

## Research Article

# Objective and Subjective Evaluation of Online Error Correction during P300-Based Spelling

**Perrin Margaux,<sup>1,2</sup> Maby Emmanuel,<sup>1,2</sup> Daligault Sébastien,<sup>3</sup>  
Bertrand Olivier,<sup>1,2</sup> and Mattout Jérémie<sup>1,2</sup>**

<sup>1</sup>Lyon Neuroscience Research Center (CRNL), INSERM, CNRS, Dycog team, 95 Boulevard Pinel, 69500 Bron, France

<sup>2</sup>Université Claude Bernard Lyon 1, 69000 Lyon, France

<sup>3</sup>CERMEP, 95 Boulevard Pinel, 69500 Bron, France

Correspondence should be addressed to Mattout Jérémie, jeremie.mattout@inserm.fr

Received 6 July 2012; Revised 15 October 2012; Accepted 24 October 2012

Academic Editor: Surjo R. Soekadar

Copyright © 2012 Perrin Margaux et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Error potentials (ErrP) are alterations of EEG traces following the subject's perception of erroneous feedbacks. They provide a way to recognize misinterpreted commands in brain-computer interfaces (BCI). However, this has been evaluated online in only a couple of studies and mostly with very few subjects. In this study, we implemented a P300-based BCI, including not only online error detection but also, for the first time, automatic correction. We evaluated it in 16 healthy volunteers. Whenever an error was detected, a new decision was made based on the second best guess of a probabilistic classifier. At the group level, correction did neither improve nor deteriorate spelling accuracy. However, automatic correction yielded a higher bit rate than a respelling strategy. Furthermore, the fine examination of interindividual differences in the efficiency of error correction and spelling clearly distinguished between two groups who differed according to individual specificity in ErrP detection. The high specificity group had larger evoked responses and made fewer errors which were corrected more efficiently, yielding a 4% improvement in spelling accuracy and a higher bit rate. Altogether, our results suggest that the more the subject is engaged into the task, the more useful and well accepted the automatic error correction.

## 1. Introduction

A brain-computer interface (BCI) is a system that connects the brain to a computer directly and avoids the need for peripheral nerve and muscle activities to execute user's actions. A major aim of BCI research is to allow patients with severe motor disabilities to regain autonomy and communication abilities [1]. This raises the crucial challenge of achieving a reliable control by measuring and interpreting brain activity on the fly. Due to the highly complex, noisy, and variable nature of brain signals, especially those obtained with noninvasive recordings using scalp EEG, the computer sometimes misinterprets the signals and makes a decision that does not match the user's intention. In this context, it is highly relevant to look for a way to detect and correct errors. One way to tackle this issue is to appeal to the hybrid BCI approach [2], where it has been shown that BCI

performance could be improved by supplementing the first-order brain signal with second-level information to aid the primary classifier and to improve the final decision or BCI output [3]. This complementary signal can be either of a cerebral origin or of a very different nature [2].

Along that line, a couple of recent studies have proposed to use error-related brain signals in BCI applications. It has been established for years that the brain produces specific evoked responses in case of errors. The error-related negativity (ERN or Ne, [4, 5]) and error positivity [6, 7] are phase locked to the motor response in alternative forced-choice tasks, whereas the feedback-related negativity (FRN or feedback ERN) is produced in response to negative feedbacks [8] (see [9] for a review on ERN and FRN). In a human-computer interface, the ERN has already been used to detect errors online [10]. In BCI, an FRN like signal has been observed in response to erroneous feedback [11]. Ferrez and

del R. Millan used the term “interaction error potential,” as the decision results from the interaction between the user and the machine. Indeed, either one or both may be responsible for the error [12].

In particular, the well-known P300-speller has been used to compare classifiers in their performance to discriminate between correct and incorrect trials [13–15], in order to achieve real-time automatic error detection. In [15], the authors compared several classifiers offline, using a three-fold cross-validation procedure. The best classification was obtained with an LDA applied on polynomial coefficients. It was then used to evaluate the putative improvement due to ErrP detection, by estimating the ensuing reduction in the number of trials needed to properly spell a letter. They found that four out of five subjects could benefit from error detection, provided that the accuracy of the P300-Speller would remain below 75%. The same group was the first to test online automatic error detection in a P300-Speller BCI. The ErrP was detected with a 68% specificity (the probability of detecting a correct trial) and a 62% sensitivity (the probability of detecting an error trial), in two out of three subjects. However, this was not sufficient to improve the information transfer rate [13]. Finally, in a recent P300-Speller study, healthy and motor-impaired participants increased their bit-rate by 0.52 (in bits/trial) using online error detection during copy spelling [14]. However, these studies implemented and evaluated automatic error detection but not automatic error correction. In other words, they could eventually suppress a wrong letter by detecting the ensuing ErrP but did not attempt to immediately replace this letter by another highly probable one [13, 14]. In the current study, we evaluated both error detection and correction, where correction was based on the second best guess of a probabilistic classifier.

In a previous experiment, we tested automatic error detection offline [16]. We achieved a very high specificity (above 0.9) and a fairly good sensitivity (up to 0.6), which yielded a significant improvement in offline spelling accuracy in about half of the participants considering an automatic correction based on the second best guess of the classifier. Importantly, it turned out that, for about 50% of the error trials, the second best guess of the classifier corresponded to the true target. Interestingly, this good correction rate (GCR) correlated with the spelling accuracy over subjects, suggesting that more attentive subjects would produce more distinguishable feedback response signals and should be more prone to benefit from automatic correction.

In the current study, we implemented and tested automatic error correction online, in a fairly large group of subjects. To ensure a high error rate in most participants, we used settings that render the spelling fast and challenging. We also evaluated subjective perception of error correction in each participant by means of a questionnaire. Importantly, the group would clearly split in-between participants with low and high ErrP detection specificity. We thus also report and discuss the difference between those two groups, both at the psychological and neurophysiological levels.

The paper is organized as follows. First, Section 2 describes the experimental design in details, including the online OpenViBE scenario that we implemented for spelling

and error detection and correction. It also describes the evaluation procedure we used to analyze data offline. Results are exposed in the next section and discussed in the final section of the paper.

## 2. Methods

*2.1. General Principle of the P300-Speller.* The P300-Speller is a BCI paradigm developed to restore communication in locked-in patients [17]. The P300 signal is an EEG positive deflection that occurs approximately 300 ms after stimulus onset and is typically recorded over centro-parietal electrodes. This response is evoked by attention to rare stimuli in a random series of stimulus events (the oddball paradigm) [18] and is even stronger when the subject is instructed to count the rare stimuli [19]. It can be used to select items displayed on a computer screen [17, 20]. In practice, all possible items are displayed while the user focuses his attention (and gaze) onto the target item. Groups of items are successively and repeatedly flashed, but only the group that contains the target will elicit a P300 response. Correct spelling thus relies on both the user’s attentional state and the ability for the BCI to detect the P300 response online.

*2.2. Participants.* Sixteen healthy subjects took part in this study (7 men, mean age =  $28.2 \pm 5.1$  (SD), range 20–37). They all signed an informed consent approved by the local Ethical Committee and received monetary compensation for their participation. All participants reported normal or corrected-to-normal vision. All subjects had no previous experience with the P300-Speller paradigm or any other BCI application.

### 2.3. Experimental Design

*2.3.1. Setup.* Participant’s brain activity was simultaneously recorded with 56 passive Ag/AgCl EEG sensors (VSM-CTF compatible system) and 275 MEG axial gradiometers (VSM-CTF Omega 275). However, only 32 EEG electrodes were used for the online and offline analysis reported in this paper. The EEG electrode placement followed the extended 10–20 system (see Figure 1). Their signals were all referenced to the nose. The ground electrode was placed on the shoulder and impedances were kept below 10 k $\Omega$ . Signals were sampled at 600 Hz.

