

Task-dependent signal variations in EEG error-related potentials for brain–computer interfaces

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2013 J. Neural Eng. 10 026024

(<http://iopscience.iop.org/1741-2552/10/2/026024>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 155.210.155.42

The article was downloaded on 26/03/2013 at 15:28

Please note that [terms and conditions apply](#).

Task-dependent signal variations in EEG error-related potentials for brain–computer interfaces

I Iturrate^{1,2}, L Montesano^{1,2} and J Minguez^{1,2,3}

¹ Instituto de Investigación en Ingeniería de Aragón (I3A), Edificio I+D+i, Mariano Esquillor, E-50018 Zaragoza, Spain

² Departamento de Informática e Ingeniería de Sistemas (DIIS), Universidad de Zaragoza, Maria de Luna 1, E-50018 Zaragoza, Spain

³ Bit&Brain Technologies SL, E-50018 Zaragoza, Spain

E-mail: iturrate@unizar.es, montesano@unizar.es and jminguez@unizar.es

Received 30 October 2012


Accepted for publication 26 February 2013

Published 26 March 2013

Online at stacks.iop.org/JNE/10/026024

Abstract

Objective. A major difficulty of brain–computer interface (BCI) technology is dealing with the noise of EEG and its signal variations. Previous works studied time-dependent non-stationarities for BCIs in which the user's mental task was independent of the device operation (e.g., the mental task was motor imagery and the operational task was a speller). However, there are some BCIs, such as those based on error-related potentials, where the mental and operational tasks are dependent (e.g., the mental task is to assess the device action and the operational task is the device action itself). The dependence between the mental task and the device operation could introduce a new source of signal variations when the operational task changes, which has not been studied yet. The aim of this study is to analyse task-dependent signal variations and their effect on EEG error-related potentials. **Approach.** The work analyses the EEG variations on the three design steps of BCIs: an electrophysiology study to characterize the existence of these variations, a feature distribution analysis and a single-trial classification analysis to measure the impact on the final BCI performance. **Results and significance.** The results demonstrate that a change in the operational task produces variations in the potentials, even when EEG activity exclusively originated in brain areas related to error processing is considered. Consequently, the extracted features from the signals vary, and a classifier trained with one operational task presents a significant loss of performance for other tasks, requiring calibration or adaptation for each new task. In addition, a new calibration for each of the studied tasks rapidly outperforms adaptive techniques designed in the literature to mitigate the EEG time-dependent non-stationarities.

 Online supplementary data available from stacks.iop.org/JNE/10/026024/mmedia

(Some figures may appear in colour only in the online journal)

1. Introduction

EEG brain–computer interfacing (BCI) is an emergent technology developed to provide a communication channel between human and device using only brain activity, with proved utility in a wide range of clinical and non-clinical applications (see [1, 2] for reviews). BCI always relies on the fact that the user performs a mental task, which presents

associated brain patterns that are measured on the EEG and decoded to obtain the intention of the user. Subsequently, these intentions are used to operate a device performing given operational tasks, which can or cannot be related to the decoded mental task. A major difficulty of BCIs is to deal with the non-stationary behaviour and noise of the EEG [3, 4]. Most of efforts in this context have been devoted to those BCIs where the mental task and the operational task are not related [5–10].

This study moves in a different direction, focusing on BCIs where the mental and operational tasks are related, and thus task-dependent signal variations could exist.

A great portion of the existing BCIs relies on mental tasks whose associated brain patterns are independent of operational tasks. For instance, many BCIs have used (i) self-regulation of rhythms' amplitudes in the temporal domain (e.g., slow cortical potentials [11]); (ii) changes in the frequency power spectrum (such as μ - and β -rhythms [12], motor imagery of body limbs [13] or performance of cognitive mental tasks [14]); or (iii) attendance to visual stimuli (e.g., visual P300 potentials [15], or steady-state visual evoked potentials [16]). In the aforementioned BCIs, the mental task is decoupled from the operational task and the signal variations are assumed to be independent of the device operation [10]. Consequently, these BCIs generalize among different operational tasks. For instance, slow cortical potentials have been used for spelling devices or controlling 2D cursors [17, 11]; motor-imagery BCIs have been used to operate 2D cursors, wheelchairs or mobile robots [13, 12, 18]; and P300 BCIs have been used to operate spelling devices, wheelchairs or mobile robots, among others [15, 19, 20]. However, even though the signal variations are independent of the operational task, the time-dependent non-stationarities of the EEG lead to changes in the feature distributions and thus on the performance of the BCI [10]. The difficulty is to achieve a robust classification from the calibration phase to the feedback phase, along the feedback phase, and/or between different sessions. The techniques developed to date focus on the feature extraction or the classification process: either relying on finding time-invariant features of the EEG to control the BCI [5, 6] or adapting the classifier with supervised techniques incorporating labelled examples of subsequent sessions [7, 8] or with unsupervised techniques [10, 9].

However, in other types of BCIs the mental task is coupled with the operational task, such as those based on the error-related negativity (ERN) ERP component [21] or the associated error-related potentials [22]. The principle of these BCIs is to detect the occurrence of an error during the device operation in the brain patterns. Although these BCIs are rather insensitive to time-dependent non-stationarities [23, 24], the operational task-dependent signal variations are natural in this context, as the monitoring of different devices or processes leads to different error-related potentials (e.g., user's own errors [21], interaction errors [22], feedback errors [25, 24] or observation errors committed either by another person [26] or by virtual or real devices [27, 28, 23], namely the observation error-related potentials). In previous works, the latter potentials have been studied by their difference average, and characterized by three main ERP components (an N2, a P3 and an N4) [22]. Mainly, these potentials have been used for the correction of incorrect commands delivered by the BCI [29, 30], or as feedback for virtual devices performing learning tasks [27, 23]. These BCIs always require re-calibration as, depending on the size of the variation of the error-related potentials, they might not generalize between small variations of the task and among different tasks. This is a large drawback for the practical deployment of these BCIs, as calibration is a

boring, tedious and tiring process that could last approximately 30–45 min before using the BCI [27, 28, 23].

This paper studies the presence of signal variations in observation error-related potentials, where the mental task is coupled with the operational task. The analysis spanned the three design steps of BCIs: an electrophysiology study to characterize the existence of these variations, a feature analysis that showed how these variations affect the feature distributions and a classification analysis to measure the impact on the final BCI performance.

2. Methods

2.1. Data recording

The EEG signals were recorded with a gTec system (2 synchronized gUSBamp amplifiers) with 32 electrodes distributed according to an extended 10/20 international system (FP1, FP2, F7, F8, F3, F4, T7, T8, C3, C4, P7, P8, P3, P4, O1, O2, AF3, AF4, FC5, FC6, FC1, FC2, CP5, CP6, CP1, CP2, Fz, FCz, Cz, CPz, Pz and Oz), with the ground on FPz and the reference on the left earlobe. The EEG signals were digitized with a sampling frequency of 256 Hz, power-line notch-filtered, common-average-reference (CAR) filtered and band-pass filtered at [0.5, 10] Hz. The data acquisition was developed under the BCI2000 platform [31].

2.2. Experimental design

Ten subjects (eight males and two females, mean age 27.20 ± 4.08 years) participated in the experiments. The participants were comfortably seated 1 m away from a computer screen displaying all the information related to the experiments. Two experimental protocols were designed. In each protocol, the subject monitored the execution of a task where a virtual device had to reach a given goal. The motion of the device could be correct (towards the goal) or erroneous (away from the goal). Each task (denoted operational task, OT) consisted of a set of subtasks where the goal location changed but the device movements were fixed (see figure 1). The subjects were asked to assess the device movements as erroneous or non-erroneous. The participants were allowed to freely move their eyes, while blinking and muscular movements were constrained to the resting periods. Each experiment took place in one session, and the time period between sessions was 3 ± 6 days.

