Double ErrP Detection for Automatic Error Correction in an ERP-Based BCI Speller

Aniana Cruz, Gabriel Pires, *Member, IEEE*, and Urbano J. Nunes, *Senior Member, IEEE*

***Abstract*— Brain–computer interface (BCI) is a useful device for people with severe motor disabilities. However, due to its low speed and low reliability, BCI still has a very limited application in daily real-world tasks. This paper proposes a P300-based BCI speller combined with a double error-related potential (ErrP) detection to automatically correct erroneous decisions. This novel approach introduces a second error detection to infer whether wrong automatic correction also elicits a second ErrP. Thus, two single-trial responses, instead of one, contribute to the final selection, improving the reliability of error detection. Moreover, to increase error detection, the evoked potential detected as target by the P300 classifier is combined with the evoked error potential at a feature-level. Discriminable error and positive potentials (response to correct feedback) were clearly identified. The proposed approach was tested on nine healthy participants and one tetraplegic participant. The online average accuracy for the first and second ErrPs were 88.4% and 84.8%, respectively. With automatic correction, we achieved an improvement around 5% achieving 89.9% in spelling accuracy for an effective 2.92 symbols/min. The proposed approach revealed that double ErrP detection can improve the reliability and speed of BCI systems.**

***Index Terms*— Double error-related potentials (ErrP), automatic error correction, brain-computer interface (BCI), P300 ERP, electroencephalogram (EEG), speller.**

I. Introduction

BRAIN-COMPUTER interfaces (BCIs) translate subject's intention expressed through brain signals into commands to control external applications such as spelling and robotic devices [1]. BCI is independent of neuromuscular activity, thus it can enhance the quality of life of individuals with severe motor impairments by allowing them to restore communication and control skills [2]–[7]. Efforts have been undertaken to increase the performance of BCI systems by exploiting: (a) different neural signals such as motor imagery [8], steady-state visual evoked potentials (SSVEP) [9], P300 event related potential [6], and slow cortical potentials [10]; (b) different signal processing and classification methods [9], [11], [12]; and (c) different designs of paradigms [6], [13], [14]. Despite the many efforts that have been made to create a reliable and usable BCI, current BCIs still have low transfer rates and low reliability in recognition of subject's intent [15]. Usually, the improvement of the BCI reliability is achieved at the expense of a decrease on transfer rate. Therefore, BCI research strives for ways to improve both simultaneously. Low transfer rate and errors are demotivating factors to the use of BCI [16]. The detection of error-related potentials (ErrPs) has been proposed aiming to improve BCI reliability [17]–[21]. ErrP is an event related potential generated when wrong actions are perceived and was first reported in choice reaction tasks [22], [23]. These studies showed that ErrPs were elicited when incorrect selections were made by the subject. ErrP also known as 'response ErrP' signal is characterized by a negative potential over the fronto-central region that appears between 50 and 100 ms after the incorrect selection, called error negativity (N_E) followed by a positive potential over parietal regions that occurs between 200 and 500 ms after the incorrect response, termed error-positivity (P_E) [22]. Vidal and colleagues used a choice reaction time task, namely a variant of a go/no go task, to verify whether N_E was related to error detection process, and they concluded that N_E might reflect emotional reaction instead of error response, once it can also be observed after a correct response [24]. More recently, Ferrez and Millán described a new kind of ErrP generated when subject interacts with a BCI system, thus they termed it as 'Interaction ErrP'. The Interaction ErrP was observed when the subject realized that the BCI feedback was different from his/her intent, that is, the error was caused by the interface. The main characteristics of this ErrP were a negative peak around 250 ms, a positive peak around 350 ms and a second negative peak around 450 ms after the BCI feedback [17], contrasting with the waveform observed in motor reaction tasks. However, it has been shown that the characteristics of ErrPs, such as amplitude and latency are much dependent of the task [25], [26]. For example, in [25] it was concluded that the amplitude of ErrPs is sensitive to motivation and involvement with the task. Thus, subjects usually exhibit a larger ErrP amplitude when errors represent greater importance.

Manuscript received March 5, 2017; revised July 31, 2017; accepted August 20, 2017. Date of publication September 21, 2017; date of current version January 8, 2018. This work was supported in part by the Ph.D. Scholarship of Aniana Cruz, granted by the Portuguese foundation for science and technology under SFRH/BD/111473/2015, in part by FEDER through COMPETE 2020, Portugal 2020 program under Grant UID/EEA/00048/2013, and in part by the Project AMS-HMI12: RECI/EEI-AUT/0181/2012. *(Corresponding author: Gabriel Pires.)*

A. Cruz is with the Institute for Systems and Robotics, University of Coimbra, 3030-290 Coimbra, Portugal (e-mail: anianabrito@isr.uc.pt).

G. Pires is with the Institute for Systems and Robotics, University of Coimbra, 3030-290 Coimbra, Portugal, and also with the Electrical Engineering Department, Polytechnic Institute of Tomar, 2300-313 Tomar, Portugal (e-mail: gpires@isr.uc.pt).

U. J. Nunes is with the Institute for Systems and Robotics, University of Coimbra, 3030-290 Coimbra, Portugal, and also with the Department of Electrical and Computer Engineering, University of Coimbra, 3030-290 Coimbra, Portugal (e-mail: urbano@isr.uc.pt).

Digital Object Identifier 10.1109/TNSRE.2017.2755018

The idea of using automatic detection of ErrP to improve the performance of BCI has been well accepted by the BCI research community. Automatic ErrP detection has been successfully used in BCIs based on sensorimotor rhythms and on event related potentials (ERP). Chavarriaga *et al.* [27] review the use of ErrP in BCI over the last decade. Here, we review only works using ErrP to improve the performance of P300 BCIs, which are directly related with our approach in this paper. The first research reporting an online P300 speller system with an integrated ErrP-based correction was developed in [28]. Although they demonstrated that it was possible to detect single trial ErrP with an accuracy around 60%, it did not lead to an improvement in the overall classification performance. This outcome is mostly because they did not tune the ErrP classifier to a low false positive rate. False positives imply a decrease in transfer rate, since correctly detected targets become incorrect. Combaz *et al.* [29] tested a standard speller matrix with nine participants to analyze the difference between correct and incorrect feedbacks, and explored the possibility of classifying these feedback responses. They showed that the integration of ErrP classifiers into the P300 Speller system could produce a theoretical (assuming perfect ErrP detection) improvement around 15%. In Schmidt *et al.* [30] the online recognition of ErrP was used in the Center Speller [13] and was tested with twelve participants. They obtained a mean accuracy of single-trial ErrPs around 89% and an increase in the spelling speed of about 49%. Spüler *et al.* [31] observed that young, elderly and motor-impaired individuals (participants with ALS and Duchenne muscular dystrophy) presented similar ErrPs and that the ErrP-based error correction system substantially increased the transfer rate (0.44, 0.73 and 0.35 bit/trial respectively).

While the approaches in previous works used the ErrP detection only to delete the wrong characters, Margaux *et al.* [32] and Zeyl *et al.* [21] extended the concept also allowing automatic error correction. While in [32] the error correction system (ECS) did not show improvement, Zeyl *et al.* [21] improved the selection by 13.67% for 2.54 effective symbols per minute. The contribution of automatic error detection to increase the BCI performance depends on the accuracy of error detection, which must be achieved from a single trial. An accuracy of the error detector lower than the P300 accuracy will lead instead to a degradation of the performance. Therefore, approaches to enhance error detection are essential to make error detection useful.