*2.3.2. Stimuli.* We used a standard  $6 \times 6$  matrix of items for stimulation [17], which we combined with our own implementation of the P300-speller in OpenViBE [21, 22]. We moved away from the traditional row and column way of flashing items, by adapting the pseudorandom stimulation procedure proposed by Townsend et al. [23]. Townsend and collaborators showed that this improves the spelling accuracy significantly since it minimizes the risk for the subject or patient to be distracted by the letters adjacent to the target. Therefore, we constructed six pairs of groups of six letters, each pair being associated with a particular item of the matrix. This item then belongs to two groups

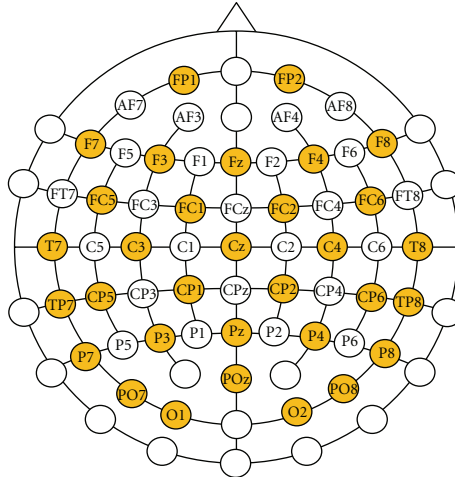


FIGURE 1: EEG channel montage. Only yellow channels were used for the online and offline analysis reported in this paper.

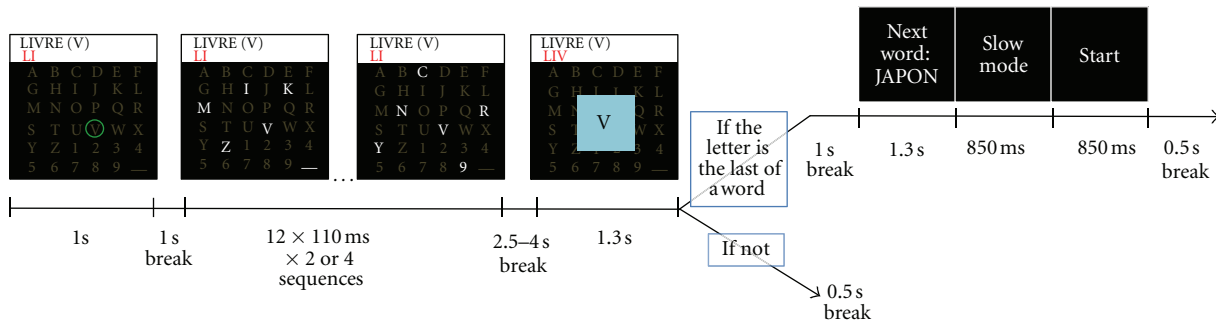


FIGURE 2: The course of one trial. The target letter is indicated by a green circle, for 1 second. Letters are flashed by group of 6 letters. After a delay, the letter chosen by the P300 classifier is displayed for 1.3 seconds. If the letter is the last of the word, a 4.5 second break occurs during which a display indicates the next word and the difficulty level (slow or fast spelling mode). If not, there is a short 0.5 s break before the next green circle appears.

only and those two groups have only this single item in common (see Figure 2). In other words, a unique pair of groups of nonadjacent letters is defined in order to replace the original pair of row and column that was associated with a unique possible target [24]. This way of flashing letters also minimizes the probability that a letter will be flashed twice in a row, which minimizes the risk for the user to miss one target presentation. The flash duration was equal to 60 ms and we set the stimulus onset asynchrony (SOA) to 110 ms. We chose these small parameter values in order to make the trials as short as possible. This has a twofold advantage. It enables us to acquire more trials per subject and it makes the spelling more difficult [22]. Both aspects are essential to generate enough error trials and study error detection and correction.

**2.3.3. One Trial.** We call a trial the succession of stimulations and observations that are needed to select one item (see Figure 2). Each trial is thus made of several sequences depending on the spelling condition. A sequence of stimulations corresponds to the successive flashing of all the groups, once and in a pseudorandom order. The longer the trial (i.e.,

the more sequences per trial), the more observations to rely on for the BCI to find the target.

We used two spelling conditions: a fast, more error-prone condition, made of short (2-sequence long) trials and a slower, less risky one, made of four-sequence long trials. These two modes are fairly fast and challenging, which ensures the recording of many trials per subject, among which enough error trials for subsequent analysis.

Since we used copy spelling, each trial started with the presentation of the current target, both at the top of the screen and within the matrix, using a green circle for 1 second. There was no break in-between sequence. At the end of each trial, 2.5 to 4 seconds after the last flash, the feedback was displayed in a blue square at the middle of the screen for 1.3 seconds. It was simultaneously written on the top of the screen (see Figure 2). This large presentation at the center of the visual field was made to favor clear single-trial responses to feedbacks. Participants were explicitly instructed to wait for the feedback at the middle of the screen and not to blink during feedback presentation.

In the session including automatic correction (see below), the second best guess of the classifier was used to

replace the current feedback, in an orange square, whenever an ErrP would have been detected. In this case, the new feedback was presented for 1 second and the item was also simultaneously corrected at the top of the screen (see Figure 2). After a 0.5 second break, the new target for the next trial was presented.

*2.3.4. Full Procedure.* The experiment was divided into five parts.

- (1) *Installation and Instructions.* After having read about the experiment and signed an informed consent form, each subject was prepared for EEG/MEG data acquisition. During preparation, the subject was asked to read the task instructions. Finally, a couple of typical trials were presented so that the subject would be familiar with stimulus presentation before starting the actual experiment.
- (2) *Speller Training.* The aim of the first session was to gather training data in order to set the supervised algorithms subsequently used in the test phase, for individual feature selection and classification. Precisely, those data were used to both compute the individual spatial filters and class parameters. In this training session, subjects were all required to successively copy spell the same 36 items (the whole matrix). Each item or trial was spelled using 3 flash-sequences. The session lasted about 10 minutes and was interleaved with short breaks after the 12th and 24th items. No feedback was provided.
- (3) *Speller Testing and Error Detection Training.* After training, the subjects had to go through four true spelling sessions, lasting approximately 12 minutes each. Each session was made of twelve 5-letter words. Subjects received feedback after each letter and words were separated by a 4.5 second break. Within a word, letters were spelled using the same number of sequences (either 2 or 4) and short- and long-lasting words were counterbalanced within sessions. The responses to feedbacks over those four sessions were used to train the feature extraction and classification algorithms for subsequent error detection.
- (4) *Speller and Error Correction Testing.* In the fifth spelling session, participants had to spell twenty 5-letter words that they had chosen themselves before the beginning of the whole experiment. Still, it consisted in copy spelling since those words were entered in the computer by the experimenter before the actual session. All letters were spelled using two sequences only (fast mode). This last session lasted 17 minutes approximately. Importantly, whenever an error was detected, automatic correction applied.
- (5) *Debrief.* After recording, participants had to fill in a questionnaire. Using marks between 1 and 10, they had to respond to questions about their perception of the performance of the machine, the difficulty of the task, the quality, and usefulness of the correction.