2.2.1. Operational task 1 (OT1, figure 1(a)). The screen displayed the virtual device as a blue circle in the centre of the screen (rest position), and also three rectangles indicating possible destinations, arranged in a triangle-like shape equidistant to the rest position. The goal was marked with a green rectangle. The device could perform three possible actions: move over one of the three rectangles. After the motion, the device returned to the rest position. The users were instructed to assess the actions as non-erroneous when it moved to the goal (green rectangle) or as erroneous otherwise. The three subtasks corresponded to the three possible goal locations (denoted subtasks OT1.Up, OT1.Left and OT1.Right). The session was recorded in 36 runs. Each run consisted of 50

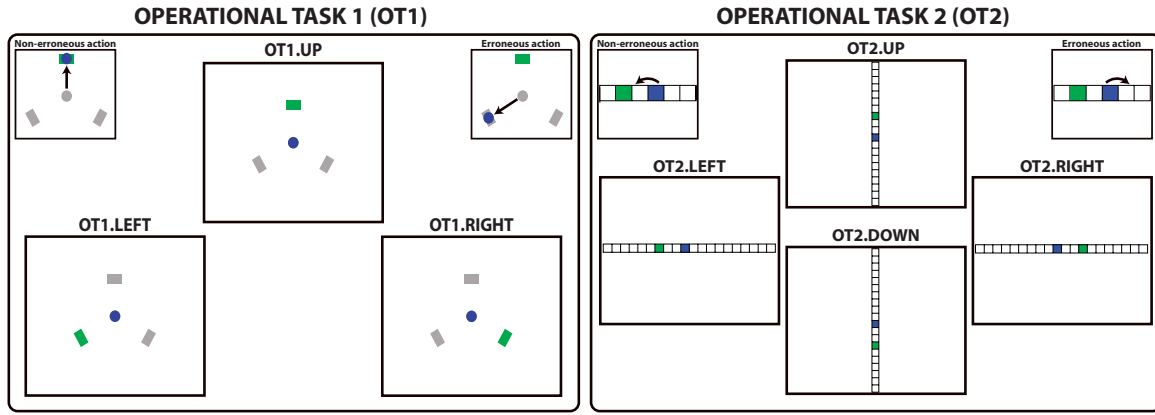


Figure 1. Schematic illustration of the operational tasks (OTs). Each OT consisted of a virtual device (blue circle or blue square) that had to move to the goal location (marked in green). For each OT, subtasks were defined by changing the goal location. Examples of non-error and error movements are shown for each task.

actions with the goal fixed at one position, with a 20% probability of not moving to the goal. For each subtask, 120 error and 480 non-error potentials were acquired, i.e. a total of 360 error and 1440 non-error potentials for the operational task. This session lasted 3 h.

2.2.2. Operational task 2 (OT2, figure 1(b)). The screen displayed the virtual device as a blue square and the goal location as a green square, located on a horizontal or vertical grid composed of 20 positions. The device could perform two possible actions: move one position left or right in the horizontal grid, or up or down in the vertical grid. The users were instructed to assess the actions as non-erroneous when the device moved towards the green square and as erroneous when it moved in the opposite direction. When the device reached the goal, the goal was moved three positions away along the grid (randomly left or right for the horizontal grid, or up or down for the vertical grid). The four subtasks corresponded to the relative position of the goal location with respect to the virtual device: OT2.Left, OT2.Right, OT2.Up and OT2.Down. The session was recorded in 24 runs. Each run consisted of 50 actions with a fixed grid (either the horizontal or the vertical grid), with a 20% probability of not moving towards the goal. For each subtask, 60 error and 240 non-error potentials were acquired, i.e. a total of 240 error and 960 non-error potentials for the operational task. The session lasted 2 h. The protocol was based on [22].

2.3. Analysis of error-related potentials and their task-dependent variations

The observation error-related potentials are characterized by three distinct components of the difference between the error and the non-error grand averages: an N2, a P3 and an N4 component [22]. Regarding the N2 and P3 components, several studies suggest that they may actually be the error-related negativity (ERN or N_e) and the following positivity (P_e) [21], but there is still an open discussion about it [32]. Regarding the N4 component, previous studies have suggested that its generation could be due to a visual semantic mismatch [33]. Nonetheless, for the observation error-related potentials, the

three components originate in the anterior cingulate cortex (ACC, Brodmann areas 24 and 32) and the pre-supplementary motor area (pre-SMA, Brodmann area 6) [22], suggesting an activation of an error-processing system in the brain [25].

The electrophysiology of the error-related potentials and their associated variations were studied through the analysis of the raw EEG, and with a filtered EEG where those components not originated in the brain sources involved in the error-related potentials were removed. The filter eliminated several types of artefacts, including electromyographic activity (such as that provoked by scalp and neck muscles), ocular activity (such as eye movements and blinks) and brain activity not originated within the error processing brain sources [34] (such as spatial attention components [35]). The filter has two main steps: (i) application of independent component analysis (ICA) [36] and (ii) isolation of the independent brain components related to error processing, with a posterior re-projection of this information to the sensor space. Note that while ICA techniques have been widely used for the characterization of brain sources [37–39] and the removal of artefacts [40], the filter proposed herein focused on the isolation of the brain process of interest (see [41] for a similar approach for P300 classification).

The ICA spatial filter is a statistical model defined as $\mathbf{x} = \mathbf{A}\mathbf{s}$, where \mathbf{x} are the input data, and \mathbf{A} and \mathbf{s} are the mixing matrix and independent components estimated by maximizing the temporal independence among the components, respectively. Each column vector \mathbf{a}_i of \mathbf{A} is the spatial pattern associated with the component s_i . While there are many ways to compute the ICA model [36], its computation has two difficulties in the EEG context: the number of independent components to estimate [39] and the non-reliable nature of the estimation process [42]. The number of components d was estimated using m -fold cross-validation principal component analysis (PCA) in the sensor space [43] (the number of folds was fixed to $m = 5$ in the experiments, leading to $d \simeq 15$ dimensions retained from the original 32 channels). The ICASSO technique was used to address the non-reliability of ICA [42]. ICASSO estimates N ICA filters using the FastICA algorithm [36] under changes of the initial conditions, and then performs clustering on the obtained estimations (N was fixed

empirically to 100). Once the ICA model was computed, the DipFit source localization [39] was used to estimate the neural source of each component. Those independent components whose brain source was in the ACC or the pre-SMA were selected, as these areas are believed to be the main generators of error-processing brain activity [21, 25, 38]. For each subject and task/subtask, the number of components selected was between one and four, which were re-projected back to the sensor space to obtain the filtered data.

The analysis of the shape and timing of the potentials (with and without the filter) for each task and subtask was carried out through the computation of time-locked averaged potentials for the error and non-error potentials in channel FCz, the difference average (error minus non-error averages) [22, 28] and by an r^2 discriminability test [1]. A topographic interpolation of the potentials was obtained at the time of the main peaks of the difference average. A source localization analysis was also performed with sLoreta [44] at the N4 component of the error averages of each operational task [22]. Additionally, the peak amplitudes and latencies of the most prominent negativity were extracted from single-trial signals as the minimum value within the time window [320, 600] ms in channel FCz. The latency-sorted single-trial potentials were plotted as a colour-encoded image with a smoothing window of 50 trials.

Finally, to assess the statistical differences among tasks/subtasks of the error-related potentials, one-way within-subjects ANOVAs (factor: tasks or subtasks) were conducted over the latencies and amplitudes of each component (N2, P3 and N4) of the difference average, averaged from channels Fz, FCz and Cz [45, 46]. When needed, the Geisser–Greenhouse correction was applied to the data to ensure sphericity.

2.4. Analysis of the impact of task-dependent signal variations

Once the existence of signal variations was studied, its impact on BCI performance was analysed at two levels: (i) changes in the feature distributions used for error detection and (ii) the corresponding classification accuracy. This analysis was performed using the filtered EEG data to avoid the influence of activity not originated in the error-processing brain areas (e.g. artefacts).

