

Latency Correction of Error Potentials Between Different Experiments Reduces Calibration Time for Single-Trial Classification

Iñaki Iturrate, Ricardo Chavarriaga, Luis Montesano, Javier Minguez, José del R. Millán

Abstract—One fundamental limitation of EEG-based brain-computer interfaces is the time needed to calibrate the system prior to the detection of signals, due to the wide variety of issues affecting the EEG measurements. For event-related potentials (ERP), one of these sources of variability is the application performed: Protocols with different cognitive workloads might yield to different latencies of the ERPs. In this sense, it is still not clear the effect that these latency variations have on the single-trial classification. This work studies the differences in the latencies of error potentials across three experiments with increasing cognitive workloads. A delay-correction algorithm based on the cross-correlation of the averaged signals is presented, and tested with a single-trial classification of the signals. The results showed that latency variations exist between different protocols, and that it is feasible to re-use data from previous experiments to calibrate a classifier able to detect the signals of a new experiment, thus reducing the calibration time.

I. INTRODUCTION

EEG-based brain-computer interfaces (BCIs) rely on classifiers that are trained during a calibration phase, to build a translation algorithm that transforms EEG features into the control signals for a device (see [1] for a review). One fundamental limitation of current BCI technology is the duration of the calibration phase, as it represents a long period before the usage of the device. This calibration is user- and application-specific due to the wide variety of issues that affect the EEG measurements.

On one hand, for asynchronous BCIs (i.e., those not relying on external cues, usually based on motor imagery of body limbs), these issues are session-dependent –both within one session and among different sessions–, such as the intrinsic non-stationarity nature of the EEG [2], or the motivation of the user [3]; and they are user-dependent, such as the EEG spectral power of specific frequency bands during resting periods [4]. Some studies have designed machine learning techniques to cope with this variability in the EEG signals to either reduce the calibration time [5] or improve the classifier performance [2].

On the other hand, synchronous BCIs (i.e., those relying on external cues, usually based on event-related potentials, ERP) need to deal with other sources of variability in the EEG, which are observed in the amplitude and the latency of the ERP components. For instance, early components (< 200 ms

after the stimulus presentation) of the ERP can be affected by factors such as the spatial attention [6]; the arousal or the valence [7], and the stimuli contrast [8]. In turn, late components (> 200 ms) of the ERP are affected by the probability of appearance of the expected stimulus [8]; the inter-stimulus interval [9]; user-dependent factors such as the age and the cognitive capabilities [10], and cognitive aspects such as the stimulus evaluation time (i.e., the amount of time required to perceive and categorize a stimulus) [8], [11]. Although studies have demonstrated that existing machine learning techniques are rather robust to small variations in the amplitudes of the ERPs [12], it is still not clear the effect that the latency variations of the ERP components have on the single-trial classification. In fact, if the effect of the latency variations of the ERPs is large on the single trial classification, this could explain why in BCI it is always necessary to build a new classifier from scratch for each new application (as the ERPs are highly dependent [6]–[11] on the experiment). We hypothesize that by dealing with the latency variations among ERPs it could be possible to re-use information from previous experiments to train a classifier for a new experiment, thus reducing the calibration time.

This paper describes an algorithm to deal with the latency differences between ERPs of different experiments, which allows for the use of ERPs from one protocol to be used for training classifiers for a new protocol. The algorithm was tested on observation error potentials (ErrP) [12], in three different experiments with increasing cognitive workloads (including an experiment involving a real robotic arm). The results illustrate how the different protocols affect the latency of the potentials. Furthermore, we show that the use of the delay correction allowed for the reduction of the calibration time by using a classifier trained with the potentials from a previous experiment to detect the potentials from a new one.