*2.4. Feature Extraction.* We used similar preprocessing steps and feature extraction algorithms to process the responses evoked by flashes (for spelling) and feedbacks (for error detection), respectively. Raw data were first downsampled to 100 Hz and bandpass filtered between 1 and 20 Hz, online. 100 Hz here corresponds to a good compromise between the need to sample above 60 Hz (the so-called engineer Nyquist frequency to avoid aliasing) and the advantage of reducing the dimension of the data for online processing. Feature extraction then consisted in linear spatial filtering, whose effect is to reduce the dimension of the data as well as to maximize the discriminability between the two classes (i.e., between targets and nontargets, during spelling, and between error and correct feedbacks, during error detection) [25]. The xDAWN algorithm provides orthogonal linear spatial filters that can be learned from training samples [25]. Based on our previous studies [16, 22], we used a five-dimensional filter for both spelling and error detection. As mentioned above, those filters were learned based on the first session, for subsequent spelling, and based on the first four spelling sessions which included feedbacks, for error detection in the last session. For online spelling, the features consisted in the spatially filtered epochs from 0 to 600 ms after flash onset. The evoked response thus included both the P300 and the early visual response [22, 26, 27]. For online error detection, the features consisted in the spatially filtered epochs from 200 to 600 ms after feedback onset [28].

*2.5. Feature Classification.* We used a mixture of multi-dimensional Gaussian model as a classifier. The model parameters (i.e., the mean and variance of each of the two Gaussians) were learned from the same training samples as the parameters of the xDAWN algorithms, for the spelling and error detection task, respectively. Importantly, we assumed conditional independence in time and space between features (naïve Bayes hypothesis) [16, 22]. This makes the real-time computation of the posterior probability of each new feature very efficient. This is particularly relevant for the spelling part, since it enables the BCI to update its posterior probability or belief about the target location, after each new observation or flash.

Indeed, for spelling, all items are initially assumed to be equiprobable targets. At each new observation, this belief is updated following Bayes rule, by optimally combining the data likelihood and priors. The obtained posteriors then furnish the prior for the next observation, in a Markovian fashion.

For error detection, a decision was made after each single feedback observation. Hence Bayes rule was applied once per trial only. We used individual priors based on each subject's averaged error rate, as given by the first four spelling and ErrP training sessions.

*2.6. Evaluation of Online Error Detection.* To evaluate ErrP classification, we use the common confusion matrix for a two-class problem (see Figure 3). It involves the estimation of the following complementary measures (reported in percent in the results section):



		Outcome		Feedback (speller outcome)	
		Error	Correct	Incorrect	Correct
Predicted	Correct/Error	TP	FP	ErrP classifier recognises an error. The second best-guess from the classifier replaces the incorrect letter (good case).	ErrP classifier recognises an error. The second best-guess from the classifier replaces the actually correct letter (worst case).
	Correct/Correct	FN	TN		

(a)

(b)

FIGURE 3: Confusion matrix for a general binary problem (a) and in the context of error correction (b), adapted from [14]. TP: true positive; FP: false positive; FN: false negative; TN: true negative.

- (i) sensitivity =  $TP/(TP + FN)$ , that is, the capacity to correctly detect errors;
- (ii) specificity =  $TN/(TN + FP)$ , that is, the capacity to correctly detect correct trials;
- (iii) accuracy =  $(TP + TN)/(TP + TN + FP + FN)$ , that is, the global efficacy of the classifier.

2.7. *Evaluation of Online Error Correction.* We computed three quantitative measures to evaluate error correction, both at the single subject and group levels.

We denote the first one by  $\theta$ . It is the percentage of error trials for which the classifier's second best guess corresponded to the actual target. Note that  $\theta$  is independent of error detection and only measures how well the classifier might help correcting errors automatically. We estimated  $\theta$  based on the whole five spelling sessions.

The second measure evaluates automatic error correction on the very last session. It is the good correction rate (GCR) and corresponds to the percentage of detected true error trials that were appropriately corrected. While  $\theta$  is an offline (predictive) measure, GCR is an online (true BCI behavioral) measure of performance.

Finally, we estimated the individual gain in spelling accuracy due to automatic error correction in the last session. It is simply the difference between the observed accuracy and the one that would have been observed with no online correction.

A commonly used although imperfect measure of BCI performance is the bit rate [29], originally derived from [30]. It can be computed from the following formula:

$$br = \log_2(M) + (1 - p) \cdot \log_2 \frac{1 - p}{M - 1}, \quad (1)$$

where  $M$  is the number of classes and  $p$  is the accuracy of the P300 classifier. We report br in bits per minute and use it to compare the spelling accuracy with and without online correction as well as the accuracy that would have been observed if error detection would have been simply followed by the opportunity to spell the letter again. To estimate the latter, we consider that the letter would be spelled again once, with an accuracy corresponding to the one observed online for each subject. Importantly, the respelling of a letter includes a short instruction indicating that the user should focus onto the same target again [14].

For each of the above parameters, we also report the obtained values at the population level. Importantly, since the subjects did not show the same amount of errors, computing the average values for TP, FN, and other parameters would yield a bias estimate of what could be predicted at the population level. Therefore, we rather report the values obtained by concatenating all the trials over subjects, that is, by considering the so-called metasubject. We refer to such quantities at the population level as metavalues. As an example, the metasensitivity corresponds to the sensitivity of the metasubject with mTP true positive and mFN false negative (mTP and mFN being the sum of TP and FN over all the 16 participants).

Finally, the spelling accuracy using automatic correction was related to the specificity, the sensitivity, and the GCR using the following simple formula:

$$Pc = P \cdot \text{Spec} + (1 - P) \cdot \text{Sens} \cdot \text{GCR}, \quad (2)$$

where  $P$  indicates the spelling accuracy in the absence of correction and  $Pc$  the spelling accuracy after automatic correction. The correction becomes useful as soon as  $Pc > P$ , which for a given initial spelling accuracy and GCR yields the following limit condition on error detection sensitivity and specificity:

$$\text{Spec} > 1 - \frac{(1 - p) \cdot \text{Sens} \cdot \text{GCR}}{p}. \quad (3)$$

2.8. *Additional Offline Analysis.* Subjects clearly separated between a low (below 0.75) and a high (above 0.85) ErrP detection specificity group. We compared those two groups in terms of performance, responses to questionnaire, and electrophysiological measures. Because of the small sample size in each group, we used the nonparametric Mann-Whitney test for statistical inference.

The electrophysiological responses we considered for quantitative comparisons are the differences between averaged responses to target and non-target stimuli, or to correct and incorrect feedbacks, respectively. These were computed from the downsampled and bandpass filtered data (see Section 4). We compared the amplitudes and latencies of the negative and positive peaks of these differential responses.

For the responses to feedback, only the (last) session with correction was considered. We computed the difference between responses to correct and incorrect feedbacks on

		Outcome	
		Error	Correct
Predicted	Error	23.7%	7.6%
	Correct	14%	54.7%

FIGURE 4: Confusion matrix for ErrP detection (group results).

central electrode Cz, as in [14]. We typically observe a first negative component (between 250 and 450 ms) followed by a positive component (between 350 and 550 ms), which we denote by neg-ErrP and pos-ErrP, respectively.

For the responses to flashes, we used the data from all sessions. We computed the difference between the averaged responses to target and non-target stimuli and selected the channels exhibiting the maximum absolute differences at the group level. We thus focused our comparison on channel P7 for a negative peak difference in time window 150–270 ms and on channel P8 for positive peak difference in time window 250–500 ms. These two components correspond to the N1 and P300 responses, respectively. N1 is known to be associated with automated stimulus processing that is affected by early attentional processes [31] and to be preponderating at parieto-occipital sites [27].

For each participant, both the amplitudes and latencies of the peaks of the above-defined components were used for subsequent analysis and statistical comparisons. Note that for technical reasons, the electrophysiological data of one participant (S08) could not be saved during the experiment. Therefore, all the results relying on offline evoked potential analysis were obtained from the other 15 subjects.

### 3. Results

**3.1. Performance in Spelling.** In fast mode (including the last session), the online spelling accuracy was  $64\% \pm 21$  (SD). In slow mode, it was  $80\% \pm 18$  (SD). Accounting for the delay between two trials (5.8 s), this corresponds to rates of  $4.52 \pm 1.2$  (SD) and  $4.31 \pm 1$  correct letters per minute, respectively. The information transfer rate is higher in fast mode, meaning that the loss in accuracy is compensated by the speed increase.