This paper presents a P300-speller that extends the previous approaches with two contributions:

1) **Double ErrP Detection** - if an ErrP is detected when the spelled symbol is shown to the user, the symbol is deleted and replaced by a new one which has the second highest target score. If this second feedback elicits again an ErrP, the first spelled symbol is re-selected, otherwise the second feedback is chosen. Thus, two single-trial responses contribute to the final selection, resulting in increased reliability in detecting errors;
2) **P300 and ErrP Feature Level Concatenation** - the error detector combines the ERP elicited by the feedback with the ERP detected as target. The rationale is that an erroneous target is usually correlated with the waveform of the ERP selected as target, thereby providing useful information for error detection.

We found that these two contributions offered a positive enhancement of 5% attaining 89.9% accuracy for an effective 2.92 symbols per minute, tested in a P300 speller [6].

II. Methodology

A. EEG Recording

Nine healthy participants (S1-S9) and one tetraplegic participant (P1) with medullar injury (C4/C5 level) with ages between 24 and 43 years old, participated in the study. Five participants (S1, S2, S5, S8 and S9) had previous experience with P300 BCI, and the others had never used a BCI before. The tetraplegic participant has slight hand movements allowing him to interact with the computer and control a powered wheelchair. Each participant signed an informed consent which included the description and purpose of the research, the experimental procedure, the potential risks and the permission to publish the results. EEG was recorded with a g.USBamp bioamplifier, from electrodes Fz, Cz, C3, C4, CPz, Pz, P3, P4, PO7, PO8, POz and Oz according to the international extended 10-20 standard system. The electrodes were referenced to the right ear lobe and the ground was located at AFz. The EEG signals were acquired with active Ag/AgCl electrodes, except for participant S2 who used passive electrodes. The EEG signals were pre-processed using a notch filter at 50 Hz and a band-pass filter with lower cutoff frequency of 1 Hz and a higher cutoff frequency of 10 Hz and sampled at 256 Hz.

B. LSC Speller

The BCI speller used in this study is the lateral single character speller (LSC) [6] designed in our research lab, in which letters flash individually and randomly to obtain an oddball paradigm. LSC minimizes some effects usually occurring in the classical row-column speller paradigm (e.g., distractors, high target probability, low target-to-target interval, eyestrain) (see details in [6]). On average, participants usually reach higher accuracies with LSC than with the classical row-column speller, for the same number of repetitions. For that reason it is usually our preferred BCI-speller. LSC contains the letters of the alphabet, and the 'spc' and 'del' symbol, which are spatially arranged as shown in Fig. 1. Symbols flash individually and the inter-stimulus interval (ISI) coincides with the stimulus duration. Each symbol is highlighted once in each round. The highlight time of the stimuli is 75 ms. The set of all rounds is referred to as a trial. The number of rounds (N_{rep}) within a trial is selected based on the user's performance. Data segments (epochs) of one second are recorded for each event from stimulus onset, and then classified as target or non-target (standard). The symbol detected as target is fed back to the user. The overall time of one trial (*TT*) is defined by

$$TT = N_{rep} \cdot N_s \cdot ISI + CT + ITI \qquad (1)$$

where N_s is the number of symbols, CT is the time associated with the last flash of the trial, and ITI is the inter-trial

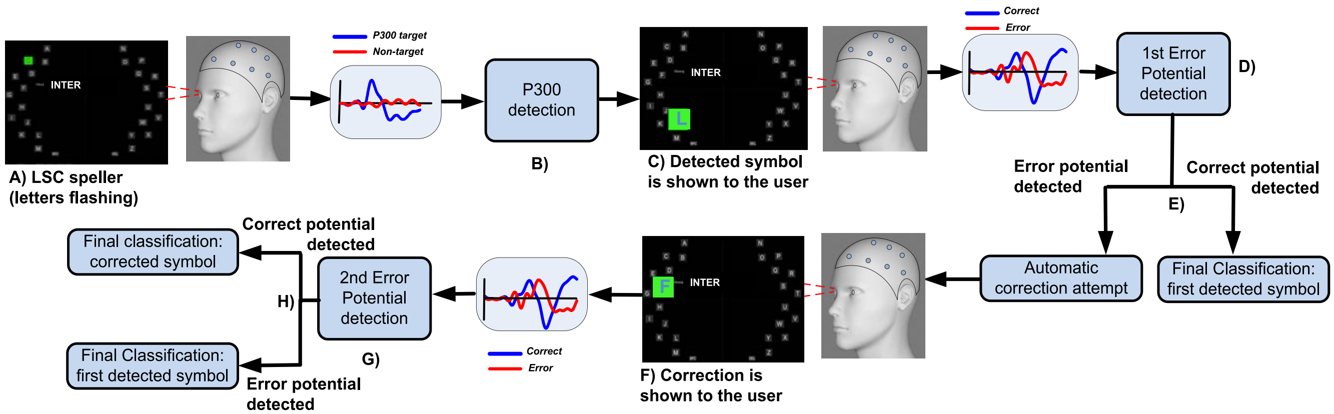

Fig. 1. Example of the P300-ErrP online operation. A) Screenshot of the LSC speller. Letters are flashed according to the oddball paradigm; B) P300 classification is applied to detect the target event; C) Detected letter is shown to the user; D) ErrP classifier is applied; E) If an error potential is detected then an auto-correction is performed; F) The new letter is shown; G) ErrP classifier is applied again; H) If the system detects an ErrP, the letter is changed to the initial P300 detection, otherwise the correction is confirmed. See a demonstrative video in [33].

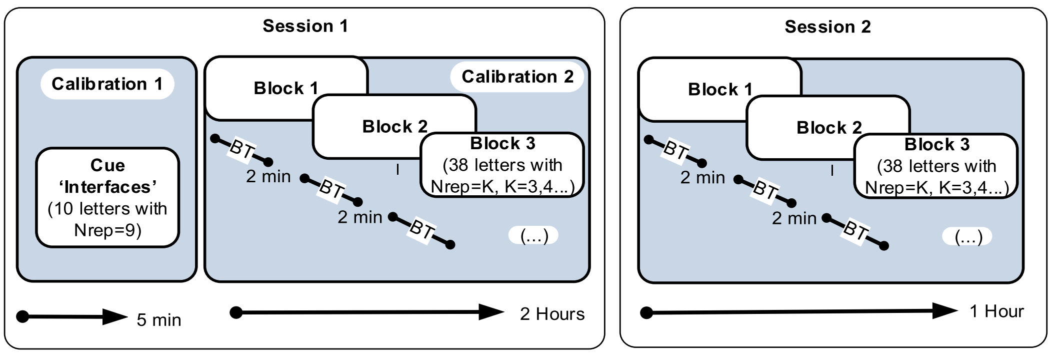

Fig. 2. Schematic representation of data acquisition sessions. Session 1 included calibration 1 to gather data associated to target and standard events (approx. 5 min), and calibration 2 to gather ErrP and correct ERP data from online feedback (approx. 2 hours). Each block corresponded to a sentence of 38 characters, with an interval of 2 min between blocks. In session 2, users wrote the sentences during one hour using the online error detection and correction BCI.

interval ($N_s = 28$, $ISI = 75\ ms$, $CT = 1$ s and $ITI = 4$ s). The feedback procedure of the original LSC speller was slightly adapted to accommodate error detection during online experiments, as explained in section II-B-1. Participants were seated in front of a computer screen adjusted at a distance of about 70 cm. They were instructed to be relaxed and to mentally count the number of target flashes. The experiment comprised two sessions which included three phases as detailed next: ***calibration 1***, ***calibration 2*** and final *online operation*.

1) Data Acquisition for Calibration: In session 1, participants made two calibrations to acquire labeled data associated with target and standard events, and with correct and error feedback responses (see Fig. 2). In ***calibration 1***, participants were asked to attend the 10 letters of the word "INTERFACES" which were successively provided at the center of the screen. For each letter, all symbols flashed 9 times (9 complete rounds), lasting about 5 minutes. The acquired dataset was composed of 90 target epochs and 2430 non-target epochs, which were used to train the P300 classifier to be used in online operation of calibration 2.