2.4.1. Feature extraction. Previous studies have demonstrated that amplitude values of the error-related potentials from several fronto-central channels are suitable features for their discrimination (error versus non-error) [22, 28, 23]. In this study, features are constructed as linear combinations of amplitudes of channels and time points that best separate these two classes [47]. Given a set of n labelled trials of the two classes, for each trial, eight fronto-central channels (Fz, FC1, FCz, FC2, C1, Cz, C2, and CPz) within a time window of [200, 800] ms were subsampled to 64 Hz and concatenated as a vector of 312 features. The feature vectors of all trials were normalized, and then decorrelated using PCA, retaining 95%

of the explained variance. The k most discriminant features were selected based on a robust variant of the Fisher score [9]:

$$FS(f^i) = \frac{|\text{med}(f_1^i) - \text{med}(f_2^i)|}{\text{med}_{\text{ad}}(f_1^i) + \text{med}_{\text{ad}}(f_2^i)}, \quad (1)$$

with $\text{med}(f_j^i)$ and $\text{med}_{\text{ad}}(f_j^i)$ being the median and the median absolute deviation of feature f^i for class $j \in \{1, 2\}$, respectively. The number of features to retain k was determined by a tenfold cross validation.

The effect of signal variations was measured by the statistical significance between the feature distributions for each class, between operational tasks (inter-task) and between subtasks (intra-task). One-way within-subjects ANOVAs (factor: tasks or subtasks) were conducted on the single-trial features of each class to assess the statistical significance. In addition, the inter/intra-task similarity of the features' distributions was quantified by the Kullback–Leibler (KL) divergence. The KL divergence from $P \sim \mathcal{N}(\mu_P, \Sigma_P)$ to $Q \sim \mathcal{N}(\mu_Q, \Sigma_Q)$ is

$$D_{\text{KL}}(P||Q) = \frac{1}{2} \left(\text{tr}(\Sigma_Q^{-1} \Sigma_P) + \mathbf{v}^T \Sigma_Q^{-1} \mathbf{v} - \ln \left(\frac{|\Sigma_P|}{|\Sigma_Q|} \right) - k \right) \quad (2)$$

with $\mathbf{v} = (\mu_Q - \mu_P)$. High values of $D_{\text{KL}}(P||Q)$ entail large differences between distributions. The KL divergences were computed using the $k = 10$ most discriminant features (according to (1)) to compare the intra/inter features distributions.

2.4.2. Single-trial classification. The classifier used in the analysis was a regularized version of linear discriminant analysis (LDA) [48]. The LDA discriminant function $D(\mathbf{f})$ is the hyperplane that maximally separates the feature distributions corresponding to two classes: $D(\mathbf{f}) = \mathbf{w}^T \mathbf{f} + b$, where \mathbf{f} is the feature vector to be classified, and \mathbf{w} and b are, respectively, the normal vector to the hyperplane and the corresponding bias computed by $\hat{\mu} = \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)$, $\mathbf{w} = \tilde{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1)$, $b = -\mathbf{w}^T \hat{\mu}$, where $\hat{\mu}_j$ is the sample mean of class j , $\hat{\mu}$ is the sample global mean and $\tilde{\Sigma}$ is the regularized sample covariance matrix (shared for both classes).

Regularization aims to minimize the covariance estimation error $E = |\Sigma - \tilde{\Sigma}|$, with Σ being the real covariance matrix, by penalizing very large and very small eigenvalues. The regularized covariance matrix was computed by $\hat{\Sigma} = (1 - \gamma(n))\hat{\Sigma} + \gamma(n)\nu I$, where n was the number of trials used for training, $\gamma(n) \in [0, 1]$ the regularization factor (whose value can be computed numerically [48]), $\hat{\Sigma}$ the sample covariance matrix and $\nu = \text{tr}(\hat{\Sigma})/k$ the average eigenvalue of $\hat{\Sigma}$, with k being the number of diagonal elements of $\hat{\Sigma}$.

To tackle the signal variations, the classifier was adapted based on a sequential process in which labelled examples of the new task were used to modify the discriminant function of the LDA classifier [9]. Namely, given a new example of class j at time t , $\mathbf{f}_j(t)$, the mean $\hat{\mu}_j$ was updated using an exponential moving average:

$$\hat{\mu}_j(t) = (1 - \alpha)\hat{\mu}_j(t-1) + \alpha\mathbf{f}_j(t). \quad (3)$$

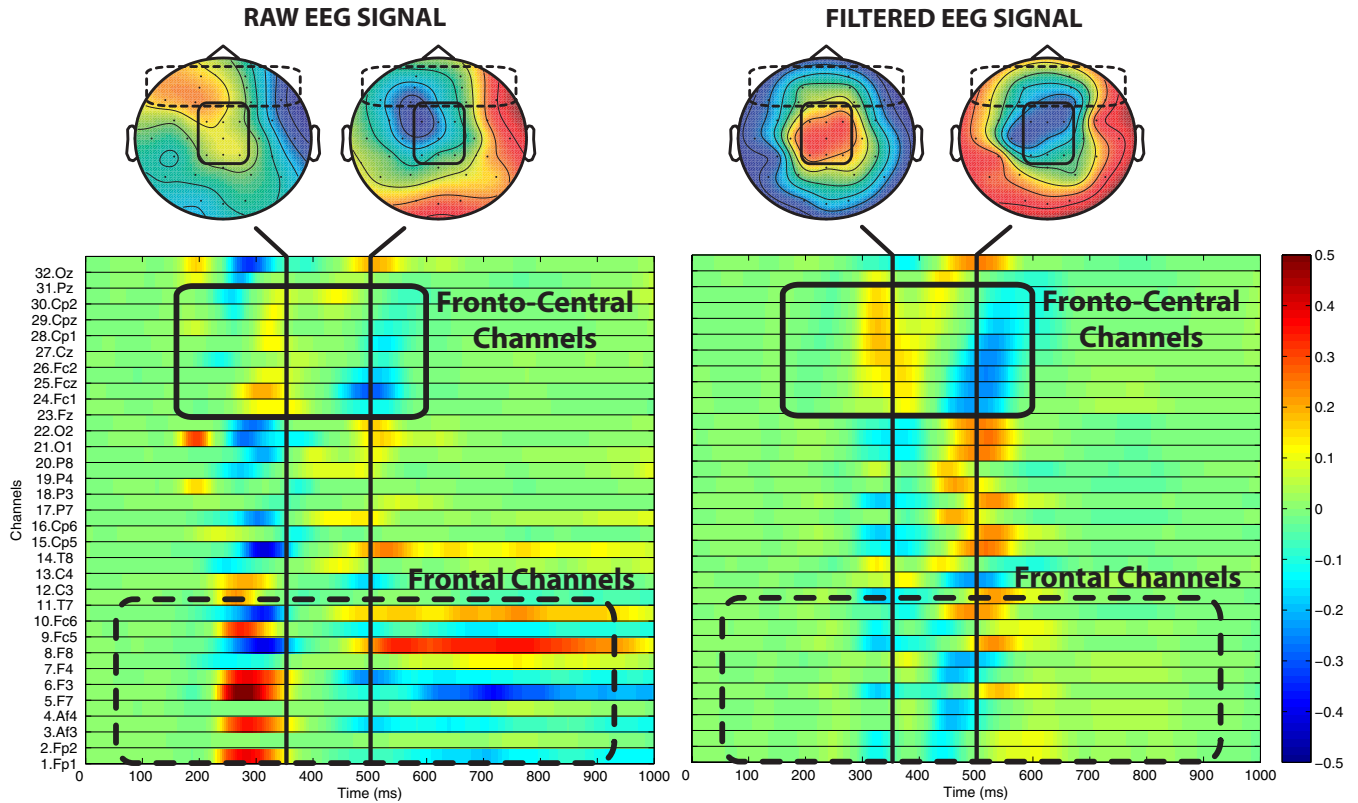


Figure 2. Signed r^2 discriminability test of non-error versus error, performed on the subtask OT1. Right, when not filtering the signal (left) and when filtering the signal (right). The x-axis represents the time (from 0 to 1000 ms) and the y-axis represents each recorded EEG channel. Topographic interpolation of the r^2 is shown at 350 and 500 ms. The solid boxes mark the position of fronto-central channels, whereas the dashed boxes mark the position of frontal channels. When not filtering the signals, most of the discriminability comes from frontal channels with the sign reversed on the left and right hemispheres. When filtering the signals, most of the discriminability comes from fronto-central channels.