**3.2. Performance in Error Detection.** The metavalues for the percent of true positive, false positive, false negative, and true negative are shown in Figure 4. The individual sensitivity and specificity values are given in Table 1. At the group level, the (meta) sensitivity, specificity, and accuracy obtained 63%, 88%, and 78%, respectively.

Interestingly, the spelling accuracy correlates with ErrP detection specificity over subjects ( $P < 0.01$ ,  $r = 0.68$ , Figure 5(a)).

What is very much striking is the split into two groups according to the individual specificities in error detection. Six

subjects have quite low specificities, below 75%, while specificities for the other 10 subjects rise above 85% (Figure 5(a)).

**3.3. Performance in Automatic Error Correction.** Individual  $\theta$  and gains are given in Table 1. At the group level, the GCR obtains 34%, meaning that for one hundred well-detected errors, thirty-four have been well corrected. The metavalues for  $\theta$  is 36%, which is very close to the GCR. Over subjects,  $\theta$  correlates with global spelling accuracy ( $P < 0.0001$ ,  $r = 0.87$ ). Precisely, the less the errors during initial spelling, the higher the probability to effectively correct for those errors (Figure 5(b)).

$\theta$  also correlates with specificity ( $P < 0.01$ ,  $r = 0.72$ ) and accuracy of ErrP detection ( $P < 0.01$ ,  $r = 0.63$ , Figure 5(c)), as well as with responses to questions Q1 (“How well did you control the machine?”,  $P < 0.001$ ,  $r = 0.75$ ) and Q2 (“Did the machine perform well?”,  $P < 0.05$ ,  $r = 0.60$ ).

Over the whole group, the spelling accuracy raised by 0.5% due to automatic error correction, relative to a 62% initial accuracy (i.e., without correction). However, interindividual variability proved quite large. Automatic error correction yielded an improvement in 50% of the subjects (with a maximum gain of 12%), while it caused a degradation of spelling accuracy in 37.5% of the subjects (with a maximum drop of 19%).

Table 2 shows the bit rates corresponding to the three compared spelling modes: no correction, predicted correction based on automatic error detection, and hypothetical respelling and online automatic correction. Over subjects, the information transfer rate decreases when we move from automatic correction to automatic detection and respelling. Moreover, the highest bit rate is obtained for the no correction case. However, when restricting the comparison to the high specificity group, the highest bit rate is obtained during automatic correction.

Figure 6 provides a graphical representation of individual performance in terms of sensitivity and specificity in error detection. It also summarizes the behavior of the different groups, by displaying the metaspeaking accuracy corresponding to the whole group, the low specificity group and the high specificity group, respectively. For those three groups, it further emphasizes the boundaries given by (3), that is, the minimum required trade-off between specificity and sensitivity. Precisely, given the observed GCR and spelling accuracy (in the absence of any correction) for each group, each boundary represents the limit above which the automatic correction becomes fruitful. The more a subject lays above the boundary, the higher the expected increase in spelling accuracy (e.g., S10 had a 12% gain of accuracy). Conversely, the more a subject lays below the boundary, the larger the expected drop of spelling accuracy (e.g., S01 had a 19% drop of accuracy).

**3.4. Self-Evaluation Questionnaires.** Questions and averaged ratings are reported in Table 3. Individual responses for Q1, Q2, and Q8 are detailed in Table 1. Interestingly, the group average answer to Q2 equals 6.6, which is very close to the observed global spelling accuracy of 68.5%. Moreover, the

TABLE 1: Individual and group performance in spelling, error detection, and correction. The first three columns indicate the subject number and demographical information such as gender and age. The next two columns indicate the spelling accuracy (percent of correct letters) for the whole five sessions (mixing words spelled with 2 and 4 sequences) and for the fifth session alone (the one dedicated to online evaluation of error correction, without considering the online correction), respectively. The next four columns indicate the individual sensitivity, specificity,  $\theta$ , and gain in spelling accuracy obtained in the last session, reported in percent. The last four columns show the individual responses to the most relevant questions (see Table 2 for a full description). The last three lines provide results at the group level (first line) and after having distinguished between low specificity (second line) and high specificity (third line) performers.

Subjects	Dem. info.		Spelling accuracy		ErrP detection		Correction		Questionnaire				
	Sex	Age (year)	Sessions 1 to 5	Session 5 only	Sens.	Spec.	$\theta$	Gain	Q1	Q2	Q8	Did you prefer with or without correction?	
1	m	34	82%	80%	35%	69%	47%	-19%	5	5	2	without	
3	m	25	32%	25%	72%	68%	25%	1%	1	1	2	without	
4	f	31	44%	36%	75%	69%	24%	-2%	6	5	1	without	
6	f	20	42%	31%	81%	61%	23%	0%	4	8	8	with	
12	m	23	66%	56%	48%	70%	39%	-9%	7	6	4	without	
14	f	23	52%	50%	70%	72%	35%	-1%	4	6	5	without	
2	f	32	65%	56%	68%	93%	39%	7%	5	7	6	with	
5	m	28	70%	72%	46%	96%	46%	3%	7	7	5	with	
7	f	37	93%	89%	27%	97%	44%	-3%	8	8	1	without	
8	f	29	90%	87%	62%	97%	59%	0%	5	7	2	without	
9	f	26	72%	65%	69%	88%	41%	1%	6	7	7	with	
10	f	22	84%	79%	90%	99%	58%	12%	7	8	7	without	
11	m	26	91%	88%	25%	95%	61%	-2%	8	8	4	without	
13	m	23	55%	41%	59%	90%	40%	11%	5	6	9	with	
15	m	27	63%	50%	34%	94%	45%	6%	7	8	8	with	
16	f	25	94%	91%	67%	99%	50%	3%	8	8	9	with	
	Summary		Metasubject										Summary
All	7 m; 9 f	26.9	69%	62%	63%	88%	36%	1%	5.8	6.6	5.0	44% with	
Spec < 0.75	3 m; 3 f	26	53%	46%	69%	69%	29%	-5%	4.5	5.2	3.7	17% with	
Spec > 0.85	4 m; 6 f	27.5	78%	72%	56%	95%	45%	4%	6.6	7.4	5.8	60% with	

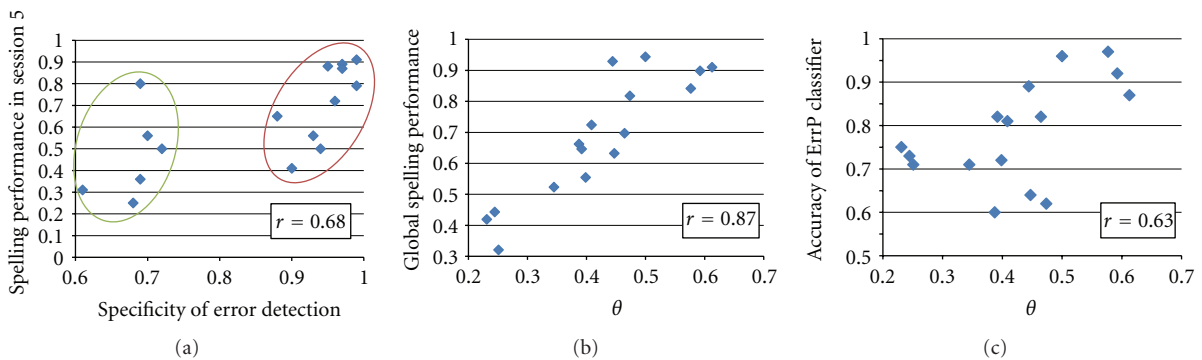


FIGURE 5: (a) Correlation between the spelling accuracy (rate of correctly spelled letters) in session 5 (without considering the correction) and specificity of error detection. The green circle includes participants with low specificity while the red one includes participants with high specificity. (b) Global spelling accuracy as a function of  $\theta$ . (c) Accuracy of error detection as a function of  $\theta$ . Each diamond represents a participant, and  $R$  values represent correlation coefficients.

response to Q2 correlates with global spelling accuracy over individuals ( $P < 0.05$ ,  $r = 0.61$ ). And it also correlates with responses to Q1 ( $P < 0.01$ ,  $r = 0.74$ ). Answers to Q1 and Q2 further correlates significantly with error detection specificity (resp.  $P < 0.05$ ,  $r = 0.59$ , and  $P < 0.01$ ,  $r = 0.62$ ),

as well as with the observed gain in spelling accuracy due to automatic correction ( $P < 0.05$ ,  $r = 0.54$ ). The gain in spelling accuracy also correlates positively with the subject's answer about the usefulness of automatic correction (Q8,  $P < 0.05$ ,  $r = 0.61$ ).

