly available at the website of the EU-project BNCI Horizon 2020 (<http://bnci-horizon-2020.eu/database/data-sets>)

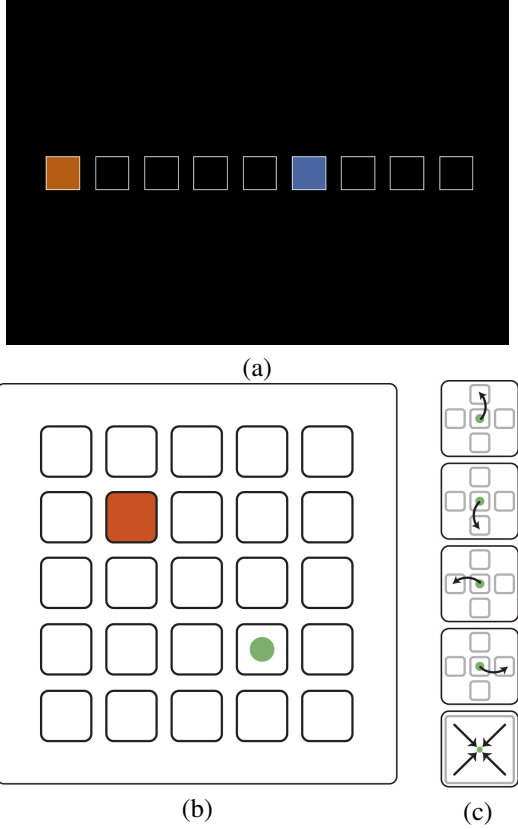


Fig. 1. Experimental protocols. The cursor moves in discrete steps among several positions towards an intended target location (marked in red). The subject is asked to monitor whether the cursor’s actions are correct or erroneous for reaching the target. (a) Experiment 1. Cursor (blue square) moves horizontally in a 1D grid. (b) Experiment 2. The green cursor moves in a 2D working space. (c) Schematic representation of the 5 possible actions corresponding to four directions of movement and a goal-reaching action.

TABLE I

EXPERIMENTAL CONDITIONS. EXPERIMENT 1. TIME DIFFERENCE (IN DAYS) BETWEEN THE TWO EXPERIMENTAL RECORDINGS. EXPERIMENT 2. DIFFERENT INTER-STIMULUS-INTERVALS (ISI).

		EXPERIMENT 1					
Subject		1	2	3	4	5	6
Days		51	50	54	211	628	643

		EXPERIMENT 2		
Condition		Baseline	Fast	Overlapping
ISI [ms]		1500	800	600

(mean age  $26.5 \pm 3.4$ ) participated in the experiment that took place in a single session. Each subject tested three conditions (c.f. Table I) defined by the time between two consecutive cursor movements (inter-stimulus interval, ISI) ranging from 600 to 1500 ms.

We recorded 64 EEG channels following the 10/10 system using a Biosemi ActiveTwo amplifier. Signal was recorded at 512 Hz, and 2048 Hz for experiments 1 and 2, respectively. As in previous studies we bandpass filtered the signal in the

frequency range [1 10] Hz using a FIR filter. For the analysis of event-related potentials (ERPs), signals were segmented into epochs corresponding to trials for both error and correct movements. Each epoch comprised the activity from 500 ms before the cursor movement onset ( $t=0$ ) until 1000 ms after.

### III. ASSESSING CLASSIFIER GENERALISATION

We assess how three pre-processing methods affect the generalisation capabilities across conditions for both experiments (recording day and cursor speed for experiments 1 and 2, respectively). In the first case we followed the same approach as in previous studies [12], [13]. In this case features were extracted from the temporal signal within the time window [200 800] ms, downsampled to 64 Hz. In this case, only the signals from eight fronto-central electrodes (Fz, FC1, FCz, FC2, C1, Cz, C2, and CPz) were taken into account. The initial feature vectors (8 channels \* 25 time samples) were then normalised and decorrelated using PCA [14]. To reduce the problem dimensionality we kept the components that explained 95% of the signal variance in the training set.