$\alpha \in [0, 1]$ is the update parameter (fixed to $\alpha = 0.05$ [9]) and the initial values for $\hat{\mu}_j$ were obtained using data of another task or subtask. The corresponding discriminant function was recomputed using the new mean $\hat{\mu}_j$ to update the w and b parameters accordingly.

The impact of signal variations was analysed firstly without adaptation, and then with supervised adaptation. Firstly, the classifier was trained with examples of one task and tested with another one (inter-task) or trained with examples of one subtask and tested with the other subtasks (intra-task). The results were then compared with the performance of the baseline classifier computed using a tenfold cross-validation scheme for each task (subtask) separately. Secondly, the adaptation was evaluated against baseline classifier performance as a function of the number of trials used to train/adapt the classifier (i.e. calibration time). The adaptive classification results were obtained using the train-test sets as follows: the classifier was initially trained using the train dataset and then, the test dataset was split into two subsets (D_1 with 300 examples and D_2 with the remaining). For each trial at time t , the classifier was updated using (3) and tested on the D_2 dataset. The results were then compared with the performance of baseline classifiers built using trials $[1 \dots t]$ of D_1 and tested on D_2 . This process was repeated ten times to reduce variability in the results for adaptive and baseline

classifiers while shuffling trial positions, and then averaging the obtained accuracies.

3. Results

3.1. Electrophysiology of potentials and their signal variations

The analysis comprises the raw EEG data along with the filtered EEG data (see subsection 2.3). The proposed filter eliminated 90% of the ICA components (which were not estimated within the ACC or pre-SMA). The majority of these components were ocular artefacts such as eye movements (estimated in frontal areas such as Brodmann areas 10, 11 or 38) and brain activity estimated either in parieto-occipital/occipital areas (Brodmann areas 17, 18 and 19) or the posterior cingulate cortex (PCC, Brodmann areas 23 and 31). Although these components contributed to the EEG, they were not originated in the main error-processing areas and thus were eliminated by the filter. Artefact correlation with the error potentials within subtasks but not within tasks is an effect worthy of mention, which might affect feature extraction and classification analysis of the generalization study. Figure 2 displays an example with the raw and filtered EEG for the subtask OT1. Right. Without filtering, a signed r^2 discriminability test indicated that the most

discriminant features were on frontal channels (originated by lateral eye movements); on the other hand, after filtering, the most discriminant features were due to fronto-central activations (originated by error-related potentials). Note that without filtering, the most discriminant features may greatly discriminate the potentials within the subtask OT1.Right (due to lateral eye movements), but would not generalize for the other subtasks or task as they involved different eye movements.

For both operational tasks and subtasks, the average difference of the raw/filtered potentials for both conditions presented a small negative deflection approximately at 250 ms (an N2 component) and prominent positive and negative peaks (P3 and N4 components) at approximately 300 and 500 ms, respectively (in agreement with the r^2 test, figure 3). The topographical scalp maps at these last two peaks showed fronto-central activations for the two operational tasks.

When not filtering the data, source estimations for OT1 (at 500 ms of the error average) were in the paracentral lobule (Brodmann Area 5), whereas for OT2 (at 450 ms of the error average) were in the ACC (Brodmann area 24). In contrast, when filtering the data the potentials from both operational tasks were estimated within the ACC, agreeing with previous studies on error-related potentials [38, 22, 23, 27]. These results indicated that the use of the ICA filter was advisable for the isolation of the error-processing brain activity.

Regarding the inter-task analysis, visual inspection revealed that the shape of the averaged potentials differed between the two operational tasks (figure 3). In OT1, the averaged error-related potentials presented two positive peaks at 200 and 300 ms, whereas in OT2 the first positive peak was smaller, with a more prominent peak at 280 ms. This difference was also appreciated in the sorted single-trial error-related potentials. Additionally, the ANOVA analyses reported statistical differences on the latency of the three main components both for the raw ($F_{1,9} = 9.574, p = 0.01$; $F_{1,9} = 24.469, p = 0.001$; and $F_{1,9} = 48.442, p = 1 \times 10^{-4}$ for the N2, P3 and N4 components) and filtered data ($F_{1,9} = 7.789, p = 0.02$; $F_{1,9} = 24.970, p = 0.001$; and $F_{1,9} = 28.809, p = 0.0005$). For the amplitude of the components, statistical differences were found only for the N4 component of the filtered data ($F_{1,9} = 15.66, p = 0.003$). These results indicated the existence of signal variations in the error-related potentials between operational tasks affecting mainly the latency of their main components.

Regarding the intra-task analysis, visual inspection revealed that the shape of the averaged potentials was very similar among subtasks (figure 3). The ANOVA analyses reported no statistical differences ($p > 0.05$) except for the N2 component of OT1 for the raw data ($F_{2,18} = 12.021, p = 0.0005$). These results indicated that on average, the components did not change among subtasks of the same task.

3.2. Features analysis

Regarding the inter-task analysis, a visual inspection of the features showed that only the best feature (f^1) reflected similar patterns between OT1 and OT2, whereas the other features

presented different spatio-temporal combinations (see figure 4 top-middle for representative examples). In the intra-task case, the features were very similar between subtasks. For instance, the best features (f^1 to f^3 in the figure) were almost equal among subtasks, while the worst feature f^{10} presented greater variations.

For each class (non-error and error), feature distributions were significantly different for inter- and intra-tasks (ANOVA test, $p < 0.001$ in all the cases). For the inter-task case, KL divergences were 0.64 ± 0.34 and 1.71 ± 0.46 for non-error and error, respectively, while for the intra-tasks of OT1 and OT2 the divergences were 0.64 ± 0.51 and 1.06 ± 0.28 , and 0.44 ± 0.19 and 1.92 ± 0.60 . For all tasks and subtasks, the non-error KL divergences were significantly lower than the error KL divergences (unpaired one-tailed t -test, $p < 1 \times 10^{-4}$). The inter and intra-task KL divergences of the error distributions for OT1 were significantly different (unpaired two-tailed t -test, $t_{38} = 5.36, p = 4 \times 10^{-6}$), but the inter/intra-task divergences of the non-error distributions were not ($t_{38} = 0.04, p = 0.97$). For OT2, inter/intra-task KL divergences were significantly different for the non-error distributions ($t_{68} = 2.69, p = 0.01$), but no significant differences were found for the inter/intra-task divergences of the error distributions ($t_{68} = -1.05, p = 0.30$). In summary, feature distributions changed significantly between tasks and among subtasks, and the error distributions changed significantly more than the non-error distributions. Furthermore, the features varied significantly more when changing the task than when changing the subtask.

3.3. Classification

3.3.1. Analysis without adaptation. The tenfold accuracies of non-error and error potentials were, on average, 89.29% and 78.00% for OT1 and 86.64% and 73.00% for OT2, and 89.38% and 77.97% for subtasks of OT1 and 84.06% and 72.33% for subtasks of OT2 (see figure 5). All baseline classifiers were above the chance level.

Regarding the inter-task generalization results (figure 5, left), when training with OT1 and testing with OT2 there was a significant average increase of 6.76% (one-tailed paired t -test, $t_{18} = 1.86, p = 0.04$) in the detection of non-error potentials, but a significant decrease of 21.54% ($t_{18} = -4.19, p = 0.0003$) for error potentials. As can be seen, the standard deviation was also increased compared to the tenfold accuracies. This indicated that the accuracy drops varied substantially from subject to subject, with subjects having large drops, and others having almost no accuracy decrease. When training with OT2 and testing with OT1, the accuracies presented significant decreases of 6.44% and 12.44% ($t_{18} = -1.83, p = 0.04$ and $t_{18} = -2.75, p = 0.007$). As with the previous case, the standard deviations increased, and thus the drops varied substantially from subject to subject.