TABLE 2: Bits per minute for the three compared spelling modes: no correction at all, automatic detection and predicted correction due to respelling, and automatic correction (online results).

Subjects	Spelling mode (bits/min)		
	No correction	Error detection and respelling	Error detection and automatic correction (online)
S01	21.75	16.27	13.55
S03	3.25	3.01	3.27
S04	6.00	5.45	5.14
S06	4.69	5.45	4.37
S12	12.23	9.68	8.88
S14	10.20	8.97	9.39
S02	12.23	13.01	14.24
S05	18.29	18.68	19.23
S07	26.10	25.63	24.43
S08	25.08	25.48	24.79
S09	15.51	15.34	15.38
S10	21.30	23.51	26.59
S11	25.58	24.78	24.40
S13	7.42	7.82	10.43
S15	10.20	10.30	11.97
S16	27.15	28.07	28.61
Metasubjects			
All	13.06	12.70	12.81
Spec < 0.75	9.03	7.71	7.37
Spec > 0.85	18.21	18.77	19.56

TABLE 3: Questionnaire and averaged ratings over the whole group ( $N = 16$ ). Ratings ranged between 1 (most negative response) and 10 (most positive response).

Questions	Averaged ratings $\pm$ SD
Q1. How well did you control the machine?	5.8 $\pm$ 1.9
Q2. Did the machine perform well?	6.6 $\pm$ 1.8
Q3. Was the task difficult?	4.6 $\pm$ 2.5
Q4. Was the task tiring?	5.9 $\pm$ 2.3
Q5. Was the task motivating?	7.9 $\pm$ 1.4
Q6. Could you predict errors?	4.9 $\pm$ 3.1
Q7. How much did errors upset you?	7.3 $\pm$ 2.2
Q8. Was the automatic error correction useful?	5.0 $\pm$ 2.8

Finally, to the last (binary) question: “Did you prefer the spelling with or without correction?” a short majority reported a preference for no correction (56%). However, when distinguishing between the low and high specificity groups, we find that 83% of the former preferred without correction whereas 60% of the latter preferred with correction.

**3.5. Electrophysiology.** The grand average feedback-related responses in correct and incorrect trials as well as their difference are depicted in Figure 7. Scalp topographies are also represented for each condition and for the two latencies

corresponding to the negative and positive peaks of the difference. The negative peak average latency is 350 ms  $\pm$  49 (SD) after feedback onset, while the positive peak average latency is 480 ms  $\pm$  51 (SD).

The pos-ErrP amplitude correlates with the accuracy of error detection ( $P < 0.01$ ,  $r = 0.64$ , Figure 8(a)), while the neg-ErrP amplitude correlates with the sensitivity of error detection ( $P < 0.01$ ,  $r = -0.70$ , Figure 8(b)). Hence the larger the difference between responses to correct and incorrect trials, the more efficient the automatic error detection. Besides, we found a significant relationship between the initial spelling accuracy (without correction) and both the amplitude and the latency of the pos-ErrP. Indeed, the higher



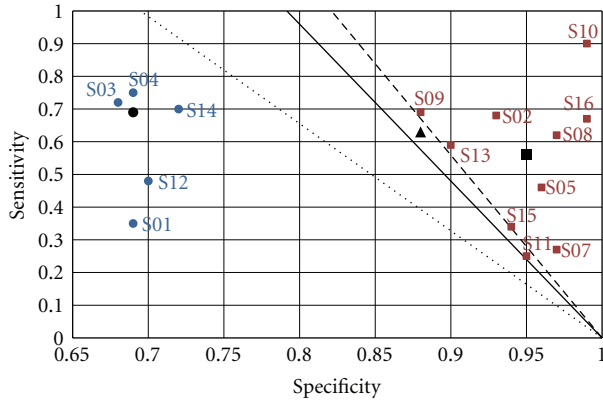


FIGURE 6: Error detection sensitivity as a function of error detection specificity. All subjects are represented as well as the global meta-subject (triangle) and the two metasubjects for the low specificity (circle) and high specificity (square) groups, respectively. The lines are the boundaries above which the automatic correction becomes fruitful, for the three groups (plain, dotted, and dashed, resp.).

the spelling accuracy, the larger ( $P < 0.05$ ,  $r = 0.58$ ) and the earlier ( $P < 0.05$ ,  $r = -0.55$ ) the positivity.

Similarly, the group average responses to target and non-target stimuli as well as the difference between the two are depicted on Figure 9. Figure 9 also shows the topographies obtained for those three responses, at the peak latency of the N1 ( $229 \text{ ms} \pm 19 \text{ (SD)}$ ) and P300 components ( $380 \text{ ms} \pm 72 \text{ (SD)}$ ).

The amplitude of the N1 component correlates with both the global spelling accuracy ( $P < 0.05$ ,  $r = -0.55$ , Figure 8(c)) and  $\theta$  ( $P < 0.01$ ,  $r = -0.72$ ). The larger the negative difference between the responses to target and non-target stimuli, the higher the spelling accuracy and the  $\theta$ . This corroborates what has been already observed in healthy subjects using the P300-speller, namely, that spelling accuracy depends on the ability to focus on the desired character, which yields a larger N1 response and provides a complementary feature to the P300 in order to achieve good classification [26].

Note that for correlations involving neg-ErrP or N1 amplitudes, values are negative since the amplitudes are negatives.

**3.6. Between-Group Differences.** The low specificity group performed spelling significantly poorer than the high specificity group (53% compared to 78%,  $P < 0.05$ ). They also benefit significantly less from automatic correction ( $P < 0.05$ ): correction improved the spelling accuracy by 4% in the high specificity group, while it degraded it by 5% in the low specificity group (Table 1). Accordingly, participants showing a low specificity tend to perceive the machine as less efficient ( $P < 0.05$ ) and felt like they had fewer control on the BCI ( $P < 0.05$ ). In agreement with their lower spelling accuracy, the subjects in the low specificity group also present a significantly lower value for  $\theta$  (0.29 compared to 0.45,  $P < 0.05$ ).

Finally, the latency of neg-ErrP peak proved shorter ( $P < 0.05$ , Figure 10(b)) and the amplitude of the N1 peak proved larger ( $P < 0.01$ , Figure 10(c)) for the high specificity group.

We found no significant differences between the two groups on other physiological parameters or on answers to questions Q3 to Q8.

## 4. Discussion

**4.1. Performance in Spelling and Error Detection.** We used our own implementation of the P300-speller BCI for the evaluation of online automatic error correction in 16 healthy volunteers. We considered two spelling conditions, namely, a slow and a fast mode, whose trial length was guided by our own previous experiments [16, 22].