Calibration 2 served to acquire labeled data when user received positive (expected symbol) and negative (wrong symbol) feedback returned by the P300 classifier. Participants had to write online several times the Portuguese sentence "ESTOU-A-ESCREVER-COM-UMA-INTERFACE-BCI" (38 characters) without either interruption or correction. This sentence is here referred to as a block. The number of repetitions per trial was adjusted individually so that all participants controlled the LSC speller with similar accuracies. Thus, N_{rep} was selected according to users' performance in ***calibration 1*** (considering a P300 offline accuracy around 90%). The sentence was written several times until a limit of two hours occurred. The duration of each block ($BT - blocktime$) varied according to N_{rep}, and is given by

$$BT = N_c \cdot TT \tag{2}$$

where $N_c = 38$ is the number of spelled characters. Between each block the participants rested two minutes. For example, if $N_{rep} = 5$ each block lasted about 10 minutes, and participants performed 10 blocks. It was expected that a long experiment would lead to a natural increase of errors. The number of errors varied across participants (the minimum number of errors was 31 and the maximum was 86). Labeled data allowed to infer the existence of ErrPs and to train the error detection classifier, as described in section II-C-2.

In session 2, the final P300-ErrP system combined the two classifiers to detect targets and wrong selections. This session was held on a different day of the first session for all participants except S9, who made the two sessions on the same

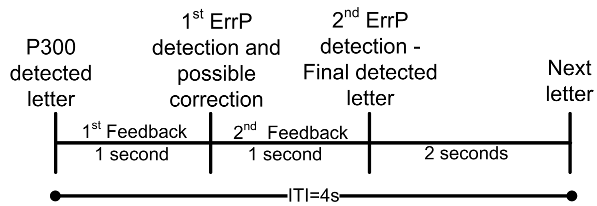

Fig. 3. Temporal diagram of the double feedback ERPs. First, the P300 system detects the target letter, 1 second after occurs the 1st ErrP detection. If an error potential is detected then an auto-correction is made. The 2nd ErrP detection occurs 1 second after. Then, users have 2 seconds to focus the new letter.

day. Two participants (S7 and S8) left the laboratory and took part only in the first phase (session 1) of the experiments. Participants were asked to spell the same sentence as in ***calibration 2***, repeating the sentence during approximately one hour. An example of the online operation of the P300-ErrP BCI is illustrated in Fig. 1 (a video is also provided in [33] for a better understanding). After the P300 classification, the detected letter is visually shown to the user (feedback is provided at the position of the letter). If the system detects an ErrP (feedback of incorrect letter) the system corrects the letter replacing it by the letter classified with the second highest classification score. The new letter is fed back to the user at the respective position. If the system detects a wrong feedback (e.g., a correct letter is changed to a wrong letter), the last letter is changed again to the first detected letter, otherwise the corrected letter is confirmed, the letter is written at the center of the screen, and the spelling system continues to the next trial. The feedbacks occur within the *ITI*, as shown in the temporal diagram of Fig. 3. Users were instructed to keep always focused on the desired letter without moving the eyes regardless of the position of the two feedbacks, until the letter was written at the center. After the letter was presented at the center, participants still had 2 seconds to focus on the next letter.

C. Feature Extraction and Classification

The 12 EEG signals are segmented into 1 second epochs, with onset on each stimulus. At the end of a trial, the P300 classifier predicts which event elicited a P300 ERP. The selected letter is fed back to the user and an epoch of 1 s is recorded immediately after. This single trial epoch is classified by the error detector, which can be applied once or twice as exemplified in Fig. 1.

1) P300 Classification: The epochs of the N_{rep} rounds are averaged for each channel. Then, a statistical spatial filter called Fisher criterion beamformer (FCB), proposed in [12] obtains a single discriminative projection from the 12 EEG channels. Spatial filtering is a common feature extraction technique in EEG-based BCIs that simultaneously allows to increase the signal-to-noise ratio (SNR) and reduce the dimension of the feature data. Considering a spatio-temporal matrix $E_{N\times L}$ representing the epochs of N channels with L time samples ($L = 256$), the spatial filter projection is obtained from $y = w_1^T E$, where w_1 is the optimal spatial filter, obtained from calibration 1, and T denotes the transpose operator (see details in [12]). The 100 most relevant features are selected using the r-square correlation method. Features are then classified by a Bayes classifier, which returns a target probability (score) for each symbol j, P^j, $j \in \{1, \cdots, N_s\}$. Consider S_1 and S_2 the symbols with the highest and second highest probabilities, respectively

$$S_1 \equiv \underset{j\in\{1,\ldots,N_s\}}{\operatorname{argmax}} \; P^j \quad (3)$$

$$S_2 \equiv \underset{j\in\{1,\ldots,N_s\}\setminus S_1}{\operatorname{argmax}} \; P^j. \quad (4)$$

The symbol S_1 is chosen as the primary target event fed back to user, and S_2 is used as a secondary target in case an error is detected.

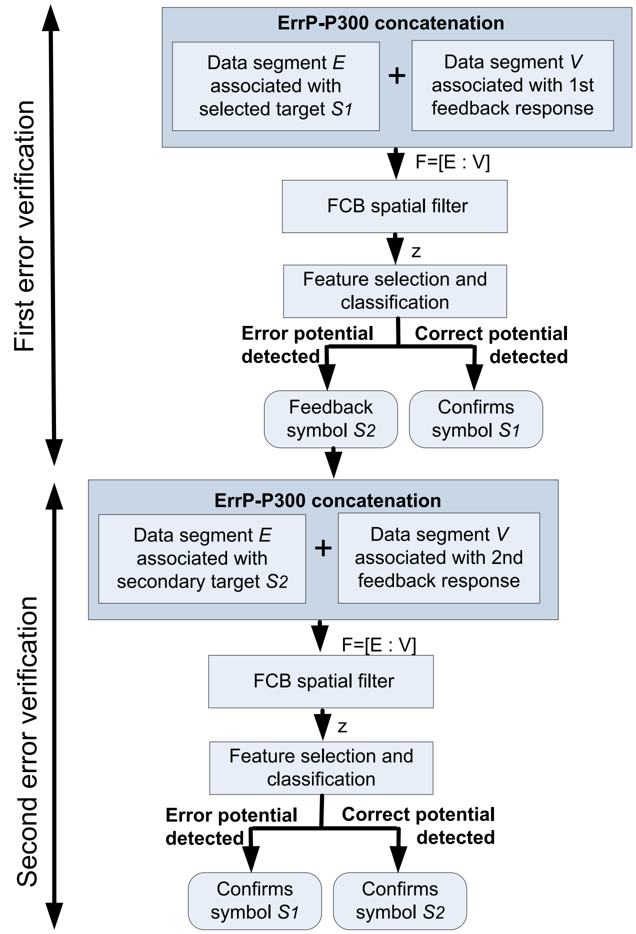

Fig. 4. Online ErrP-P300 classification algorithm with double error verification. First and second error verifications use the single trial response elicited by the feedback concatenated with the EEG data segment of the supposed target event.

