Update article

Improving BCI performance through co-adaptation: Applications to the P300-speller

Jérémie Mattout\textsuperscript{a,b,*}, Margaux Perrin\textsuperscript{a,b}, Olivier Bertrand\textsuperscript{a,b}, Emmanuel Maby\textsuperscript{a,b}

\textsuperscript{a}Brain Dynamics and Cognition Team, Lyon Neuroscience Research Center, INSERM U1028-CNRS, UMR5292, 69000 Lyon, France
\textsuperscript{b}University Lyon 1, 69000 Lyon, France

A R T I C L E   I N F O

Article history:
Received 6 October 2014
Accepted 6 October 2014

Keywords:
Brain-Computer Interfaces
P300-Speller
Error potentials
Feedback Related Negativity
Optimal stopping
Motivation

A B S T R A C T

A well-known neurophysiological marker that can easily be captured with electroencephalography (EEG) is the so-called P300: a positive signal deflection occurring at about 300 ms after a relevant stimulus. This brain response is particularly salient when the target stimulus is rare among a series of distracting stimuli, whatever the type of sensory input. Therefore, it has been proposed and extensively studied as a possible feature for direct brain–computer communication. The most advanced non-invasive BCI application based on this principle is the P300-speller. However, it is still a matter of debate whether this application will prove relevant to any population of patients. In a series of recent theoretical and empirical studies, we have been using this P300-based paradigm to push forward the performance of non-invasive BCI. This paper summarizes the proposed improvements and obtained results. Importantly, those could be generalized to many kinds of BCI, beyond this particular application. Indeed, they relate to most of the key components of a closed-loop BCI, namely: improving the accuracy of the system by trying to detect and correct for errors automatically; optimizing the computer’s speed-accuracy trade-off by endowing it with adaptive behavior; but also simplifying the hardware and time for set-up in the aim of routine use in patients. Our results emphasize the importance of the closed-loop interaction and of the ensuing co-adaptation between the user and the machine whenever possible. Most of our evaluations have been conducted in healthy subjects. We conclude with perspectives for clinical applications.

© 2015 Elsevier Masson SAS. All rights reserved.

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. BCI is still a young field that is currently maturing by borrowing from several disciplines such as engineering, computational sciences, signal processing and neurophysiology. As a matter of fact, no BCI application has yet fully succeeded in being accurate and robust enough to be used routinely in clinical applications on impaired patients.

EEG is the most popular technique for BCI applications, simply because it is non-invasive, cheap and fairly easy to use at patient’s bedside. Moreover, tremendous efforts are being put into wireless and gel free EEG nowadays. EEG-based BCI are being explored for several years and a few neurophysiological markers have proved useful and promising. Probably the most prominent application that has emerged so far is the so-called P300-speller whose aim is to enable partially or fully locked-in patients to communicate [2]. Although efficient, this application remains limited in several aspects [3]. A central limitation lies in the need for high signal-to-noise ratio in order to make an accurate decision. This yields a challenging compromise between the speed and the accuracy of the spelling. In this paper, we synthesize results from our own recent online studies that aimed at optimizing this speed-accuracy trade-off. Two complementary strategies were used.

On the one hand, we made use of additional signals from the user. These are EEG responses to the display of the machine's
decision. Interestingly, these responses evoked by the machine’s feedback reflect whether the decision is erroneous or not. We proposed to measure them online in order to implement some automatic error correction.

On the other hand, we endowed the machine with adaptive behavior in order to make it more flexible and yet able to explicitly optimize the speed-accuracy trade-off at each trial. This approach relies on evidence accumulation such that the higher the signal-to-noise ratio in the EEG command, the faster the spelling. Evaluating this strategy online, we could evaluate the effect of this optimization on the user’s performance and motivation.

In both studies, conducted with healthy subjects, online data processing was performed within the OpenViBE software environment [4].