II. METHODS

A. Data Recording

The instrumentation used to record the EEG was a gTec system with 16 active electrodes located at Fz, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, and CP4 according to the 10/10 international system. The ground and reference were placed on the forehead and the left earlobe, respectively. The EEG was digitized at 256 Hz, power-line notch filtered at 50 Hz, and zero-phase band-pass filtered at $[1, 10]$ Hz. The EEG was recorded by a custom-made C++ application running under Linux. Synchronization of the stimuli onset with the EEG was made with a hardware trigger to assure a good time-event resolution.

Iñaki Iturrate, Luis Montesano and Javier Minguez are with the I3A, DIIS, and Univ. Zaragoza, Spain. Javier Minguez is also with Bit&Brain Technologies SL, Spain. eMail: {iturrate, montesano, jminguez}@unizar.es. Ricardo Chavarriaga and José del R. Millán are with EPFL, Chair in non-invasive brain-computer interface (CNBI), Switzerland. eMail: {ricardo.chavarriaga, jose.millan}@epfl.ch. This work has been supported by Spanish projects HYPER-CSD2009-00067 and DPI2009-14732-C02-01, European project FP7-224631, CAI Programa Europa, and DGA-FSE (grupo T04).

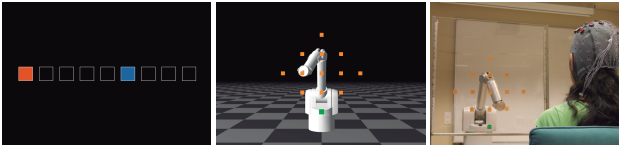


Fig. 1. Experiments performed (from left to right: experiments 1 to 3).

B. Experimental Setup

Six participants (five males and one female, mean age 27.33 ± 2.73 years) performed the experiments, always in the same order as presented below. Each participant executed each experiment in one session of ~ 2.5 hours, and the time elapsed between experiments was 17.58 ± 10.09 days.

Three experimental conditions were designed (see Figure 1) to elicit the error potentials. The experiments presented different setups (and devices) with progressively higher cognitive workloads in order to assess changes in the potentials. The goal of the device was to reach a target by moving along different positions. The device executed random actions with approximately 20% probability of performing an erroneous movement. The time between two actions was random and within the range $[1.7, 4.0]$ s. The target position was randomly changed after 100 actions. The participants were instructed to observe the device movements and evaluate them as correct when they were towards the target position, and as incorrect otherwise, thus eliciting correct and error potentials. The participants were asked to restrict eye movements and blinks to specific resting periods.

1) *Experiment 1, Virtual Moving Square* [12]: This experiment consisted of a one-dimensional space with 9 possible positions (marked by a horizontal grid), a blue square (device) and a red square (target). The device could execute two actions: move one position to the left or to the right. For each subject, approximately 600 trials were acquired.

2) *Experiment 2, Simulated Robotic Arm*: This experiment consisted of a two-dimensional space with 13 possible positions (marked in orange), a simulated robotic arm (Barrett WAM) with 7 degrees of freedom (device) [13] and a green square (target). The robot was situated behind the squares pointing at one position, and could perform four possible actions: moving one position to the left, right, up, or down. The device's movements between two positions were continuous, lasting ~ 500 ms. For each subject, approximately 800 trials were acquired.

3) *Experiment 3, Real Robotic Arm*: This experiment followed the configuration of Experiment 2 but using a real Barret WAM. The user was seated two meters away from the robot. A transparent panel was used to mark the positions, and the distance between two neighbor positions was 15 cm. For each subject, approximately 800 trials were acquired.

C. Delay Correction for Error-Related Potentials

As stated before, different experimental protocols could yield different latencies of the ERP components [8]. The objective of the delay correction algorithm was to remove latency variations of the potentials between two experimental conditions. The algorithm worked as follows: Let P and Q