2) Error Detection: The error detector uses the same classification framework of the P300 detector. Two approaches were tested, the first one, referred to as ErrP classifier, classifies the EEG data segment $V_{N\times L}$ of the single trial response elicited by the feedback. The second approach, called ErrP-P300 classifier, illustrated in Fig. 4, classifies the EEG data segment $V_{N\times L}$ of the single trial response elicited by the feedback, concatenated with the EEG data segment of the supposed target event ($E_{N\times L}$), resulting in a spatio-temporal matrix $F = [E : V]_{N\times 2L}$. Each vector of F corresponds to one channel with 512 time samples $[e_1 e_2 \ldots e_L v_1 v_2 \ldots v_L]$. It is hypothesized that the concatenation of these two data segments may improve error detection, since they correlate to

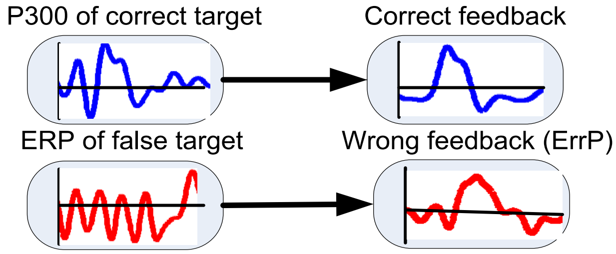

Fig. 5. Expected ERP combinations: correct feedback is related with a typical P300 ERP (true target), and a wrong feedback is related to an ERP with waveform similarities to P300 ERP but still different from this (false target).

each other, thereby helping in the challenging task of single trial classification. It is expected that a correct symbol detected as a true P300 originates a positive feedback, whereas a false target with similarities to the P300 waveform, but still distinct from this one, leads to an ErrP feedback response, as exemplified in Fig. 5. The diagram of the ErrP-P300 classifier is illustrated in Fig. 4. After concatenation, the FCB spatial filter is applied to F, obtaining the projection $z = w_2^T F$, where w_2 is the optimal spatial filter, obtained from ***calibration 2***. The 200 most relevant features selected from z with r-square correlation are classified by a Bayes classifier whose model was obtained in ***calibration 2***. This process is the same in the first and second steps of error verification, as shown in Fig. 4.

3) Double-Check Error Correction: To be effective, the accuracy of the error detection must be higher than the P300 detection, otherwise the false positives will introduce a degradation of the overall performance. Moreover, the error correction rests on the premise that in case of error, the second event with the highest score is the target. This assumption may of course not be true, and therefore a second error feedback verification can avoid a performance degradation due to both wrong correction and wrong detection. The theoretical overall accuracy of the system with error detection and correction, but without using the second feedback is given by

$$\begin{aligned} P_f &= P_{p3} \cdot \frac{TN}{TN+FP} + \left(1-P_{p3}\right) \cdot \frac{TP}{TP+FN} \cdot P_{cor} \\ &= P_{p3} \cdot Spec_1 + \left(1-P_{p3}\right) \cdot Sens_1 \cdot P_{cor} \end{aligned} \quad (5)$$

where $Spec_1$ and $Sens_1$ are the specificity and sensitivity of the first error detector, P_{p3} is the accuracy of the P300 speller (without correction), and P_{cor} is the correction rate of the errors well identified, i.e., those errors that were identified and corrected by selecting target S_2 (4). By introducing the second error checking in the system, it is given the opportunity to eliminate some False Positives, attaining the new overall accuracy

$$P'_f = P_{p3} \cdot \frac{TN + FP \cdot Sens_2}{TN+FP} + \left(1-P_{p3}\right) \cdot Sens_1 \cdot P_{cor} \cdot Spec_2 \quad (6)$$

where $Spec_2$ and $Sens_2$ are the specificity and sensitivity of the second error detector. If an ErrP is evoked when the secondary target is shown, a percentage of the False Positives may be transformed into True Negatives ($FP \times Sens_2$), although some well corrected symbols may be also altered (second term of (6)). Therefore, the second feedback yields an improvement

$$\Delta P = \frac{P_{p3} \cdot FP \cdot Sens_2}{TN+FP} - \left(1-P_{p3}\right) \cdot Sens_1 \cdot P_{cor} \cdot (1-Spec_2). \quad (7)$$

The classification procedure is schematically illustrated in Fig. 4. If an error is detected when the spelled symbol S_1 is fed back to the user, the symbol is deleted and replaced by S_2 which has the second highest target probability. The user reaction to this second symbol will provide a second verification of the existence of an error. If the new target is accompanied by a second error, it is discarded and the final classification corresponds to the first target, S_1. On the other hand, if the new target is accompanied by a non-ErrP, target S_2 is confirmed as the final classification.

D. Performance Metrics

The following metrics were used to evaluate the binary offline data obtained from calibration of the ErrP-P300 system: the sensitivity (*Sens*), specificity (*Spec*) defined respectively as $Sens = TP/(TP + FN)$ and $Spec = TN/(TN + FP)$, and the accuracy (*Acc*), where *TP*, *TN*, *FN* and *FP* are respectively the number of true positives, true negatives, false negatives and false positives.

The online performance was evaluated using the *Sens*, *Spec*, and *Acc*, the information transfer rate (*ITR*), the bandwidth of error correction systems (B_{ECS}) and the effective symbols per minute (*eSPM*), thereby facilitating a direct comparison with state-of-the-art studies. The most common is the *ITR* [34], which represents the maximum capacity of the channel in bits per minute (bpm).

$$ITR = rSPM \cdot \left[log_2((N_s) + p \cdot log_2(p) + (1-p) \cdot log_2 \frac{1-p}{N_s - 1} \right] \quad (8)$$

where p is the online accuracy of the overall system (including error detection and correction) and r*SPM* is the raw rate of symbols per minute defined as *60/TT*, where *TT* is the trial time defined in (1). For BCI systems contemplating error correction, other metrics have been proposed to measure the information transfer rate, and considering the impact of errors to convey error-free messages. We will use the bandwidth B_{ECS}, proposed in [31] and [35] in the context of spelling systems with ErrP detection, defined in bits per trial (bpt)

$$B_{ECS} = log_2 (N_s - 1) \cdot \frac{TN - FN}{TN + FP + TP + FN}. \quad (9)$$

The *eSPM*, i.e., the number of symbols per minute conveyed without errors, is also a straightforward and meaningful metric that encompasses speed and error impact [35]

$$eSPM = rSPM.(2p - 1). \quad (10)$$

III. Results

A. Offline Performance of P300 and ErrP Detection

An offline analysis was performed to infer the efficiency of the extracted features for the binary P300, ErrP and

TABLE I
OFFLINE BINARY CLASSIFICATION RESULTS FOR P300 AND ERRP FEATURES USED SEPARETLY AND CONCATENED. DATA OBTAINED FROM CALIBRATION 2

		Target detection			Error detection					
		P300			**ErrP**			**ErrP-P300**		
		Sens	Spec	Pp3	Sens	Spec	Pep	Sen	Spec	Pp3e
Subjects	**S1**	82,7	83,7	83,2	90,7	86,3	88,5	92	86,9	89,5
	S2	70,9	96,6	83,8	70,9	83,4	77,1	90,9	96,9	93,9
	S3	87,9	96,2	92,0	93,9	92,1	93,0	98,5	94,6	96,6
	S4	81,4	79,6	80,5	93,0	76,8	84,9	93	79,6	86,3
	S5	92,4	86,0	89,2	75,8	88,5	82,1	83,3	91,5	87,4
	S6	93,5	90,2	91,9	71,0	86,8	78,9	83,9	87,2	85,6
	S7	85,9	93,8	89,9	87,3	80,5	83,9	94,4	92,3	93,3
	S8	97,1	100,0	98,5	91,2	94,2	92,7	94,1	100	97,1
	S9	86,9	92,5	89,7	98,4	95,9	97,1	100	97,2	98,6
	P1	76,1	83,4	79,7	59,2	75,2	67,2	91,5	85,4	88,4
Mean		**85,5**	**90,2**	**87,8**	**83,1**	**86,0**	**84,5**	**92,2**	**91,2**	**91,7**
STD		**8,0**	**6,8**	**5,9**	**13,0**	**7,1**	**8,9**	**5,4**	**6,3**	**4,8**

Pp3 is the balanced accuracy of P300 detector, Pep is the balanced accuracy of the error detector using only ErrP features, Pp3e is the balanced accuracy of the error-detector using the concatenation of P300 and ErrP features.