In the current study, the online spelling accuracy proved fairly high. On average, participants could spell about 4.5 correct letters per minute, to be compared with 1.57 correct letters per minute in [14]. This high performance level might be partly attributable to the use of the xDAWN algorithm for spatial filtering [32] and to the departure from the row-column paradigm adapted from [23]. This initial high spelling accuracy has to be kept in mind when interpreting the outcome of error correction. Indeed, it might be that the effectiveness of error detection and correction depends upon the ongoing bit rate, which directly relies on the speed-accuracy trade-off that was targeted when choosing specific stimulation settings (i.e., short- or long-flash durations, short or long sequences).

Regarding the responses to feedback, we observed fronto-central evoked signals whose time courses resemble the ones that have been described recently in a BCI context [12, 14, 16]. As we already noted in [23], those components that are usually referred to as interaction ErrP in BCI [2], exhibit spatial and temporal patterns that strongly resemble the ones of feedback responses classically reported in cognitive neuroscience studies [8]. This suggests that although contextual modulations can be observed [33, 34], responses to feedback may present spatial and temporal components that are independent of the context but specific to the core process of learning from external outcomes. It stresses out the question of what part of the signal and underlying process is specific to a BCI type of interaction and what part is not.

As in a couple of earlier studies [13, 14], we showed that those responses to feedbacks can be detected online, from single trials. At the group level, we obtained 88% specificity and 63% sensitivity. For comparison, Dal Seno and collaborators tested two subjects and obtained an averaged specificity and sensitivity of 68% and 62%, respectively [13]. In nine healthy subjects, Spuler and collaborators report a 96% specificity and 40% sensitivity [14]. Note that in the latter study, the authors used a biased classifier to favor specificity. Indeed, a high specificity guarantees that correctly spelled letters will not be detected as errors mistakenly.

In our experiment, we did not use a biased classifier. However, specificity was higher than sensitivity for most of the subjects. This is because spelling accuracy is fairly high,

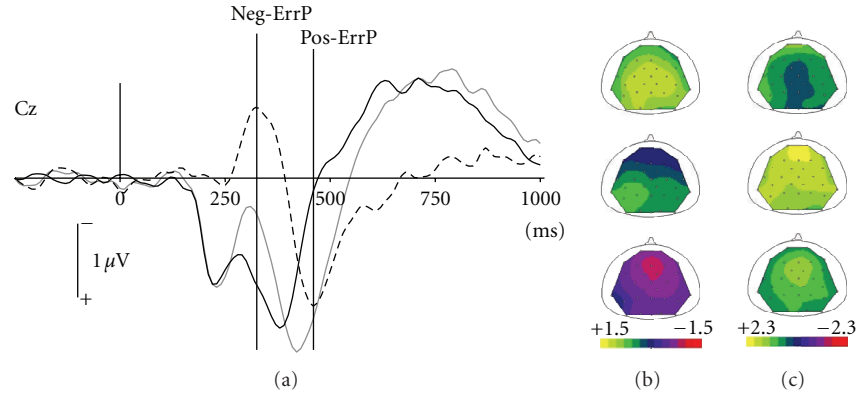


FIGURE 7: Grand average event-related dynamics ( $-250$  ms to  $1000$  ms, channel Cz) associated with responses to correct feedbacks (black solid line), responses to incorrect feedbacks (grey solid line), and the difference between the two (black dashed line) (a). Zero time corresponds to feedback onset. Topographies for correct (top), incorrect trials (middle), and their difference (bottom) correspond to the latencies of the peaks of the neg-ErrP (b) and pos-ErrP (c), respectively.

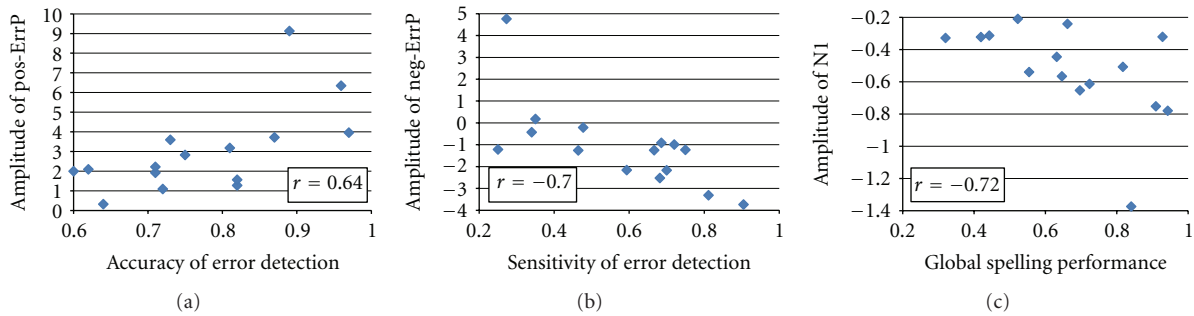


FIGURE 8: (a) Amplitude of pos-ErrP as a function of the accuracy of error detection (microvolts). (b) Amplitude of neg-ErrP as a function of the sensitivity of error detection (microvolts). (c) Amplitude of N1 as a function of the global spelling accuracy (microvolts). Each diamond represents a participant, and  $R$  values represent correlation coefficients.

which yields much more training samples for the correct than for the incorrect feedback responses. Indeed, sensitivity highly depends upon the quality of the learning of the error class. As an example, subject S07 made 13 errors only. In [14], the authors recommend the use of at least 50 samples to train the ErrP classifier. Hence, part of the interindividual variability in sensitivity is simply due to the variability in the initial spelling accuracy between subjects, which directly affected the quality of the training. However, this is not the case for specificity since as a corollary, high spelling accuracies yield a lot of training samples for the correct class. We obtained 84 training samples in the poorer performer (S03).

Nevertheless, the current results for error detection are slightly worse than the ones we obtained offline in a previous experiment [16]. This might be due to several factors. One factor might be the offline use of ICA, to clean up the signals in our previous study. Second, the use of fairly fast modes here induced smaller interval between two consecutive feedbacks, which might have diminished the expectancy for each new outcome. A known possible cause for smaller feedback responses [35, 36]. The latter effect might not be seen in patients where more sequences

would typically be implemented. However, there is certainly room for improvement. At least two lines of research are worth mentioning in that respect. One is to make use of preexisting databases to tune the algorithms and suppress individual training [14]. However, this does not seem to be a good option in patients [14]. The second option, which could be used in combination with the first one, would be to use adaptive algorithms and keep updating the individual parameters while using the BCI. Reinforcement learning methods might prove very useful in this context [35].

**4.2. Global Performance in Automatic Error Correction.** The novelty of our approach comes from the implementation and online evaluation of automatic error correction. Whenever an error is detected, we simply proposed to replace the supposedly erroneous choice by our classifier's second best guess. We evaluated the efficiency of this automatic correction through the computation of the good correction rate (GCR). At the group level, the GCR was equal to 34%. On the one hand, this is much higher than chance level ( $1/35 = 2.86\%$ ), which speaks in favor of applying this strategy for automatic error correction. On the other hand, this is too low to yield a significant gain in spelling accuracy (0.5% only).

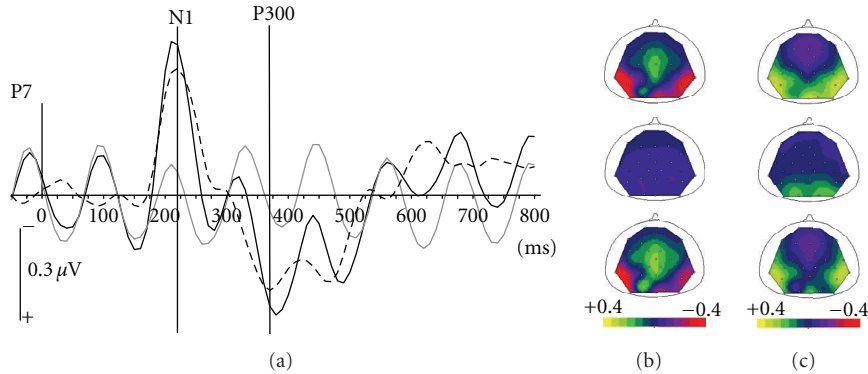


FIGURE 9: Grand average event-related dynamics ( $-50$  ms to  $800$  ms, channel P7) associated with responses to target flashes (black solid line), response to non-target flashes (grey solid line), and the difference between the two (black dashed line) (a). Zero time corresponds to feedback onset. Topographies for correct (top), incorrect trials (middle), and their difference (bottom) correspond to the latencies of the peaks of the N1 (b) and P300 (c), respectively.