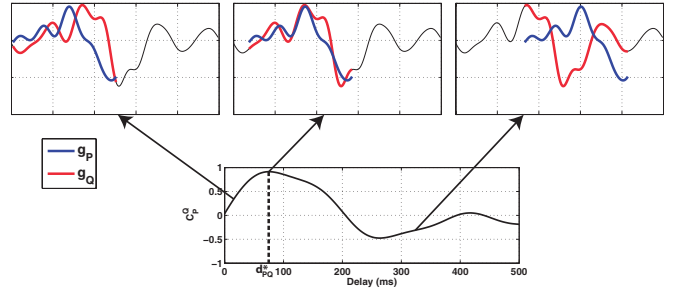


Fig. 2. Delay correction of two difference averages. The cross-correlation function of g_P and g_Q , C_P^Q , is computed by applying different delay shifts to g_Q . The delay d_{PQ}^* is computed as the maximum value of C_P^Q . Three time points of the correlation are shown as an example, where the black line represents g_Q for the window $[0, 1000]$ ms, and the blue and red lines are g_P and g_Q for $x \in [1, T]$ ms, with $T = 500$ ms.

be the datasets from both experimental conditions including trials from both error and correct potentials. Firstly, the averaged signals from both datasets were computed for the error and correct potentials, for channel FCz. Then, the difference average (error minus correct averages) was calculated for each condition, and denoted $g_P(x)$ and $g_Q(x)$. Finally, the cross-correlation of $g_P(x)$ and $g_Q(x)$ was computed as:

$$C_P^Q(d) = \frac{\sum_{x=1}^T [(g_P(x) - \overline{g_P})(g_Q(x+d) - \overline{g_Q})]}{\sqrt{\sum_{x=1}^T (g_P(x) - \overline{g_P})^2} \sqrt{\sum_{x=1}^T (g_Q(x+d) - \overline{g_Q})^2}}$$

where $x \in [1, T]$ is the time range used to perform the correlation, and d the delay. Thus, it was assumed that there was a positive delay between P and Q (i.e., the difference average was elicited earlier on P). Once the correlation was computed, the delay was calculated as $d_{PQ}^* = \max(C_P^Q)$ (note that ideally $d_{PQ}^* = -d_{QP}^*$). Figure 2 shows an example of correlation between two difference grand averages.

D. Feature Extraction

Feature extraction was based on a spatio-temporal filter [14]. The filter input was a dataset including trials from error and correct potentials, and the output were the features. The filter presented the following steps: Firstly, the input EEG data were common-average-reference (CAR) filtered. Then, 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 downsampled to 64 Hz, and concatenated to form a vector of 312 features. The feature vectors of all trials were normalized, and then decorrelated using PCA, retaining 95% of the explained variance. Finally, the k -most discriminant features were selected based on the r^2 metric [15] by a ten-fold cross-validation.

E. Methods for the Single-Trial Classification Study

The objective of the classification study was to analyze whether the delay correction allowed for the use of data from a previous experiment i to train a classifier able to detect data from another experiment j (and thus reducing the calibration time of the experiment j). Single-trial classification was carried out using a linear discriminant analysis

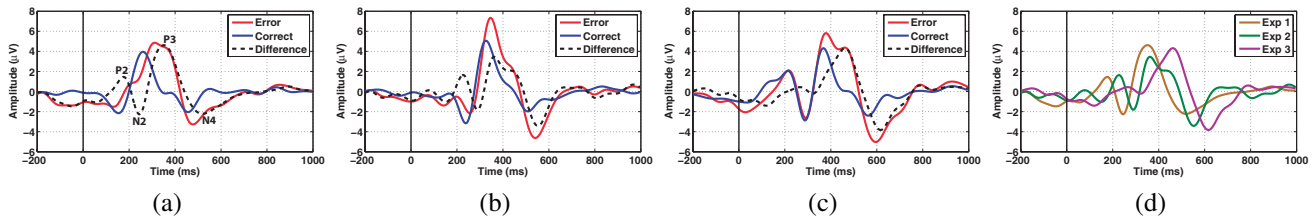


Fig. 3. (a-c) Grand averages in channel FCz, averaged for all subjects for experiments 1 to 3. (d) Difference grand average for experiments 1 to 3. Time 0 ms indicates when the device started the action.