In the remaining two cases, we compared two preprocessing methods previously used in the literature: xDAWN [15] and Canonical Correlation Analysis, CCA [16]. These methods are described in detail below. For this purpose, we computed spatial filters based on the entire 64 channel montage in the same temporal window as before. In each experiment the filters were obtained using only the data from one of the conditions. Extracted features using the three techniques were classified as erroneous or correct using PCA and regularised linear discriminant analysis (LDA). The stability of the spatial filters was assessed by computing the decoding performance achieved by using features extracted using this filter on different conditions (e.g. for the experiment 1 we estimated the parameters of the spatial filter, PCA and LDA from the data of the first recording day and tested it on data from the second day). For reference we also report the performance obtained through 10-fold cross-validation when the same condition is used for computing the decoding parameters. To account for the unbalanced conditions, we report performance in terms of the area under the sensitivity-specificity curve (AUC) [17].

#### A. Spatial filtering methods

1) *xDAWN*: This method aims at finding a projection that maximizes the signal to signal-noise ratio. Let  $x_j(t)$  denote the EEG signal recorded by  $j^{th}$  sensor at time index  $t$  and let  $X \in \mathbb{R}^{N_t \times N_s}$  be the matrix of recorded EEG signals.  $N_s$  is the number of sensors and  $N_t$  is the number of temporal samples in the recording. Let  $a_j(t)$  denote the signal for the  $j^{th}$  sensor at time index  $t$ , and let  $A \in \mathbb{R}^{N_e \times N_s}$  be the matrix of evoked signals, where  $N_e$  is the number of temporal samples in each trial.

The erroneous trials of ErrP experiments leads to the following model

$$X = DA + N \quad (1)$$

where  $D \in \mathbb{R}^{N_t \times N_e}$  is the Toeplitz matrix whose first column is defined such that  $D_{\tau_k, 1} = 1$ , where  $\tau_k$  is the stimulus onset

of  $k^{th}$  target stimulus and such that all the other elements are null.  $DA$  in eq.(1) thus represents the evoked response synchronous with stimuli of interest, and matrix  $N$  the on-going activity of the user's response.  $A$  is defined by

$$\tilde{A} = \arg \min_A \|X - DA\|^2 = (D^T D)^{-1} D^T X \quad (2)$$

where  $\cdot^T$  is the transpose operator. If there is no overlap between trials,  $\tilde{A}$  is equal to the averaged signal ( $D^T A$ ).

The next step of xDAWN modelling process is to include the  $N_f$  spatial filters  $u_i (1 \leq i \leq N_f \leq N_s)$  into the model,

$$XU = DAU + NU \quad (3)$$

where  $U \in \mathbb{R}^{N_s \times N_f}$  are the spatial filters matrix whose  $i^{th}$  column is  $u_i$ . The spatial filters are designed to maximise the signal-to-signal plus noise ratio with the generalised Rayleigh quotient.

$$\tilde{u} = \arg \max_U \frac{\text{Tr}(U^T \tilde{A}^T D^T D \tilde{A} U)}{\text{Tr}(U^T X^T X U)} \quad (4)$$

where  $\text{Tr}(\cdot)$  is the trace operator.

The final step requires to solve the optimisation problem by a combination of QR decomposition and singular value decomposition applied to  $X$  and  $D$  matrices. For further details we refer to [15]. Note that the Generalised Eigenvalue Decomposition is a common approach to solve the Rayleigh quotient optimisation problem [18]. The result is a set of filters which are sorted by the absolute value of the eigenvalues.

2) *Canonical Correlation Analysis Based Spatial Filter*: Canonical Correlation Analysis (CCA) is a multivariate statistical method that given two datasets aims at finding linear transformations that maximise the pair-wise correlation across the transformed datasets [16]. Let  $X$  and  $Y$  denote two multivariate datasets with  $p$  and  $q (\leq p)$  variables respectively,  $X = (X_1, X_2, \dots, X_p)^T$  and  $Y = (Y_1, Y_2, \dots, Y_q)^T$ . and let  $U$  and  $V$  denote their transformed datasets,

$$\begin{aligned} U &= W_x^T X = (U_1, U_2, \dots, U_q)^T \\ V &= W_y^T Y = (V_1, V_2, \dots, V_q)^T. \end{aligned} \quad (5)$$

The CCA method yields two linear projection matrices  $W_x$  and  $W_y$ , that maximize the canonical correlation  $\rho_i^*$  between the canonical variables  $U_i$  and  $V_i$ ,

$$\rho_i^* = \frac{\text{cov}(U_i, V_i)}{\sqrt{\text{var}(U_i) \text{var}(V_i)}}. \quad (6)$$