This paper is organized as follows. The Methods section starts with a short description of the P300-speller application. Then the two above approaches are introduced. Study 1 focuses on error signals during spelling through brain-computer interaction and evaluates the usefulness of such signals to correct for errors online, in an automated fashion. Study 2 introduces and validates an adaptive P300-speller and highlights the importance of co-adaptation. The results section summarizes the outcomes of these studies. In the last section, we discuss the implications and perspectives offered by those complementary studies.

2. Methods

2.1. General principle of the P300-Speller

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) [5] and is even stronger when the subject is instructed to count the rare stimuli [6]. It can be used to select items displayed on a computer screen [7]. 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 of the BCI to detect the P300 response.

We call a trial the succession of stimulations and observations that are needed to select one item. Each trial is made of several sequences, depending on the stopping criterion. A sequence of stimulations corresponds to the successive flashing of all the groups once, in a pseudo-random order. The longer the trial (i.e. the more sequences per trial), the more observations to rely on to find the target. Fig. 1 illustrates the general principle of the P300-Speller and the notion of sequence of stimulations to spell an item. Between 10 and up to 15 sequences are typically used in common implementations of the P300-Speller.

2.2. Study 1: making use of EEG error signals

We conducted this study to evaluate the benefits of automatic error correction during P300-based spelling. Error correction can be implemented in BCI thanks to EEG responses evoked by the feedback [8–11]. Indeed, such evoked responses differ depending on whether the feedback is correct or not, that is whether the item detected by the BCI is indeed the one that the user wanted to spell. This can easily be measured in the case of copy spelling, when the machine knows what is the target letter. Hence, online feedbacks can be readily labelled as correct or incorrect for subsequent analysis of feedback evoked responses.

Fig. 2 shows the typical (averaged) evoked responses for correct and incorrect feedbacks as obtained during P300-based spelling. They are typically measured on fronto-central recording sites, thus requiring more anterior electrodes than the ones needed for spelling only.

A three-step procedure is required to evaluate automatic error correction online, as follows:

- an initial phase is used to calibrate the error detection algorithm. This requires acquiring samples of responses to both correct and incorrect feedbacks while the user is spelling in the absence of automated correction, so as to learn those responses for this particular individual;
- the initial training phase enables to optimize spatial filters [12] and a probabilistic classifier [13] that can then be used to detect error signals from each feedback related response;
- once an error has been detected, the automated correction consists in replacing the presumable erroneous item with another item, without user’s intervention. Since P300-based spelling also relies on probabilistic classification, items are ranked according to their probability of being the target. A natural strategy when an error is detected is then to propose the second most probable item according to this ranking.

![Fig. 1. Illustration of the general principle of a P300-Speller BCI: when the user focuses on the target and the target is flashed (e.g. letter H), the typical EEG evoked response will exhibit a strong N1 component followed by a P300 waveform (A); when the user focuses on the target but the target is not flashed, the typical EEG evoked response should be weaker with a smaller N1 component and no P300 waveform (B). Each group of letter is flashed, one after the other. One sequence is obtained as soon as every group has been flashed once. One trial or item spelling may consists in several sequences (C).](image-url)
In this online study, we evaluated error detection and correction at both the group and the individual level. In order to promote errors, we made the spelling challenging by considering very short (2 sequence-long) and short (4 sequence-long) trials.

Sixteen healthy volunteers participated in this study. Thirty-two EEG sensors were used for both spelling and error correction. Their placement followed the extended 10–20 systems. At the end of the experiment, the subjects were asked to answer a short questionnaire about their perception of the BCI performance in terms of both spelling and error correction.

All the details about this first study can be found in [14].

2.3. Study 2: the benefits of optimal stopping

In this second study, we endowed our P300-Speller BCI with adaptive decision-making. Instead of keeping the number of sequences constant, we enabled the BCI to stop in an optimal fashion [15,16]. We wanted the BCI to be fast when it is confident about its decision and conversely, to be slow and to keep acquiring data when it is not yet clearly decided. This can be done by implementing some evidence accumulation process and an original stopping criterion that explicitly trades speed and accuracy.