Regarding the intra-task generalization results (figure 5, right), there was a general decrease in classification accuracy with respect to the tenfold accuracy. For the subtasks of OT1, significant average decreases of 5.99% and 5.58% ($t_{18} = -1.84, p = 0.04$ and $t_{18} = -2.42, p = 0.01$) were identified for non-error and error potentials, respectively.

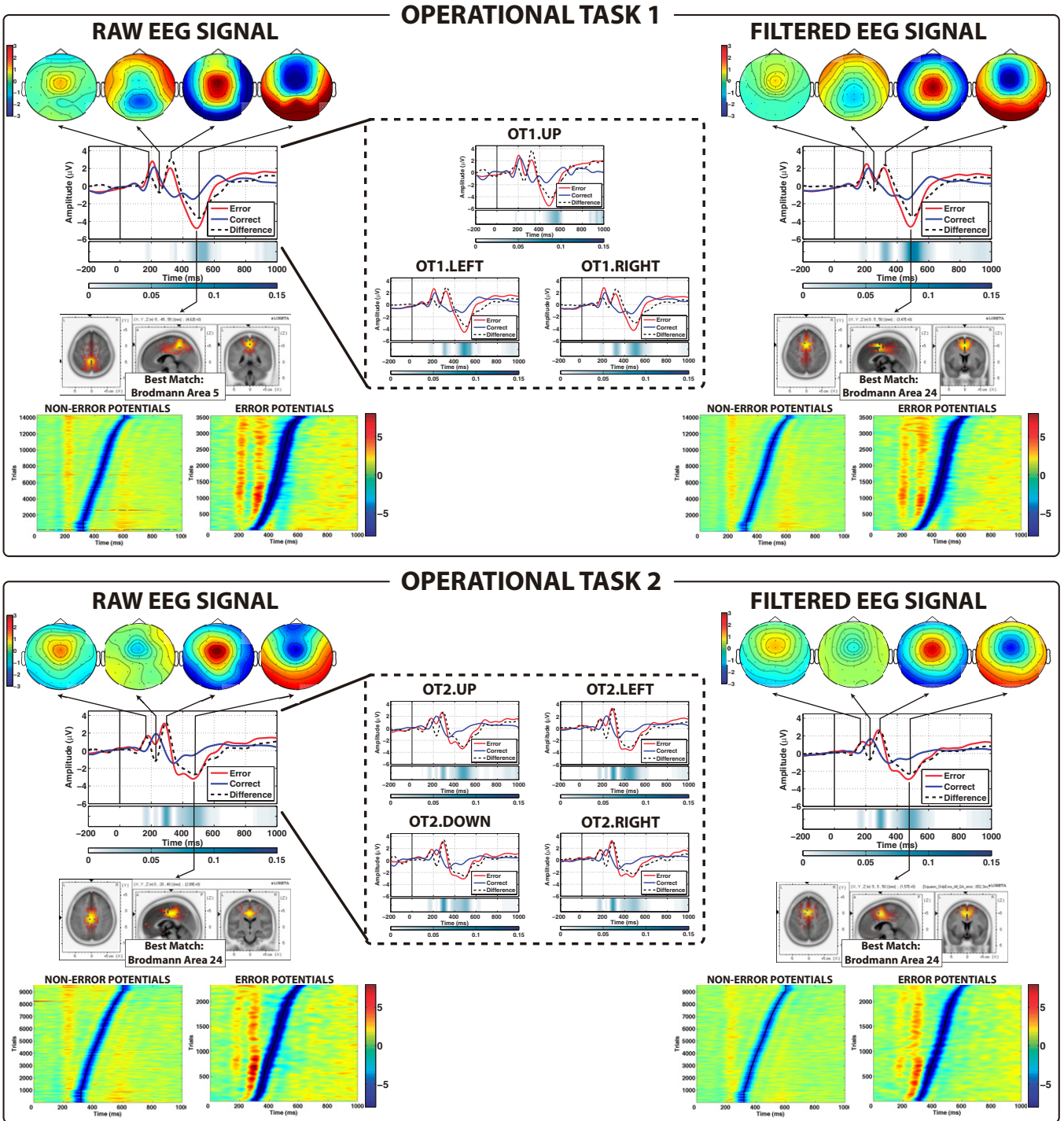


Figure 3. Time-locked grand averaged signals for the raw EEG data (left) and after data filtering (right) on channel FCz (averaged for all subjects) for OT1 (top), OT2 (down) and averages of each subtask for the raw data (centre). The time range is $[-200, 1000]$ ms with 0 being the onset of the action. Error and non-error potentials are in red and blue, respectively, and the difference averages (error minus non-error averages) are in dashed lines. The r^2 discriminability test [1] between error and non-error potentials is below each plot, where dark colours indicate high values (i.e. large differences) between the potentials in both conditions. The spatial location of each peak of the difference average is displayed as topographical scalp maps, as well as the source location of the error grand average at the most prominent negativity (500 and 450 ms for OT1 and OT2). The single-trial potentials sorted by the negative peak latencies are shown below the source localization as a colour encoded image (red and blue indicate amplitudes higher and lower than $0 \mu V$ respectively).

For the subtasks of OT2, average decreases of 1.65% and 5.84% were obtained, but they were not significant ($p > 0.05$) (see the supplementary materials (available at stacks.iop.org/JNE/10/026024/mmedia) for the accuracies obtained for each subject).

3.3.2. Analysis with adaptation. For the operational tasks, the baseline accuracies reached maximum mean accuracies of 81.26% and 78.30% for OT1 and OT2 after 300 trials. For the subtasks, the baseline reached maximum mean accuracies of 81.51% and 77.63% for OT1 and OT2 after 300 trials.