(LDA) [2]. The analysis was performed by comparing three different classifiers. The first one, denoted baseline classifier, followed the standard calibration approach of current BCIs, where the classifier was trained with a subset from the new experiment j (denoted E_j). The classifier was evaluated with training subsets of different sizes (from 10 to 200 trials with increments of 10 trials) to assess the accuracy of the classifier for different calibration times. The second classifier was trained using all data from experiment i (denoted E_i) without correcting the delay, and the same subsets of E_j as in the baseline classifier. The third classifier was trained as the previous one, but using the delay-corrected dataset E_i . To compute the delay d_{E_i, E_j}^* , the window for the cross-correlation function was fixed to $T = 500$ ms. Note that the delay was calculated using only the train datasets E_i and E_j . Then, the classifier was trained with the dataset E_i using the time window $[200, 800] - d_{E_i, E_j}^*$ ms, and the dataset E_j using the original time window of $[200, 800]$ ms (c.f. subsection II-D). These three classifiers were tested on a fixed number of trials (400 trials) from experiment j not included in the training dataset. The combinations of experiments tested were E_1E_2 , E_1E_3 , and E_2E_3 .

III. RESULTS

A. Error Potentials Analysis

Figure 3a-c shows the error, correct and difference grand averaged potentials in channel FCz for the three experiments. The difference grand averages obtained were in agreement with previous studies on ErrPs [12], with two early positive and negative peaks (denoted here P2 and N2, see Fig. 3a), and two broader positive and negative peaks (denoted P3 and N4). Despite the fact that the signals from the three experiments elicited error potentials, the time latency of the difference average changed across experiments (Fig. 3d).

A one-way within-subjects ANOVA was performed to analyze the differences among experiments in the most prominent peaks of the difference average (P3 and N4). For each component (P3 and N4), two separate ANOVAs were computed, with the dependent variable being the peak latency and the peak amplitude. When needed, the Geisser-Greenhouse correction was applied to data to assure sphericity. No statistical differences were found on the peak amplitudes for the P3 component ($F(2, 10) = 0.621, p = 0.557$) or the N4 component ($F(2, 10) = 1.376, p = 0.297$). On the other hand, statistical differences were found for the delays of both the P3 component ($F(2, 10) = 43.845, p = 0.00001$) and the N4 component ($F(1.05, 5.25) = 11.198, p = 0.018$). Thus, the main differences of the error processing activity

among experiments were due to latency variations of the components, while the amplitudes remained similar.

B. Delay Correction and Single-Trial Classification

Figure 4 displays the mean accuracy and the classifier bias (defined as the difference between error and correct accuracies) obtained for the three classifiers, for each experiment combination. The results are averaged across all subjects.

For the E_1E_2 case (Fig. 4 left), the baseline classifier increased as more trials were added to the training set, reaching a mean accuracy of 69.34% after 200 trials. The use of data from experiment 1 without correcting the delay did not improve the accuracy or time, and always produced worse accuracies (62.08% after 200 trials). In contrast, using the delay-corrected data from experiment 1, only 20 trials (around 1 minute of EEG) from experiment 2 were sufficient to reach a better accuracy (70.04%) than the baseline classifier after 200 trials. Furthermore, the accuracies of the delay-corrected classifier were always better than the baseline classifier, reaching 74.60% after 200 trials. Thus, the delay-corrected data allowed for an improvement in both calibration time and classification accuracy. Regarding the bias of the classifiers, the delay-corrected classifier achieved a very low bias even when using only a small amount of data from experiment 2. In contrast, the baseline classifier had a larger bias. After 200 trials, the bias of the delay-corrected classifier was of 10.74% versus a 25.32% for the baseline.