Importantly, this approach relies on updating after each flash (instead of each sequence) the probability for each item to be the target. This is performed using Bayesian learning, which yields an evolutionary posterior probability distribution whose entropy reflects the confidence over the current target estimation or, in other words, the accumulated evidence in favor of each item. This information theoretic criterion is convenient since it is bounded and can thus easily be used to adjust the speed-accuracy trade-off.

Eleven healthy and BCI-naive subjects took part in this study. We compared our new adaptive mode with a traditional fixed mode. In the latter, the spelling was based on five sequences, while in the adaptive condition the speed-accuracy trade-off was individually tuned so as to reach roughly the same speed (five sequences) on average.

We also considered a further optimized BCI in terms of stimulations and EEG setup.

Regarding flashes, like in study 1, we departed from the traditional row/column way of grouping items. Instead, we grouped letters in a pseudo-random fashion and in a way that prevents from flashing neighboring items [17]. This reduces errors due to distractions.

Besides, unlike in study 1, we reduced our number of EEG sensors down to 9, focusing on parieto-central, parietal, pietro-occipital and occipital sites, i.e. the back of the head where most of the relevant information come from [18].

We performed two complementary analyses out of this experiment.

The first (online) analysis simply enabled us to compare the two modes in terms of spelling speed and accuracy, as well as to ask the subjects about their preferences.

The second (offline) analysis consisted in reprocessing part of the data from both modes, but using a fixed and identical amount of evidence. In other words, instead of using an optimal stopping criterion, this second analysis consisted in a time-based criterion in order to compare modes based on the same number of trials. As a consequence, if a difference in performance between modes remains, it won’t be due to a difference in the criterion itself but to a virtuous effect of it onto the subject’s engagement or motivation.

More details about this second study can be found in [19].

3. Results

3.1. Study 1

In the initial spelling phase, where no error correction was performed yet, spelling accuracy reached 64% ± 21 (SD) in the faster mode and 80% ± 18 (SD) in the slower mode. This corresponds to an information transfer rate of 4.52 ± 1.2 (SD) and 4.31 ± 1 correct letters per minute, respectively. This is a high performance level compared for instance with 1.57 correct letters per minute in [20]. This high level of spelling accuracy is to be kept in mind when interpreting the outcome of error correction.

But prior to error correction is the error detection step. Performance in error detection is reflected by the related sensitivity (the capacity to correctly detect errors), specificity (the capacity to correctly detect correct trials) and accuracy (the global efficacy of the classifier). At the group level, we obtained 63%, 88% and 78%, respectively.

The quality of error detections impacts the performance in error correction. Over the whole group, spelling accuracy only improved by 0.5% with automatic error correction. However, inter-individual variability was 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%). Beyond the poor gain obtained on average, it is important to understand the reasons behind this high inter-subject variability for future application of automatic error correction.

Interestingly, error detection specificity correlated with spelling accuracy, over subjects ($r = 0.68$, $P < 0.01$). Moreover, spelling accuracy prior to correction was also highly correlated with the quality of the classifier’s second best guess, which we referred to as theta ($r = 0.87$, $P < 0.0001$). Theta is simply the percentage of correct second best guess in case of an error (independently of whether this error would be detected or not). On average, theta
was found equal to 34%, which happened to be very close to the observed good correction rate\(^1\) of 36%.

Moreover, theta, as well as the global spelling accuracy did correlate with the subject’s responses to question “How well did you control the machine?” \((r = 0.75, P < 0.001; r = 0.74, P < 0.01)\) and question “Did the machine perform well?” \((r = 0.6, P < 0.05; r = 0.61, P < 0.05)\).

Finally, what was also very striking is the split into two groups according to the individual specificities in error detection. Six subjects presented quite low specificities, below 75%, while specificities for the other 10 subjects rose above 85%. As reported in Table 1, the first group corresponded to good performer who also benefited more from automatic correction and reported accordingly a rather positive subjective feeling regarding the BCI performance and the usefulness of error correction. Conversely, the second group corresponded to poor performer whose performance was even degraded by the automatic correction. They reported accordingly a rather negative subjective feeling about the BCI performance and the usefulness of error correction.