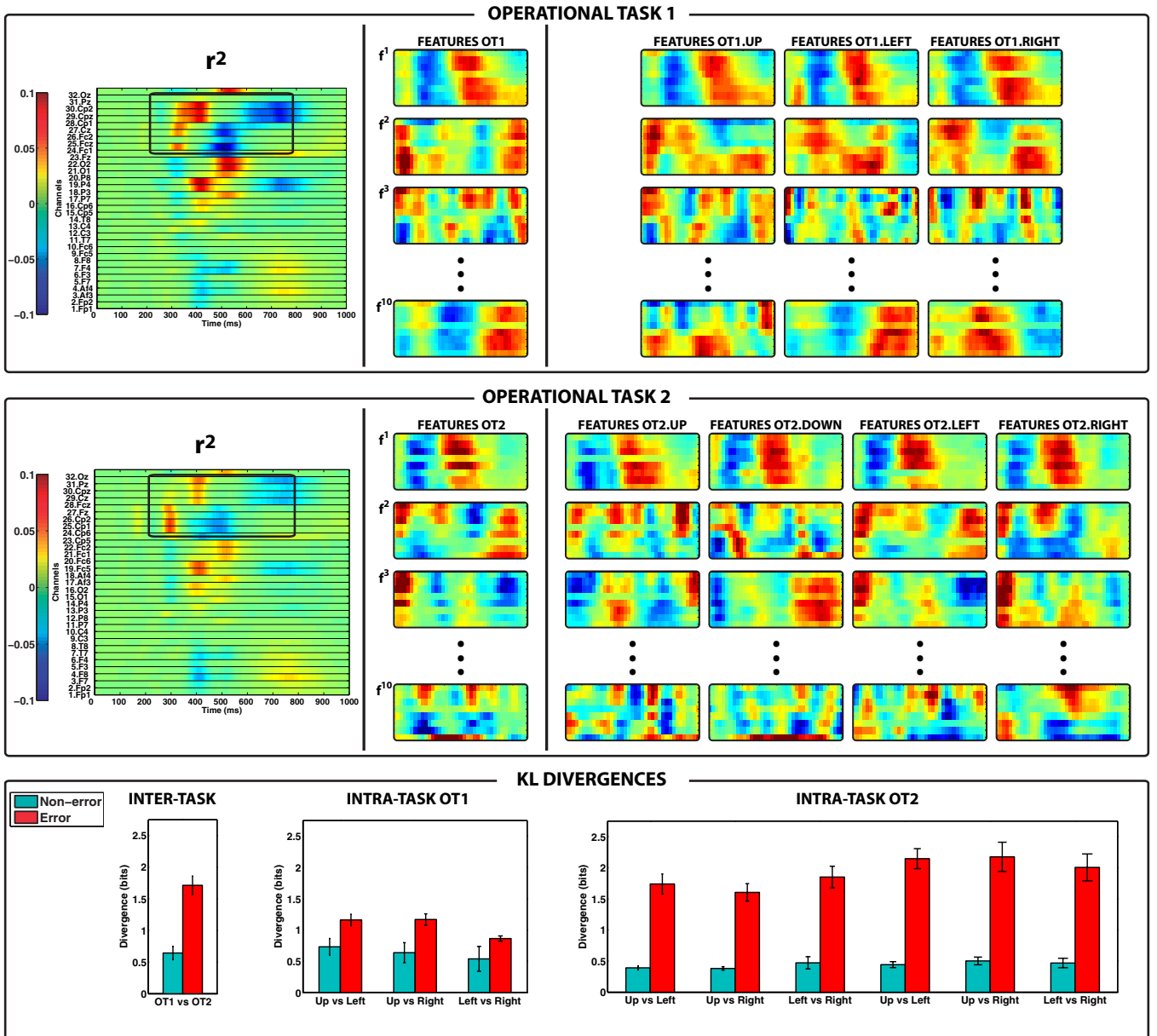


Figure 4. (Top-middle) Representative examples of the feature extraction process for each operational task. The r^2 metric (left) was used to choose the time window of [200, 800] ms in fronto-central channels and then extract the initial features, as in [28]. These features were the inputs to the spatio-temporal filter, whose outputs were the k most ($k = 10$ for the features analysis) discriminant features, each of them encoding combinations of time points and channels. The weights of some features for each task and subtask are shown as a colour encoded image (blue and red indicate negative and positive weights, respectively). (Lower part of figure) Bar plots of the KL divergences (mean \pm SEM) between the feature distributions for the inter-task and intra-task conditions (blue and red for the non-error and error distributions).

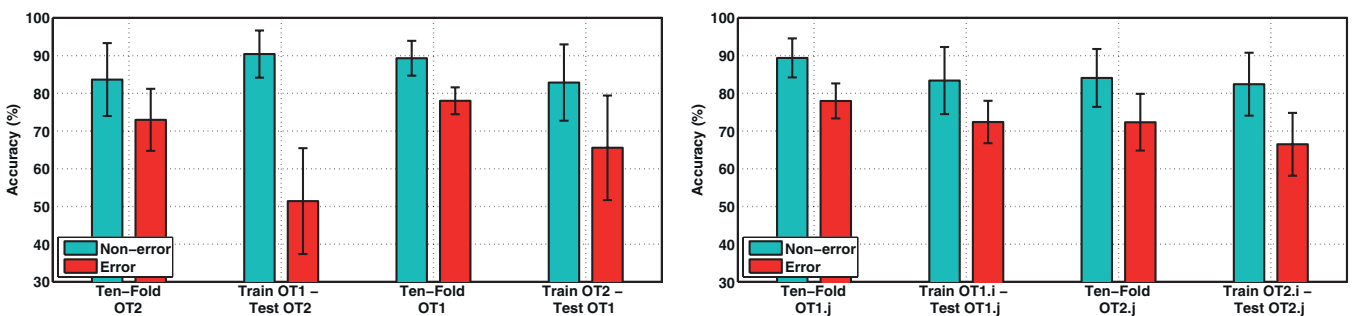


Figure 5. Mean \pm std classification accuracies averaged for all the subjects, for (left) inter-task and their baseline and (right) intra-task and their baseline. Blue and red bars indicate accuracies for non-error and error potentials, respectively.

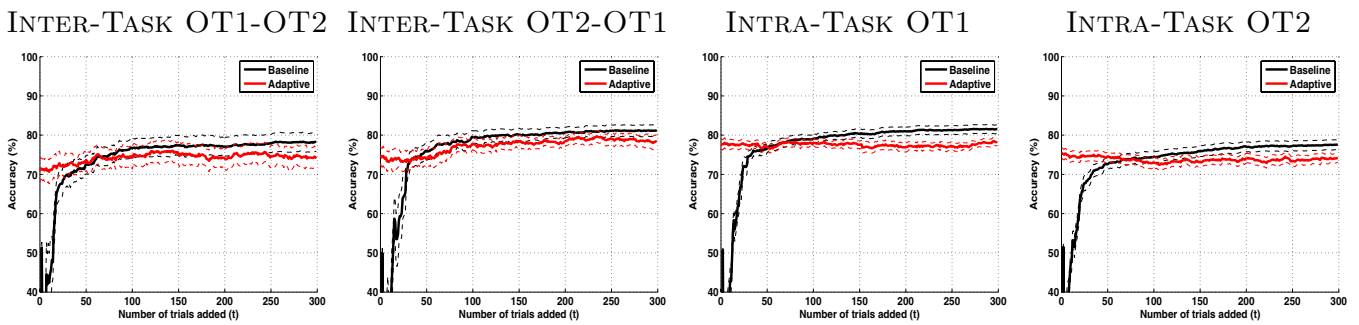


Figure 6. Mean \pm SEM classification (solid \pm dashed lines) accuracies averaged for all subjects, for the baseline and adaptive classifiers across trials. The results are shown for the inter-task (two left columns) and intra-task (two right columns) conditions. For each plot, the x -axis indicates the number of trials used to adapt the previous classifier (adaptive classifier, shown in red lines), or the number of trials used to train the classifier (baseline classifier, shown in black lines), and the y -axis represents the single-trial accuracy.

Note that the accuracy convergence was fast, since only a few examples were needed to reach high accuracies. For instance, the baseline classifier obtained accuracies of 79.49% and 76.63% (tasks) and 79.09% and 74.56% (subtasks) with 100 trials (approximately 5 min of EEG recording, see figure 6, black lines).

The adaptive classifier started with the mean accuracies obtained with the classifiers of the previous subsection (cf figure 5), and as more examples were available, the accuracy of the adaptive classifier increased (see figure 6). After 300 trials (examples), the mean accuracies were 74.07% and 78.33% for the inter-task cases, and 78.26% and 74.22% for the intra-task cases. This increase in performance was due to a reduction in accuracy differences between the two classes, more relevant in the inter-task case.

In the inter-task analysis (training with OT1 and testing with OT2 and vice versa), the adaptive classifier started with better accuracies than the baseline classifier but: (i) after 66 trials and 32 trials, the baseline classifier outperformed the adaptive classifier; and (ii) after 300 trials the adaptive classifier presented accuracies 4.23% and 2.93% lower than the baseline (figure 6, first and second columns). The intra-task analysis showed similar results (figure 6, third and fourth columns). The adaptive classifier presented worse accuracies than the baseline after 51 and 65 trials for OT1 and OT2, respectively, and a lower accuracy compared to the baseline, with drops of 3.25% and 3.41% after 300 trials.

In summary, the supervised adaptation achieved high accuracies from the beginning of the new task/subtask that were improved as more examples were available. The baseline classifier (calibrating from scratch the new task/subtask) started with lower accuracies than supervised adaptation, but rapidly outperformed the latter as the number of examples used to train the baseline classifier increased. In all the situations, less than 100 examples (five minutes of EEG recording) were sufficient to calibrate the BCI to obtain better accuracies than adaptation.

4. Discussion

This paper studied and showed the presence of signal variations in error-related BCIs, where the mental task was coupled with the operational task. The electrophysiology analysis presented

statistical differences mainly in the latencies of the three error-related potential components. There is previous evidence that the error-related potentials remain very similar between different days [24] or even months and years [23]. Thus, the authors understand that the variations of these potentials were due to the difference of operational tasks and not the fact that the experimentation was performed on different days.