ErrP-P300 classifiers. This analysis was based on the datasets collected during ***calibration 2*** (see Table I). The P300 average accuracy was 87.8%, and the ErrP average accuracy was 84.5%. Concatenating the P300 and ErrP feature vectors in the ErrP-P300 classifier, the error detection accuracy increased about 7% (paired t-test, p = 0.005) over the ErrP classifier. These offline results showed that the feature concatenation improves the error detection as initially hypothesized. These results agree with those reported in [21] and [32], where a different fusion approach, was applied that combines the classification scores of P300 and ErrP classification, instead of the feature combination used in our approach.

B. Online Performance

Session 2 used the P300 and ErrP-P300 classifier models obtained in calibrations of session 1. The online experiments were carried out by 8 participants, since participants S7 and S8 had to leave the laboratory. The number of repetitions per trial was set individually to values between 3 and 7 (average of 5.8 repetitions) according to users' offline performances attained in Session 1 (yielding offline performance around 90%). The online results, obtained after one hour spelling out letters, are presented in Table II. We opted to set the same *ITI* of 4 seconds for both conditions (with and without error correction) to keep the same raw rate of symbols, although a shorter *ITI* could have been considered in the P300-speller without error correction. The online accuracy without error correction (Pre-Acc) was 84.8%. With error detection and correction, the online accuracy (Post-Acc), which included the detection of the first and second ErrP, as well as the error correction, increased the accuracy of 5.1% to 89.9%. The online accuracy was computed as the ratio between the number of correct symbols and the total number of spelled symbols. All participants, except S6 had an improvement, however this participant had the best initial performance (95.8%). Subject S1 had the highest improvement (11%) followed by the subject S9 (10%). The paired t-test show that the improvement between the two conditions is statistically significant (paired t-test, p = 0.003). The *ITR* (3th and 4th columns) and B_{ECS} (5th and 6th columns) were for the Pre-correction condition 12.79 bpm and 3.31 bpt, respectively and for the Post-correction the enhancement was of 1.40 bpm and 0.29 bpt respectively (paired t-test, p = 0.006 and p = 0.002). The average detection of the 1st and 2nd error were 88.4% and 84.8% respectively. The discrepancy of accuracy between the two detections can be explained by the lack of generalization of the classifier, once it was trained with the responses of the 1st feedback. As will be shown in Section III-C, the correct ERP and ErrP waveforms of the 1st feedback responses differ from those of the 2nd feedback, and therefore the classifier model may be overfitted to the 1st feedback responses. In order to evaluate if there was a correlation between the speed of the speller paradigm (measured by the *rSPM*) and the detection accuracy of the ErrP, we calculated the Pearson correlation between them. Its value ($r = 0.04$) shows that the performance of ErrP detection does not correlate with the *rSPM* value, which may suggest that the ErrP detection was not influenced by the BCI speed, as also suggested in [36].

Table III shows the confusion matrix for the 1st and 2nd error detection and its sensitivity and specificity. The mean sensitivity and specificity for the 1st error detector are respectively 91.9% and 88.0%, however, there is some variability across subjects. Subjects S4 and S6 had the highest values of false positives and consequently the lowest specificity. Analyzing the errors for these two subjects in the 2nd error-detector, we verify that they present a high rate of true positives showing that most of the correct targets detected as errors in the 1st error-detector are re-corrected in the 2nd error-detector (confirmed in Table II), emphasizing the importance of the double ErrP detection. The mean sensitivity and specificity for the 2nd error-detector were respectively 85.8% and 73.4%. Fig. 6 compares the effective symbols per minute (10) with and without error correction. For the

TABLE II
ONLINE CLASSIFICATION PERFORMANCE

		Pre-Acc (%)	Post-Acc (%)	Pre-ITR (bpm)	Post-ITR (bpm)	Pre-B_{ECS} (bpt)	Post-B_{ECS} (bpt)	*Pre-eSPM*	*Post-eSPM*	Acc-$ErrP_1$ (%)	Acc-$ErrP_2$ (%)	N_{rep}
Subjects	**S1**	81,6	92,8	9,87	12,45	3,00	3,57	1,92	2,6	90,8	86,1	7
	S2	93,2	94,7	15,96	16,49	4,10	4,20	3,34	3,46	96,3	81,3	5
	S3	93,2	96,3	15,96	17,05	4,10	4,25	3,34	3,59	89,5	90,9	5
	S4	84,7	89,5	11,81	13,03	3,30	3,60	2,37	2,69	75,8	80,0	6
	S5	73,7	79,6	8,30	9,47	2,25	2,66	1,44	1,8	92,1	78,6	7
	S6	95,8	95,8	14,85	14,85	4,35	4,35	3,12	3,12	76,3	92,2	6
	S9	81,1	90,5	17,02	20,73	2,95	3,50	3,3	4,3	96,8	85,0	3
	P1	75,0	79,6	8,55	9,47	2,38	2,63	1,52	1,8	89,5	84,8	7
Mean		**84,8**	**89,9**	**12,79**	**14,19**	**3,31**	**3,60**	**2,54**	**2,92**	**88,4**	**84,8**	**5,80**
STD		**8,5**	**6,8**	**3,58**	**3,90**	**0,81**	**0,70**	**0,83**	**0,9**	**8,1**	**4,9**	**1,40**

Pre-Acc=pre-correction accuracy, Post-Acc=post-correction accuracy, Pre-ITR=pre-correction ITR, Post-ITR=post-correction ITR, pre-B_{ECS}=Pre-correction ITR with ECS, Post-B_{ECS}=post-correction ITR with ECS, Acc-$ErrP_1$ is the accuracy of first error detector, Acc-$ErrP_2$ is the accuracy of second error detector, N_{rep} is the number of repetitions used for P300 detection.

TABLE III
CONFUSION MATRIX OF THE ONLINE CLASSIFICATION FOR THE FIRST AND SECOND ERROR DETECTORS USING ERRP-P300 CLASSIFIER

		First Error detection						Second Error detection					
		TP	FP	FN	TN	Sens	Spec	TP	FP	FN	TN	Sens	Spec
Subjects	**S1**	25	11	3	113	89,3	91,1	13	3	2	18	86,7	85,7
	S2	11	5	2	172	84,6	97,2	10	2	1	3	90,9	60,0
	S3	13	20	0	157	100,0	88,7	22	1	2	8	91,7	88,9
	S4	29	46	0	115	100,0	71,4	40	1	14	20	74,1	95,2
	S5	35	7	5	105	87,5	93,8	22	3	6	11	78,6	78,6
	S6	7	44	1	138	87,5	75,8	46	3	1	1	97,9	25,0
	S9	35	5	1	149	97,2	96,8	14	0	6	20	70,0	100,0
	P1	34	12	4	102	89,5	89,5	32	6	1	7	97,0	53,8
Mean		**23,6**	**18,8**	**2,0**	**131,4**	**91,9**	**88,0**	**24,9**	**2,4**	**4,1**	**11,0**	**85,8**	**73,4**
STD		**11,6**	**16,9**	**1,9**	**26,3**	**6,1**	**9,5**	**13,2**	**1,8**	**4,5**	**7,6**	**10,5**	**25,4**

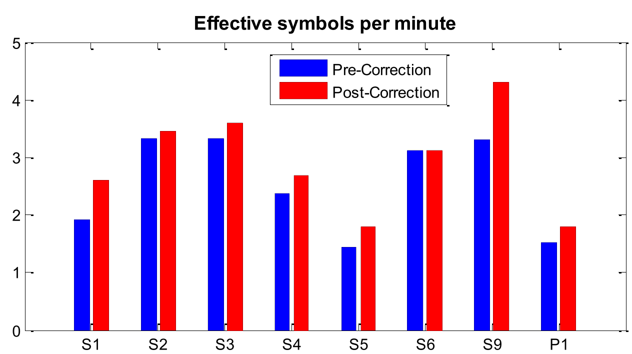

Fig. 6. Effective symbols per minute for each subject in the two conditions: without error correction (blue) and with error correction (red).