### 3.2. Study 2

In the fixed condition, the online spelling accuracy was 71% ± 16 (SD), which corresponds to a transfer rate in bits/minute of 18.8. In the adaptive condition, it was 80% ± 14 (SD), for an average of 57 ± 4 (SD) flashes, which corresponds to 24.1 bits/minute (Fig. 3A).

Wilcoxon tests revealed that both the spelling accuracy and the bit rate are significantly higher in the adaptive condition compared to the fixed condition \((P < 0.01)\) for both tests. Importantly, the number of flashes was not significantly different between the two conditions \((P = 0.1)\), it was even slightly lower in the adaptive condition.

This first online analysis reveals the better performance obtained with the adaptive condition.

Now, in the adaptive condition, the subjects knew and could effectively notice that the better they concentrate on the spelling task, the faster the (correct) spelling. This means that the adaptive mode might have triggered up subject’s motivation. Hence, part of the above online results may be attributable to an increase of the subject’s engagement into the task rather than to the adaptive capacities of the BCI system per se.

The second (offline) analysis was meant to quantify this part independently of the first and direct effect of optimizing the speed-accuracy trade-off.

---

\(^1\)The good correction rate is the percentage of detected true error trials that were appropriately corrected.

---

**Fig. 3.** Spelling performance and associated number of flashes obtained over the group in online study 2, for the fixed (red) and adaptive (blue) conditions (A). Evidence for an additional motivation effect: comparison of the performance obtained offline based on the same amount of data (2 sequences only) between the fixed (red) and adaptive (blue) conditions (B) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

**Table 1.** Separately for the group with high and low error detection specificity, respectively, this table shows the average spelling performance before error correction, the Theta value, the gain in accuracy due to automatic error correction and the averaged answers to two questions that the subjects answered at the end of the experiment.

<table>
<thead>
<tr>
<th></th>
<th>High Specificity</th>
<th>Low Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specificity</strong></td>
<td>&lt;75%</td>
<td>&gt;85%</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>46%</td>
<td>72%</td>
</tr>
<tr>
<td><strong>Theta</strong></td>
<td>29%</td>
<td>45%</td>
</tr>
<tr>
<td><strong>Gain in accuracy</strong></td>
<td>–5%</td>
<td>+4%</td>
</tr>
<tr>
<td><strong>Subjective report on machine’s performance</strong></td>
<td>4.5/10</td>
<td>6.6/10</td>
</tr>
<tr>
<td><strong>Feeling of control</strong></td>
<td>5.2/10</td>
<td>7.4/10</td>
</tr>
</tbody>
</table>

Therefore, data from both conditions were reanalyzed offline, using the same time-based stopping criterion: a decision was made after the 24 first flashes (i.e. 2 sequences). The obtained spelling accuracy proved significantly higher in the adaptive than in the fixed condition \((P < 0.01)\) (Fig. 3B). Since the number of observations was the same for both conditions in this analysis, the ensuing information transfer rate proved also significantly higher in that same condition \((P < 0.01)\). This last result emphasizes a complementary increase in performance in the adaptive mode, due to an increase in subject’s engagement into the task.

As shown on Fig. 3A, the adaptive mode outperformed the fixed mode by an average of ~10% in accuracy, where ~6% of this increase was attributable to motivation (Fig. 3B). This leaves a ~4% increase due to the method’s itself, an estimate that we could confirm with independent simulations [21].

### 4. Discussion and conclusion

In the first online study, we could show that automatic error correction can be implemented in BCI, using EEG responses evoked by the machine’s feedback. This confirms that responses to feedbacks can be detected online, from single trials. For comparison, Dal Seno et al. tested two subjects and obtained an averaged sensitivity and specificity of 62% and 68%, respectively [11]. In nine healthy subjects, Spüler et al. report a 40% sensitivity and 96% specificity, using a biased classifier to favor specificity [20]. Indeed, a high specificity is desirable in order to guarantee 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, simply because spelling accuracy is fairly high, which yields much more training samples for correct than incorrect feedback responses.
However, our approach relies on a cumbersome calibration phase and obtained results were poor, on average, so that many subjects did not benefit from it and preferred spelling without it. Nevertheless, a very interesting result is that the higher the spelling accuracy, the higher the performance in error detection and correction. At first, this might sound counter-intuitive, since the higher the spelling accuracy, the more difficult the rare error detection.