The study analysed the raw EEG as well as a filtered EEG signal that retained only EEG information whose origin was estimated in the brain areas involved in the generation of the error-related potentials. This filter played a crucial role in the analysis as, while the grand averages showed that the most prominent activity was due to errors, an r^2 test revealed the presence of activity within subtasks not associated with the error-related potentials, but correlated with the erroneous/non-erroneous actions of the device. This affected the intra-task feature extraction and classification process, in such a way that the artefactual information helped to improve the tenfold intra-task classification (with information not related to the error-processing, e.g. artefacts), but it did not generalize for other subtasks of the same task. For instance, when using raw EEG data from the separated OT1 subtasks, the tenfold classification accuracy was 92.03%/82.86% for non-error and error potentials, respectively, while using the filtered EEG signal the accuracy was 89.38%/77.97%, respectively. However, when generalization of the classifiers was tested on the other subtasks of OT1, the accuracy of the raw EEG dropped 17.00%/13.83% for error and non-error responses, while for filtered data the decrease was only 5.99%/5.58%.

The signal variations affected significantly the distributions of the features selected for classification and the final performance of the trained classifiers. The quantitative analysis (KL divergences) confirmed that the features' differences were greater for error than for non-error responses. When generalizing among operational tasks and subtasks, there was a decrease in accuracy with respect to the baseline, which was more pronounced in the inter-task than in the intra-task and more pronounced for error potentials than for non-error potentials (the decrease was not always symmetric for error and non-error potentials). Further studies might focus on understanding whether this effect was either due to the error and non-error event-related activity or dependent on the

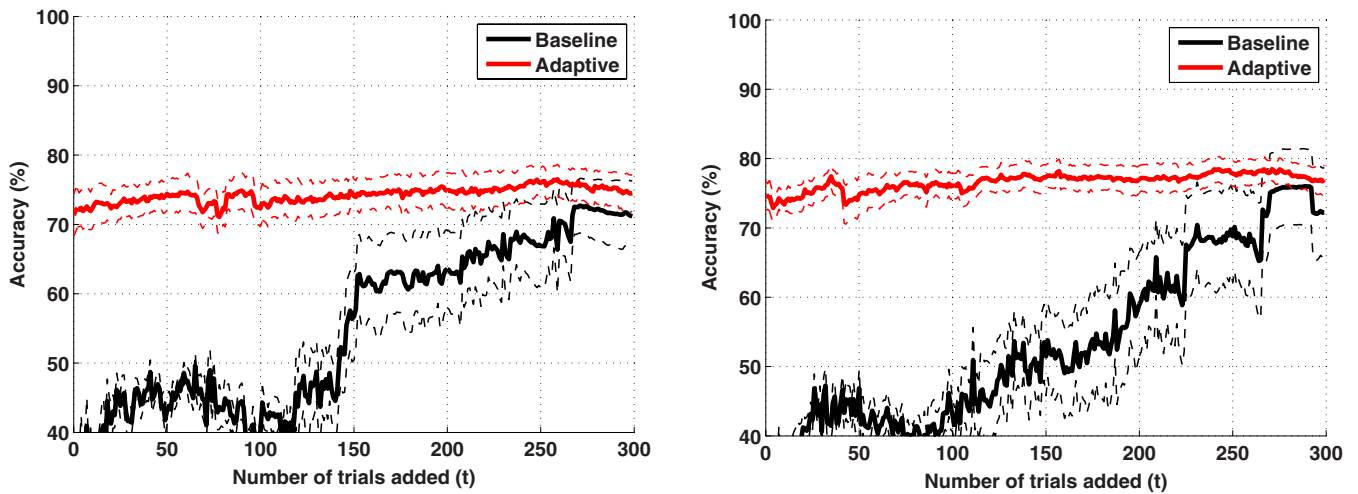


Figure 7. Mean \pm SEM classification (solid \pm dashed lines) accuracies averaged for all subjects, for the baseline and adaptive classifiers across trials using the feature extraction method as in [22, 28, 23], for the inter-task case.

Table 1. Inter-task accuracies (Mean \pm std). Features comparison.

	Tenfold OT1		Tenfold OT2	
	Features as in [22, 28, 23]	Features used in this work	Features as in [22, 28, 23]	Features used in this work
Non-error	90.91 \pm 4.05	89.29 \pm 4.62	86.21 \pm 8.38	83.64 \pm 9.67
Error	79.31 \pm 4.95	78.00 \pm 3.57	73.46 \pm 7.42	72.96 \pm 8.23
	Train OT2 – Test OT1		Train OT1 – Test OT2	
	Features as in [22, 28, 23]	Features used in this work	Features as in [22, 28, 23]	Features used in this work
Non-error	84.43 \pm 9.22	82.85 \pm 10.11	91.72 \pm 5.61	90.40 \pm 6.25
Error	64.39 \pm 11.89	65.56 \pm 13.87	50.87 \pm 15.84	51.42 \pm 14.04

experimental procedure (as the dataset presented an unbalanced number of examples for each class).

Feature analysis and classification results depend on the type of features, which in this study were computed based on a spatio-temporal filter that decorrelated signals to maximize the difference between classes. A possible question that arises is whether such feature extraction introduces differences with respect to the use of EEG amplitudes in the selected channels and time windows, as it is a widespread procedure in these BCIs [22, 28, 23]. Table 1 compares the results obtained using both types of features computed after filtering the EEG for the inter-task generalization (results of the intra-task study were similar). The classification accuracies presented no significant differences neither in the tenfold baseline case nor in the generalization one, with only slight decreases in the accuracies of around 1% when using the proposed features. Nonetheless, in the performed experiments the proposed feature extraction presented an important advantage over the use of EEG amplitudes in terms of calibration time. The use of a lower number of features (an average of 25 features versus 312) reduces substantially the dimensionality of the classifier's hypothesis space, simplifying the learning process and reducing the calibration time. In contrast, for higher dimensional spaces (as with EEG amplitudes with no feature selection), the calibration time is much higher and adaptation pays off. Figure 7 displays, for the EEG amplitudes, the same comparison between adaptation and calibration shown in figure 6 for the inter-task case. The selection of a set

of decorrelated features proposed herein reduces calibration time considerably and achieves a better performance than adaptation in less than 5 min (i.e. less than 100 examples).

In this study, the features were fixed and adaptation was performed only on the classifier. The choice of a different set of features could help reduce the impact of signal variations. For instance, several works have proposed feature extraction methods to find time-invariant features to deal with time-dependent non-stationarities [5], which could be extrapolated to find task-invariant features. However, this would require data from multiple tasks, thus increasing the calibration effort for these BCIs. Other approaches could be used to adapt the classifier to a new task without using labels (i.e. during the device operation), for instance those based on a maximum likelihood estimation of the distribution parameters [49] or predicted labels [50]. Future work by the authors will consider the possibility of combining both paradigms to jointly adapt the classifier and the features for new tasks, as a way to increase the performance of BCIs for new operational tasks while reducing or removing the calibration effort.

Acknowledgments

This work has been partially supported by Spanish Ministry of Education, Culture and Sport by projects HYPER-CSD2009-00067, DPI2009-14732-C02-01, DPI2011-25892 and by DGA-FSE (grupo T04).