The linear projection matrices  $W_x$  and  $W_y$  are selected so that their variance is equal to one, while they are uncorrelated to all other canonical variables

$$\text{var}(U_i) = \text{var}(V_i) = 1 \quad (7)$$

$$\text{cov}(U_i, V_j) = \text{cov}(U_i, V_i) = 0, \quad \forall i \neq j. \quad (8)$$

In order to obtain CCA-based spatial filters, the first step is to make the two datasets  $X$  and  $Y$  from the available data. Assuming that recordings yield  $k$  trials of EEG signal whose

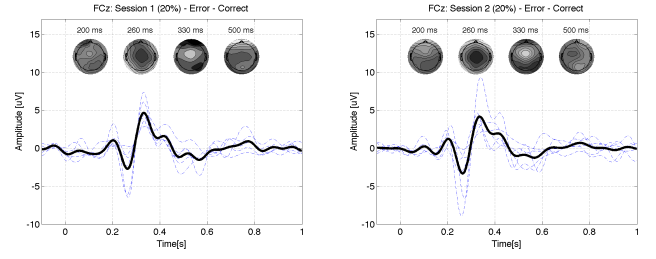


Fig. 2. Experiment 1. Grand average event related potentials error-minus-correct condition in the FCz electrode. The *thick line* corresponds to the average across subjects; *thin dashed lines* indicate individual subject averages. Each plot corresponds to a different recording session. Topographical maps of scalp activity appear in the insets (nose up); activity is color coded from black to white in the range  $[-5, 5]$  Time  $t=0$  corresponds to the feedback onset. Adapted from [11]

TABLE II  
EXPERIMENTAL PROTOCOL 1

### STANDARD

		Test	
Train		day 1	day 2
day1		87.7± 2.9	81.2± 5.8
day2		82.5± 1.8	84.7± 7.4

### xDAWN

		Test	
Train		day 1	day 2
day 1		90.7±2.0	82.9±7.6
day 2		84.6±6.7	89.7±4.8

### CCA

		Test	
Train		day 1	day 2
day 1		93.5±1.2	79.8±14.5
day 2		84.9±5.6	91.0±5.3

dimension is  $n \times m$ , where  $k$ ,  $n$  and  $m$  are the number of trials, number of channels and the number of samples in one epoch, respectively. The first dataset is composed of all trials concatenated into a new matrix  $X$  with dimensions  $n \times (m \cdot k)$ . The second dataset  $Y$  is built based on the grand average ERP over the  $k$  trials,  $R$ . Then  $R$  is replicated  $k$  times, to make a  $n \times (m \cdot k)$  matrix  $Y = [RR \dots R]$ . Finally, CCA is applied on  $X$  and  $Y$  to find  $W_x$ , which can be used as a spatial filter. See [19] for further details.

## IV. RESULTS

### A. Event-related potentials

Figures 2 and 3 show the ErrPs observed in each condition for both experiments. It can be seen that signals are rather stable across conditions. We do not report in-depth analysis of the characteristics of these patterns, since this has been reported previously [11], [13].