Pre-correction condition the average *eSPM* was 2.54 symbols per minute and for Post-correction condition it was 2.92, with a significant increase of 0.38 (paired t-test, p = 0.006). The highest *eSPM* was 4.3 attained by subject S9 who had the greatest increase between the two conditions. It is important to note that the *eSPM* was computed including the 4-seconds *ITI*. The P_{cor} (the correction rate of the errors well identified) was 44%. Replacing P_{cor} in (5) we obtain an improvement of 6.4% (second term). However, since the true negative rate (*Spec*) is not 100%, there are several false positives degrading the overall accuracy. Eq. (6) defines the performance P'_f, obtained after applying the second error detector, which allows to eliminate some of the false positives. The improvement is defined by (7), yielding a value of 8.1% over the value obtained from the first correction in (5).

C. Evoked Potentials After Correct and Wrong Feedbacks

Previous studies [17], [18] demonstrated that the most discriminative brain regions for ErrP are fronto–central channels along the midline. We focused our analysis on channels Fz and Cz. The grand average ERPs of the eight subjects regarding to correct and incorrect feedback are shown in Fig. 7. There is an ERP after the presentation of both error and correct trials. The ErrP after the incorrect feedback has a positive peak around 200 ms followed by a negative peak around 300 ms. A second positive and negative peak appears around 450 ms and 600 ms, respectively. The waveform of the ErrP response is similar to that observed in [17], [21], and [32]. The ERP after correct feedback differs from the ErrP in latency, shape and amplitude. It has only one positive peak that appears about 350 ms followed by a negative peak about 560 ms after the feedback. The correct ERPs were also reported in [32] and [21]. However, in both studies, the waveforms between ErrP and correct ERP were very similar, except for a time lag between ERPs (larger in [32]). So, our results differ

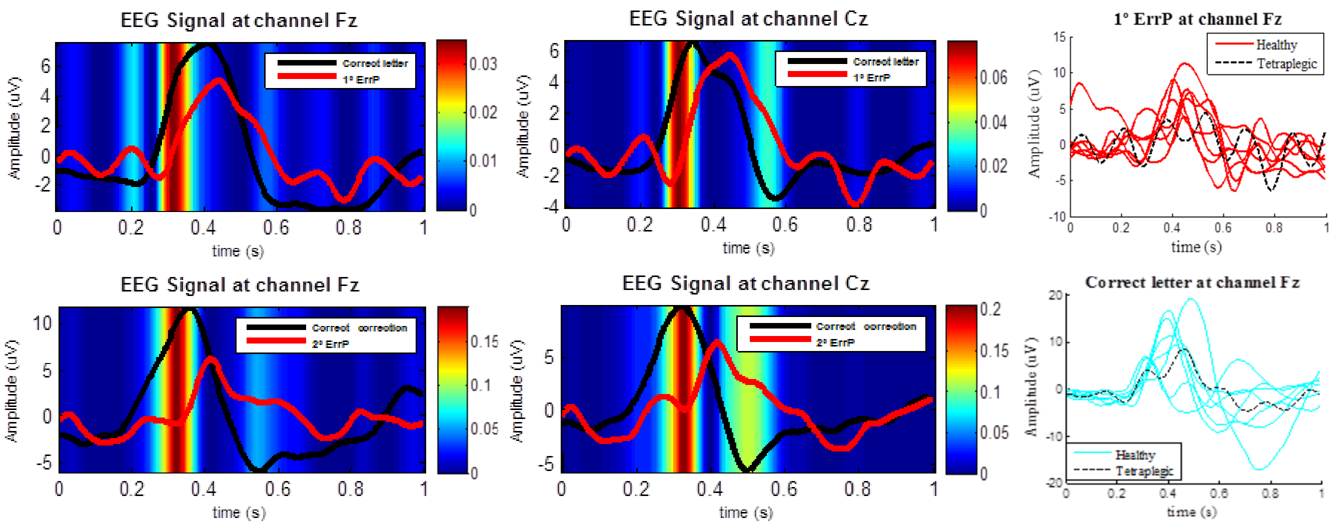

Fig. 7. Grand average of first (top row: left and middle columns) and second (bottom row: left and middle columns) responses for error and correct feedback for the channels Fz and Cz. The background of these plots are the R-square values between correct and incorrect ERPs. The R-square identifies two discriminative components around 260-350 ms and 460-550 ms. Right column (top and bottom row): 1° ErrP and ERP after correct trials from Fz electrode for each healthy participant (solid line) and the tetraplegic participant (dashed line).

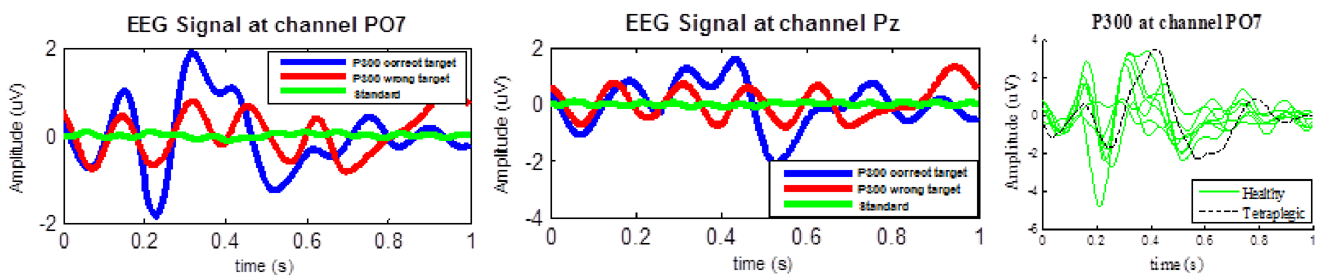

Fig. 8. Grand average of EEG signals for electrodes PO7 (left column) and Pz (middle column) for three situations: correct P300 target, ERP incorrectly detected as target and standard ERPs. Right column: P300 ERP from PO7 electrode for each healthy participant (solid line) and the tetraplegic participant (dashed line).

from previous studies considering that distinct waveforms were evoked for wrong and correct feedback. The 1st and 2nd ErrP waveform slightly differ. There is a latency difference mainly at the first positive peak and second negative peak (a difference of 70 ms and 100 ms respectively). The amplitude of the positive peak of the 1st ErrP is slightly smaller than the 2nd ErrP. On the other hand, the 1st ErrP has a stronger first negative peak amplitude and slighter second negative peak amplitude. The 2nd correct ERP has a greater amplitude than the 1st correct response. The differences between the 1st and 2nd correct ERPs may suggest that the significance of the feedback varies. For example, when the correct letter appears, it is an expected outcome, however, when the feedback is a correction of a wrong letter, it may represent greater relevance to the user since the system is correcting an error. The statistical R-square between ErrP and correct ERP identifies two clear discriminative components around 260-350 ms and 460-550 ms.