However, contrary to Visconti et al. [22], we do observe that good performers achieve better correction. This strongly suggests that the more the subject engages into the task, the higher the performance in terms of both spelling accuracy and error correction.

Indeed, attentional focus might have a twofold beneficial effect: increasing the signal-to-noise ratio of responses to feedbacks on the one hand, and increasing the relevance of the classifier's second best guess in case of an error (the Theta value), on the other hand.

This is a strong indication in favor of a possible use of the P300-speller to train subjects in their abilities to focus attention [23].

In the second online study, we proposed and validated an adaptive decision-making strategy to overcome the limitation of the traditional time-based decision criterion used in the P300-speller and BCI in general. This strategy consists in endowing the machine with some flexible, optimal reaction time. Indeed, in a way that mimics the reaction time of human beings, which relies on the amount and quality of accumulated evidence from incoming sensory inputs, the stopping criterion we implemented trades speed and accuracy. A short reaction time will be produced whenever the accumulated evidence in favor of a single choice is strong. Conversely, the reaction time should be longer whenever evidence is noisy and ambiguous, since more data will be needed to make a reliable decision. Compared to a time-based criterion, this can accommodate the slow intrinsic fluctuations of the electro-physiological signals, which might be due to fluctuations in attention. In the P300-Speller, this is particularly relevant, since sustained attention is what is required from the subject to keep performing the task efficiently.

The first significant effect we observed is that, for the same spelling duration, the user is able to spell letters more accurately. The time saved by stopping the flashes earlier, whenever possible, was efficiently reallocated to letters that required longer stimulation time in order to be accurately identified. Equivalently, given an objective in terms of accuracy, fewer flashes should be required with adaptive decision making, on average.

Secondly, a very interesting and significant effect of motivation could be observed online. Indeed, spelling accuracy was found higher for adaptive sessions than for fixed ones when these datasets were reprocessed offline with the same stopping criterion. This suggests that the subjects were on average more engaged into the task during the adaptive session, thus producing electro-physiological responses with a larger signal-to-noise ratio, which resulted in higher spelling accuracies. Indeed, the N1 and P300 responses, which are the electrophysiological responses used to identify the target, are known to reflect the participant's involvement in the task [24]. The P300 has also been shown to increase with motivation in a BCI context [25]. The fact that spelling accuracy is optimized by continuously and explicitly adapting the stimulation to the user's need appears to create a virtuous cycle by boosting the user's motivation.

These two studies illustrate different ways of improving BCI performance that are not specific to the P300-Speller. Importantly, both ways emphasize the importance of the real-time close-loop interaction between the user and the machine. Indeed, nor the responses to feedback, neither the effect of adaptive decision-making, could have been demonstrated in an offline experiment. This highlights the importance of possible virtuous co-adaptation mechanisms in BCI.

The proposed advances now need to be evaluated in patients. This is the aim of a current collaborative clinical trial in patients suffering from Amyotrophic Lateral Sclerosis (ALS) [26].

Note that complementary innovations could further improve the clinical efficacy of BCI such as the use of wireless EEG systems with few sensors. We could show in the second study that high information transfer rate could be achieved in the P300-Speller with only 9 carefully located sensors. Another promising avenue is the use of dictionaries and word prediction software in order to make the spelling faster and even more accurate, just like in modern typing.

However, one important limitation of the classical P300-Speller is that it relies on vision and eye-gaze control. Therefore, auditory-based alternatives have to be explored [27]. For now, those BCIs offer less degree of freedom than visual BCIs. Nevertheless, they might also benefit from the advances presented in this paper.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

Acknowledgments

This work was supported by the French ANR project ANR-DEFIS 09-EMER-002 CoAdapt.

References


Treder MS, Blankertz B. (C)overt attention and visual speller design in an ERP-based brain-computer interface. Behav Brain Funct 2010;6:28.