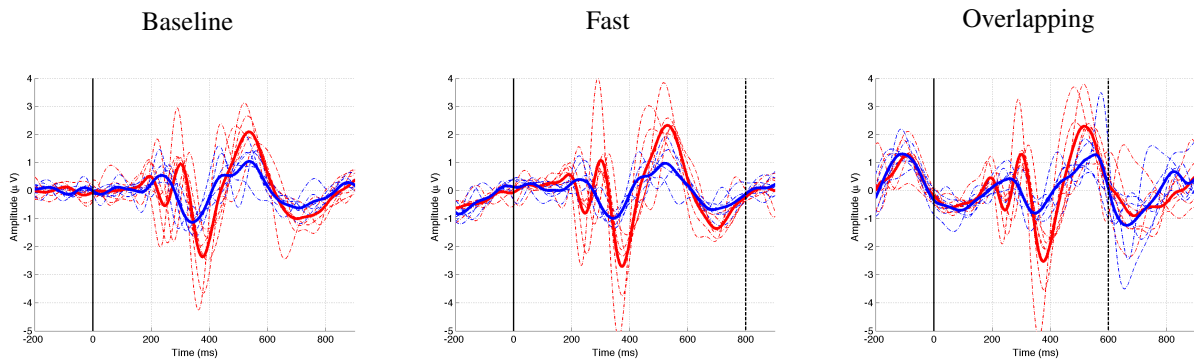


Fig. 3. Experiment 2. Grand-average event-related potentials (FCz electrode). Each dashed trace corresponds to the subject grand average ERP for both types of trials (Red: Error; Blue: Correct). Solid lines show the average across subjects.  $t=0$  corresponds to the cursor movement onset and the vertical dashed line in (b)-(c) shows the time of the next movement onset.

TABLE III  
EXPERIMENTAL PROTOCOL 2

**STANDARD**

Train	Test		
	slow	fast	overlap
slow	$72.2 \pm 7.6$	$69.9 \pm 5.2$	$69.8 \pm 6.8$
fast	$71.3 \pm 5.2$	$72.8 \pm 4.6$	$71.3 \pm 6.1$
overlap	$69.1 \pm 6.1$	$70.8 \pm 5.8$	$71.2 \pm 7.9$

**xDAWN**

Train	Test		
	slow	fast	overlap
slow	$72.1 \pm 10.5$	$73.2 \pm 7.5$	$71.0 \pm 9.6$
fast	$73.6 \pm 8.8$	$75.7 \pm 6.4$	$73.9 \pm 7.6$
overlap	$72.5 \pm 8.0$	$75.0 \pm 7.0$	$76.4 \pm 7.7$

**CCA**

Train	Test		
	slow	fast	overlap
slow	$75.7 \pm 9.4$	$75.4 \pm 6.3$	$71.6 \pm 9.4$
fast	$74.3 \pm 7.6$	$78.1 \pm 4.5$	$75.2 \pm 7.6$
overlap	$70.3 \pm 13.8$	$73.9 \pm 11.2$	$76.2 \pm 7.9$

**B. Classification Accuracy**

The decoding performance averaged over subjects is shown in Tables II, and III. In these tables, 'Train' denotes the data used to obtain the decoding parameters (spatial filter, PCA and LDA). In general, the use of spatial filters, derived using either xDAWN or CCA, yielded higher AUCs than the standard method based on 8 predefined channels. No major differences were found between the use of CCA and xDAWN-based filters. As previous study reported [20], [13], decoding performance is higher when the training and the testing conditions are the same (c.f., diagonal positions in Tables II, and III). Unsurprisingly, when the conditions differ the performance decreases.

Moreover, large inter-subject differences can be observed.

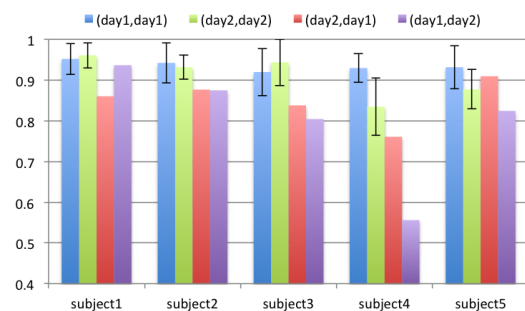


Fig. 4. Result of AUCs in experimental protocol 1 with CCA based spatial filter. Legends correspond to (data for training of spatial filter, data for testing). i.e., the First two columns correspond to the case where the spatial filter was obtained in the same condition used for the testing.

For illustration, we report individual decoding performance using CCA in both experimental protocols, Fig. 4 and 5. We can see in experiment 1 that subjects 4 shows the largest performance decrease in the (day 1, day2) condition. It is worth noticing that this is the case with almost 21 months spanning period between recording sessions (c.f., Table I). However, in the case of subject 5 AUCs remain stable across different conditions even with more than 21 months spanning days. As for the experiment 2, changes in the ISI seem to affect more some subjects than others; c.f., AUCs differences for Subject 2. In general, as the ISI difference between the training and testing condition increases the change in performance increases as well.