References

- [1] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
- [2] Millán J d R et al 2010 Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges *Front. Neurosci.* **4**
- [3] Millán J d R 2004 On the need for on-line learning in brain–computer interfaces *IJCNN: Proc. Int. Joint Conf. on Neural Networks (Budapest, Hungary)* vol 4 pp 2877–82
- [4] Vidaurre C and Blankertz B 2010 Towards a cure for BCI illiteracy *Brain Topogr.* **23** 194–8
- [5] von Büna P, Meinecke F, Király F and Müller K R 2009 Finding stationary subspaces in multivariate time series *Phys. Rev. Lett.* **103** 214101
- [6] Kawanabe M and Vidaurre C 2009 Robust common spatial filters with a maxmin approach *EMBC: Proc. Annu. Int. Conf. of the IEEE Engineering in Medicine and Biology Society (Minnesota, USA)* pp 2470–3
- [7] Perdakis S, Tavella M, Leeb R, Chavarriaga R and Millán J d R 2011 A supervised recalibration protocol for unbiased BCI *5th Int. Brain–Computer Interface Conf.* pp 8–11
- [8] Shenoy P, Krauledat M, Blankertz B, Rao R P N and Müller K R 2006 Towards adaptive classification for BCI *J. Neural Eng.* **3** 13–23
- [9] Vidaurre C, Sannelli C, Müller K R and Blankertz B 2011 Co-adaptive calibration to improve BCI efficiency *J. Neural Eng.* **8** 025009
- [10] Vidaurre C, Kawanabe M, von Büna P, Blankertz B and Müller K R 2011 Toward unsupervised adaptation of LDA for brain–computer interfaces *IEEE Trans. Biomed. Eng.* **58** 587–97
- [11] Birbaumer N, Kubler A, Ghanayim N, Hinterberger T, Perelmouter J, Kaiser J, Iversen I, Kotchoubey B, Neumann N and Flor H 2000 The thought translation device (TTD) for completely paralyzed patients *IEEE Trans. Rehabil. Eng.* **8** 190–3
- [12] Wolpaw J R and McFarland D J 2004 Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans *Proc. Natl Acad. Sci. USA* **101** 17849–54
- [13] Pfurtscheller G, Flotzinger D and Kalcher J 1993 Brain–computer interface: a new communication device for handicapped persons *J. Microcomput. Appl.* **16** 293–9
- [14] Millán J d R, Renkens F, Mouriño J and Gerstner W 2004 Noninvasive brain-actuated control of a mobile robot by human EEG *IEEE Trans. Biomed. Eng.* **51** 1026–33
- [15] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23
- [16] Middendorf M, Mcmillan G, Calhoun G and Jones K S 2000 Brain–computer interfaces based on the steady-state visual-evoked response *IEEE Trans. Rehabil. Eng.* **8** 211–14
- [17] Elbert T, Rockstroh B, Lutzenberger W and Birbaumer N 1980 Biofeedback of slow cortical potentials: I *Electroencephalogr. Clin. Neurophysiol.* **48** 293–301
- [18] Tonin L, Carlson T, Leeb R and Millán J d R 2011 Brain-controlled telepresence robot by motor-disabled people *EMBC: Proc. Annu. Int. Conf. of the IEEE Engineering in Medicine and Biology Society (Boston, USA)* pp 4227–30
- [19] Iturrate I, Antelis J, Kübler A and Minguez J 2009 Non-invasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation *IEEE Trans. Robot.* **25** 614–27
- [20] Escolano C, Antelis J and Minguez J 2012 A telepresence mobile robot controlled with a noninvasive brain–computer interface *IEEE Trans. Syst. Man. Cybern. B* **42** 793–804
- [21] Falkenstein M, Hoormann J, Christ S and Hohnsbein J 2000 ERP components on reaction errors and their functional significance: a tutorial *Biol. Psychol.* **51** 87–107
- [22] Ferrez P W and Millán J d R 2008 Error-related EEG potentials generated during simulated brain–computer interaction *IEEE Trans. Biomed. Eng.* **55** 923–9
- [23] Chavarriaga R and Millán J d R 2010 Learning from EEG error-related potentials in noninvasive brain–computer interfaces *IEEE Trans. Neural Syst. Rehabil. Eng.* **18** 381–8
- [24] López-Larraz E, Iturrate I, Montesano L and Minguez J 2010 Real-time recognition of feedback error-related potentials during a time-estimation task *EMBC: Proc. Annu. Int. Conf. of the IEEE Engineering in Medicine and Biology Society (Buenos Aires, Argentina)* pp 2670–3
- [25] Nieuwenhuis N, Holroyd C B, Mola N and Coles M G H 2004 Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance *Neurosci. Biobehav. Rev.* **28** 441–8
- [26] van Schie H T, Mars R B, Coles M G H and Bekkering H 2004 Modulation of activity in medial frontal and motor cortices during error observation *Nature Neurosci.* **7** 549–54
- [27] Iturrate I, Montesano L and Minguez J 2010 Robot reinforcement learning using EEG-based reward signals *ICRA: Proc. IEEE Int. Conf. on Robotics and Automation (Anchorage, USA)* pp 4822–29
- [28] Iturrate I, Montesano L and Minguez J 2010 Single trial recognition of error-related potentials during observation of robot operation *EMBC: Proc. Annu. Int. Conf. of the IEEE Engineering in Medicine and Biology Society (Buenos Aires, Argentina)* pp 4181–4
- [29] Blankertz B, Dornhege G, Schafer C, Krepki R, Kohlmorgen J, Müller K R, Kunzmann V, Losch F and Curio G 2003 Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis *IEEE Trans. Neural Syst. Rehabil. Eng.* **11** 127–31
- [30] Spüler M, Bensch M, Kleih S, Rosenstiel W, Bogdan M and Kübler A 2012 Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI *Clin. Neurophysiol.* **123** 1328–37
- [31] Schalk G, McFarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 BCI2000: a general-purpose brain–computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43
- [32] Gruendler T O J, Ullsperger M and Huster R J 2011 Event-related potential correlates of performance-monitoring in a lateralized time-estimation task *PloS one* **6** e25591
- [33] Connie C et al 2009 Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, p300, and n400 *Clin. Neurophysiol.* **120** 1883–908
- [34] Delorme A, Palmer J, Onton J, Oostenveld R and Makeig S 2012 Independent EEG sources are dipolar *PloS one* **7** e30135
- [35] Li L, Yao D and Yin G 2009 Spatio-temporal dynamics of visual selective attention identified by a common spatial pattern decomposition method *Brain Res.* **1282** 84–94
- [36] Hyvärinen A, Karhunen J and Oja E 2001 *Independent Component Analysis* (New York: Wiley-Interscience)
- [37] Makeig S, Westerfield M, Jung T P, Enghoff S, Townsend J, Courchesne E and Sejnowski T J 2002 Dynamic brain sources of visual evoked responses *Science* **295** 690–4

- [38] Debener S, Ullsperger M, Siegel M, Fiehler K, Von Cramon D Y and Engel A K 2005 Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring *J. Neurosci.* **25** 11730
- [39] Onton J and Makeig S 2006 Information-based modeling of event-related brain dynamics *Prog. Brain Res.* **159** 99–120
- [40] Fatourechhi M, Bashashati A, Ward R K and Birch G E 2007 EMG and EOG artifacts in brain computer interface systems: a survey *Clin. Neurophysiol.* **118** 480–94
- [41] Xu N, Gao X, Hong B, Miao X, Gao S and Yang F 2004 BCI competition 2003-data set IIB: enhancing P300 wave detection using ICA-based subspace projections for BCI applications *IEEE Trans. Biomed. Eng.* **51** 1067–72
- [42] Himberg J, Hyvärinen A and Esposito F 2004 Validating the independent components of neuroimaging time series via clustering and visualization *NeuroImage* **22** 1214–22
- [43] Diana G and Tommasi C 2002 Cross-validation methods in principal component analysis: a comparison *Stat. Methods Appl.* **11** 71–82
- [44] Pascual-Marqui R D 2002 Standardized low resolution brain electromagnetic tomography (sLORETA): technical details *Methods Find Exp. Clin. Pharmacol.* **24** 5–12
- [45] Luck S J 2005 *An Introduction to the Event-Related Potential Technique* (Cambridge, MA: MIT Press)
- [46] Handy T C 2005 *Event-Related Potentials: A Methods Handbook* (Cambridge, MA: MIT Press)
- [47] Iturrate I, Montesano L, Chavarriaga R, Millán J d R and Minguez J 2011 Spatio-temporal filtering for EEG error related potentials *5th Int. Brain–Computer Interface Conf.* pp 12–5
- [48] Blankertz B, Lemm S, Treder M, Haufe S and Müller K R 2010 Single-trial analysis and classification of ERP components: a tutorial *NeuroImage* **56** 814–25
- [49] Duda R O, Hart P E and Stork D G 2000 Unsupervised learning and clustering *Pattern Classification* (New York: Wiley)
- [50] Lu S, Guan C and Zhang H 2009 Unsupervised brain computer interface based on intersubject information and online adaptation *IEEE Trans. Neural Syst. Rehabil. Eng.* **17** 135–45