However, the meta-bit-rate is slightly better with correction than with detection only. This highlights the efficiency of even poor automatic error correction compared to sole error detection which requires cumbersome additional spelling. In order to improve the GCR, one promising option is to use priors from a dictionary in order to bias the classifier [36]. It might improve both the initial spelling accuracy and the ensuing GCR.

*4.3. The Importance of Interindividual Differences.* Beyond results at the group level, we observe a large interindividual variability. Specificity of error detection enabled us to clearly distinguish between two groups of participants: the low ( $<75\%$ ,  $N = 6$ ) and the high ( $>85\%$ ,  $N = 10$ ) specificity group. Our analysis revealed that those two groups differ not only in terms of specificity but also in terms of electrophysiological responses, initial spelling accuracy,  $\theta$  values, and spelling accuracy gain as well as in their subjective perception of the BCI experience.

One obvious possible explanation is that the high specificity group corresponds to subjects who were more engaged into the task, which yields electrophysiological responses with large signal-to-noise ratio. Indeed, the N1 and P300 responses are known to reflect the participant's involvement in the task ([27, 37], see also [38] for a review on the P300 and [31] for a review on the N1). The P300 has been shown to increase with motivation in a BCI context [39] as well as the ensuing feedback-related negativity (the neg-ErrP) [16, 33]. As an expected consequence of a higher signal-to-noise ratio, the spelling accuracy increases as well as the  $\theta$ . We indeed observe that physiological amplitudes in response to flashes, spelling accuracy, and  $\theta$  are strongly and positively correlated. Similarly, the larger the ErrP, the more efficient the error detection, especially in terms of specificity. This is in agreement with the correlations we observed between the spelling accuracy,  $\theta$ , and specificity and the perception of the BCI by the user (Q1 and Q2). Indeed, a high  $\theta$  associated with an efficient ErrP classification yields a significant improvement due to automatic correction.

All the correlations and group differences we observed are coherent with the hypothesis of a role of task engagement. However, we should remain cautious in our interpretations. This assumption requires further in-depth investigations. Indeed, we could not show any significant correlation between objective measures and the subjective responses to motivation-related question (viz. Q5 and Q7).

Nevertheless, our conclusion can be refined by looking closely at the results in the high specificity group ( $N = 10$ ). In this group, only 2 subjects experienced a drop of spelling accuracy due to automatic error correction. Accordingly, 6 subjects reported a preference in favor of a spelling including automatic correction. Six subjects showed a higher bit rate when using the automatic correction compared to no correction. And 7 out of the 10 participants obtained a higher bit rate with automatic correction compared to automatic detection with predicted respelling. Three out of the 4 participants, who reported preference for spelling with no correction, did not improve their spelling accuracy with correction. Interestingly, the last one (S10) who obtained the largest improvement ( $+12\%$ ), surprisingly, reported verbally that efficient correction required too much attentional effort. All those results demonstrate the putative usefulness of automatic error correction but highlight the fact that it should certainly be used in a different way depending on the user's profile.

#### *4.4. Requirements for Automatic Correction to Be Relevant.*

In theory, any initial spelling accuracy could benefit from automatic correction, provided that the sensitivity, the specificity, and the GCR are high enough. All studies concur in stating that specificity is a primary requirement, in order to avoid the highly frustrating situation of automatically discarding a correctly spelled item. But contrary to Visconti and colleagues [15], our results show that the better the spelling accuracy, the more relevant the correction. At first, this might sound counter-intuitive, since the higher the spelling accuracy, the more difficult the rare error detection. However, we do observe that good performers achieve

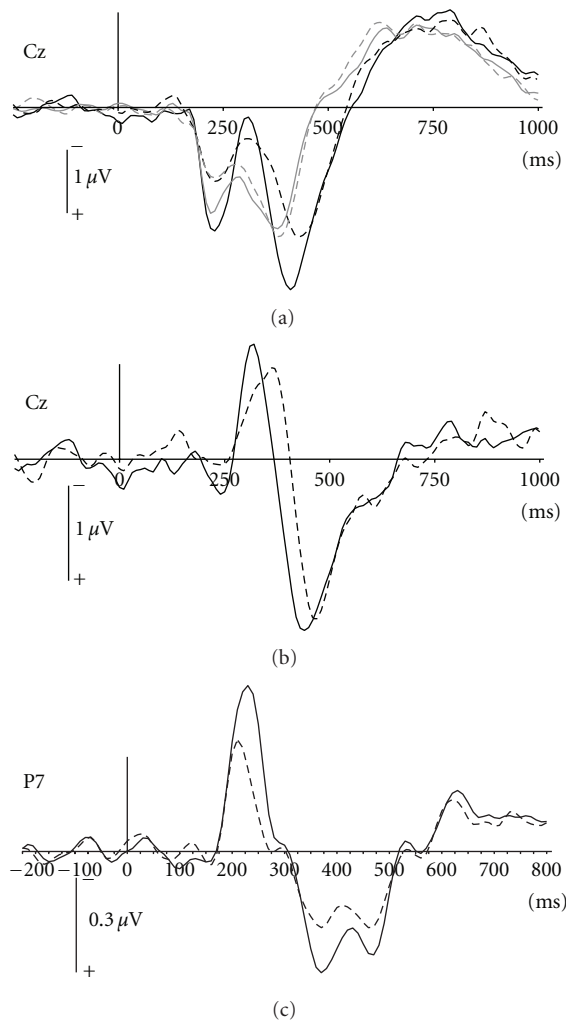


FIGURE 10: Grand average EEG event-related dynamics for group with a high specificity in solid line ( $N = 10$ ) and group with a low specificity in dashed line ( $N = 6$ ). (a) Responses to bad (grey) and good feedbacks (black) in channel Cz. (b) Difference of bad and good feedback responses in channel Cz. (c) Difference of target and non-target responses in channel P7.

better correction. Indeed, we show that both ErrP detection accuracy and  $\theta$  (the subsequent ability to automatically correct for errors) are closely related to the spelling accuracy. In other words, the higher the spelling accuracy, the higher the performance in error detection and correction. As also supported by larger averaged ErrP and P300 responses, these results suggest that the more the subject engages into the task, the higher the performance in terms of both spelling accuracy and error correction. This is a strong indication in favor of a possible use of the P300-speller to train subjects in their abilities to focus attention [40].

## 5. Conclusion

The BCI presented here is the first online P300-Speller employing automatic correction by another item when an

ErrP is detected. Our results are competitive in terms of ErrP detection, although they could probably be improved using adaptive training and a biased classifier. The automatic correction could also be improved, possibly by using the information from a probabilistic dictionary. However, it proved already relevant in terms of bit rate, compared to classical automatic ErrP detection alone. It also proved significant in most of the participants who had the best initial spelling accuracy. Importantly, the correction needs to be adjusted to each participant, depending on their initial spelling accuracy and preference.

## Acknowledgment

This work is supported by the French ANR project ANR-DEFIS 09-EMER-002 CoAdapt. We thank Dominique Morlet for technical help.