In the first experimental protocol, a 3 (Preprocessing: Standard, xDAWN, CCA)  $\times$  2 (recording day) ANOVA showed a significant effect of the pre-processing method ( $F_{2,24}=4.77$ ,  $p < 0.018$ ). A post-hoc Tukey test showed that only CCA had significantly higher AUCs than the standard method ( $p < 0.05$ ). Moreover, AUCs were significantly higher when the spatial filter was computed on the data from the same day as the testing data (two-way ANOVA,  $F_{2,54}=18.07$ ,  $p < 0.001$ ). No significant difference was found between CCA and xDAWN.

On the other hand, in the second experiment a two-way

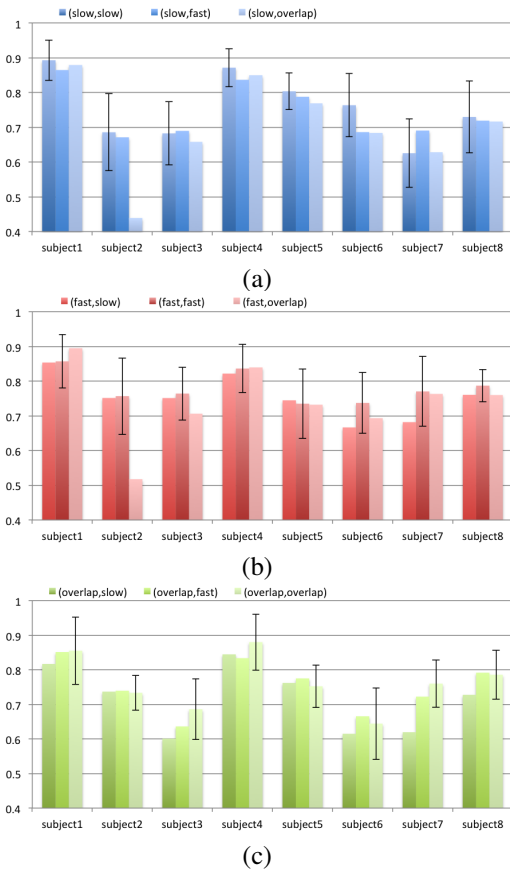


Fig. 5. Result of individual AUCs in experimental protocol 2 with CCA based spatial filter. Legend of this figure is (data for training of spatial filter, data for testing). (a) Individual AUCs with spatial filter trained by using slow trials. (b) Individual AUCs with spatial filter trained by using fast trials. (c) Individual AUCs with spatial filter trained by overlap trials.

ANOVA did not show any significant effect of the preprocessing when testing in the same condition ( $F_{2,64}=2.21$ ,  $p=0.12$ ). It is natural since our previous study reported that there was no significant difference between ErrPs with different ISI, when testing in the same condition. Post-hoc analysis showed again that CCA and xDAWN do not differ significantly, but both outperform the standard approach (Tukey test,  $p<0.001$ ).

For each spatial filtering technique, we performed three additional one-way ANOVA tests to assess the effect of change of temporal demand. Each test corresponds to the condition in which the spatial filter was obtained (i.e., slow, fast, and overlap). No significant differences were found in any of the tests when the spatial filter was computed using CCA. In contrast, for xDAWN-based filters significant effects were found when the spatial filter was obtained in the fast condition ( $F_{2,45}=9.02$ ,  $p<0.01$ ). Post-hoc analysis showed differences for the AUCs obtained when testing in the slow and overlap datasets ( $p<0.01$ ). Similarly, for filters obtained in the overlap condition ( $F_{2,45}=5.92$ ,  $p<0.01$ ), differences were found between the (overlap,overlap) and (overlap,slow) cases (post-hoc analysis,  $p<0.01$ ). This suggests that CCA may be more robust to changes across conditions, although

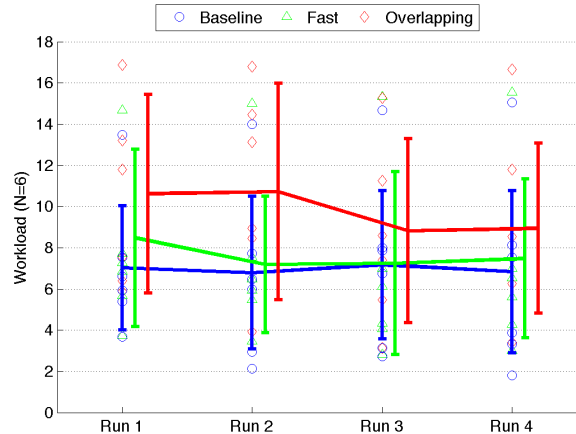


Fig. 6. User workload levels in experiment 2 (mean  $\pm$  SD across subjects) for each condition and run as indicated by NASA-TLX [13].

more extended tests are required to corroborate this finding.