For the E_1E_3 case (Fig. 4 center), the baseline classifier reached an accuracy of 70.88% after 200 trials. Again, when the delay was not corrected, the performance of the classifier was worse, reaching 60.91% after 200 trials. On the other hand, the delay-corrected classifier obtained better accuracies than the baseline classifier for a low number of trials (for instance, after 50 trials, the improvement with respect to the baseline was 5.29%). However, after 200 trials the accuracies of the delay-corrected and baseline classifiers were similar. Thus, the delay-corrected data allowed for an improvement in the calibration time but not in the classification accuracy. Nonetheless, the bias of the delay-corrected classifier was always lower than the bias of the baseline classifier (e.g., 14.86% versus 41.34% for 50 trials).

For the E_2E_3 case (Fig. 4 right), the delay-corrected classifier also outperformed the baseline classifier. After 90 trials, the delay-corrected classifier reached a 69.59% accuracy. In this case, correcting the delay reduced the calibration time to half of the baseline classifier time. However, in this case the best results were obtained when not correcting the delay, indicating that the classifier was not affected by delay

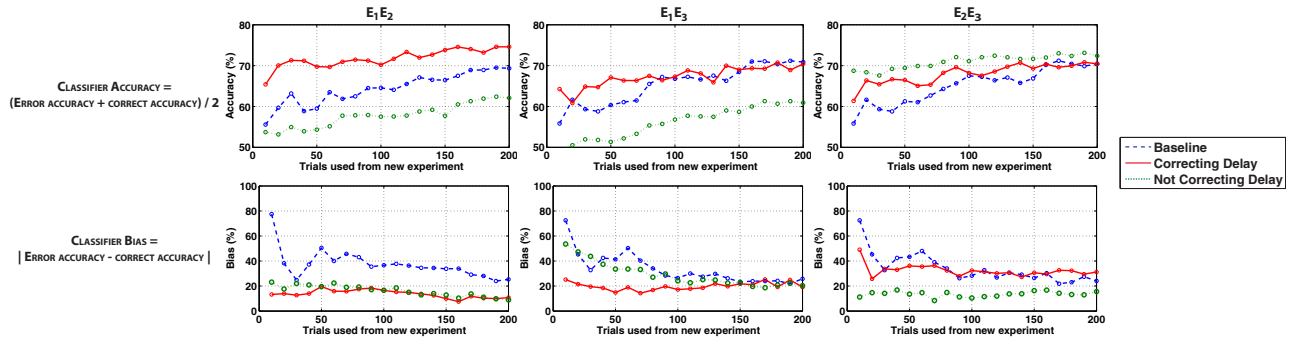


Fig. 4. Mean accuracy (Top) and classifier bias (Bottom) when correcting the delay from E_1E_2 , E_1E_3 and E_2E_3 . The x-axis represents the number of trials used for the E_j dataset. Blue-dashed, green-dotted and red-solid lines represent, respectively, the results for the baseline classifier, the classifier trained without correcting the delay, and the classifier trained when correcting the delay.

differences as in the previous cases. For this case, the bias of the delay-corrected and the baselines classifiers remained similar ($\sim 26\%$) with more than 90 trials.

Finally, Table I shows the estimated delays between experiments when using 200 trials from E_j . The delay between the experiments was always positive, indicating that all subjects needed more time to evaluate the stimulus when moving to more complex experiments. However, there were substantial variations across subjects. For instance, for the E_2E_3 case, the delays ranged from 23.44 ms to 89.84 ms.

TABLE I
DELAY BETWEEN EXPERIMENTS (MS)

	s1	s2	s3	s4	s5	s6	mean
$d_{E_1E_2}^*$	82.03	54.69	31.25	70.31	46.88	89.84	62.50 ± 22.23
$d_{E_1E_3}^*$	152.34	113.28	101.56	121.09	66.41	140.63	115.89 ± 30.43
$d_{E_2E_3}^*$	89.84	62.50	54.69	50.78	23.44	82.03	60.55 ± 23.79

IV. CONCLUSIONS AND FUTURE WORK

ERP-based BCI systems need a calibration phase to train the system due to the large variability of the EEG. A fundamental issue in these systems is the time needed for this calibration phase, since it is necessary to build a classifier from scratch for every new application. This paper has shown that a source of variability of the ERPs are the latency variations among experiments with different cognitive workloads, as confirmed by the statistical analysis (c.f. Section III-A). In this sense, a delay-correction algorithm based on cross-correlation was designed to remove these latency variations, allowing for the re-use of data from previous experiments to reduce the calibration time during a new one.