## References

- [1] B. H. Dobkin, "Brain-computer interface technology as a tool to augment plasticity and outcomes for neurological rehabilitation," *Journal of Physiology*, vol. 579, part 3, pp. 637–642, 2007.
- [2] J. D. Millan, R. Rupp, G. R. Müller-Putz et al. et al., "Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges," *Frontiers in Neuroscience*, vol. 4, article 161, 2010.
- [3] T. O. Zander, C. Kothe, S. Welke, and M. Roetting, "Utilizing secondary input from passive brain-computer interfaces for enhancing human-machine interaction," in *Foundations of Augmented Cognition Neuroergonomics and Operational Neuroscience*, pp. 759–771, San Diego, Calif, USA, 2009.
- [4] J. W. Gehring, B. Goss, M. G. Coles, D. E. Meyer, and E. Donchin, "A neural system for error detection and compensation," *Psychological Science*, vol. 4, pp. 385–390, 1993.
- [5] M. Falkenstein, J. Hohnsbein, and J. Hoormann, "Event-related potential correlates of errors in reaction tasks," *Electroencephalography and Clinical Neurophysiology*, vol. 44, pp. 287–296, 1995.
- [6] J. Hohnsbein, M. Falkenstein, and J. Hoormann, "Error processing in visual and auditory choice reaction tasks," *Journal of Psychophysiology*, vol. 3, p. 32, 1989.
- [7] M. Falkenstein, J. Hohnsbein, J. Hoormann, and L. Blanke, "Effects of errors in choice reaction tasks on the ERP under focused and divided attention," in *Psychophysiological Brain Research*, C. H. M. Brunia, A. W. K. Gaillard, and A. Kok, Eds., pp. 192–195, Tilburg, The Netherlands, 1990.
- [8] W. H. R. Miltner, C. H. Braun, and M. G. H. Coles, "Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection," *Journal of Cognitive Neuroscience*, vol. 9, no. 6, pp. 788–798, 1997.
- [9] S. Nieuwenhuis, C. B. Holroyd, N. Mol, and M. G. H. Coles, "Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance," *Neuroscience and Biobehavioral Reviews*, vol. 28, no. 4, pp. 441–448, 2004.
- [10] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Response error correction—a demonstration of improved human-machine performance using real-time EEG monitoring," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 173–177, 2003.



- [11] G. Schalk, J. R. Wolpaw, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: presence of an error potential," *Clinical Neurophysiology*, vol. 111, no. 12, pp. 2138–2144, 2000.
- [12] P. W. Ferrez and J. J. Del, "Error-related EEG potentials generated during simulated brain-computer interaction," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 923–929, 2008.
- [13] B. Dal Seno, M. Matteucci, and L. Mainardi, "Online detection of P300 and error potentials in a BCI speller," *Computational Intelligence and Neuroscience*, vol. 2010, Article ID 307254, 5 pages, 2010.
- [14] M. Spuler, M. Bensch, S. Kleih, W. Rosenstiel, M. Bogdan, and A. Kubler, "Online use of error-related potentials in healthy users and people with severe motor impairment increases performance of a P300-BCI," *Clinical Neurophysiology*, vol. 123, no. 7, pp. 1328–1337, 2012.
- [15] G. Visconti, B. Dal Seno, M. Matteucci, and L. Mainardi, "Automatic recognition of error potentials in a P300-based brain-computer interface," in *Proceedings of the 4th International Brain-Computer Interface Conference*, 2008.
- [16] M. Perrin, E. Maby, R. Bouet, O. Bertrand, and J. Mattout, "Detecting and interpreting responses to feedback in BCI," in *Proceedings of the 5th International Brain-Computer Interface Workshop & Training Course*, pp. 116–119, Graz, Austria, 2011.
- [17] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.
- [18] M. Fabiani, D. Karis, and E. Donchin, "P300 and recall in an incidental memory paradigm," *Psychophysiology*, vol. 23, no. 3, pp. 298–308, 1986.
- [19] N. K. Squires, K. C. Squires, and S. A. Hillyard, "Two varieties of long latency positive waves evoked by unpredictable auditory stimuli in man," *Electroencephalography and Clinical Neurophysiology*, vol. 38, no. 4, pp. 387–401, 1975.
- [20] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 174–179, 2000.
- [21] Y. Renard, F. Lotte, G. Gibert et al., "OpenViBE: an open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments," *Presence*, vol. 19, no. 1, pp. 35–53, 2010.
- [22] E. Maby, G. Gibert, P. E. Aguera, M. Perrin, O. Bertrand, and J. Mattout, "The OpenViBE P300-speller scenario: a thorough online evaluation," in *Proceedings of the Human Brain Mapping Conference*, 2010.
- [23] G. Townsend, B. K. LaPallo, C. B. Boulay et al., "A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109–1120, 2010.
- [24] H. Cecotti and B. Rivet, "One step beyond rows and columns flashes in the P300 speller: a theoretical description," *International Journal of Bioelectromagnetism*, vol. 13, no. 1, pp. 39–41, 2011.
- [25] H. Cecotti, B. Rivet, M. Congedo et al., "A robust sensor-selection method for P300 brain-computer interfaces," *Journal of Neural Engineering*, vol. 8, no. 1, Article ID 016001, 2011.
- [26] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk, "Does the 'P300' speller depend on eye gaze?" *Journal of Neural Engineering*, vol. 7, no. 5, Article ID 56013, 2010.
- [27] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based brain-computer interface," *Behavioral and Brain Functions*, vol. 6, article 28, 2010.
- [28] E. Lopez-Larraz, I. Iturrate, L. Montesano, and J. Minguez, "Real-time recognition of feedback error-related potentials during a time-estimation task," in *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '10)*, pp. 2670–2673, 2010.
- [29] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks et al., "Brain-computer interface technology: a review of the first international meeting," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164–173, 2000.
- [30] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication: Presence of an Error Potential*, University of Illinois Press, Chicago, Ill, USA, 1949.
- [31] R. Naatanen and T. Picton, "The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure," *Psychophysiology*, vol. 24, no. 4, pp. 375–425, 1987.
- [32] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: application to brain-computer interface," *IEEE Transactions on Bio-Medical Engineering*, vol. 56, no. 8, pp. 2035–2043, 2009.
- [33] N. Yeung, C. B. Holroyd, and J. D. Cohen, "ERP correlates of feedback and reward processing in the presence and absence of response choice," *Cerebral Cortex*, vol. 15, no. 5, pp. 535–544, 2005.
- [34] R. B. Mars, E. R. A. De Bruijn, W. Hulstijn, W. Miltner, and M. Coles, "What if I told you: 'You were wrong'? brain potentials and behavioral adjustments elicited by performance feedback in a time-estimation task," in *Errors, Conflicts, and the Brain Current Opinions on Performance Monitoring*, MPI of Cognitive Neuroscience, pp. 129–134, Leipzig, Germany, 2004.
- [35] A. Rachez, T. Proix, E. Maby, J. Mattout, and E. Dac c, "Direct policygradient for online learning in BCI," *International Journal of Bioelectromagnetism*, vol. 13, no. 1, pp. 52–53, 2011.
- [36] T. M. Vaughan, D. J. McFarland, G. Schalk et al., "The wadsworth BCI research and development program: at home with BCI," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 229–233, 2006.
- [37] A. Datta, R. Cusack, K. Hawkins et al., "The P300 as a marker of waning attention and error propensity," *Computational Intelligence and Neuroscience*, vol. 2007, Article ID 93968, 9 pages, 2007.
- [38] T. W. Picton, "The P300 wave of the human event-related potential," *Journal of Clinical Neurophysiology*, vol. 9, no. 4, pp. 456–479, 1992.
- [39] S. C. Kleih, F. Nijboer, S. Halder, and A. K bler, "Motivation modulates the P300 amplitude during brain-computer interface use," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1023–1031, 2010.
- [40] E. Maby, M. Perrin, O. Bertrand, G. Sanchez, and J. Mattout, "BCI could make old two-player games even more fun: a proof of concept with, 'Connect Four,'" *Advances in Human-Computer Interaction*. In press.





# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