For each ISI, we performed three additional two-way ANOVA (3: Preprocessing  $\times$  3: ISI) tests to assess the effect of change of temporal demand. Each test corresponds to the condition in which the decoding parameters was obtained (i.e., slow, fast, and overlap). No significant differences were found in any of the tests (slow:  $F_{4,63}=0.63$ ,  $p=0.54$ , fast:  $F_{4,63}=0.02$ ,  $p=0.98$ , overlap:  $F_{4,63}=1.27$ ,  $p=0.29$ ). These results suggest that temporal demand does not affect the decoding AUCs.

## V. CONCLUSION

In this work we evaluate the capacity of different techniques to yield spatial filters that maximise ErrP features over different experimental conditions (recording sessions and stimulus presentation rate). Results support previous studies on the stability of the EEG correlates of error monitoring. Our results show that although decoding performance decrease across conditions, the use of spatial filters yield results comparable to those obtained with standard methods within the same conditions (see Tables II and III). This is particularly important for practical BMI systems where frequent need of decoder recalibration may hinder the user acceptance.

Designing a BMI classifier involves optimisation of both preprocessing techniques and classification parameters. In order to avoid parameter overfitting, it is necessary to divide the available calibration data in order to leave apart samples that can be to properly assess the decoding performance (i.e. testing data). This requires careful evaluation of performance using methods such as nested cross-validation. Since spatial filters may not require constant recalibration as shown in this work, the need for recalibration of all parameters is relaxed. As such, given the task demands, only the classifier may need to be updated when the experimental conditions change. Therefore we suggest the use of spatial filtering techniques, in combination with adaptation techniques for improving the generalisation of ErrP-based BMIs (e.g., [12]). Future work

should also assess whether these techniques are also beneficial for achieving transfer learning across subjects.

In experiment 2 the temporal demand of the task has influence in the user perceived workload (c.f., Fig 6, [13]). Therefore the capability of generalising across conditions is particularly important, as subjects can start using the system in the condition that induces the lower workload (i.e., slow), and as they get more used to the protocol the presentation rate could be increased without need for system updating. However, the decoding performance decreases as the difference between ISIs in the training and testing conditions get bigger. This can be explained by the fact that some late components of previous stimulus may overlap with the current ERP. Theoretically the xDAWN filter should be robust to overlapping effects, as long as they correspond to a linear combination of components. Performance decrease in this case may suggest non-linear effects in the case of shorter ISIs. Further study of this will be subject to future work as well as the assessment of other techniques to overcome overlapping effects [21].

Furthermore we compare different spatial filtering techniques. This shows how these techniques can help to enhance features that are mainly correlated to error processing in general, and are less modulated by other factors and process. The comparison at the level of single-trial decoding can be a valuable tool to analyse these signals also for studies on cognitive neurophysiology, complementing standard studies based on average analysis [22].

#### ACKNOWLEDGMENT

This work was supported by the Swiss National Science Foundation NCCR Robotics (<http://www.nccr-robotics.ch>). I.I. acknowledges support from the "EPFL Fellows" fellowship programme co-funded by Marie Curie, FP7 Grant agreement no. 291771. F.I. was funded by the TOBITATE! Young Ambassador Program. This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