The waveform of the standards, true P300 targets and false selected targets are also analyzed. Their grand-averages at channels Pz and PO7 are plotted in Fig. 8. The results show that the waveform associated with the wrong targets is different from that of the standard events, approaching the P300 morphology, but still distinct from this one, as hypothesized from Fig. 5. This finding confirms our expectations and rationale to combine these features to build the error classifier. For comparison of individual waveforms, last column of Fig. 7 and Fig. 8 shows respectively for each participant the waveforms of the ERPs associated with error and correct feedbacks at channel Fz and the waveforms of the P300 ERP at channel PO7.

IV. DISCUSSION

A. Significance of the Online Results

In this paper, we present a new approach based on double ErrP detection for automatic error correction applied in a P300-based BCI speller. The approach shows the possibility of using ErrPs in a closed-loop human-computer interaction, allowing the user to change or confirm system decisions. Therefore, the proposed framework can be useful in many BCI applications beyond BCI spellers, in particular in human-machine collaborative systems, such as robotics, driving assistance or emergency situations. The approach presented here extends our previous LSC paradigm [6] with this additional level of interaction, which can be used in different ways, namely as an ECS to increase the reliability of the system (the purpose of the current study); in the selection of samples to adapt the calibration model during online operation; or even as a secondary communication channel. To accommodate the double ErrP detection, the only change that was made in the LSC paradigm [6] was extending the ITI from 2.5 to 4 seconds, which slightly decreased the raw SPM. To our knowledge, this is the first study using a second ErrP to improve error correction. The online experiments with this approach showed an improvement of 5 percentage points, attaining 89.9%, which is considerable attending that the initial classification accuracy was already high, around 85% (typically, error correction systems have higher improvements in subjects with low BCI performance [31]). The online error detector used the concatenation of the target ERP with the feedback response, an approach whose offline results showed a 7% improvement compared to the error detector using only the feedback response.

Table IV compares the achieved online results with previous studies. We started with the highest initial spelling accuracy of all studies and yet there was an improvement for all subjects, except S6 who maintained the initial score. Dal Seno et al. [28] obtained no improvement, Margaux et al. [32] had a very small gain (only 0.5%). Zeyl *et al.* [21] had the highest online accuracy and the greatest gain in spelling accuracy, but the initial classification accuracy was much lower than in our study. We obtained the highest effective symbols per minute (2.92) and the greatest information transfer rate (14.19 bpm).

The tetraplegic participant (P1) had results similar to able-bodied subjects, and had an improvement around 5% on spelling performance. Thus the system presented here might be a viable alternative for this participant and for other BCI target users. However, a thoroughly experimental analysis is planned as future work in order to validate clinically the approach.

As verified from the online results, the overall improvement of the double error approach relies on many factors, namely the rate of correction P_{cor} and the accuracy of error detection. The specificity of the 1st error detection has an important impact as can be inferred from (5). Despite of an average error detection of 88.4%, the levels of specificity varied across subjects. Those with lower specificity values had a performance degradation after the first feedback correction. The second feedback, allowed to overcome this degradation, by re-correcting the false positives, thus contributing to an overall improvement of 5.1% (Table II). In future versions of the system, the classifier will need to be individually biased for a high specificity.

B. ErrP Morphology

Both correct and wrong feedbacks evoked ERPs, in agreement with previous studies [21], [32], but here we have obtained a greater waveform distinction between them. The R- square analysis showed a strong discrimination ($r^2 = 0.08$ and $r^2 = 0.26$ for the first and second ErrP, respectively) between wrong and correct ERPs, which is of great importance for single trial classification. The 1st and 2nd feedback ErrP responses exhibited slight differences in terms of latency and amplitude. The amplitude of the positive peak of the 2nd correct ERP is greater than that of the 1st, which may be explained by the higher expectation of the user, due to the importance of correcting a wrong letter. The amplitude and latency of ErrPs have been related to user's motivation and workload [25], [26], and it is also known that the amount of attentional resources have an effect in the morphology of ERPs [37].

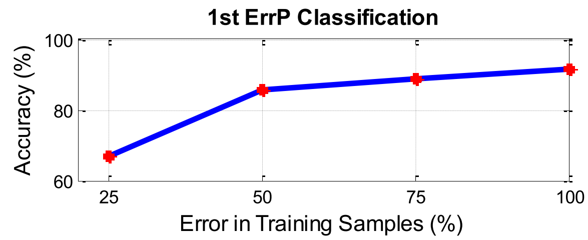

Fig. 9. Classification of 1st ErrP using datasets of **calibration 2** with different percentages of error-trials (25, 50, 75 and 100%).

C. Limitations and Further Improvements

The implementation of a P300 BCI with error detection requires the calibration of the P300 and ErrP classifiers. Although the P300 calibration was quite fast (around 5 minutes), the ErrP calibration was longer so that a sufficient number of error-trials could be acquired under conditions similar to those expected during online operation, i.e., 10 to 20% error-rate and a long operation time. Performing a calibration for each session is something to avoid, since it limits the usability of the BCI system [38], [39]. Therefore we have validated the system using the calibration of the first session. The classifiers were trained in one day and the validation session was conducted on a different day, relying that the P300 and ErrP features remained almost the same between experiments. Although the classifiers presented a good generalization to ERPs variability across the two sessions, a small performance decay occurred, which was expected given that the classification is made at a single trial level (case of ErrP) and for a few number of repetitions (case of P300). The generalization of calibration models across sessions has a major impact on classification performance, particularly relevant when single trial classification is required, and it plays an important role in real-world applications to avoid recalibration procedures and performance loss [36], [40]. Doing the calibration only once mitigates the impact of a long calibration session. Even so, other calibration approaches may be considered in the future to reduce the calibration time, for example, by forcing a higher probability of errors (e.g., decreasing the number of repetitions of target events or using sham errors), thus obtaining a larger amount of error-trials in less time, as reported in [30]. To analyze the effect of the amount of error-trials (directly associated to the calibration time), used for training the calibration model, on the classification performance, we split the calibration set in four subsets of consecutive samples, including respectively 25, 50, 75 and 100% of all error-samples, and we computed offline the classification accuracy using cross-validation. The results depicted in Fig. 9 show that from 50% of the total error-samples, the ErrP classification accuracy approximates the maximum. Therefore, it seems possible to combine several strategies to reduce the calibration time.

To further improve the proposed approach we need also to improve the correction rate (P_{cor}). Other methods beyond the use of the classifier's second best guess can be researched to determine the right target.

TABLE IV
ONLINE RESULTS OBTAINED FROM PREVIOUS BCI SPELLER SYSTEM USING ERROR DETECTION AND OUR PROPOSED SYSTEM

Author	N	ΔBecs	Δacc	Acc	Becs	ITR	eSPM	Pcor
Dal Seno et al. [28]	3	None	-	75.0*	-	7.64*	2.00*	-
Spüller et al. [31]	23	0.52	-	-	2.34	-	-	-
Margaux et al. [32]	16	0.54*	0.5	62.5	1.37*	-	-	34.0
Zeyl et al. [21]	11	-	13.7	92.6	-	8.79*	2.54*	-
This study	**8****	**0.29**	**5.1**	**89.9**	**3.60**	**14.19**	**2.92**	**44.2**

N=number of participants; Acc is the final accuracy (ratio between the number of correct symbol and total number of spelled symbols) denoted as p in eq. (8) and (10) to compute respectively ITR and eSPM; 'Δ 'stands for difference achieved with error correction; *value computed from data provided in paper; **10 participants enrolled the study, but only 8 performed session 2; '-' not reported and unable to compute with provided data.