The proposed technique has been tested with an LDA classifier. The use of delay-corrected potentials always reduced the calibration time and produced a lower bias between the two classes. Furthermore, the technique always achieved similar or better accuracies than the classifier calibrated following the standard procedure. The delay correction was crucial for a successful re-use of data from other experiments, since the results without delay correction were worse than or similar to other approaches.

As future work, the authors plan to test the delay correction algorithm on additional ERPs (such as P300 potentials)

and during online experiments.

V. ACKNOWLEDGEMENTS

The authors would like to thank the participants of the experiments, and M. Tavella, A. Biasucci, S. Dégallier and E. Sauser for their comments and help developing the protocols.

REFERENCES

- [1] J.d.R. Millán, R. Rupp, GR Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, et al., "Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges," *Frontiers in neuroscience*, vol. 4, 2010.
- [2] C. Vidaurre, M. Kawanabe, P. von Büchau, B. Blankertz, and K.R. Müller, "Toward Unsupervised Adaptation of LDA for Brain-Computer Interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 587–597, 2011.
- [3] J.d.R. Millán, "On the need for On-line learning in Brain-Computer Interfaces," in *International Joint Conference on Neural Networks*, 2004, vol. 4, pp. 2877–2882.
- [4] B. Blankertz, C. Sannelli, S. Halder, E. M Hammer, A.a Kübler, K.R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303–9, July 2010.
- [5] C. Vidaurre, C. Sannelli, K.R. Müller, and B. Blankertz, "Co-adaptive calibration to improve BCI efficiency," *Journal of Neural Engineering*, vol. 8, no. 2, pp. 025009, Apr. 2011.
- [6] L. Li, D. Yao, and G. Yin, "Spatio-temporal dynamics of visual selective attention identified by a common spatial pattern decomposition method," *Brain research*, vol. 1282, pp. 84–94, 2009.
- [7] J.K. Olofsson, S. Nordin, H. Sequeira, and J. Polich, "Affective picture processing: an integrative review of ERP findings," *Biological psychology*, vol. 77, no. 3, pp. 247–65, Mar. 2008.
- [8] S.J. Luck, *An introduction to the event-related potential technique*, The MIT Press, 2005.
- [9] E Sellers, D. Krusienski, D. McFarland, T. Vaughan, and J. Wolpaw, "A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance," *Biological psychology*, vol. 73, no. 3, pp. 242–52, Oct. 2006.
- [10] J. Polich, "On the relationship between EEG and P300: Individual differences, aging, and ultradian rhythms," *International journal of psychophysiology*, vol. 26, no. 1-3, pp. 299–317, 1997.
- [11] M. Kutas, G. McCarthy, and E. Donchin, "Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time," *Science*, vol. 197, no. 4305, pp. 792, 1977.
- [12] R. Chavarriaga and J.d.R. Millán, "Learning from EEG error-related potentials in noninvasive brain-computer interfaces," *IEEE Trans. Neural Syst. and Rehab.Eng.*, vol. 18, no. 4, pp. 381–388, 2010.
- [13] E. Sauser, "Robottoolkit," Available online: lasa.epfl.ch/RobotToolKit.
- [14] I. Iturrate, L. Montesano, R. Chavarriaga, J.d.R. Millán, and J. Mínguez, "Spatio-temporal filtering for EEG error related potentials," in *5th Int. Brain-Computer Interface Conference*, 2011.
- [15] I. Iturrate, J. Antelis, A. Kübler, and J. Mínguez, "Non-invasive brain-actuated wheelchair based on a P300 neurophysiological protocol and automated navigation," *IEEE Transactions on Robotics*, vol. 25, no. 3, pp. 614–627, 2009.