#### REFERENCES

- [1] M. Ullsperger, A. G. Fischer, R. Nigbur, and T. Endrass, "Neural mechanisms and temporal dynamics of performance monitoring." *Trends Cogn Sci*, vol. 18, no. 5, pp. 259–267, May 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.tics.2014.02.009>
- [2] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial." *Biol Psychol*, vol. 51, no. 2-3, pp. 87–107, Jan 2000.
- [3] T. U. Hauser, R. Iannaccone, P. Stämpfli, R. Drechsler, D. Brandeis, S. Walitza, and S. Brem, "The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization." *Neuroimage*, vol. 84, pp. 159–168, Jan 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.neuroimage.2013.08.028>
- [4] P. W. Ferrez and J. d. R. Millán, "Error-related EEG potentials generated during simulated brain-computer interaction." *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 923–929, 2008.
- [5] R. Chavarriaga, A. Sobolewski, and J. d. R. Millán, "Errare machinale est: the use of error-related potentials in brain-machine interfaces." *Front Neurosci*, vol. 8, p. 208, 2014. [Online]. Available: <http://dx.doi.org/10.3389/fnins.2014.00208>
- [6] P. W. Ferrez and J. d. R. Millán, "Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy," in *Proc 4th Intl. Brain-Computer Interface Workshop and Training Course*, Graz, Austria, September 2008.
- [7] M. Spüler, M. Bensch, S. Kleih, W. Rosenstiel, M. Bogdan, and A. Kübler, "Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI." *Clin Neurophysiol*, vol. 123, no. 7, pp. 1328–1337, Jan 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.clinph.2011.11.082>
- [8] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. d. R. Millán, "Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control." *Scientific Reports*, vol. 5, p. 13893, 2015. [Online]. Available: <http://dx.doi.org/10.1038/srep13893>
- [9] R. Chavarriaga, X. Perrin, R. Siegwart, and J. d. R. Millán, "Anticipation- and error-related EEG signals during realistic human-machine interaction: A study on visual and tactile feedback," in *34th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'12)*, 2012, pp. 6723–6726.
- [10] H. Zhang, R. Chavarriaga, Z. Khalilardali, L. Gheorghie, I. Iturrate, and J. d. R. Millán, "EEG-based decoding of error-related brain activity in a real-world driving task." *J Neural Eng*, vol. 12, no. 6, p. 066028, Nov 2015. [Online]. Available: <http://dx.doi.org/10.1088/1741-2560/12/6/066028>
- [11] R. Chavarriaga and J. d. R. Millán, "Learning from EEG error-related potentials in noninvasive brain-computer interfaces." *IEEE Trans Neural Syst Rehabil Eng*, vol. 18, no. 4, pp. 381–388, Aug 2010. [Online]. Available: <http://dx.doi.org/10.1109/TNSRE.2010.2053387>
- [12] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. d. R. Millán, "Latency correction of error potentials between different experiments reduces calibration time for single-trial classification," *J Neural Engineering*, vol. 11, p. 036005, 2014.
- [13] R. Chavarriaga, I. Iturrate, Q. Wannebroucq, and J. d. R. Millán, "Decoding fast-paced error-related potentials in monitoring protocols." *Conf Proc IEEE Eng Med Biol Soc*, vol. 2015, pp. 1111–1114, Aug 2015. [Online]. Available: <http://dx.doi.org/10.1109/EMBC.2015.7318560>
- [14] D. Talmi, R. Atkinson, and W. El-Deredy, "The feedback-related negativity signals salience prediction errors, not reward prediction errors." *J Neurosci*, vol. 33, no. 19, pp. 8264–8269, May 2013. [Online]. Available: <http://dx.doi.org/10.1523/JNEUROSCI.5695-12.2013>
- [15] B. Rivet, A. Souloumiatic, 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, Aug 2009. [Online]. Available: <http://dx.doi.org/10.1109/TBME.2009.2012869>
- [16] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.
- [17] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861 – 874, 2006.
- [18] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, Dec 2000.
- [19] M. Spüler, A. Walter, W. Rosenstiel, and M. Bogdan, "Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data." *IEEE Trans Neural Syst Rehabil Eng*, vol. 22, no. 6, pp. 1097–1103, Nov 2014. [Online]. Available: <http://dx.doi.org/10.1109/TNSRE.2013.2290870>
- [20] I. Iturrate, L. Montesano, and J. Minguez, "Shared-control brain-computer interface for a two dimensional reaching task using EEG error-related potentials." *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, pp. 5258–5262, 2013. [Online]. Available: <http://dx.doi.org/10.1109/EMBC.2013.6610735>
- [21] M. D. Burns, N. Bigdely-Shamlo, N. J. Smith, K. Kreutz-Delgado, and S. Makeig, "Comparison of averaging and regression techniques for estimating event related potentials." *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, pp. 1680–1683, 2013. [Online]. Available: <http://dx.doi.org/10.1109/EMBC.2013.6609841>
- [22] M. X. Cohen, "Comparison of different spatial transformations applied to EEG data: A case study of error processing," *International Journal of Psychophysiology*, 2014.