V. CONCLUSION

A new error detection and correction system based on double ErrP detection was here proposed. Promising results were achieved with a significant increase of online accuracy, information transfer rate, and effective SPM validated with able bodied participants and a tetraplegic participant. Therefore, the integration of error detection into the BCI system might provide a preponderant solution significantly improving BCI reliability to enable its use in clinical settings. Further improvements of the current system could be obtained by enhancing the generalization of the ErrP classifier across sessions, and by researching new approaches to increase the correction rate.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002.

[2] A. Kübler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw, and N. Birbaumer, "Brain–computer communication: Unlocking the locked in.," *Psychol. Bull.*, vol. 127, no. 3, pp. 358–375, May 2001.

[3] J. N. Mak and J. R. Wolpaw, "Clinical applications of brain-computer interfaces: Current state and future prospects," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 187–199, Jan. 2009.

[4] E. W. Sellers, T. M. Vaughan, and J. R. Wolpaw, "A brain-computer interface for long-term independent home use," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 5, pp. 449–455, Oct. 2010.

[5] G. Townsend *et al.*, "A novel P300-based brain–computer interface stimulus presentation paradigm: Moving beyond rows and columns," *Clin. Neurophysiol.*, vol. 121, no. 7, pp. 1109–1120, Jul. 2010.

[6] G. Pires, U. Nunes, and M. Castelo-Branco, "Comparison of a row-column speller vs. a novel lateral single-character speller: Assessment of BCI for severe motor disabled patients," *Clin. Neurophysiol.*, vol. 123, no. 6, pp. 1168–1181, Jun. 2012.

[7] L. Bi, X.-A. Fan, and Y. Liu, "EEG-based brain-controlled mobile robots: A survey," *IEEE Trans. Human-Machine Syst.*, vol. 43, no. 2, pp. 161–176, Mar. 2013.

[8] F. Pichiorri *et al.*, "Brain–computer interface boosts motor imagery practice during stroke recovery," *Ann. Neurol.*, vol. 77, no. 5, pp. 851–865, 2015.

[9] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface," *J. Neural Eng.*, vol. 12, no. 4, p. 046008, 2015.

[10] N. Birbaumer *et al.*, "The thought translation device (TTD) for completely paralyzed patients," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 190–193, Jun. 2000.

[11] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 44, pp. E6058–E6067, 2015.

[12] G. Pires, U. Nunes, and M. Castelo-Branco, "Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis," *J. Neurosci. Methods*, vol. 195, no. 2, pp. 270–281, Feb. 2011.

[13] M. S. Treder, N. M. Schmidt, and B. Blankertz, "Gaze-independent brain–computer interfaces based on covert attention and feature attention," *J. Neural Eng.*, vol. 8, no. 6, p. 66003, Oct. 2011.

[14] T. Kaufmann, S. M. Schulz, A. Köblitz, G. Renner, C. Wessig, and A. Kübler, "Face stimuli effectively prevent brain–computer interface inefficiency in patients with neurodegenerative disease," *Clin. Neurophysiol.*, vol. 124, no. 5, pp. 893–900, 2013.

[15] A. Kübler, E. M. Holz, E. W. Sellers, and T. M. Vaughan, "Toward independent home use of brain-computer interfaces: A decision algorithm for selection of potential end-users," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 3, pp. S27–S32, 2015.

[16] C. Zickler *et al.*, "A brain-computer interface as input channel for a standard assistive technology software," *Clin. EEG Neurosci.*, vol. 42, no. 4, pp. 236–244, Oct. 2011.

[17] P. W. Ferrez and J. del R. Millán, "Error-related EEG potentials generated during simulated brain–computer interaction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 923–929, Mar. 2008.

[18] G. Schalk, J. R. Wolpaw, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: Presence of an error potential," *Clin. Neurophysiol.*, vol. 111, no. 12, pp. 2138–2144, 2000.

[19] T. Zeyl, E. Yin, M. Keightley, and T. Chau, "Improving bit rate in an auditory BCI: Exploiting error-related potentials," *Brain-Comput. Interfaces*, vol. 3, no. 2, pp. 75–87, 2016.

[20] T. Zeyl, E. Yin, M. Keightley, and T. Chau, "Partially supervised P300 speller adaptation for eventual stimulus timing optimization: Target confidence is superior to error-related potential score as an uncertain label," *J. Neural Eng.*, vol. 13, no. 2, p. 026008, 2016.

[21] T. Zeyl, E. Yin, M. Keightley, and T. Chau, "Adding real-time Bayesian ranks to error-related potential scores improves error detection and auto-correction in a P300 speller," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 46–56, Jan. 2016.

[22] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial," *Biol. Psychol.*, vol. 51, nos. 2–3, pp. 87–107, 2000.

[23] W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, and E. Donchin, "A neural system for error detection and compensation," *Psychol. Sci.*, vol. 4, no. 6, pp. 385–390, 1993.

[24] F. Vidal, T. Hasbroucq, J. Grapperon, and M. Bonnet, "Is the 'error negativity' specific to errors?" *Biol. Psychol.*, vol. 51, nos. 2–3, pp. 109–128, 2000.

[25] G. Hajcak, J. S. Moser, N. Yeung, and R. F. Simons, "On the ERN and the significance of errors," *Psychophysiology*, vol. 42, no. 2, pp. 151–160, 2005.

[26] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. del R. Millán, "Latency correction of event-related potentials between different experimental protocols," *J. Neural Eng.*, vol. 11, no. 3, p. 036005, 2014.

[27] R. Chavarriaga, A. Sobolewski, and J. del R. Millán, "Errare machinale est: The use of error-related potentials in brain-machine interfaces," *Front. Neurosci.*, vol. 8, Jul. 2014, Art. no. 208.

[28] B. Dal Seno, M. Matteucci, and L. Mainardi, "Online detection of P300 and error potentials in a BCI speller," *Comput. Intell. Neurosci.*, vol. 2010, Nov. 2010, Art. no. 307254.

[29] A. Combaz, N. Chumerin, N. V. Manyakov, A. Robben, J. A. K. Suykens, and M. M. Van Hulle, "Towards the detection of error-related potentials and its integration in the context of a P300 speller brain–computer interface," *Neurocomputing*, vol. 80, pp. 73–82, Mar. 2012.

[30] N. M. Schmidt, B. Blankertz, and M. S. Treder, "Online detection of error-related potentials boosts the performance of mental typewriters," *BMC Neurosci.*, vol. 13, no. 1, p. 19, 2012.

[31] 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, 2012.

[32] P. Margaux, M. Emmanuel, D. Sébastien, B. Olivier, and M. Jérémie, "Objective and subjective evaluation of online error correction during P300-based spelling," *Adv. Hum.-Comput. Interact.*, vol. 2012, Oct. 2012, Art. no. 578295.

[33] *LSC-Double ErrP Detection Video*. Accessed: Sep. 1, 2017. [Online]. Available: http://home.isr.uc.pt/~gpires/ErrP/

[34] J. R. Wolpaw *et al.*, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 73–164, Jun. 2000.

[35] B. D. Seno, M. Matteucci, and L. T. Mainardi, "The utility metric: A novel method to assess the overall performance of discrete brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 1, pp. 20–28, Feb. 2010.

[36] F. Iwane, R. Chavarriaga, I. Iturrate, and J. del R. Millán, "Spatial filters yield stable features for error-related potentials across conditions," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 000661–000666.

[37] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007.

[38] F. Lotte, "Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces," *Proc. IEEE*, vol. 103, no. 6, pp. 871–890, Jun. 2015.

[39] D. Wu, "Online and offline domain adaptation for reducing BCI calibration effort," *IEEE Trans. Hum.-Mach. Syst.*, vol. 47, no. 4, pp. 550–563, Aug. 2017.

[40] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Müller, "Single-trial analysis and classification of ERP components—A tutorial," *Neuroimage*, vol. 56, no. 2, pp. 814–825, 2011.